frontiers | Research Topics

Optics and machine vision for marine observation

Edited by
Hong Song, Ran Liao, Rizwan Ali Naqvi and Surui Xie

Published in
Frontiers in Marine Science



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-3329-1
DOI 10.3389/978-2-8325-3329-1

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Optics and machine vision for marine observation

Topic editors

Hong Song — Zhejiang University, China

Ran Liao — Tsinghua University, China

Rizwan Ali Naqvi — Sejong University, Republic of Korea

Surui Xie — University of Houston, United States

Citation

Song, H., Liao, R., Naqvi, R. A., Xie, S., eds. (2023). *Optics and machine vision for marine observation*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3329-1

Table of contents

06 Editorial: Optics and machine vision for marine observation
Hong Song, Syed Raza Mehdi, Mengjie Wang, Ran Liao, Rizwan Ali Naqvi and Surui Xie

10 Submarine Optical Fiber Sensing System for the Real-Time Monitoring of Depth, Vibration, and Temperature
Zhengyong Liu, Shengqi Zhang, Chengkun Yang, Weng-Hong Chung and Zhaohui Li

21 PSS-net: Parallel semantic segmentation network for detecting marine animals in underwater scene
Yu Hwan Kim and Kang Ryoung Park

34 Robust segmentation of underwater fish based on multi-level feature accumulation
Adnan Haider, Muhammad Arsalan, Jiho Choi, Haseeb Sultan and Kang Ryoung Park

51 Underwater image restoration through regularization of coherent structures
Usman Ali and Muhammad Tariq Mahmood

59 Single underwater image enhancement based on differential attenuation compensation
Yunting Lai, Zhuang Zhou, Binghua Su, Zhe Xuanyuan, Jialin Tang, Jinghui Yan, Wanxin Liang and Jiongjiang Chen

77 RACE-SM: Reliability and adaptive cooperation for efficient UWSNs using sink mobility
Irfan Ahmad, Taj Rahman, Inayat Khan, Salman Jan, Shahrulniza Musa and M. Irfan Uddin

93 Underwater object detection algorithm based on attention mechanism and cross-stage partial fast spatial pyramidal pooling
Jinghui Yan, Zhuang Zhou, Dujuan Zhou, Binghua Su, Zhe Xuanyuan, Jialin Tang, Yunting Lai, Jiongjiang Chen and Wanxin Liang

108 UCRNet: Underwater color image restoration *via* a polarization-guided convolutional neural network
Haofeng Hu, Yizhao Huang, Xiaobo Li, Liubing Jiang, Li Che, Tiegen Liu and Jingsheng Zhai

119 RMP-Net: A structural reparameterization and subpixel super-resolution-based marine scene segmentation network
Jiongjiang Chen, Jialin Tang, Shounan Lin, Wanxin Liang, Binghua Su, Jinghui Yan, Dujuan Zhou, Lili Wang, Yunting Lai and Benxi Yang

133 UMOTMA: Underwater multiple object tracking with memory aggregation
Zhicheng Hao, Jun Qiu, Haimiao Zhang, Guangbo Ren and Chang Liu

145 A dual stream hierarchical transformer for starvation grading of golden pomfret in marine aquaculture
Kaijian Zheng, Renyou Yang, Rifu Li, Liang Yang, Hao Qin and Ziyun Li

158 Multiscale attention-based detection of tiny targets in aerial beach images
Shurun Gao, Chang Liu, Haimiao Zhang, Zhehai Zhou and Jun Qiu

169 The analysis and design of deep-sea lighting field based on spectral transfer function
Xiangqian Quan, Yucong Wei, Kaibin Liu and Bo Li

180 Deep learning-based marine big data fusion for ocean environment monitoring: Towards shape optimization and salient objects detection
Sulaiman Khan, Inam Ullah, Farhad Ali, Muhammad Shafiq, Yazeed Yasin Ghadi and Taejoon Kim

191 Multi-scale ship target detection using SAR images based on improved Yolov5
Muhammad Yasir, Liu Shanwei, Xu Mingming, Sheng Hui, Md Sakaouth Hossain, Arife Tugsan Isiacik Colak, Dawei Wang, Wan Jianhua and Kinh Bac Dang

212 Single underwater image enhancement based on adaptive correction of channel differential and fusion
Zefeng Zhao, Zhuang Zhou, Yunting Lai, Tenghui Wang, Shujie Zou, Haohao Cai and Haijun Xie

227 Deep focus-extended darkfield imaging for *in situ* observation of marine plankton
Tao Chen, Jianping Li, Wenqi Ma, Guannan Guo, Zhenyu Yang, Zhenping Li and Jianping Qiao

241 Energy-efficient clustering protocol for underwater wireless sensor networks using optimized glowworm swarm optimization
Salil Bharany, Sandeep Sharma, Naif Alsharabi, Elsayed Tag Eldin and Nivin A. Ghamry

257 Laser-induced breakdown spectroscopy instrument and spectral analysis for deep-ocean Fe-Mn crusts
Guang Yang, Guanyu Chen, Zixuan Cai, Xiangqian Quan and Yang Zhu

266 Investigating the rate of turbidity impact on underwater spectral reflectance detection
Hong Song, Syed Raza Mehdi, Zixin Li, Mengjie Wang, Chaopeng Wu, Vladimir Yu Venediktov and Hui Huang

274 Identification of marine oil spill pollution using hyperspectral combined with thermal infrared remote sensing
Junfang Yang, Yabin Hu, Jie Zhang, Yi Ma, Zhongwei Li and Zongchen Jiang

288 Real-time detection of deep-sea hydrothermal plume based on machine vision and deep learning
Xun Wang, Yanpeng Cao, Shijun Wu and Canjun Yang

301 An underwater image enhancement model for domain adaptation
Xiwen Deng, Tao Liu, Shuangyan He, Xinyao Xiao, Peiliang Li and Yanzhen Gu

311 An early warning model for starfish disaster based on multi-sensor fusion
Longyu Li, Tao Liu, Hui Huang, Hong Song, Shuangyan He, Peiliang Li, Yanzhen Gu and Jiawang Chen

frontiers | Frontiers in Marine Science

TYPE Editorial
PUBLISHED 14 August 2023
DOI 10.3389/fmars.2023.1256183

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Oliver Zielinski,
Leibniz Institute for Baltic Sea Research (LG), Germany

*CORRESPONDENCE
Hong Song
hongsong@zju.edu.cn

RECEIVED 10 July 2023
ACCEPTED 02 August 2023
PUBLISHED 14 August 2023

CITATION
Song H, Mehdi SR, Wang M, Liao R, Naqvi RA and Xie S (2023) Editorial: Optics and machine vision for marine observation. *Front. Mar. Sci.* 10:1256183. doi: 10.3389/fmars.2023.1256183

COPYRIGHT
© 2023 Song, Mehdi, Wang, Liao, Naqvi and Xie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Optics and machine vision for marine observation

Hong Song[1]*, Syed Raza Mehdi[1], Mengjie Wang[1], Ran Liao[2], Rizwan Ali Naqvi[3] and Surui Xie[4]

[1]Department of Ocean Engineering, Ocean College, Zhejiang University, Zhoushan, Zhejiang, China, [2]Shenzhen International Graduate School, Tsinghua University, Shenzhen, China, [3]Department of Intelligent Mechatronics Engineering, Sejong University, Seoul, Republic of Korea, [4]Department of Civil and Environmental Engineering, University of Houston, Houston, TX, United States

KEYWORDS

underwater optics, underwater imaging, image enhancement, image processing, machine learning, marine observation, object recognition

Editorial on the Research Topic
Optics and machine vision for marine observation

The aquatic ecosystem of the planet makes up a sizable amount of 71% of its surface that contain numerous living forms and an abundance of organic and inorganic resources throughout this enormous area (Issac and Kandasubramanian, 2021). Scientists and researchers have long been enthralled by the immense and enigmatic expanse of the marine ecosystems. The ocean's intricate ecosystems, diverse marine life, and the profound impact they have on our planet make understanding and monitoring these environments crucial (Maximenko et al., 2019). Both anthropogenic and natural activities have significantly increased recently, causing ecological problems in the marine environment (Huang et al., 2023). To successfully address and mitigate the resulting ecological mutilations, these disturbances call for the development of quick monitoring and mitigation mechanisms. As a result, the scientific community has been forced to explore numerous routes to push the limits of marine observation.

Underwater ecosystems have been mostly shrouded in darkness due to light attenuation, hindering comprehensive observation and data collection. But improvements in optics have fundamentally altered our capacity to perceive the underwater environment. Advances in high-resolution image capture, video recording, and spectral data acquisition have been made possible by cutting-edge imaging technology like underwater cameras, spectrometers, and hyperspectral sensors (Song et al., 2021a; Shahani et al., 2021). Through the study of species' behavior, distribution, and interactions, hidden ecosystems are revealed and scientists are able to explore marine habitats in new detail.

Automated analysis of underwater imagery has been made possible by machine vision techniques used in conjunction with optics. Computers can now extract complex traits and accurately categorize marine organisms thanks to deep learning techniques, a subset of machine learning that has revolutionized image processing and pattern identification. There are many new possibilities for marine surveillance now that machine vision systems, optics, and deep learning approaches have been combined. Automation, data analysis, and real-time monitoring are just a few advantages that machine vision and deep learning algorithms together offer. The topic of marine species tracking and identification is one of

the most notable applications (Chuang et al., 2016). Massive volumes of underwater imagery may be quickly analyzed using deep learning algorithms, which can then accurately and automatically identify and classify aquatic organisms (Song et al., 2020; Song et al., 2021b). These developments are essential for following migration patterns, evaluating the health of marine populations, and spotting possible threats to biodiversity. Machine vision and deep learning speed up research efforts by reducing the time-consuming and labor-intensive process of manual identification, enabling scientists to make educated conclusions about conservation measures and policy-making.

In conjunction with machine vision algorithms, remote sensing systems can monitor changes in ocean currents, sea surface temperature, and the spread of dangerous algal blooms (Son et al., 2015). For studying climate patterns, predicting weather occurrences, and reducing the possible effects of natural disasters on coastal communities, these real-time measurements are crucial. Additionally, the monitoring of human activities and their effects on marine habitats is made easier by the integration of optics, machine vision, and deep learning. Machine vision systems can monitor and identify potential pollution, illicit fishing, and habitat devastation (Mehdi et al., 2022; Yasir et al., 2023).

Understanding and maintaining a close eye on the dynamics and health of oceans depends heavily on marine observation. We can employ machine vision, which focuses on creating algorithms and systems for understanding visual data, and optics, which deals with the study and manipulation of light, to better observe and understand marine ecosystems. For this purpose, the Research Topic "*Optics and machine vision for marine observation*" focuses to explore the intersection of optics, machine vision, and deep learning technologies and their applications in making the field of marine observation more effective. It provides a collection of recent findings, developments, and innovative strategies related to underwater sensors, imaging systems, computer vision algorithms, and data analysis techniques that leverage optics and machine vision technologies for various aspects of marine observation. The Research Topic explores the transformative potential of optics and machine vision and their applications in contributing to the advancements of marine observation systems. The Research Topic is comprised of 24 articles, collectively representing the contribution of 118 authors (Table 1).

A wide domain of research is involved in the development and implementation of optics and machine vision for marine observation, including optical sensors and monitoring systems, image processing, deep learning techniques, deep-sea illumination, spectral image analysis, etc. Several researchers address the development of underwater monitoring methods based on optical fiber sensing for real-time study of

TABLE 1 Summary of chapters published in this Research Topic.

DOI	Title	Keywords	Authors
10.3389/fmars.2022.922669	Submarine optical fiber sensing system for the real-time monitoring of depth, vibration, and temperature	optical fiber sensing, fiber Bragg grating, submarine real-time monitoring, environmental monitoring, optical fiber sensor	Liu Z, Zhang S, Yang C, Chung W-H and Li Z
10.3389/fmars.2022.1003568	PSS-net: Parallel semantic segmentation network for detecting marine animals in underwater scene	detecting marine animal, underwater scene, protective colors, PSS-net, attention technique	Kim YH and Park KR
10.3389/fmars.2022.1010565	Robust segmentation of underwater fish based on multi-level feature accumulation	artificial intelligence, marine environment, underwater computer vision, fish segmentation, EFS-net and MFAS-net	Haider A, Arsalan M, Choi J, Sultan H and Park KR
10.3389/fmars.2022.1024339	Underwater image restoration through regularization of coherent structures	underwater images, image restoration, robust regularization, coherent structures, optimization problem	Ali U and Mahmood MT
10.3389/fmars.2022.1047053	Single underwater image enhancement based on differential attenuation compensation	underwater image, image enhancement, contrast stretching, differential attenuation compensation, machine vision	Lai Y, Zhou Z, Su B, Xuanyuan Z, Tang J, Yan J, Liang W and Chen J
10.3389/fmars.2022.1030113	RACE-SM: Reliability and adaptive cooperation for efficient UWSNs using sink mobility	UWSNs, energy efficient routing, routing protocols, sink mobility, cooperative routing	Ahmad I, Rahman T, Khan I, Jan S, Musa S and Uddin MI
10.3389/fmars.2022.1056300	Underwater object detection algorithm based on attention mechanism and cross-stage partial fast spatial pyramidal pooling	Underwater Object detection, ACFP-YOLO, YOLOv7, attention, SPPFCSPC	Yan J, Zhou Z, Zhou D, Su B, Xuanyuan Z, Tang J, Lai Y, Chen J and Liang W
10.3389/fmars.2022.1031549	UCRNet: Underwater color image restoration *via* a polarization-guided convolutional neural network	Polarization, polarimetric imaging, scattering media, imaging recovery, physical imaging	Hu H, Huang Y, Li X, Jiang L, Che L, Liu T and Zhai J
10.3389/fmars.2022.1032287	RMP-Net: A structural reparameterization and subpixel super-resolution-based marine scene segmentation network	submarine exploration, underwater scene, RMP-Net, structural re-parameterization, multiscale fusion	Chen J, Tang J, Lin S, Liang W, Su B, Yan J, Zhou D, Wang L, Lai Y and Yang B

(Continued)

TABLE 1 Continued

DOI	Title	Keywords	Authors
10.3389/fmars.2022.1071618	UMOTMA: Underwater multiple object tracking with memory aggregation	artificial intelligence, underwater multiple object tracking, marine environment, long-short term memory, vision transformer	Hao Z, Qiu J, Zhang H, Ren G and Liu C
10.3389/fmars.2022.1039898	A dual stream hierarchical transformer for starvation grading of golden pomfret in marine aquaculture	neural network, transformer, starvation grading, marine image processing, behavior recognition	Zheng K, Yang R, Li R, Yang L, Qin H and Li Z
10.3389/fmars.2022.1073615	Multiscale attention-based detection of tiny targets in aerial beach images	tiny object detection, multiscale attention, feature pyramid network, attention mechanism, unmanned aerial vehicle	Gao S, Liu C, Zhang H, Zhou Z and Qiu J
10.3389/fmars.2022.1058201	The analysis and design of deep-sea lighting field based on spectral transfer function	underwater optics, underwater imaging, underwater lighting, marine observation, object recognition	Quan X, Wei Y, Liu K and Li B
10.3389/fmars.2022.1094915	Deep learning-based marine big data fusion for ocean environment monitoring: Towards shape optimization and salient objects detection	data fusion, marine big data, ocean environment, underwater saliency detection, underwater image processing	Khan S, Ullah I, Ali F, Shafiq M, Ghadi YY and Kim T
10.3389/fmars.2022.1086140	Multi-scale ship target detection using SAR images based on improved Yolov5	synthetic aperture radar (SAR), ship identification, artificial intelligence, deep learning (DL), YOLOv5S, SAR ship detection dataset (SSDD), AirSARship	Yasir M, Shanwei L, Mingming X, Hui S, Hossain MS, Colak ATI, Wang D, Jianhua W and Dang KB
10.3389/fmars.2022.1058019	Single underwater image enhancement based on adaptive correction of channel differential and fusion	underwater image processing, image enhancement, histogram stretching, color correction, fusion	Zhao Z, Zhou Z, Lai Y, Wang T, Zou S, Cai H and Xie H
10.3389/fmars.2023.1074428	Deep focus-extended darkfield imaging for *in situ* observation of marine plankton	underwater imaging, deep learning, focus extension, ocean observation, marine plankton	Chen T, Li J, Ma W, Guo G, Yang Z, Li Z and Qiao J
10.3389/fmars.2023.1117787	Energy-efficient clustering protocol for underwater wireless sensor networks using optimized glowworm swarm optimization	UWSN, energy, GSO, routing, clustering, bio-inspired	Bharany S, Sharma S, Alsharabi N, Tag Eldin E and Ghamry NA
10.3389/fmars.2023.1135058	Laser-induced breakdown spectroscopy instrument and spectral analysis for deep-ocean Fe-Mn crusts	deep ocean, marine spectral analysis, marine resource exploration, laser-induced breakdown spectroscopy, marine Fe-Mn crusts, parameter optimization	Yang G, Chen G, Cai Z, Quan X and Zhu Y
10.3389/fmars.2023.1031869	Investigating the rate of turbidity impact on underwater spectral reflectance detection	spectral reflectance, spectral imaging, turbidity, spectral detection, liquid crystal tunable filters, spectral features, underwater scattering	Song H, Mehdi SR, Li Z, Wang M, Wu C, Venediktov VY and Huang H
10.3389/fmars.2023.1135356	Identification of marine oil spill pollution using hyperspectral combined with thermal infrared remote sensing	marine oil spills, hyperspectral remote sensing, thermal infrared remote sensing, oil pollution types identification, deep learning	Yang J, Hu Y, Zhang J, Ma Y, Li Z and Jiang Z
10.3389/fmars.2023.1124185	Real-time detection of deep-sea hydrothermal plume based on machine vision and deep learning	hydrothermal plume, deep-sea, real-time, object detection, deep learning, transfer learning, YOLOv5	Wang X, Cao Y, Wu S and Yang C
10.3389/fmars.2023.1138013	An underwater image enhancement model for domain adaptation	underwater image, image enhancement, underwater dataset, domain adaptation, deep learning	Deng X, Liu T, He S, Xiao X, Li P and Gu Y
10.3389/fmars.2023.1167191	An early warning model for starfish disaster based on multi-sensor fusion	starfish disaster, multi-sensor fusion, early-warning model, self-supervised model, feature selection	Li L, Liu T, Huang H, Song H, He S, Li P, Gu Y and Chen J

environmental parameters (Liu et al.), and water quality observation based on multi-sensor fusion for early warning of starfish disaster (Li et al.). Two studies discuss underwater sensor networks and protocols for explorations of underwater resources through efficient data collection (Ahmad et al.; Bharany et al.). Numerous papers deliver improved techniques and applications of deep learning for underwater object detection while several studies highly concentrated on underwater image enhancement in support of algorithm development, validation, and verification. Multiple papers explore the applications of deep learning for underwater object detection (fish classes, and organic and inorganic submarine objects: Yan et al.; Khan et al.; hydrothermal plumes detection: Wang et al.) and image segmentation (fish: Kim and Park; Haider et al.; Chen, J. et al.). One study proposes an advanced trajectory tracking mechanism for underwater fish classes including multi-object detection (Hao et al.). Another study proposes and assesses a starvation grading model for fish class based on image processing and CNN that can benefit the field of fisheries (Zheng et al.). For aerial-based monitoring of coastal areas, a paper suggests small size objects detection technique based on CNN (Gao et al.). Papers based on spectral technologies address a range of topics including deep-sea illumination to compensate light attenuation (Quan et al.), effects of turbidity on spectral imaging (Song et al.), and spectral imaging based deep-sea mineral exploration (Yang, G. et al.). In the

field of marine observation, remote sensing provides valuable insights into the state of marine environment. Two papers contributed to the field of ocean remote sensing using hyperspectral imaging and CNNs for the detection of ships (Yasir et al.), and the classification of oil spills (Yang, J. et al.). Several contributions in the field of underwater image processing include image restoration (Ali and Mahmood; color restoration: Hu et al.), and image enhancement (Lai et al.; Zhao et al.; Chen, T. et al.; Deng et al.).

The ability to monitor and understand the marine environment has changed dramatically as a result of the merging of optics and machine vision technologies with marine research. These developments have given scientists the tools they need to solve the urgent ecological problems that are being caused by both natural events and human activity. These potent tools can be used by researchers to gain insightful knowledge of the marine ecosystem, facilitating well-informed decision-making and efficient mitigation measures. As a result, the limits of scientific understanding in marine science are being widely pushed, advancing our comprehension of this complex field to unprecedented heights.

Author contributions

HS: Data curation, Investigation, Supervision, Writing – original draft, Writing – review & editing. SM: Data curation, Investigation, Writing – original draft, Writing – review & editing. MW: Data curation, Investigation, Writing – original draft, Writing – review & editing. RL: Data curation, Investigation, Writing – original draft, Writing – review & editing. RN: Data curation, Investigation, Writing – original draft, Writing – review & editing. SX: Data curation, Investigation, Writing – original draft, Writing – review & editing.

Acknowledgments

We extend our appreciation to the researchers and scientists who provided this Research Topic with their insightful knowledge. We also express our sincere gratitude to the devoted reviewers for their meticulous consideration and constructive criticism. Their efforts were crucial in determining the success of this Research Topic.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Chuang, M.-C., Hwang, J.-N., Ye, J.-H., Huang, S.-C., and Williams, K. (2016). "Underwater fish tracking for moving cameras based on deformable multiple kernels," in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 47. 2467–2477.

Huang, H., Cai, H., Qureshi, J. U., Mehdi, S. R., Song, H., Liu, C., et al. (2023). Proceeding the categorization of microplastics through deep learning-based image segmentation. *Sci. Total Environ.* 896, 165308. doi: 10.1016/j.scitotenv.2023.165308

Issac, M. N., and Kandasubramanian, B. (2021). Effect of microplastics in water and aquatic systems. *Environ. Sci. pollut. Res.* 28, 19544–19562. doi: 10.1007/s11356-021-13184-2

Maximenko, N., Corradi, P., Law, K. L., Van Sebille, E., Garaba, S. P., Lampitt, R. S., et al. (2019). Toward the integrated marine debris observing system. *Front. Mar. Sci.* 6, 447. doi: 10.3389/fmars.2019.00447

Mehdi, S. R., Raza, K., Huang, H., Naqvi, R. A., Ali, A., and Song, H. (2022). Combining deep learning with single-spectrum UV imaging for rapid detection of HNSs spills. *Remote Sens.* 14 (3), 576. doi: 10.3390/rs14030576

Shahani, K., Song, H., Mehdi, S. R., Sharma, A., Tunio, G., Qureshi, J., et al. (2021). Design and testing of an underwater microscope with variable objective lens for the study of benthic communities. *J. Mar. Sci. Appl.* 20, 170–178. doi: 10.1007/s11804-020-00185-9

Son, Y. B., Choi, B.-J., Kim, Y. H., and Park, Y.-G. (2015). Tracing floating green algae blooms in the Yellow Sea and the East China Sea using GOCI satellite data and Lagrangian transport simulations. *Remote Sens. Environ.* 156, 21–33. doi: 10.1016/j.rse.2014.09.024

Song, H., Mehdi, S. R., Huang, H., Shahani, K., Zhang, Y., Raza, K., et al. (2020). Classification of freshwater zooplankton by pre-trained convolutional neural network in underwater microscopy. *Int. J. Adv. Comput. Sci. Appl.* 11 (7), P 252–258. doi: 10.14569/IJACSA.2020.0110733

Song, H., Mehdi, S. R., Wu, C., Li, Z., Gong, H., Ali, A., et al. (2021a). Underwater spectral imaging system based on liquid crystal tunable filter. *J. Mar. Sci. Eng.* 9 (11), 1206. doi: 10.3390/jmse9111206

Song, H., Mehdi, S. R., Zhang, Y., Shentu, Y., Wan, Q., Wang, W., et al. (2021b). Development of coral investigation system based on semantic segmentation of single-channel images. *Sensors* 21 (5), 1848. doi: 10.3390/s21051848

Yasir, M., Zhan, L., Liu, S., Wan, J., Hossain, Md S., Colak, A. T. I., et al. (2023). Instance segmentation ship detection based on improved Yolov7 using complex background SAR images. *Front. Mar. Sci.* 10, 1113669. doi: 10.3389/fmars.2023.1113669

frontiers | Frontiers in Marine Science

ORIGINAL RESEARCH
published: 24 June 2022
doi: 10.3389/fmars.2022.922669

Submarine Optical Fiber Sensing System for the Real-Time Monitoring of Depth, Vibration, and Temperature

Zhengyong Liu[1,2]*, *Shengqi Zhang*[1], *Chengkun Yang*[1], *Weng-Hong Chung*[3] *and Zhaohui Li*[1,2]

[1] Guangdong Provincial Key Laboratory of Optoelectronic Information Processing Chips and Systems, School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou, China, [2] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, China, [3] Photonics Research Center, Department of Electrical Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong, SAR, China

Ocean observation becomes increasingly important as the ocean climate changes diversely and the marine disasters (such as tsunamis, typhoon, and earthquakes) occur frequently, which typically requires widespread and reliable monitoring techniques. In such a scenario, this paper presents a submarine optical fiber sensing system to realize real-time monitoring of the environmental parameters. The system consists of an undersea optical interrogation module together with multiple fiber Bragg grating (FBG)-based sensors, particularly for the measurement of depth, vibration, and temperature. The experimentally demonstrated sensitivities of the pressure, temperature, and vibration sensors are -1.993 nm/MPa, 0.08 nm/°C, and 0.139 nm/g (g = 9.8 m/s^2), corresponding to the resolutions of 0.25 kPa, 0.006°C, and 0.004 g, respectively, based on the interrogation resolution of ~0.5 pm. To verify the feasibility and reliability of the proposed submarine sensing system, a prototype was developed and a proof test under the sea was conducted in an area close to Pearl River Estuary in China. The achieved results from the sea test show promising accuracy that is comparable to the commercially available electric-based sensors. Good characteristics of the surface water wave were observed by conducting the fast Fourier transform of the measured depth change, which shows a dominant frequency of ~0.25 Hz. The system provides the flexibility of replacing various optical fiber sensors easily and the capability of real-time monitoring in a remote way. The demonstrated submarine sensing system could find potential applications in real-time monitoring of the undersea ecosystem and the environmental evolution where multiparameter sensing is in demand.

Keywords: optical fiber sensing, fiber bragg grating, submarine real-time monitoring, environmental monitoring, optical fiber sensor

OPEN ACCESS

Edited by:
Surui Xie,
University of California, San Diego, United States

Reviewed by:
Yong Zhao,
Northeastern University, China
Carlos Marques,
University of Aveiro, Portugal

****Correspondence:***
Zhengyong Liu
liuzhengy@mail.sysu.edu.cn

Specialty section:
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

Received: *18 April 2022*
Accepted: *25 May 2022*
Published: *24 June 2022*

Citation:
Liu Z, Zhang S, Yang C, Chung W-H and Li Z (2022) Submarine Optical Fiber Sensing System for the Real-Time Monitoring of Depth, Vibration, and Temperature. Front. Mar. Sci. 9:922669. doi: 10.3389/fmars.2022.922669

INTRODUCTION

Ocean observation becomes vitally important due to the close relationship between marine ecology and human life. The variation of environmental parameters such as temperature, depth, salinity, and vibration in a specific region is a key vector to evaluate the change in the local ocean environment, which typically influences the balance of the undersea ecosystem and the climate.

Full characterization of these physical parameters in a wide range could help to establish the geophysical model in an attempt to understand the evolution of the ocean (Favali and Beranzoli, 2006). Some specific chemical sensing system can even contribute to investigate the substance cycling and the ecosystem around the hydrothermal plume (Gartman and Findlay, 2020). As extreme weather occurs frequently on the sea and causes disasters to coastal areas, it is of great significance to have real-time monitoring of the undersea environmental change. To achieve this, basically, a full monitoring network is required to map the physical parameters and build an effective forecast mechanism.

In recent years, various approaches have been proposed to realize monitoring of the submarine environment. There are several well-developed cabled seafloor observatories established to monitor the multiparameters under the sea, for instance, the NEPTUNE in Canada, MARs in the United States, ESONET in Europe, S-NET in Japan, etc. All of these undersea observatories are based on sophisticated measurement systems on the seabed and connected by submarine optical cables. Among them, the high-precision sensing system plays an important and irreplaceable role. Basically, most sensing system integrates various electrical sensors, which provide high sensitivity to certain measurands. The sensing performance depends on the materials and mechanical structure of the sensors. Since 2012, a Joint Task Force (JTF) on SMART cable systems chaired by Howe et al. (2019) was established to realize ocean and climate monitoring and disaster warning utilizing the submarine telecommunication optical cables. The main feature of such a system is to integrate individual electrical temperature, vibration, and pressure sensors into the repeater, which therefore works as a multiparameter sensing element on the seabed. Owing to the wide and long-distance network of the submarine cables, full characterization of the ocean environment and dynamics could be achieved (Howe et al., 2022).

In addition to electrical sensing approaches, optical fiber sensing technique becomes increasingly feasible in monitoring of marine information due to its advantages of low cost, compactness, immunity to electromagnetic interference and corrosion, long-distance capability, and easy compatibility to the optical communication system. In 2018, Marra et al. (2018) reported on laser interferometry *via* submarine optical cables and achieved monitoring of earthquakes. The vibration caused the slight change in the optical phase of the fiber-based interferometer. Such submarine monitoring system is beneficial to the use of existing optical cables. A similar approach was proposed by Zhan et al. (2021) in 2021, where the state of the polarization (SOP) was monitored to retrieve the seismic wave occurring along the optical cable. The advantage of such interferometry is its high sensitivity to microseism, which however lacks precise localization. To further achieve localization of a particular event, the distributed acoustic sensing (DAS) system was proposed using existing submarine dark fibers, which is proven efficient to monitor undersea vibrations caused by earthquakes, ocean waves, and microseisms (Jousset et al., 2018; Ajo-Franklin et al., 2019; Williams et al., 2019; Walter et al., 2020). This distributed approach is promising to construct a cable-based sensing network easily and realize monitoring of ocean dynamics, as existing submarine optical cables can be employed in an ultralow cost. One possible shortage of it may be the sensing capability of only vibration, which could be complemented by various sensors.

To achieve the multiparameter measurement, various optical fiber sensors of high performance have been investigated. By employing the architecture of fiber Bragg grating (FBG) or fiber interferometry, temperature, pressure, vibration, and even chemical parameters could be measured with high precision and resolution. The sensing systems consisting of those sensors have been employed successfully in structural health monitoring in industries of railway (Liu et al., 2017), oil and gas (Qiao et al., 2017), biomedical engineering (Najafzadeh et al., 2020), and civil engineering (Wu et al., 2021). To achieve high sensing performance, sensing configurations based on the detection of either the spectral wavelength (Leitão et al., 2021) or the optical intensity (Leal-Junior et al., 2018; Leal-Junior et al., 2019) are proposed. The proper enhancement of the demodulation techniques provides an improvement of the sensing accuracy. Regarding ocean information monitoring, the sensing performance in terms of sensitivity, accuracy, and limit of detection typically varies with the sensor structures (Min et al., 2021). Undersea pressure is one key parameter to obtain depth information, which is relevant to the sea surface height and tsunamis (Yu, 2015). Optical fiber pressure sensors have been developed with various sensitivities from ~0.004 nm/MPa (Xu et al., 1993) up to ~50 nm/MPa (Liu et al., 2018) by measuring the optical spectral shift with respect to pressure change. By varying the structure of the fiber interferometer, simultaneous measurement of temperature and pressure or salinity in a single sensor can be achieved (Zhao et al., 2019; Zhao et al., 2022a). A similar fiber interferometer was also reported by micromachining an in-line cavity to realize temperature and salinity sensing (Flores et al., 2019). Basically, optical fiber sensors provide full flexibility in measuring seawater parameters. However, most optical fiber sensors are demonstrated in laboratory tests and few are employed to monitor submarine environmental changes due to harsh conditions and requirements of strict sensing systems.

As a demonstration of the sea test, Wang et al. (2020) in 2020 developed an FBG sensor array along the marine optical cable and laid the cable undersea to monitor temperature and depth. The FBG sensors worked efficiently and provided profiling with high spatial resolution owing to the densely distributed pressure and temperature sensors. In this sensing scheme, the signal demodulation system is on the boat, meaning that the submarine part is passive. In another verification of the sea test, researchers from Ireland developed a compact optical fiber sensing system containing an FBG-based sensor and managed to measure temperature,

depth, and salinity (Duraibabu et al., 2017). The sensing system was mounted on a remotely operated vehicle (ROV). From these field tests, it can be seen that the undersea optical fiber sensors show high potential in multiparameter monitoring of the marine environment. However, the capability of the sensors and the relevant undersea sensing system is still under investigation, especially to improve the sensitivity and accuracy of the specific sensors as well as the compatibility of the sensing system with the optical communication system.

In this paper, we report a submarine optical fiber sensing system integrated with pressure, temperature, and vibration sensors to realize real-time monitoring of the undersea environment. The sensing system contains an undersea interrogation module to demodulate the optical sensing signal. Various optical fiber sensors could be connected to the system to establish the local monitoring network and eventually acquiring multiple parameters widely in real time. The demonstrated sensitivities of the pressure, temperature, and vibration sensors are -1.993 nm/MPa, 0.08 nm/°C, and 0.139 nm/g (g = 9.8 m/s^2), corresponding to the resolutions of 0.25 kPa, 0.006°C, and 0.004 g, respectively, if taking the interrogation resolution of 0.5 pm into account. The proposed optical fiber sensing system was proof-tested under the sea, and the monitored results are comparable to the commercial electrical sensors installed on-site. In contrast to the conventional underwater electric-based sensors, the novelty of the proposed submarine sensing system is that the system has flexible capability of multiplexing various FBG-based sensors by a simple connection and the measurands obtained by the sensors can be monitored in real-time *via* optical fiber cable. The proposed submarine sensing system could find potential applications in real-time monitoring of the undersea ecosystem and the environmental evolution where multiparameter sensing is highly required.

MATERIALS AND METHODS

Principle and Development of the Optical Fiber Sensors

The proposed temperature, pressure, and vibration sensors are developed based on the principle of FBG, which basically is sensitive to the perturbations caused by the external strain and temperature change. The strain can be induced by any mechanical stress exerted on the fiber such as pressure and vibration. In principle, FBG works as a filter to reflect back a specific wavelength propagating in the fiber core if a broadband incident light is launched into the optical fiber. As illustrated in **Figure 1A**, there is a dominating peak showing in the reflection spectrum, indicating that the wavelength at the peak is totally reflected while the other wavelengths transmit continuously to the end. The reflected wavelength, also called Bragg wavelength (λ_B), is determined *via* the phase-matching condition as expressed by (Erdogan, 1997):

$$\lambda_B = 2n_{eff}\Lambda \tag{1}$$

where n_{eff} is the effective refractive index of the fundamental mode propagating in the fiber core, and Λ is the pitch of the FBG. By fabricating the FBGs with different pitches in a strand of fiber, the refection spectrum shows various peaks, each of which corresponds to a certain FBG, as illustrated in **Figure 1B**. To inscribe the FBG in the core of the optical fiber, the standard phase-mask technique with 248-nm UV laser scanning (Cheng et al., 2003; Pospori et al., 2017) was utilized in this work. Such

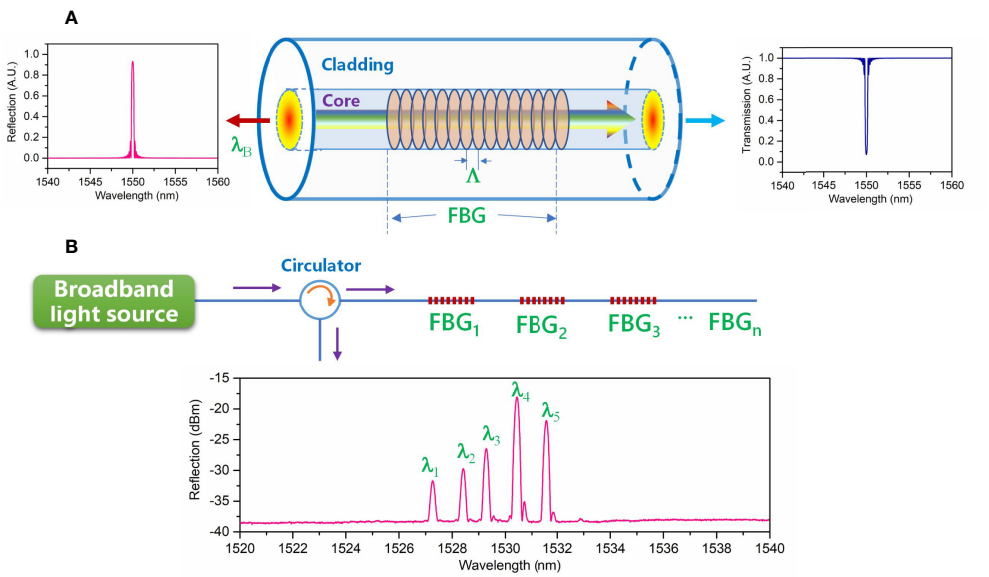


FIGURE 1 | Schematic illustration of **(A)** working principle of the optical fiber Bragg grating (FBG) and **(B)** multiplexing principle of various FBG-based sensors.

inscription method has been demonstrated to be reliable and repeatable during volume production, also suitable for the grating inscription of polymer optical fibers (Marques et al., 2013; Marques et al., 2018). After inscription, the refractive index of the core is modulated by the UV laser depending on the pitch of the phase mask used. The grating length depends on the design of the sensor, varying from a few millimeters to 10 mm.

Basically, the Bragg wavelength of FBG is a function of the refractive index of the fiber core and the grating pitch. Owing to the thermo-optic effect and photoelastic effect, the effective index and pitch can be altered by the external stress or temperature. The relationship between the Bragg wavelength change and the external measurand (i.e., temperature change: ΔT, strain: ε) can be deduced as:

$$\Delta\lambda_B = \lambda_B[\eta_1(1-p_e)\epsilon + (\eta_2 \cdot \alpha + \beta)\Delta T] \quad (2)$$

where p_e is the photoelastic coefficient, α is the coefficient of thermal expansion (CTE) of the fiber core, and β is the thermo-optic coefficient. In Eq. 2, η_1 is the factor to describe the amount of external stress caused by pressure or vibration transferred to the strain and η_2 is the transferring coefficient of thermal expansion, which are related to the particular packaging materials used to embed the FBG. These two coefficients could be determined through the calibration measurement of the sensors. Typically, the photoelastic effect is avoided to develop the temperature sensor so that only temperature change induces the Bragg wavelength shift, whereas the temperature effect cannot be reduced in the design of pressure and vibration sensors, which can only be compensated.

Interrogation Design of the Sensing System

The interrogation is necessary to demodulate the optical sensing signal that includes the change of measurands. In this work, a homemade interrogation system was designed and developed. **Figure 2** shows the schematic layout of the interrogation system, including a superluminescent diode (SLD, Denselight DL-CS5169A), laser driver, interrogation module (Ibsen I-MON 512), microprocessor, power controller, optical transceiver, circulator, and coupler. The coupler is utilized to split the optical signal into four paths, each of which is able to connect to the FBG-based sensors. The wavelength range of the SLD is over 80 nm covering from 1,510 to 1,590 nm. The minimum power of the SLD is 16 mW, which is enough for the four output ports split by the coupler. The temperature and current of the SLD are controlled by the driver. An optical circulator is used to send the incident light from port 1 to port 2 and circulate the reflection light from the FBG sensors to port 3. Regarding the demodulation method of the interrogation module, it employs an array of 512 photodiodes (PDs) spaced in a line to detect the intensity of the input light. There is a blazed grating and mirror in the module that can diffract the broadband light to various components of light with different wavelengths, as illustrated in the inset of **Figure 2**. After accurate calibration, each PD receives the intensity of a certain wavelength. The resolution of the hardware is about 170 pm. The reflected spectral peak of one FBG typically covers a wavelength bandwidth of about 500–700 pm, meaning that about 4–5 PDs are occupied to profile the peak spectrum. As FBG is typically apodized with a Gaussian profile during fabrication (Hill and Meltz, 1997), a Gaussian peak fit algorithm is utilized in the interrogation module to achieve precise peak detection according to the arrayed intensities from the illuminated PDs. Therefore, the fitted resolution of the interrogation could reach ~0.5 pm. By detecting the peaks in the reflection spectrum of FBGs, all of the measurands can be monitored and acquired remotely in real time *via* the optical fiber cable and the transceiver with a speed of 100 Mbps. The power and fiber cables provide the function of electricity and communication, which could enable the submarine sensing system work independently or jointly with other systems in a network.

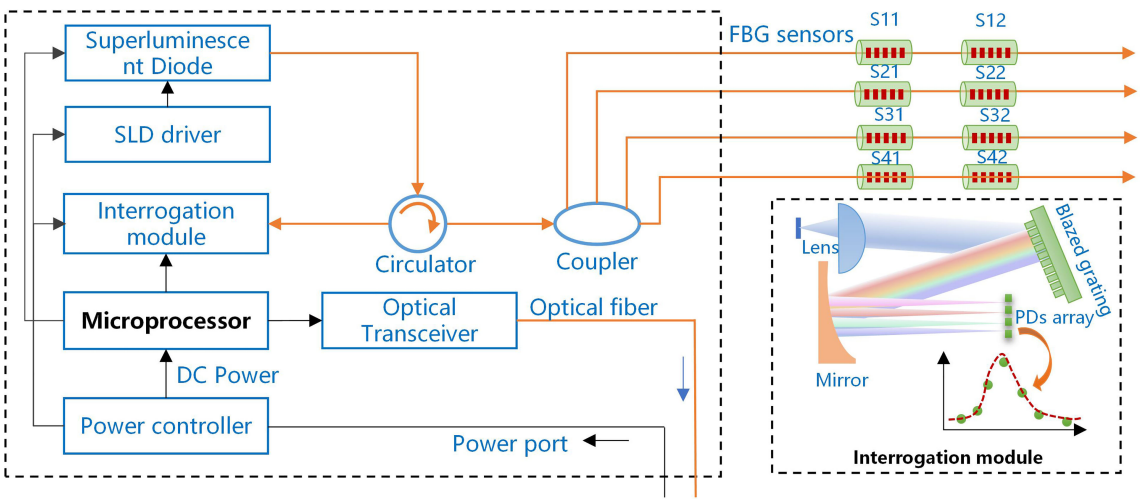


FIGURE 2 | Schematic layout of the proposed interrogation system.

To realize and demonstrate the feasibility of the proposed submarine optical fiber sensing system, a prototype of the submarine interrogation system was developed based on the design illustrated in **Figure 2**. As shown in **Figure 3**, the entire system was designed in a cylinder made of stainless steel in an attempt to withstand high hydraulic pressure. The height, inner diameter, and outer diameter of the main body is 566, 120, and 132 mm, respectively, which however could be optimized to a smaller size by rearranging the modules inside. The main body was tested in a pressurized chamber of 15 MPa over 3 h, and no leakage was found, meaning that the system could be employed under the sea for over 1,000 m if leaving a pressure margin of one-third. Regarding the sensing capability, it is possible to carry various FBG-based sensors by connecting to the optical fiber connectors. The entire system possesses the external connection of optical fiber cable to implement the fast and real-time communication from the remote station, for instance, on the shore or the ship.

Proof Test Under the Sea

To verify the feasibility and performance of the proposed submarine optical fiber sensing system, a proof test under the sea was carried out, as illustrated in **Figure 4**. **Figure 4A** shows the test route employed to create three different conditions in terms of depth, acceleration, and temperature, i.e., (D_1, A_1, T_1), (D_2, A_2, T_2), and (D_3, A_3, T_3). The test was carried out in a specific area close to Pearl River Estuary in China, as shown in **Figure 4B**. **Figures 4C, D** are the photos captured during the sea test, and the submarine optical fiber sensing system is powered by an external battery. There is a commercial Conductivity, Temperature, Depth (CTD) device installed close to the system to provide a comparison to the monitored results.

RESULTS

Calibration Results of the Optical Fiber Sensors

To test the proposed submarine optical fiber sensing system, highly sensitive pressure, temperature, and vibration sensors are developed based on the sensing principle of FBG described in *Principle and Development of the Optical Fiber Sensors*. The pressure sensor is utilized to monitor the depth determined by the relationship of $h = P/pg$, where P is the water pressure measured by the sensor, p is the mass density of the sea water, typically in the range of 1.02~1.07 g/cm^3, and g is the gravitational acceleration equal to 9.8 m/s^2. The pressure sensor is designed in a cylindrical tube of stainless steel, and the FBG with a Bragg wavelength of ~1,540 nm is packaged inside. As schematically illustrated in **Figure 5A**, the FBG is aligned in the center of the cylindrical tube and prestressed at one end. Once the water pressure is exerted on the top, the prestressed FBG gets released to some extent according to the value of pressure, resulting in a blue shift of the Bragg wavelength. The performance of the pressure sensitivity was characterized by placing the sensor in a hydraulic pressure chamber and increasing the pressure from 0 kPa to 0.5 MPa. For each state of the pressure, the Bragg wavelength was recorded. **Figure 5A** plots the response of the pressure sensor in terms of Bragg wavelength with respect to the applied pressure. From the experimentally measured results, the pressure sensitivity is -1.993 nm/MPa. The maximum calibrated pressure is up to 0.5 MPa, meaning that the depth variation could reach ~50 m.

Similarly, the temperature sensor was characterized by placing it in an environment with the temperature controlled precisely.

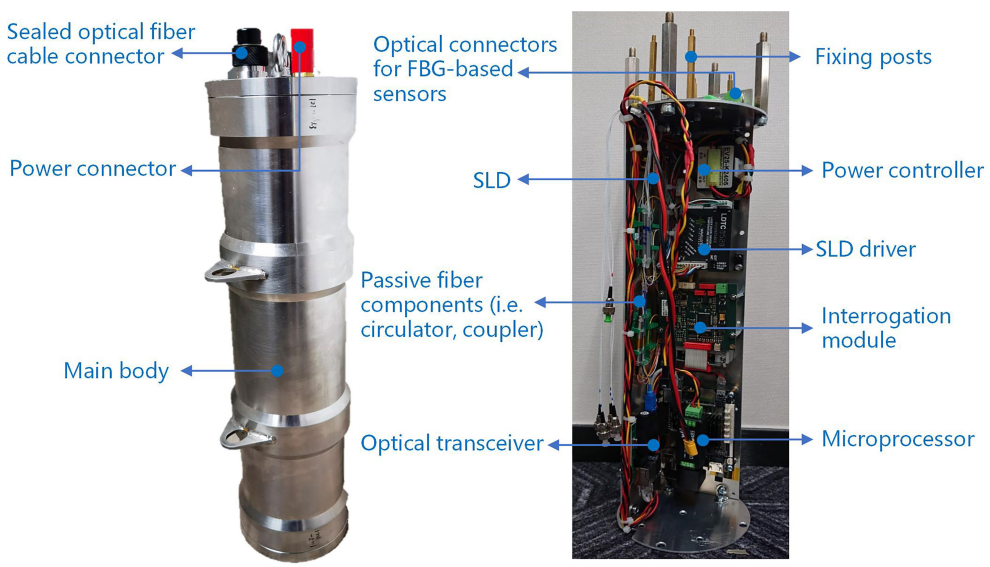


FIGURE 3 | Captured image of the prototype of the developed submarine interrogation system.

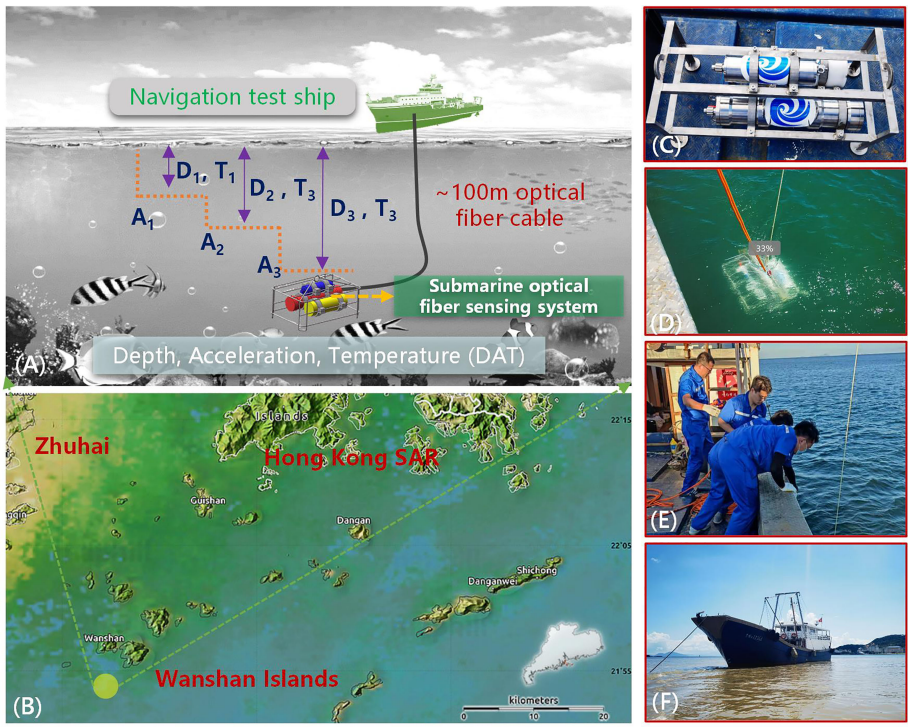


FIGURE 4 | Illustration of the sea test conducted to verify the submarine optical fiber sensing system in a specific area close to Pearl River Estuary in China, where panel **(A)** is the test route employed to introduce three different conditions in terms of depth, acceleration, and temperature, panel **(B)** shows the test area near Wanshan Islands, and panels **(C–F)** are the photos captured during the field test.

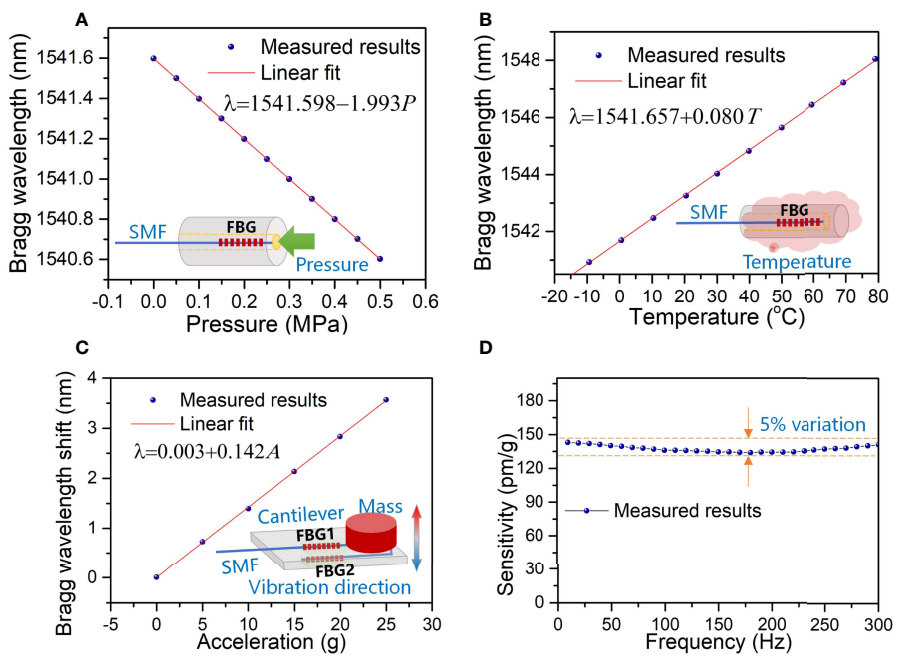


FIGURE 5 | Experimentally measured responses of the pressure, temperature, and vibration sensors, where **(A)** and **(B)** are the Bragg wavelength as a function of pressure and temperature change, **C** shows the wavelength shift with respect to the applied acceleration of the vibration, and **(D)** plots the frequency response of the vibration sensor.

The sensing FBG is packaged and glued inside a cylindrical tube of stainless steel. By proper alignment, the FBG is absence of external strain during measurement, so that only the temperature induces a corresponding wavelength shift. As shown in **Figure 5B**, the calibrated temperature sensitivity is 0.08 nm/°C. Given that the resolution of the interrogation module is ~0.5 pm, the measurement resolution of the pressure and temperature can be ~0.25 kPa (corresponding to a depth variation of ~0.024 m) and 0.006°C, respectively. The relationship between the Bragg wavelength change and the measurands shows a good linearity, and the fitting R^2 exceeds 0.999.

Regarding the vibration sensor, its sensing performance was investigated by applying various accelerations to the sensor in a fixed vibration direction. As schematically illustrated in the inset of **Figure 5C**, two FBGs are adhered on the top and bottom of the cantilever, which experience compression and tension strain, respectively, under a certain degree of vibration. The difference in the Bragg wavelengths of these two FBGs is utilized to characterize and calibrate the acceleration. After packaging the FBGs, the sensor was placed on a vibration shaker to induce the desired acceleration to the sensor. The actual acceleration was calibrated by an electric-based accelerometer. The shift between these two Bragg wavelengths vs. acceleration is plotted in **Figure 5C**, where the sensitivity is calculated to be 0.142 nm/g (g = 9.8 m/s^2) and measured acceleration is up to 25 g. The FBG of the vibration sensor is packaged on a cantilever with an inertial mass at the top to induce periodical strain to the grating together with the vibration. The vibration at various frequencies was measured and the corresponding sensitivity was characterized, which is shown in **Figure 5D**. It can be observed that the response could remain nearly constant (within 5% variation) in a wide frequency range, and the sensor shows the sensing capability in small frequencies less than 10 Hz.

Real-Time Monitoring Results of the Depth, Acceleration and Temperature During the Sea Test

The proposed submarine optical fiber sensing system integrated with the pressure, vibration, and temperature sensors was tested and verified under the sea, as demonstrated in **Figure 4**. The entire system was laid down to three depths, i.e., ~5, ~10, and ~15 m under the sea level. The actual depth was also measured by the commercial CTD device. At each stage, the system was hanged on for over 1 h and then moved to the next level. The monitored results are shown in **Figure 6**, where the depth and temperature measured by the commercial CTD are plotted as well for comparison. From the results, it can be observed that the FBG-based sensors are able to measure the depth and temperature accurately, and small variations compared to the commercial CTD device are found. As for the acceleration, the sudden spikes clearly indicate the movement of the system to the next stage. Since there is no electrical vibration sensor installed on the commercial CTD device, no comparison is given. However, the moving posture of the sensing system is possible to be retrieved by the two acceleration sensors placed orthogonally, which managed to capture the slight vibration caused by the movement of the positioning or the undersea water. In this particular test lasting for a time duration of ~5 h, the first depth (~5.4 m) was reached at 13:43 on the testing day, and the monitored temperature by the FBG sensor dropped to 28.9°C. At this depth level, the temperature showed some minor fluctuations due to the non-uniform temperature distribution

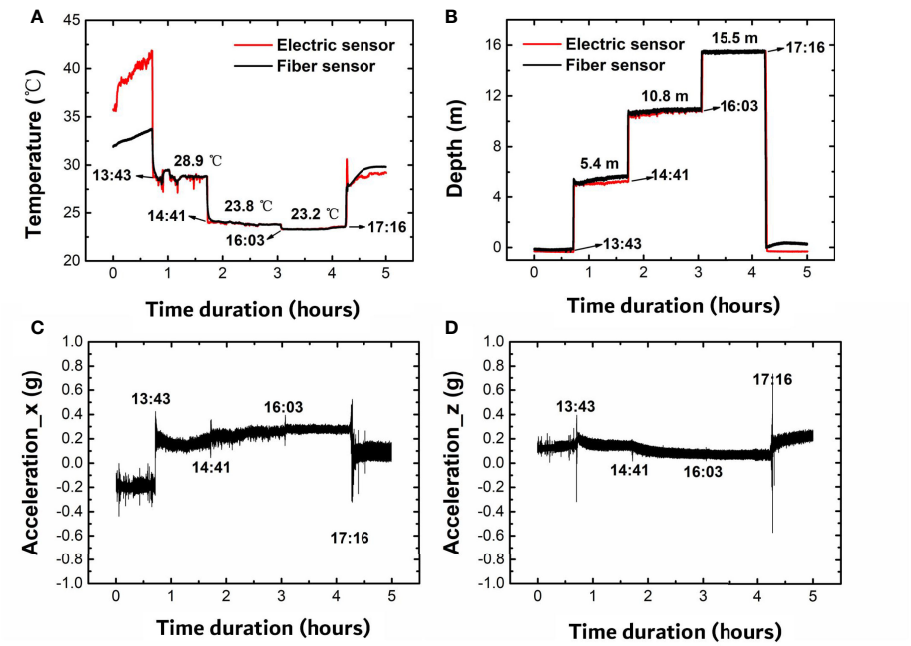


FIGURE 6 | Measured results of the **(A)** temperature, **(B)** depth, and **(C, D)** acceleration by the submarine optical fiber sensing system under the sea.

of the water. At 14:41, the second depth (~10.8 m) was reached and the temperature further dropped to ~23.8°C. As the depth was increased to ~15.5 m, occurring at 16:03, the reduction of the temperature was small, staying at about 23.2°C. The temperature distribution decreasing with the depth is similar to the general surface layer of the sea as reported previously (Duraibabu et al., 2017).

Dynamics of the Surface Water Wave

During the sea test, the submarine optical fiber sensing system was also maintained at a certain depth (~15.4 m) for over 12 h to test the long stability and to retrieve the dynamics of the surface water wave according to the monitored pressure. **Figure 7** plots the monitored depth, temperature, and acceleration in a time duration over 12 h, starting at 18:09 of the testing day and ending at 05:50 the next day. The test was conducted overnight, during which the temperature basically decreases gradually with time. This is confirmed from the measured temperature as illustrated in **Figure 7C**, which shows the sea water temperature at the depth of ~15.4 m under the sea drops from ~25°C to the lowest point of ~22.8°C. The measured acceleration during this time duration keeps still, since the entire system was maintained at the same depth, except the starting and ending moments, where the vibration was caused during the movement of sinking and taking out. By taking a closer view of the depth change with time shown in the inset of **Figure 7A**, a periodic depth change can be observed clearly, which is supposed to be the result of the surface water wave. The fast Fourier transform of the depth change over time is plotted in **Figure 7B**, showing a prominent peak at the frequency of ~0.25 Hz with a signal-to-noise ratio of >20 dB. This frequency component falls into the range of ocean surface waves observed by the GPS-tracked buoys (Herbers et al., 2012). Thus, it is possible to employ the proposed submarine optical fiber sensing system to realize the real-time investigation of the surface water wave.

DISCUSSION

To realize environmental monitoring in real time, various techniques have been developed and employed, including the seabed cabled observatories and distributed optical fiber sensing systems. Those systems are typically established in a large-scale area, and high cost is required. In those monitoring systems, the multiparameter sensors with high performance are in high demand. As demonstrated in this work, a submarine optical fiber sensing interrogation system is proposed and various optical fiber sensors could be integrated with the system. Since the optical fiber sensors are pluggable, the proposed system is flexible to conduct different measurements of multiple parameters. This feature allows it to monitor the sophisticated undersea environment with the ease of replacing the desired sensors.

To further investigate the repeatability of the sensor performance, new batches of the pressure, temperature, and acceleration sensors were fabricated in a similar approach as the sensors presented above have been installed in the system and being under monitoring of the field test. The sensing responses were measured three times, as displayed in **Figure 8**, where the error bar presents the standard deviation of the repeatable measurements. The averaged wavelength deviation

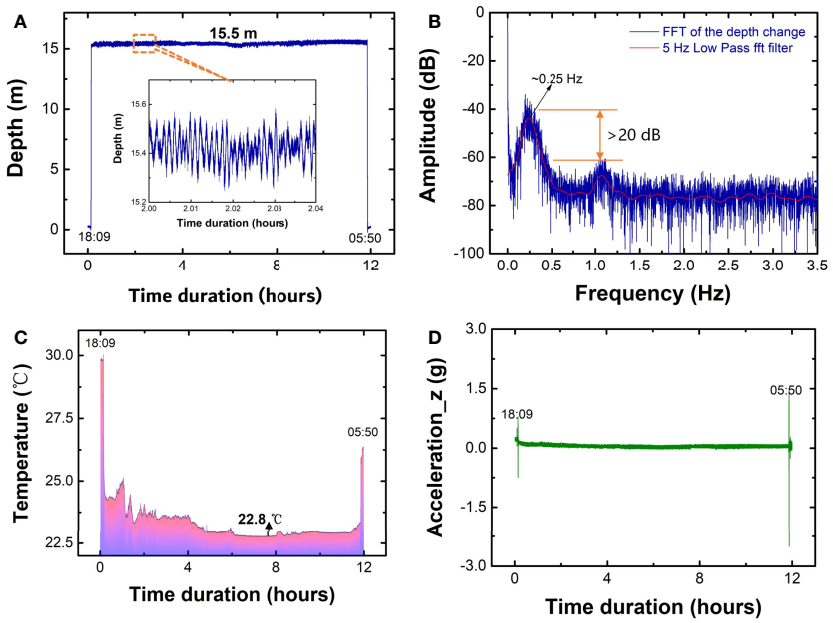


FIGURE 7 | Measured results of the depth, temperature, and acceleration in a time duration over 12 h under the sea. **(A**, **C**, and **D)** are the depth, temperature, and acceleration monitored in real time, and **(B)** is the fast Fourier transform of the depth change in the time domain.

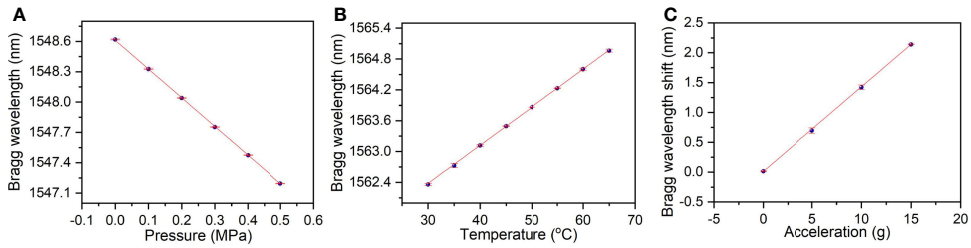


FIGURE 8 | Measured responses of the **(A)** pressure, **(B)** temperature, and **(C)** acceleration sensors fabricated in the same way, where the error bar represents the standard deviation of the repeatable measurements.

of the pressure, temperature, and acceleration measurement is ~0.004, ~0.01, and ~0.02 nm, respectively, corresponding to a measurement accuracy of ~2 kPa, ~0.1°C, and ~0.14 g. It is worth noting that the measurement accuracy could be enhanced further by establishing more precise calibration and demodulating the FBG spectra using deep learning method (Cao et al., 2022), which we are currently working on.

As a proof test under the sea, preliminary results show good verification of the sensor system, and the pressure sensor is able to depict the curve of the depth during sinking while the temperature is monitored precisely at various positions. With regard to the vibration, a highly sensitive accelerometer is capable of recording the abrupt perturbation during the test. It is worth indicating that the resolution and limit of detection (LOD) of the sensors are not the best optimized during the verification test. Proper improvement in terms of the sensor structure, especially the novel packages with enhanced sensitivity, could be further investigated. With this regard, the simultaneous measurement of temperature, depth, and salinity is possible using a single optical fiber sensor. Prior examples can be found by the interferometric configurations, such as the FBG integrated with Fabry–Pérot interferometer (Liu et al., 2021), Fabry–Pérot interferometer with inline microcavity (Flores et al., 2019), Mach–Zehnder interferometer with core diameter mismatch structure (Selokar and Giraldi, 2021), Fabry–Pérot interferometer with U-shaped defect (Zhao et al., 2022b), and no-core fiber structure (Zhao et al., 2022a). All of these sensing structures show good sensitivity and discrimination of the multiple parameters based on a single structure. However, to realize the *in situ* monitoring under the sea, the stability and accuracy of the sensors need further investigation together with the submarine sensing system, as demonstrated in this work. With the integration of the interferometry-based fiber sensors, better algorithms such as deep learning method could be employed in the microprocessor of the proposed submarine interrogation system illustrated in **Figure 2**, which eventually retrieves the sensing signal in a smart approach. In a recent study (Cao et al., 2022), we have demonstrated that the spectrum of the FBG-based sensors can be directly demodulated by the deep convolutional neural network (DCNN) model, and good accuracy can be maintained; even the resolution of the hardware is reduced. This approach allows for the development of a compact and low-cost submarine sensing system, which is under investigation.

CONCLUSION

In conclusion, to achieve real-time monitoring of the environmental parameters under the sea, a submarine optical fiber sensing system is proposed. The principle of FBG-based sensors and the sensing system have been described. Basically, the proposed interrogation system is feasible to various sensors. As a demonstration of the sensing capability, the FBG-based pressure, acceleration, and temperature sensors are designed and utilized, respectively, for the measurement of the depth, vibration, and temperature. In the experiment, the achieved sensitivities of the pressure, temperature, and vibration sensors are -1.993 nm/MPa, 0.08 nm/°C, and 0.139 nm/g (g = 9.8 m/s^2), meaning that the respective resolution can reach ~0.25 kPa, ~0.006°C, and ~0.004 g if taking the interrogation resolution of ~0.5 pm into account. A prototype of the submarine sensing system integrated with the FBG-based sensors was designed and manufactured. The proof test under the sea based on the prototype shows good capability of real-time monitoring of multiple parameters under the sea. The measured results during the sea test are comparable to the commercial CTD device installed at the same position, which exhibits the feasibility and accuracy of the proposed submarine sensing system. From the results of the sea test, the monitored depth change demonstrated a dominant frequency of ~0.25 Hz close to that of the surface water wave, which is promising for the long-term investigation of the dynamics of the ocean surface wave. The submarine optical fiber sensing system is able to integrate with various fiber sensors, and novel demodulation algorithms could be employed to realize the simultaneous measurement of multiple parameters with higher accuracy.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors without undue reservation.

AUTHOR CONTRIBUTIONS

ZYL conceived the idea of the work and wrote the article. SZ and CY conducted the experimental tests. ZYL and W-HC designed and manufactured the hardware of the optical fiber sensing system. ZYL and ZHL organized the sea test and supervised the project. All of the authors discussed the results and revised the article.

FUNDING

The authors would like to acknowledge the financial support from the National Natural Science Foundation of China under Grant Nos. 61905096, 61975250, and U2001601, also from the Science and Technology Planning Project of Guangzhou (202102020868), the Program of Marine Economy Development Special Fund (Six Marine Industries) under the Department of Natural Resources of Guangdong Province (Project No. GDNRC [2021] 33), and the Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (No. 311021011).

ACKNOWLEDGMENTS

The authors would like to acknowledge Prof. Bo Liu from Nankai University and Prof. Xinwei Wang from the Institute of Semiconductors, Chinese Academy of Sciences, for the assistance during the sea test.

REFERENCES

Ajo-Franklin, J. B., Dou, S., Lindsey, N. J., Monga, I., Tracy, C., Robertson, M., et al. (2019). Distributed Acoustic Sensing Using Dark Fiber for Near-Surface Characterization and Broadband Seismic Event Detection. *Sci. Rep.* 9, 1–14. doi: 10.1038/s41598-018-36675-8

Cao, Z., Zhang, S., Xia, T., Liu, Z., and Li, Z. (2022). Spectral Demodulation of Fiber Bragg Grating Sensor Based on Deep Convolutional Neural Networks. *J. Light. Technol.* 40, 1–1. doi: 10.1109/JLT.2022.3155253

Cheng, W., Wang, L., and Hsieh, C. (2003). Phase Masks Fabricated by Interferometric Lithography for Working in 248 Nm Wavelength. *Microelectron. Eng.* 67–68, 63–69. doi: 10.1016/S0167-9317(03)00060-1

Duraibabu, D. B., Leen, G., Toal, D., Newe, T., Lewis, E., and Dooly, G. (2017). Underwater Depth and Temperature Sensing Based on Fiber Optic Technology for Marine and Fresh Water Applications. *Sens. (Switzerland).* 17, 7–10. doi: 10.3390/s17061228

Erdogan, T. (1997). Fiber Grating Spectra. *J. Light. Technol.* 15, 1277–1294. doi: 10.1109/50.618322

Favali, P., and Beranzoli, L. (2006). Seafloor Observatory Science: A Review. *Ann. Geophys.* 49, 515–567. doi: 10.4401/ag-3125

Flores, R., Janeiro, R., and Viegas, J. (2019). Optical Fibre Fabry-Pérot Interferometer Based on Inline Microcavities for Salinity and Temperature Sensing. *Sci. Rep.* 9, 9556. doi: 10.1038/s41598-019-45909-2

Gartman, A., and Findlay, A. J. (2020). Impacts of Hydrothermal Plume Processes on Oceanic Metal Cycles and Transport. *Nat. Geosci.* 13, 396–402. doi: 10.1038/s41561-020-0579-0

Herbers, T. H. C., Jessen, P. F., Janssen, T. T., Colbert, D. B., and MacMahan, J. H. (2012). Observing Ocean Surface Waves With GPS-Tracked Buoys. *J. Atmos. Ocean. Technol.* 29, 944–959. doi: 10.1175/JTECH-D-11-00128.1

Hill, K. O., and Meltz, G. (1997). Fiber Bragg Grating Technology Fundamentals and Overview. *J. Light. Technol.* 15, 1263–1276. doi: 10.1109/50.618320

Howe, B. M., Angove, M., Aucan, J., Barnes, C. R., Barros, J. S., Bayliff, N., et al. (2022). SMART Subsea Cables for Observing the Earth and Ocean, Mitigating Environmental Hazards, and Supporting the Blue Economy. *Front. Earth Sci.* 9. doi: 10.3389/feart.2021.775544

Howe, B. M., Arbic, B. K., Aucan, J., Barnes, C. R., Bayliff, N., Becker, N., et al. (2019). SMART Cables for Observing the Global Ocean: Science and Implementation. *Front. Mar. Sci.* 6. doi: 10.3389/fmars.2019.00424

Jousset, P., Reinsch, T., Ryberg, T., Blanck, H., Clarke, A., Aghayev, R., et al. (2018). Dynamic Strain Determination Using Fibre-Optic Cables Allows Imaging of Seismological and Structural Features. *Nat. Commun.* 9, 2509. doi: 10.1038/s41467-018-04860-y

Leal-Junior, A. G., Díaz, C. R., Marques, C., Pontes, M. J., and Frizera, A. (2019). Multiplexing Technique for Quasi-Distributed Sensors Arrays in Polymer Optical Fiber Intensity Variation-Based Sensors. *Opt. Laser. Technol.* 111, 81–88. doi: 10.1016/j.optlastec.2018.09.044

Leal-Junior, A. G., Frizera, A., Marques, C., Sanchez, M. R. A., dos Santos, W. M., Siqueira, A. A. G., et al. (2018). Polymer Optical Fiber for Angle and Torque Measurements of a Series Elastic Actuator's Spring. *J. Light. Technol.* 36, 1698–1705. doi: 10.1109/JLT.2017.2789192

Leitão, C., Leal-Junior, A., Almeida, A. R., Pereira, S. O., Costa, F. M., Pinto, J. L., et al. (2021). Cortisol AuPd Plasmonic Unclad POF Biosensor. *Biotechnol. Rep.* 29, e00587. doi: 10.1016/j.btre.2021.e00587

Liu, Z., Htein, L., Lee, K.-K., Lau, K.-T., and Tam, H.-Y. (2018). Large Dynamic Range Pressure Sensor Based on Two Semicircle-Holes Microstructured Fiber. *Sci. Rep.* 8, 65. doi: 10.1038/s41598-017-18245-6

Liu, Y., Jing, Z., Liu, Q., Li, A., Lee, A., Cheung, Y., et al. (2021). All-Silica Fiber-Optic Temperature-Depth-Salinity Sensor Based on Cascaded EFPIs and FBG for Deep Sea Exploration. *Opt. Exp.* 29, 23953–23966. doi: 10.1364/OE.432943

Liu, S.-Y., Tam, H.-Y., Lee, K.-K., and Cheng, K.-C. (2017). Optical Fibre Networks Facilitate Shift to Predictive Maintenance. *Int. Railw. J.* 57, 38–41.

Marques, C. A. F., Bilro, L. B., Alberto, N. J., Webb, D. J., and Nogueira, R. N. (2013). Narrow Bandwidth Bragg Gratings Imprinted in Polymer Optical Fibers for Different Spectral Windows. *Opt. Commun.* 307, 57–61. doi: 10.1016/j.optcom.2013.05.059

Marques, C. A. F., Min, R., Junior, A. L., Antunes, P., Fasano, A., Woyessa, G., et al. (2018). Fast and Stable Gratings Inscription in POFs Made of Different Materials With Pulsed 248 Nm KrF Laser. *Opt. Exp.* 26, 2013–2022. doi: 10.1364/oe.26.002013

Marra, G., Clivati, C., Luckett, R., Tampellini, A., Kronjäger, J., Wright, L., et al. (2018). Ultrastable Laser Interferometry for Earthquake Detection With Terrestrial and Submarine Cables. *Science (80-.).* 361, 486–490. doi: 10.1126/science.aat4458

Min, R., Liu, Z., Pereira, L., Yang, C., Sui, Q., and Marques, C. (2021). Optical Fiber Sensing for Marine Environment and Marine Structural Health Monitoring: A Review. *Opt. Laser. Technol.* 140, 107082. doi: 10.1016/j.optlastec.2021.107082

Najafzadeh, A., Serandi Gunawardena, D., Liu, Z., Tran, T., Tam, H., Fu, J., et al. (2020). Application of Fibre Bragg Grating Sensors in Strain Monitoring and Fracture Recovery of Human Femur Bone. *Bioengineering* 7, 98. doi: 10.3390/bioengineering7030098

Pospori, A., Marques, C. A. F., Bang, O., Webb, D. J., and André, P. (2017). Polymer Optical Fiber Bragg Grating Inscription With a Single UV Laser Pulse. *Opt. Exp.* 25, 9028–9038. doi: 10.1364/OE.25.009028

Qiao, X., Shao, Z., Bao, W., and Rong, Q. (2017). Fiber Bragg Grating Sensors for the Oil Industry. *Sensors* 17, 429. doi: 10.3390/s17030429

Selokar, T., and Giraldi, M. T. R. (2021). All-Fiber Sensors for Salinity and Temperature Simultaneous Measurement. *Opt. Quant. Electron.* 53, 1–17. doi: 10.1007/s11082-020-02678-x

Walter, F., Gräff, D., Lindner, F., Paitz, P., Köpfli, M., Chmiel, M., et al. (2020). Distributed Acoustic Sensing of Microseismic Sources and Wave Propagation in Glaciated Terrain. *Nat. Commun.* 11, 1–10. doi: 10.1038/s41467-020-15824-6

Wang, L., Wang, Y., Wang, J., and Li, F. (2020). A High Spatial Resolution FBG Sensor Array for Measuring Ocean Temperature and Depth. *Photo. Sens.* 10, 57–66. doi: 10.1007/s13320-019-0550-0

Williams, E. F., Fernández-Ruiz, M. R., Magalhaes, R., Vanthillo, R., Zhan, Z., González-Herráez, M., et al. (2019). Distributed Sensing of Microseisms and Teleseisms With Submarine Dark Fibers. *Nat. Commun.* 10, 5778. doi: 10.1038/s41467-019-13262-7

Wu, P., Tan, D., Chen, W., Malik, N., and Yin, J. (2021). Novel Fiber Bragg Grating-Based Strain Gauges for Monitoring Dynamic Responses of Celtis Sinensis Under Typhoon Conditions. *Measurement* 172, 108966. doi: 10.1016/j.measurement.2021.108966

Xu, M., Reekie, L., Chow, Y., and Dakin, J. P. (1993). Optical in-Fibre Grating High Pressure Sensor. *Electron. Lett.* 29, 398–399. doi: 10.1049/el:19930267

Yu, K. (2015). Tsunami-Wave Parameter Estimation Using GNSS-Based Sea Surface Height Measurement. *IEEE Trans. Geosci. Remote Sens.* 53, 2603–2611. doi: 10.1109/TGRS.2014.2362113

Zhan, Z., Cantono, M., Kamalov, V., Mecozzi, A., Müller, R., Yin, S., et al. (2021). Optical Polarization–Based Seismic and Water Wave Sensing on Transoceanic Cables. *Science (80-.)* 371, 931–936. doi: 10.1126/science.abe6648

Zhao, Y., Wu, Q., and Zhang, Y. (2019). Simultaneous Measurement of Salinity, Temperature and Pressure in Seawater Using Optical Fiber SPR Sensor. *Measurement* 148, 106792. doi: 10.1016/j.measurement.2019.07.020

Zhao, J., Zhao, Y., and Cai, L. (2022a). Hybrid Fiber-Optic Sensor for Seawater Temperature and Salinity Simultaneous Measurements. *J. Light. Technol.* 40, 880–886. doi: 10.1109/JLT.2021.3121674

Zhao, Y., Zhao, J., Wang, X., Peng, Y., and Hu, X. (2022b). Femtosecond Laser-Inscribed Fiber-Optic Sensor for Seawater Salinity and Temperature Measurements. *Sens. Actuat. B. Chem.* 353, 131134. doi: 10.1016/j.snb.2021.131134

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Liu, Zhang, Yang, Chung and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 07 September 2022
DOI 10.3389/fmars.2022.1003568

Check for updates

OPEN ACCESS

EDITED BY
Rizwan Ali Naqvi,
Sejong University, South Korea

REVIEWED BY
Eui Chul Lee,
Sangmyung University, South Korea
E. Kim,
Yonsei University, South Korea

*CORRESPONDENCE
Kang Ryoung Park
parkgr@dongguk.edu

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 26 July 2022
ACCEPTED 15 August 2022
PUBLISHED 07 September 2022

CITATION
Kim YH and Park KR (2022) PSS-net: Parallel semantic segmentation network for detecting marine animals in underwater scene.
Front. Mar. Sci. 9:1003568.
doi: 10.3389/fmars.2022.1003568

COPYRIGHT
© 2022 Kim and Park. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

PSS-net: Parallel semantic segmentation network for detecting marine animals in underwater scene

Yu Hwan Kim and Kang Ryoung Park*

Division of Electronics and Electrical Engineering, Dongguk University, Seoul, South Korea

Marine scene segmentation is a core technology in marine biology and autonomous underwater vehicle research. However, it is challenging from the perspective of having a different environment from that of the conventional traffic segmentation on roads. There are two major challenges. The first is the difficulty of searching for objects under seawater caused by the relatively low-light environment. The second problem is segmenting marine animals with protective colors. To solve such challenges, in previous research, a method of simultaneously segmenting the foreground and the background was proposed based on a simple modification of the conventional model; however, it has limitations in improving the segmentation accuracy. Therefore, we propose a parallel semantic segmentation network to solve the above issues in which a model and a loss are employed to locate the foreground and the background separately. The training task to locate the foreground and the background is reinforced in the proposed method by adding an attention technique in a parallel model. Furthermore, the final segmentation is performed by aggregating two feature maps obtained by separately locating the foreground and the background.The test results using an open dataset for marine animal segmentation reveal that the proposed method achieves performance of 87%, 97.3%, 88%, 95.2%, and 0.029 in the mean intersection of the union, structure similarities, weighted F-measure, enhanced-alignment measure, and mean absolute error, respectively. These findings confirm that the proposed method has higher accuracy than the state-of-the-art methods. The proposed model and code are publicly available *via* Github[1].

KEYWORDS

detecting marine animal, underwater scene, protective colors, PSS-net, attention technique

1 https://github.com/Kimyuhwanpeter/marine_segmentation

1. Introduction

In studies on marine life using autonomous underwater vehicles, its accurate detection is essential (Xu et al., 2021; Kandimalla et al., 2022). However, detecting marine life has more challenging issues than detecting terrestrial animals caused by the difficultly in observing objects under sea because of lesser light underwater than that on land. Furthermore, many marine animals have protective colors (Ditria et al., 2020; Panaïotis et al., 2022). To solve these problems, previous studies have been actively conducted on convolutional neural network (CNN)-based object detection (Pedersen et al., 2019; Li et al., 2021; Berg et al., 2022; Chen et al., 2022). Object detection can be classified into two types: box-based and pixel-based object detection (where latter is semantic segmentation). Box-based detection is very efficient in detecting the position of a box containing an object. However, there is a limitation in that it cannot detect exact regions of the object. Pixel-based object detection, unlike box-based object detection, can detect exact regions of an object in pixel units. Therefore, this study focused on the semantic segmentation of marine animals.

Previous studies have concentrated on CNN model modifications for the same purpose (Islam et al., 2020; Li et al., 2021; Zhang et al., 2021; Chen et al., 2022). A representative CNN model modification is to apply an attention mechanism (Zhang et al., 2021; Chen et al., 2022).

For example, the WaterSNet (Chen et al., 2022) used an attention fusion block (AFB) module, which employs an attention mechanism. The AFB module is adopted to utilize global context information. In addition, the WaterSNet was designed to extract multiscale features using a receptive field block (RFB) module. Moreover, Koch et al. (2015) reduced the impact of water degradation diversity by basically designing a model with a Siamese structure (Koch et al., 2015) and simultaneously using a random style adaption (RSA) module. Finally, they improved the performance of marine animal segmentation by combining RSA, RFB, and AFB modules. Next, Li et al. (2021) designed the ECD-Net by combining various modules, similar to WaterSNet. Specifically, first, an interactive feature enhancement module (IFEM) was proposed to consider the relationships between the features in the encoder. Furthermore, cascade decoder modules (CDMs), which integrate the features of the cross-layer, were developed to improve the performance of marine animal segmentation (Li et al., 2021). Based on the above, the combination of various modules in the WaterSNet and the ECD-Net is effective in improving the model performance.

A dual pooling-aggregated attention network called the DPANet was proposed (Zhang et al., 2021). A residual network (ResNet) is used as the backbone of a DPANet model (He et al., 2016). The feature maps output by this backbone model are input into a pooling-aggregated position attention module (PPAM) as well as a pooling-aggregated channel attention module (PCAM). The outputs of these modules are finally combined for marine animal segmentation. The existing position attention modules (Huang et al., 2019; Fu et al., 2020) do not consider fine-grained contextual detail information. However, the PPAM overcomes this drawback and generates spatial attention maps more efficiently than the previous researches. The existing channel attention modules (Fu et al., 2020) have limitations in various feature representations because they do not focus on the fine-grained inter-channel relationship. However, the PCAM improves various feature representation functions using the interdependent relationships between the channel maps in a computationally friendly manner (Zhang et al., 2021). Another research proposed the SUIM-Net (Islam et al., 2020) and introduced two versions: SUIM-Net$_{RSB}$ and SUIM-Net$_{VGG}$. First, the SUIM-Net$_{RSB}$, in which a residual skip block (RSB) is added to the decoder, is the first model designed to utilize the benefits of skip connections (Ronneberger et al., 2015) and residual learning (He et al., 2016). Furthermore, this design is aimed at ensuring real-time processing while accomplishing good segmentation performance (Islam et al., 2020). In comparison, the SUIM-Net$_{VGG}$ model was designed only to improve the segmentation performance using the visual geometry group (VGG) (Simonyan and Zisserman, 2015). Finally, the SUIM-Net$_{VGG}$ model was chosen as the final proposed model in this study to focus more on the segmentation performance.

In addition, the effects of various image quality enhancement techniques for underwater change detection on the segmentation algorithm have been investigated previously (Radolko et al., 2016; Radolko et al., 2017). Although research on CNN-based segmentation has not been conducted, experiments related to various existing segmentation algorithms have been reported (KaewTraKulPong and Bowden, 2001; Zivkovic, 2004; Zivkovic and Heijden, 2006; Radolko and Gutzeit, 2015).

The above previous studies generally focused on improving marine segmentation performance using conventional models. Specifically, they improved the segmentation performance by extracting features of the foreground and the background simultaneously. However, this method has the disadvantage of not ensuring the segmentation performance according to the ratio of the background and the foreground. Considering this aspect, the parallel semantic segmentation network (PSS-Net) proposed in this study not only extracts the background and foreground features separately but also improves the segmentation performance for the background and the foreground by designing the corresponding loss functions. Furthermore, two segmentation maps are created during segmentation by dividing an image into the foreground and the background. The PSS-Net was designed to fuse two feature maps to obtain the final segmentation result. Section 2 describes the PSS-Net in detail. In this study, the MAS3K dataset (Li et al., 2021) is used for the experiments because the object

segmentation problem is more challenging than with other datasets (SUIM (Islam et al., 2020) and underwater change detection (Radolko et al., 2016). Moreover, the object segmentation problem is more challenging with the MAS3K dataset because it has more diverse entities of marine animals. The contributions of this study are as follows:

-This study proposed the PSS-Net, which is expected to achieve high performance in detecting marine animals by separately learning their foreground and background regions.

-For the PSS-Net proposed in this study, two models are designed in parallel. Moreover, an attention mechanism and separate losses (object and background losses) are established between the two models to extract the features of the foreground and background regions more efficiently.

-The PSS-Net used feature fusion segmentation to fuse the foreground and background feature maps extracted from the two models to improve the segmentation accuracy. The proposed PSS-Net and algorithm are publicly available *via* Github[1].

The remainder of this paper is organized as follows. Section 2 describes the proposed method, and Section 3 presents the experimental results. Section 4 discusses the results, and Section 5 draws the conclusions.

2. Materials and methods

2.1. Model architecture of PSS-net

This section describes the PSS-Net proposed in this study, and its overall structure is shown in Figure 1.

As can be seen in Figure 1, the U-net (Ronneberger et al., 2015) is used as the backbone and VGG-16 (Simonyan and Zisserman, 2015) as the encoder of our proposed PSS-Net. The existing conventional models for semantic segmentation (Islam et al., 2020; Li et al., 2021; Zhang et al., 2021; Chen et al., 2022) train both the background and foreground from the feature map extracted from one model. However, the PSS-Net proposed in this study extracts the background and foreground feature maps from different models respectively, and trains the background and foreground based on them. Moreover, the layers between the models focus on the background and foreground components, helping to predict the final background and foreground regions.

2.2. Loss for PSS-net

Previously, a loss considering both the background and foreground segmentation was used. However, for the loss of the PSS-Net proposed in this study, background and foreground losses were designed to correspond to the structure of the model described in Section 2.1. This was because the class ratio of the foreground and the background is imbalanced, and overfitting may occur. Previously, the focal loss (Lin et al., 2017), as expressed in Equation (1), was used to solve this problem.

$$FOCAL_{loss} = \begin{cases} -\frac{1}{MN}\sum_{i=0}^{M}\sum_{j=0}^{N}\alpha \cdot \left(1-\frac{1}{1+e^{y_{ij}}}\right)\cdot \ln\left(\frac{1}{1+e^{y_{ij}}}\right) & (class=1) \\ -\frac{1}{MN}\sum_{i=0}^{M}\sum_{j=0}^{N}(1-\alpha)\cdot \left(\frac{1}{1+e^{y_{ij}}}\right)\cdot \ln\left(1-\frac{1}{1+e^{y_{ij}}}\right) & (class=0) \end{cases} \quad (1)$$

where M and N denote the height and width of the final feature map, respectively, and i and j denote the pixel coordinates, respectively. The pixel value of the feature map is expressed as y. Equation (1) is commonly derived from the final feature map extracted from one model. However, the focal loss in Equation (1) is primarily focused on solving the class imbalance problem, and it is ineffective in extracting the

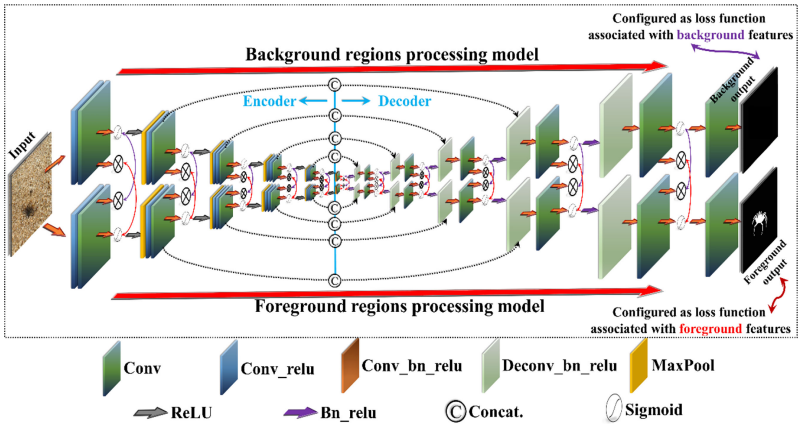


FIGURE 1
Structure of proposed PSS-Net.

features of the object boundary region. Therefore, in this study, we designed a loss where the background and foreground training tasks can be performed independently of each other. First, the loss related to the foreground consists of a weighted binary cross entropy loss and a pixel perception loss (Pang et al., 2020). Equation (2) expresses the weighted binary cross entropy loss, and it was used because it was deemed effective in predicting the boundary features of object regions.

$$L_{wc} = \frac{\sum_{i,j=0}^{M,N}(1+5\alpha_{i,j}) \cdot \sum_{x,y=0}^{M,N}(g_{x,y} \cdot \ln(p_{x,y}) + (1-g_{x,y}) \cdot \ln(1-p_{x,y}))}{\sum_{i,j=0}^{M,N} 5 \cdot \alpha_{i,j}} \quad (2)$$

where $g_{x,y}$ and $p_{x,y}$ denote the probability values that pass through the sigmoid function for the ground truth image and the predicted feature map, respectively. $\alpha_{i,j}$ is calculated using the surrounding pixels (the pixel values for a specific region are used during the calculation). In addition, a weighted intersection of union (wIoU)-based pixel perception loss was used to predict object regions excluding boundaries, and it is expressed in Equation (3).

$$L_{wi} = 1 - \frac{\sum_{x=0}^{M}\sum_{y=0}^{N}(g_{x,y} \cdot p_{x,y}) \cdot (1+5 \cdot \alpha_{x,y})}{\sum_{x=0}^{M}\sum_{y=0}^{N}((g_{x,y} + p_{x,y} - g_{x,y} \cdot p_{x,y})(1+5 \cdot \alpha_{x,y}))} \quad (3)$$

Equations (2) and (3) were used as the object (foreground) loss, as expressed in Equation (4).

$$L_{object} = L_{wc} + L_{wi} \quad (4)$$

The dice loss was used as the loss for the background regions, as expressed in Equation (5).

$$L_{dice} = 1 - \frac{2\sum_{x=0}^{M}\sum_{y=0}^{N}(g_{x,y} \cdot p_{x,y})}{\sum_{x=0}^{M}\sum_{y=0}^{N}(g_{x,y} + p_{x,y})} \quad (5)$$

Based on Equation (5), the original dice loss only considers the case in which the ground truth label is 1 (the ground truth pixels for the object and background are 1 and 0, respectively). However, Equation (6), which is a modified version of Equation (5), was used in this study to consider the dice for the background (when the ground truth pixel is zero).

$$L_{background} = 1 - \frac{2 \cdot \sum_{x=0}^{M}\sum_{y=0}^{N}((1-g_{x,y}) \cdot (1-p_{x,y}))}{\sum_{x=0}^{M}\sum_{y=0}^{N}((1-g_{x,y}) + (1-p_{x,y}))} \quad (6)$$

Equation (7) expresses the loss used in this study using Equations (4) and (6).

$$L_{total} = L_{object} + \beta \cdot L_{background} \quad (7)$$

β is the ratio of the foreground and background components of an input image. It is calculated by accumulating the number of the foreground and background pixels of the input image separately and converting each accumulated number of pixels to a number less than 1 using the softmax function. We refer to the method used in a previous study (Kim and Park, 2022) for this calculation. However, in some cases, the input image may have no background or foreground component. Therefore, in this study, the loss was finally designed as expressed in Equation (8), to account for these cases.

$$L_{total} = \begin{cases} L_{object} + \beta \cdot L_{background} (if\ O_{ratio} = True\ and\ B_{ratio} = True) \\ \beta \cdot L_{background}\ (if\ O_{ratio} = False\ and\ B_{ratio} = True) \\ L_{object}\ (if\ O_{ratio} = True\ and\ B_{ratio} = False) \end{cases} \quad (8)$$

where O_{ratio} and B_{ratio} denote the frequencies of the background and foreground pixels in the input image. In Equation (8), if O_{ratio} and B_{ratio} are each zero, *False* label is assigned to them. On the other hand, if O_{ratio} and B_{ratio} are nonzero, respectively, *True* label is assigned to them. Finally, the effect of the attention between the models, as described in Section 2.1, is maximized owing to these background and foreground losses, which aids in improving the segmentation performance. The causes for this performance improvement are as follows. In this study, the ground truth pixel for the background was set as zero, and the ground truth pixel for the foreground was set as one. Consequently, the $p_{x,y}$ related to the background trains close to zero, and the $p_{x,y}$ of the foreground trains close to one. Therefore, the features that provide attention to the background and foreground region processing focus more on the background and foreground regions.

2.3. Feature fusion-based segmentation of PSS-net

Figure 2 shows the feature fusion-based segmentation method of the PSS-Net proposed in this study. Moreover, to explain the feature fusion-based segmentation process, the outputs before predicting zero (background) and one (foreground) are called the background final feature map and the foreground final feature map, respectively, in this study. In addition, the output after predictions is called the foreground final segmentation map.

As shown in Figure 2, pixel multiplication-based fusion is performed on the respective final feature maps (background and foreground final feature maps) obtained from the model that deals the background region processing and the model that deals with the foreground regions processing. This process is the same as the background final feature map-based attention operation for the foreground final feature map. Consequently, the final segmentation results are obtained by predicting the foreground and the background. The reason for fusing these two feature maps is to improve the segmentation results of the objects.

3. Results

3.1. Experimental dataset

In this study, experiments were conducted using the MAS3K dataset (Li et al., 2021), which is an open database. This dataset

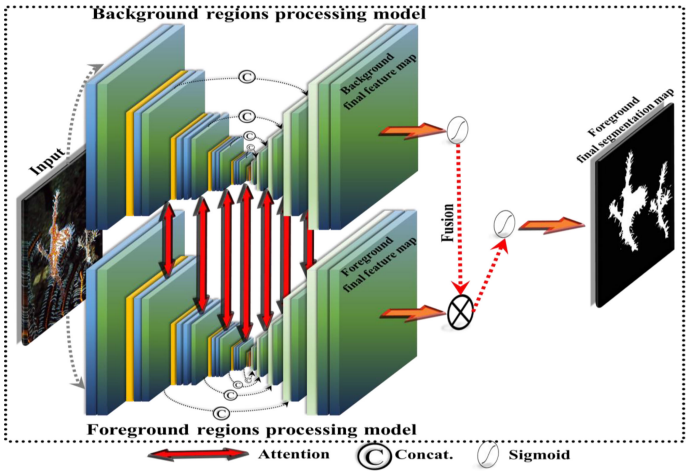


FIGURE 2
Feature fusion-based segmentation of PSS-Net.

consists of 1,588 camouflaged, 1,322 common, and 193 background images. We divided the MAS3K dataset into two cases in this study and conducted the experiments. Table 1 briefly describes cases 1 and 2 of the MAS3K dataset.

In case 1, the numbers of images in the train, validation, and test subsets are divided into the ratio of 6:2:2, including the background images, similarly to in a previous study (Li et al., 2021). Moreover, in Case 2, images in the dataset are included, and it excludes the background images, which are divided into only the train and test subsets similar to in a previous study (Chen et al., 2022). In addition, the resolution of the input images was 352 × 352 pixels for both Cases 1 and 2. We conducted experiments with Cases 1 and 2, as listed in Table 1, to conduct a reasonable performance evaluation using the same method (Cases 1 and 2) used in the existing studies.

3.2. Implementation detail

The experiments were performed on a desktop computer using Ubuntu 20.04 with CPU Intel® Core™ i7-9700F CPU (Intel® Core™ i7-9700F CPU, accessed on 25 July 2022). It contains 16 GB of RAM and NVIDIA GeForce RTX 3070 graphics processing unit (GPU) card (NVIDIA GeForce RTX 3060, accessed on 25 July 2022). We used TensorFlow 2.6.0 (TensorFlow, accessed on 25 July 2022) for this implementation.

3.3. Training for proposed method

The proposed PSS-Net was trained using the adaptive moment estimation (Adam) optimizer (Kingma and Ba, 2014). Table 2 lists the hyperparameters used to train the PSS-Net. The same hyperparameters were used in the experiments for Cases 1 and 2.

Figure 3 shows the training and validation losses and accuracy graphs of the PSS-Net.

As shown in Figure 3(A), the loss graph with the training data converges to a sufficiently small value as the epoch increased, whereas the accuracy graph with the training data converged to a sufficiently large value. This result indicates that the PSS-Net proposed in this study is sufficiently trained on the training data. In addition, as shown in Figure 3(B), the loss and accuracy graphs with the validation data in the experiment of case 1 converge to sufficiently small and large values, respectively, as the epoch increases. This result confirms that the PSS-Net proposed in this study was not overfitted with the training dataset.

TABLE 1 Descriptions of two cases in MAS3K datasets for training, validation, and testing.

Case	Train	Validation	Test	Total
1	1,962	568	573	3,103
2	1,769	–	1,141	2,910

TABLE 2 Hyperparameters used for training of PSS-Net.

Hyperparameters	Learning rate	beta_1	beta_2	Batch size	Epochs
Values	10^{-4}	0.9	0.99	10	200

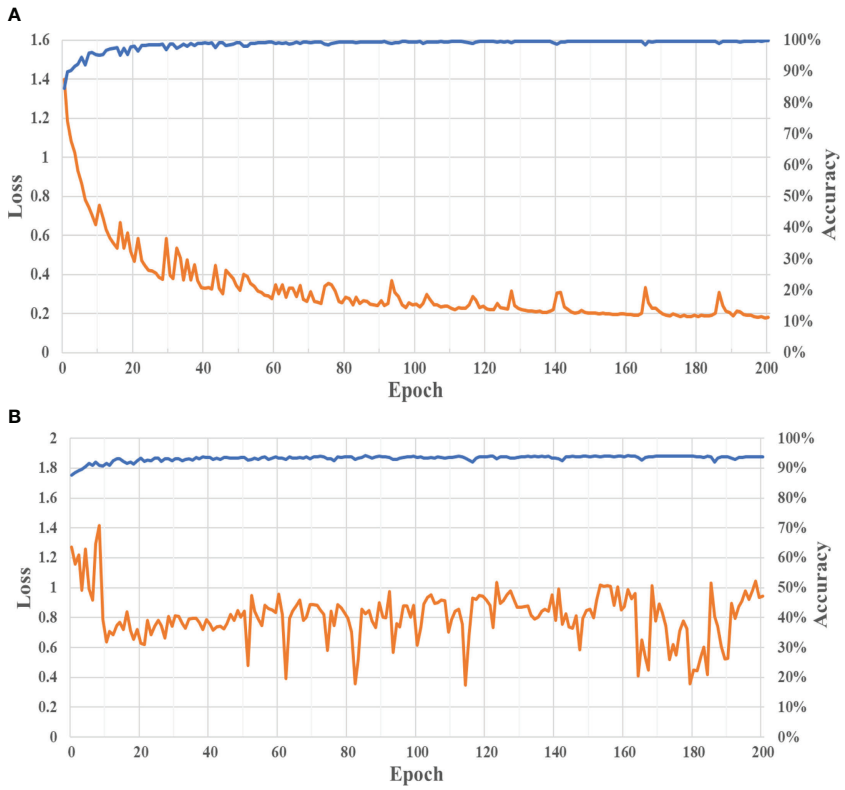


FIGURE 3
Graphs of losses and accuracies with **(A)** training and **(B)** validation data. (Orange and blue lines represent loss and accuracy, respectively).

3.4. Testing for proposed method

3.4.1. Evaluation metrics

The testing accuracy was measured using the mean intersection of union ($mIoU$), F_β^ω (weighted precision and recall) (Margolin et al., 2014), $mE_\varnothing$ (combination of local pixel values and image-level average values) (Fan et al., 2018), S_α object-aware and region-aware structure similarities) (Cheng and Fan, 2021), and mean absolute error (MAE). To use these metrics, precision and recall need to be measured, as expressed in Equations (9) and (10), respectively. TP, FN, and FP in Equations (9) and (10) denote the numbers of true positive, false negative, and false positive, respectively.

$$RECALL = \frac{TP}{TP + FN} \tag{9}$$

$$PRECISION = \frac{TP}{TP + FP} \tag{10}$$

F_β^ω $mE_\varnothing$ and S_α are expressed in Equations (11)–(13), respectively.

$$F_\beta^\omega = \frac{(1 + \beta^2) \cdot PRECISION^\omega \cdot RECALL^\omega}{\beta^2 \cdot PRECISION^\omega \cdot RECALL^\omega} \tag{11}$$

$$mE_\varnothing = \frac{2 \cdot \varphi_G \circ \varphi_Y}{\varphi_G \circ \varphi_G + \varphi_Y \circ \varphi_Y}, \quad (\varphi_G = G - \mu_G \cdot \mathbb{A}, \quad \varphi_Y = Y - \mu_Y \cdot \mathbb{A}) \tag{12}$$

$$S_\alpha = \alpha S_o + (1 - \alpha) S_r \tag{13}$$

Equations (11) and (13) are the equations used in (Margolin et al., 2014) and (Cheng and Fan, 2021), respectively. In addition, β 0.3 and α of 0.5 were used based on (Margolin et al., 2014) and (Cheng and Fan, 2021). In Equation (12), Y and G denote the foreground map and the ground truth image, respectively. Furthermore, μ_Y and μ_G are the averages of Y and G respectively. $\mathbb{A}$ is a matrix in which all element values are equal to one, and its size is equal to Y and G respectively. Specifically, φ_G and φ_Y denote the foreground map of the ground truth image and the foreground map of the prediction, respectively. In Equation (13), S_o and S_r denote the similarity to the object structure and the similarity to the

structure of the boundary region (structural similarity), respectively. Moreover, high *mIoU*, F_β^ω $mE_\varnothing$ and S_α values imply good performance. Conversely, a small *MAE* value implies a good performance.

3.4.2 Testing on case 1 with MAS3K dataset

3.4.2.1. Ablation studies

Ablation studies on case 1 of Table 1 with the MAS3K dataset were conducted first. Experiments were conducted by dividing the feature map fusion method, which is presented in Figure 2, into five cases. The first method is "testing with feature maps fusion using addition and average (Method 1)." The second method is "testing with feature maps fusion using multiplication without sigmoid function (Method 2)." The third and fourth methods are "testing with feature maps fusion using addition (Method 3)" and "testing only foreground regions model (Method 4)," respectively. Finally, the fifth method is "testing with feature maps fusion using attention based on sigmoid function (proposed method) (Method 5)." Table 3 reports the experimental results for the five methods described above.

As can be seen from Table 3, the "feature maps fusion using attention based on sigmoid function" used in the PSS-Net proposed in this study exhibits higher segmentation accuracy than the other feature map fusion methods. This result is attributed to the fact that the combination of the attention for each layer and the last feature attention in the PSS-Net contributes to the improvement in the segmentation performance. Figure 4 shows sample segmentation images obtained from the PSS-Net.

Figure 5 shows sample attention feature maps in the PSS-Net. Specifically, it shows the feature maps or GradCAM (Selvaraju et al., 2017) images extracted from the third–fifth convolutional layers of the encoder and the fourth–sixth convolutional layers of the decoder in Figure 1, respectively. In Figure 5, the values that are close to red represent important features, whereas those close to blue represent unimportant features. It can be seen that the PSS-Net proposed in this study appropriately extracts important features for the foreground and background segmentation.

3.4.2.2. Comparison of proposed PSS-net and state-of-the-art methods

In the present experiments, the performances of the proposed PSS-Net and state-of-the-art methods—U-Net++ (Zhou et al., 2018), PiCANet (Liu et al., 2018), BASNet (Qin et al., 2019), CPDNet (Wu et al., 2019), PoolNet (Liu et al., 2019), EGNet (Zhao et al., 2019), SCRN (Wu et al., 2019), U2-Net (Qin et al., 2020), SINet (Fan et al., 2020), and ECD-Net (Li et al., 2021)—were compared. We referred to the results presented in a previous paper (Li et al., 2021) for the performance benchmark of these methods. Table 4 compares the PSS-Net proposed in this study and the existing state-of-the-art methods.

As summarized in Table 4, the proposed PSS-Net is confirmed to exhibit a higher segmentation accuracy than the existing state-of-the-art methods. In particular, the *mIoU*, S_α F_β^ω and $mE_\varnothing$ performance of the proposed PSS-Net are approximately 16%, 12%, 11%, and 5% higher than those of the second-best method, respectively. In addition, the *MAE* of the proposed PSS-Net is 0.007 lower than that of the second-best method. Therefore, the proposed PSS-Net improves the detection of the regions of the object and the background compared to the existing state-of-the-art methods.

3.4.3. Testing on case 2 with MAS3K dataset

3.4.3.1. Ablation studies

Experiments for case 2 with the MAS3K dataset were conducted. First, an ablation experiment was performed similar to the five experiments resulting from dividing the feature maps fusion method into five cases, as discussed in Section 3.4.2.1. Table 5 summarizes the results of the ablation experiments related to case 2.

Based on Table 5, the "feature maps fusion using attention based on a sigmoid function" used in this study exhibits a higher segmentation accuracy than the other feature map fusion methods for case 2. This result is attributed to the fact that the feature fusion segmentation proposed in this study is effective in improving the performance. Figure 6 shows sample segmentation result images obtained by the PSS-Net proposed in this study.

TABLE 3 Comparative accuracies of testing with feature maps fusion using addition and average, multiplication, addition, and attention (proposed method) on MAS3K dataset (case 1).

Method	*mIoU*	S_α	F_β^w	$mE_\varnothing$	*E*
1	0.867	0.971	0.879	0.951	0.029
2	0.868	0.971	0.879	0.951	0.029
3	0.867	0.970	0.874	0.951	0.029
4	0.869	0.972	0.880	0.952	0.029
5 (PSS-Net)	0.870	0.973	0.880	0.952	0.029

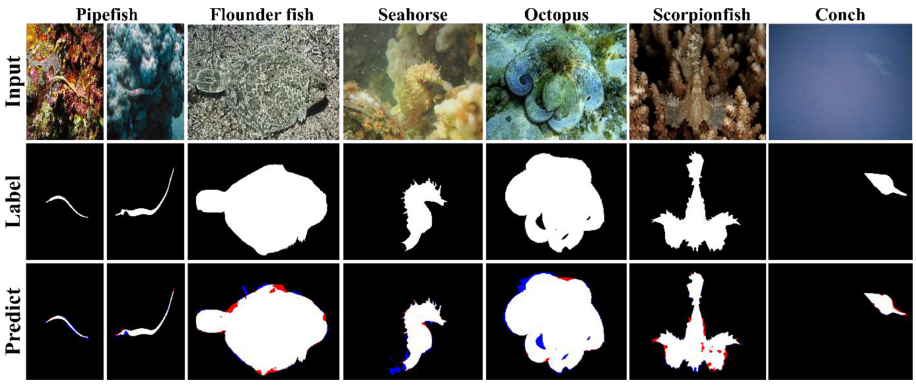


FIGURE 4
Input and label (ground-truth) images, and predict images obtained using PSS-Net with MAS3K dataset (case 1). TP, FP, FN, and TN are expressed in white, blue, red, and black, respectively.

Figure 7 shows a sample attention feature map in the PSS-Net. It can be found that the PSS-Net appropriately extracts important features for the foreground and background segmentation, even in images in which the background and the foreground are similar.

3.4.3.2. Comparison of proposed PSS-Net and state-of-the-art methods

In this section, we present the comparison of the experimental results of the PSS-Net proposed in this study with those of the BASNet (Qin et al., 2019), SCRN (Wu et al., 2019), SINet (Fan et al., 2020), U2-Net (Qin et al., 2020), SINet-V2 (Fan et al., 2021), C2FNet (Sun et al., 2021), and WaterSNet (Chen et al., 2022). We refer to the results of a previous study (Chen et al., 2022) for the performance benchmarking of these methods. Table 6 summarizes the comparison of the PSS-Net proposed in this study and the existing state-of-the-art methods.

Based on Table 6, even though the existing methods slightly outperform the PSS-Net in terms of the F_β^ω $mE_\varnothing$ and *MAE*, the PSS-Net shows much higher segmentation accuracies than the existing state-of-the-art methods in terms of *mIoU* and S_α However $mE_\varnothing$ measures the accuracy by correcting the average values in the foreground map and the ground truth image, respectively, as expressed in Equation (12). Hence, it has a disadvantage in that the performance is affected by the average value of the foreground map and the ground truth image. In addition, because F_β^ω β (0.3 based on (Margolin et al., 2014) and (Cheng and Fan, 2021) considers *RECALL* to be more important than *PRECISION*, as expressed in Equation (11), it has a disadvantage in that the *FPs* are improperly considered in the performance evaluation, as expression in Equation (10). Moreover, as written in Equation (13), S_α does not simply represent the *MAE* between the predicted image and the ground truth image. It is a measure that represents the

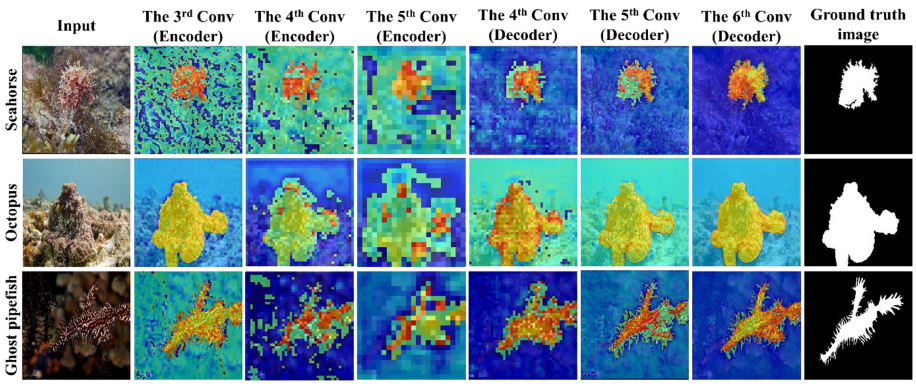


FIGURE 5
Examples of attention feature map with MAS3K dataset (case 1).

TABLE 4 Comparison of proposed PSS-Net and state-of-the-art methods with MAS3K dataset (case 1).

Method	*mIoU*	S_α	F_β^w	mE_ϕ	*E*
U-Net++	0.429	0.692	0.522	0.754	0.099
PiCANet	0.611	0.805	0.680	0.859	0.051
BASNet	0.511	0.732	0.572	0.791	0.076
CPDNet	0.590	0.810	0.636	0.829	0.057
PoolNet	0.604	0.799	0.685	0.867	0.045
EGNet	0.596	0.806	0.677	0.853	0.047
SCRN	0.649	0.830	0.686	0.863	0.047
U2Net	0.541	0.776	0.615	0.802	0.058
SINet	0.652	0.830	0.678	0.874	0.044
ECD-Net	0.711	0.850	0.766	0.901	0.036
PSS-Net (proposed method)	0.870	0.973	0.880	0.952	0.029

TABLE 5 Comparative accuracies of testing with feature maps fusion using addition and average, multiplication, addition, and using attention (proposed method) in MAS3K dataset (case 2).

Method	*mIoU*	S_α	F_β^w	mE_ϕ	*E*
1	0.815	0.965	0.783	0.894	0.045
2	0.815	0.965	0.783	0.894	0.044
3	0.815	0.964	0.783	0.894	0.045
4	0.816	0.966	0.784	0.894	0.044
5 (PSS-Net)	0.816	0.966	0.784	0.895	0.044

respective structure similarity-based accuracy of a detected object and the boundary region. Hence, it can be inferred that S_α along with the *mIoU*, shows more accurate detection results than the other metrics.

3.4.4. Processing time

In this section, we discuss the inference time of the PSS-Net proposed in this study. First, Table 7 lists the measured inference time of the PSS-Net on a desktop and an embedded system (Jetson TX2 board) (Jetson TX2 embedded system, accessed on 25 July 2022) as shown in Figure 8. Jetson TX2 includes an NVIDIA Pascal™-family GPU (256 CUDA cores) with 8 GB of GPU memory.

Based on Table 7, the inference times for one image on the desktop and the Jetson embedded system are 6.43 ms and 38.61 ms, respectively. These inference times imply processing speeds of 155.5 frames per second (fps) (1000/6.43) and 25.9 fps (1000/38.61), respectively. We employed the PSS-Net proposed in this

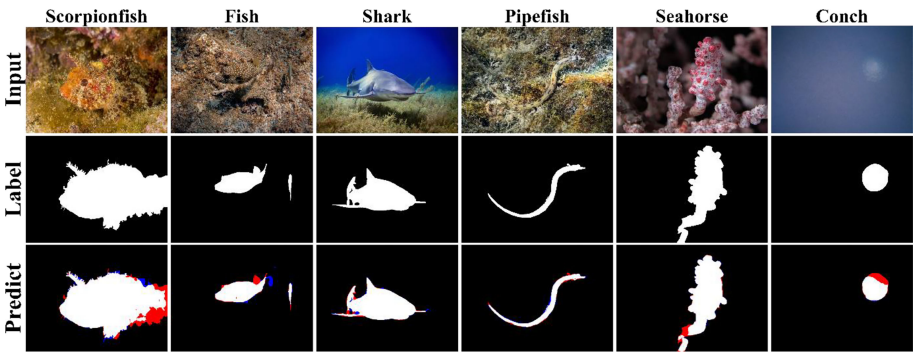


FIGURE 6
Result images of PSS-Net with MAS3K dataset (case 2). TP, FP, FN, and TN are expressed in white, blue, red, and black, respectively.

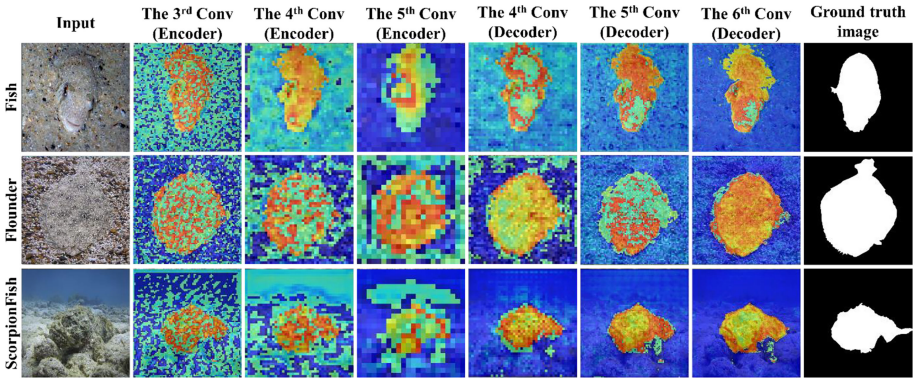


FIGURE 7
Examples of attention feature maps using MAS3K dataset (case 2).

TABLE 6 Comparison of proposed PSS-Net and state-of-the-art methods with MAS3K dataset (case 2).

Method	*mIoU*	S_α	F_β^w	$mE_\varnothing$	*E*
BASNet	0.678	0.820	0.748	0.869	0.044
SCRN	0.690	0.832	0.762	0.884	0.038
SINET	0.657	0.815	0.745	0.885	0.039
U2-Net	0.651	0.809	0.722	0.851	0.047
SINet-V2	0.561	0.757	0.648	0.826	0.061
C2F-Net	0.717	0.844	0.781	0.903	0.036
WaterSNet	0.739	0.856	0.804	0.913	0.032
PSS-Net (proposed method)	0.816	0.966	0.784	0.895	0.044

study in the Jetson embedded system because an underwater vehicle is generally used in marine segmentation. In addition, because a Jetson embedded system is extensively used as an edge computing device that is attached to an underwater vehicle, the proposed PSS-Net was used in the Jetson embedded system to examine whether it can perform edge computing. Finally, based on Table 7, it is confirmed that the PSS-Net proposed in this study can operate in a limited embedded system.

4. Discussion

For the ECD-Net (Li et al., 2021), which exhibits the second-best performance in Table 4, an IFEM is developed to extract the main features effectively when the features are compressed in the encoder. In addition, CDMs that can integrate features are designed to improve the final marine animal segmentation performance. In the WaterSNet (Chen et al., 2022), which exhibits the second-best performance in terms of the *mIoU* and S_α in Table 6 and slightly outperforms the PSS-Net in terms of the F_β^ω $mE_\varnothing$ and *MAE*, the AFB module is used to utilize the global context information. In addition, an RFB module is employed to extract multiscale features. This method is considered to use an attention mechanism. Furthermore, two models are used to reduce the overfitting, and a Siamese structure is used to allow various learning.

TABLE 7 Inference times of PSS-Net on desktop and Jetson embedded system.

Environment	Inference time (per one image)
Desktop	6.43 ms
Jetson embedded system	38.61 ms

In contrast, the core of the PSS-Net proposed in this study can be summarized in three points. First, the proposed PSS-Net is a model designed with parallel CNNs that can be separately trained on the foreground and background of marine animal images. Second, the PSS-Net proposed in this study is a model that reinforces the foreground and background learning by adding an attention mechanism between the parallel CNNs. This mechanism allows focusing on the foreground and background learning tasks without conflicts. Third, when testing is performed, the foreground and background feature maps extracted from

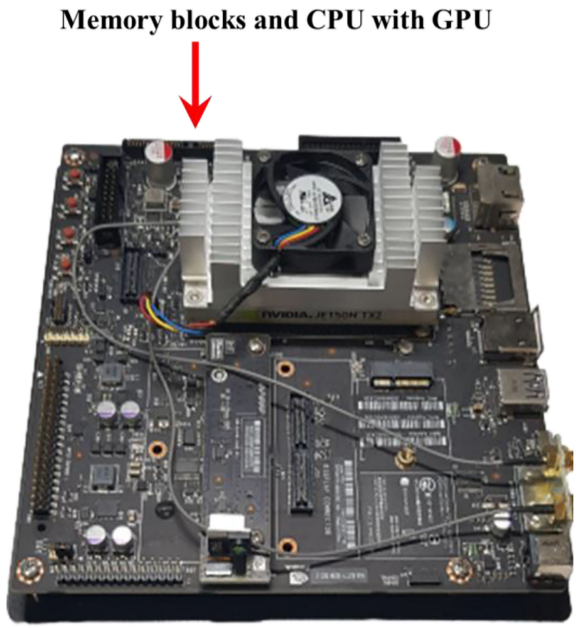


FIGURE 8
Jetson TX2 board.

the parallel CNNs are fused to perform the final segmentation. Moreover, the fusion proposed in this study is regarding the attention mechanism. Thus, it is confirmed that the PSS-Net proposed in this study shows improved marine animal segmentation performance. However, the detection error of the proposed PSS-Net increases when an object has complex boundaries or the distinctiveness of the object is very low compared to background due to protective colors, as shown in Figure 9.

As shown in Sections 3.4.2 and 3.4.3, the accuracies of case 1 experiments are higher than those of case 2 experiments. That is because the number of training data of case 1 is larger than that of case 2 as shown in Table 1, which makes the PSS-Net more robust to various data in case 1. In addition, the number of

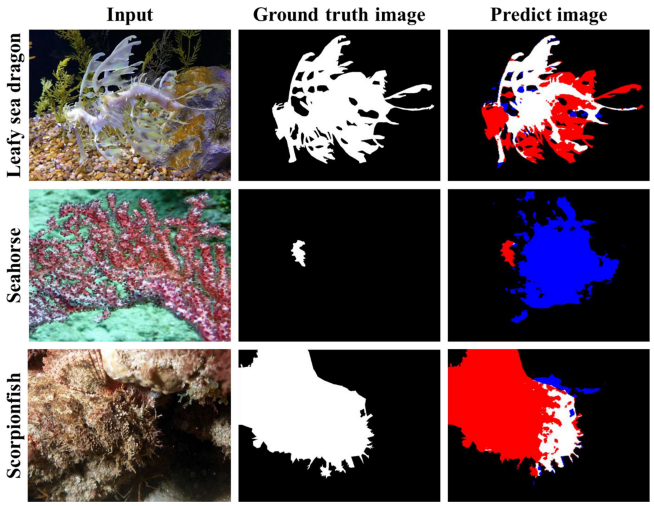


FIGURE 9
Examples of error cases for complex objects. TP, FP, FN, and TN are denoted in white, blue, red, and black colors, respectively.

testing data of case 1 is smaller than that of case 2 as shown in Table 1, which can reduce the testing error in case 1.

5. Conclusions

In this study, we established the PSS-Net for marine animal segmentation. In addition, the PSS-Net was trained on the foreground and the background separately to detect marine animals accurately. This mechanism was possible because the PSS-Net was designed by connecting the foreground and background models in parallel. Moreover, an attention mechanism was connected between the parallel CNNs, and the foreground and background learning tasks were reinforced by separate losses (object and background losses). Finally, the background feature maps were fused to the foreground feature maps to perform the final segmentation when testing. Subsequently, ablation studies were conducted on this fusion. The results confirmed that the attention-based fusion proposed in this study has a high segmentation performance. Furthermore, we verified that the improved marine animal segmentation performance of the proposed PSS-Net by conducting comparison experiments with existing state-of-the-art methods. The segmentation results of case 1 experiments using the MAS3K open dataset reveal that the proposed method achieves performance of 87%, 97.3%, 88%, 95.2%, and 0.029 in the $mIoU$, S_α F_β^ω $mE_\varnothing$ and MAE, respectively. In addition, those of case 2 experiments using the MAS3K open dataset reveal that the proposed method achieves performance of 81.6%, 96.6%, 78.4%, 89.5%, and 0.044 in the $mIoU$, S_α F_β^ω $mE_\varnothing$ and MAE, respectively. However, it was found that the detection error of the proposed PSS-Net increased for cases where the object has complex boundaries and protective colors.

In future research, we plan to perform the segmentation of marine animal with severely complex boundaries and protective colors based on the segmentation approach of object within the rectangular region roughly detected by a conventional CNN detector. In addition, we aim to expand the results of this study to investigate how to improve the semantic segmentation performance for multiclass marine animals.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: (MAS3K database) https://github.com/LinLi-DL/MAS.

Ethics statement

Ethical review and approval was not required for the animal study because in our study, we used an open database of MAS3K dataset which can be freely used for academic purpose according to the instructions of database provider. Therefore, ethical review and approval was not required for this study.

Author contributions

Y.H.K designed the method with experiments and wrote the first draft of the manuscript supervised by K.R.P. K.R.P critically reviewed the initial manuscript and provided helpful input. All authors contributed to the article and approved the submitted version.

Funding

This research was supported in part by the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (MSIT) through the Basic Science Research Program (NRF-2021R1F1A1045587), in part by the NRF funded by the MSIT through the Basic Science Research Program (NRF-2022R1F1A1064291), and in part by the MSIT, Korea, under the ITRC (Information Technology Research Center) support program (IITP-2022-2020-0-01789) supervised by the IITP (Institute for Information and Communications Technology Planning and Evaluation).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Berg, P., Maia, D. S., Pham, M.-T., and Lefèvre, S. (2022). Weakly supervised detection of marine animals in high resolution aerial images. *Remote Sens.* 14, 1–17. doi: 10.3390/rs14020339

Chen, R., Fu, Z., Huang, Y., Cheng, E., and Ding, X. (2022). "A robust object segmentation network for underwater scenes," in *Proc. IEEE int. conf. acoust. speech signal process*(Singapore: IEEE), 2629–2633.

Cheng, M.-M., and Fan, D.-P. (2021). Structure-measure: A new way to evaluate foreground maps. *Int. J. Comput. Vis.* 129, 2622–2638. doi: 10.1007/s11263-021-01490-8

Ditria, E. M., Lopez-Marcano, S., Sievers, M., Jinks, E. L., Brown, C. J., and Connolly, R. M. (2020). Automating the analysis of fish abundance using object detection: Optimizing animal ecology with deep learning. *Front. Mar. Sci.* 7. doi: 10.3389/fmars.2020.00429

Fan, D.-P., Gong, C., Cao, Y., Ren, B., Cheng, M.-M., and Borji, A. (2018). "Enhanced-alignment measure for binary foreground map evaluation," in *Proc. int. joint conf. artif. intell* (Stockholm, Sweden: Stockholmsmässan), 698–704. Available at: https://doi.org/10.24963/ijcai.2018/97.

Fan, D.-P., Ji, G.-P., Cheng, M.-M., and Shao, L. (2021). Concealed object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* doi: 10.1109/TPAMI.2021.3085766

Fan, D.-P., Ji, G.-P., Sun, G., Cheng, M.-M., Shen, J., and Shao, L. (2020). "Camouflaged object detection," in *Proc. conf. comput. vis. pattern recognit*(Seattle, Online, USA: IEEE), 2777–2787.

Fu, J., Liu, J., Jiang, J., Li, Y., Bao, Y., and Lu, H. (2020). Scene segmentation with dual relation-aware attention network. *IEEE Trans. Neural Netw. Learn. Syst.* 32, 2547–2560. doi: 10.1109/TNNLS.2020.3006524

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proc. conf. comput. vis. pattern recognit* (Las Vegas, Nevada, USA: IEEE), 770–778.

Huang, Z., Wang, X., Huang, L., Huang, C., Wei, Y., and Liu, W. (2019). "CCNet: Criss-cross attention for semantic segmentation," in *Proc. int. conf. comput. vis*(Seoul, Korea (South: IEEE), 603–612.

Intel® Core™ i7-9700F CPU (2022). Available at: https://ark.intel.com/content/www/us/en/ark/products/193738/intel-core-i79700f-processor-12m-cache-up-to-4-70-ghz.html (Accessed July 25, 2022).

Islam, M. J., Edge, C., Xiao, Y., Luo, P., Mehtaz, M., Morse, C., et al. (2020). Semantic segmentation of underwater imagery: Dataset and benchmark. *arXiv*. 1–9. doi: 10.48550/arXiv.2004.01241

Jetson TX2 embedded system (2020). Available at: https://developer.nvidia.com/embedded/jetson-tx2 (Accessed July 25, 2022).

KaewTraKulPong, P., and Bowden, R. (2001). An improved adaptive background mixture model for realtime tracking with shadow detection. *Proc. Adv. Video Surveill. Syst.*, 135–144. doi: 10.1007/978-1-4615-0913-4_11

Kandimalla, V., Richard, M., Smith, F., Quirion, J., Torgo, L., and Whidden, C. (2022). Automated detection, classification and counting of fish in fish passages with deep learning. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.823173

Kim, Y. H., and Park, K. R. (2022). MTS-CNN: Multi-task semantic segmentation-convolutional neural network for detecting crops and weeds. *Comput. Electron. Agric.* 199, 107146. doi: 10.1016/j.compag.2022.107146

Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv* 1–15. doi: 10.48550/arXiv.1412.6980

Koch, G., Zemel, R., and Salakhutdinov, R. (2015). "Siamese Neural networks for one-shot image recognition," in *Proc. int. conf. mach. learn* (Lille, France: ICML).

Li, L., Dong, B., Rigall, E., Zhou, T., Dong, J., and Chen, G. (2021). Marine animal segmentation. *IEEE Trans. Circuits Syst. Video Technol.* 32, 2303–2314. doi: 10.1109/TCSVT.2021.3093890

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). "Focal loss for dense object detection," in *Proc. int. conf. comput. vis* (Venice, Italy: IEEE), 2999–3007. doi: 10.1109/ICCV.2017.324

Li, L., Rigall, E., Dong, J., and Chen, G. (2021). "MAS3K: An open dataset for marine animal segmentation," in *Proc. Symp. Benchmarking, Meas. Optim*, Vol. 12614. doi: 10.1007/978-3-030-71058-3_12

Liu, N., Han, J., and Yang, M.-H. (2018). "PiCANet: Learning pixel-wise contextual attention for saliency detection," in *Proc. conf. comput. vis. pattern recognit* (Salt Lake City, UT, USA: IEEE), 3089–3098.

Liu, J.-J., Hou, Q., Cheng, M.-M., Feng, J., and Jiang, J. (2019). "A simple pooling-based design for real-time salient object detection," in *Proc. conf. comput. vis. pattern recognit* (Long Beach, CA, USA: IEEE), 3912–3921.

Margolin, R., Zelnik-Manor, L., and Tal, A. (2014). "How to evaluate foreground maps," in *Proc. conf. comput. vis. pattern recognit*(Columbus, OH, USA: IEEE).

NVIDIA (2022). Available at: https://www.nvidia.com/ko-kr/geforce/graphics-cards/30-series/rtx-3060-3060ti/ (Accessed July 25, 2022).

Panaïotis, T., Caray–Counil, L., Woodward, B., Schmid, M. S., Daprano, D., Tsai, S. T., et al. (2022). Content-aware segmentation of objects spanning a large size range: Application to plankton images. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.870005

Pang, Y., Zhao, X., Zhang, L., and Lu, H. (2020). "Multi-scale interactive network for salient object detection," in *Proc. conf. comput. vis. pattern recognit* (Seattle, Online, USA: IEEE), 9413–9422.

Pedersen, M., Haurum, J. B., Gade, R., and Moeslund, T. B. (2019). "Detection of marine animals in a new underwater dataset with varying visibility," in *Proc. conf. comput. vis. pattern recognit* (Long Beach, CA: IEEE), 18–26.

Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O. R., and Jagersand, M. (2020). U2-net: Going deeper with nested u-structure for salient object detection. *Pattern Recognit.* 106, 1–12. doi: 10.1016/j.patcog.2020.107404

Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M., and Jagersand, M. (2019). "BASNet: Boundary-aware salient object detection," in *Proc. conf. comput. vis. pattern recognit* (Long Beach, CA: IEEE), 7479–7489.

Radolko, M., Farhadifard, F., and Lukas, U. F. (2016). "Dataset on underwater change detection," in *Proc. oceans Monterey* (Monterey, California, USA: IEEE). doi: 10.1109/OCEANS.2016.7761129

Radolko, M., Farhadifard, F., and Lukas, U. V. (2017). "Change detection in crowded underwater scenes - *via* an extended Gaussian switch model combined with a flux tensor pre-segmentation," in *Proc. int. joint conf. comput. vis, img. comput. theory appl.* (Setúbal Portugal: SCITEPRESS)

Radolko, M., and Gutzeit, E. (2015). "Video segmentation *via* a Gaussian switch background model and higher order Markov random fields," in *Proc. int. joint conf. comput. vis, img. comput. theory appl* (Berlin, Germany: SCITEPRESS), 537–544.

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. int. conf. med. comput. comput.-ASSIST. interv* (Munich, Germany: Springer), 234–241. Available at: https://doi.org/10.48550/arXiv.1505.04597.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). "Grad-cam: visual explanations from deep networks *via* gradient-based localization," in *Proc. int. conf. comput. vis* (Venice, Italy: IEEE).

Simonyan, K., and Zisserman, A. (2015). "Very deep convolutional networks for large-scale image recognition," in *Proc. conf. comput. vis. pattern recognit* (San Diego, CA, USA: IEEE), 1–14. Available at: https://doi.org/10.48550/arXiv.1409.1556.

Sun, Y., Chen, G., Zhou, T., Zhang, Y., and Liu, N. (2021). "Context-aware cross-level fusion network for camouflaged object detection," in *Proc. int. joint conf. artif. intell* (Montreal-themed Virtual Reality), 1025–1031. (CA, USA: IJCAI). Available at: https://doi.org/10.48550/arXiv.2105.12555.

TensorFlow (2022). Available at: https://www.tensorflow.org/ (Accessed July 25, 2022).

Wu, Z., Su, L., and Huang, Q. (2019a). "Cascaded partial decoder for fast and accurate salient object detection," in *Proc. conf. comput. vis. pattern recognit*(Long Beach, CA, USA: IEEE). Available at: https://doi.org/10.48550/arXiv.1904.08739.

Wu, Z., Su, L., and Huang, Q. (2019b). "Stacked cross refinement network for edge-aware salient object detection," in *Proc. int. conf. comput. vis* (Seoul, Korea (south: COEX), 8779–8788.

Xu, G., Xie, W., Dong, C., and Gao, X. (2021). Application of three deep learning schemes into oceanic eddy detection. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.672334

Zhang, W., Wu, C., and Bao, Z. (2021). DPANet: Dual pooling-aggregated attention network for fish segmentation. *IET Comput. Vis.* 16, 67–82. doi: 10.1049/cvi2.12065

Zhao, J.-X., Liu, J.-J., Fan, D.-P., Cao, Y., Yang, J.-F., and Cheng, M.-M. (2019). "EGNet: Edge guidance network for salient object detection," in *Proc. int. conf. comput. vis* (Seoul, Korea (south: COEX), 8779–8788.

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J. (2018). "UNet++: A nested u-net architecture for medical image segmentation," in *Proc. deep learning med. img* (Granada, Spain: Springer), 3–11.

Zivkovic, Z. (2004). "Improved adaptive Gaussian mixture model for background subtraction," in *Proc. int. assoc. pattern recognit* (Cambridge, UK: IEEE).

Zivkovic, Z., and Heijden, F. (2006). Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognit. Lett.* 27, 773–780. doi: 10.1016/j.patrec.2005.11.005

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 06 October 2022
DOI 10.3389/fmars.2022.1010565

Check for updates

OPEN ACCESS

EDITED BY
Hong Song,
Zhejiang University, China

REVIEWED BY
Nitin Agarwala,
National Maritime Foundation, India
Jamil Hussain,
Kyung Hee University, South Korea

*CORRESPONDENCE
Kang Ryoung Park
parkgr@dongguk.edu

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 03 August 2022
ACCEPTED 20 September 2022
PUBLISHED 06 October 2022

CITATION
Haider A, Arsalan M, Choi J, Sultan H and Park KR (2022) Robust segmentation of underwater fish based on multi-level feature accumulation. *Front. Mar. Sci.* 9:1010565. doi: 10.3389/fmars.2022.1010565

COPYRIGHT
© 2022 Haider, Arsalan, Choi, Sultan and Park. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Robust segmentation of underwater fish based on multi-level feature accumulation

Adnan Haider, Muhammad Arsalan, Jiho Choi, Haseeb Sultan and Kang Ryoung Park*

Division of Electronics and Electrical Engineering, Dongguk University, Seoul, South Korea

Because fish are vital to marine ecosystems, monitoring and accurate detection are crucial for assessing the potential for fisheries in these environments. Conventionally, fish-related assessment is conducted manually, which makes it labor-intensive and time-consuming. In addition, the assessments are challenging owing to underwater visibility limitations, which leads to poor detection accuracy. To overcome these problems, we propose two novel architectures for the automatic and high-performance segmentation of fish populations. In this study, the efficient fish segmentation network (EFS-Net) and multi-level feature accumulation-based segmentation network (MFAS-Net) are the base and final networks, respectively. In deep convolutional neural networks, the initial layers usually contain potential spatial information. Therefore, the EFS-Net employs a series of convolution layers in the early stage of the network for optimal feature extraction. To boost segmentation accuracy, the MFAS-Net uses an initial feature refinement and transfer block to refine potential low-level information and subsequently transfers it to the deep stages of the network. Moreover, the MFAS-Net employs multi-level feature accumulation that improves pixel-wise prediction for fish that are indistinct. The proposed networks are evaluated using two publicly available datasets, namely DeepFish and semantic segmentation of underwater imagery (SUIM), both of which contain challenging underwater fish segmentation images. The experimental results reveal that mean intersection-over-unions of 76.42% and 92.0% are attained by the proposed method for the DeepFish and SUIM datasets, respectively; these values are higher than those by the state-of-the-art methods such as A-LCFCN+PM and DPANet. In addition, high segmentation performance is achieved without compromising the computational efficiency of the networks. The MFAS-Net requires only 3.57 million trainable parameters to be fully trained. The proposed model and the complete code will be made available[1].

KEYWORDS

artificial intelligence, marine environment, underwater computer vision, fish segmentation, EFS-net and MFAS-net

1 https://github.com/AdnanHaider72/Fish-segmentation.

1 Introduction

Fish are the subjects of interest in the marine fishing and aquaculture industries. In a marine environment, fish habitat monitoring has become an essential function in research for sustainable fisheries. Fish monitoring primarily includes shape and size assessment, which helps in analyzing the fisheries' potential (Laradji et al., 2021). Moreover, fish measurements and observations are useful for feed, stock, and growth analysis. Conventionally, manual methods are employed for fish monitoring and measurements. These include ruler and echosounder-based measurements that are time-consuming, error-prone, and tedious (Zhang et al., 2022). Therefore, traditional fish assessment methods should be replaced with automated methods for a more efficient process.

In recent years, artificial intelligence has made significant contributions to automating manual processes (Arsalan et al., 2022c; Haider et al., 2022a; Mahmood et al., 2022b). In particular, the combination of deep learning with computer vision has enabled complex problems to be solved using multimedia-based learning (Owais et al., 2021; Sultan et al., 2021; Arsalan et al., 2022a). Moreover, convolutional neural networks (CNNs) supported by machine vision have been used in automated marine observation. Automatic machine vision-based marine observation can contribute substantially to developing an intelligent decision support system (Laradji et al., 2020). In particular, machine vision for fish observation is a noteworthy step toward modern and intelligent fisheries. However, limited research work and databases are available for underwater machine vision methods because of the underwater imaging limitations and constraints. Existing deep feature-based methods for underwater fish segmentation have limitations in segmentation performance and computational efficiency. Underwater, the segmentation of objects in general and small objects in particular, is challenging because of the visibility constraints.

Existing automatic fish segmentation methods have not delivered high segmentation accuracy with the required small number of trainable parameters. To overcome these problems, two novel architectures were designed to segment fish in an underwater environment. The efficient fish segmentation network (EFS-Net) and multi-level feature accumulation-based segmentation network (MFAS-Net) respectively are the base and final networks of this study. In CNNs, layers close to the image input layer carry potential low-level information (Hosseinzadeh Kassani et al., 2022). In EFS-Net, a series of convolution layers is set at the start of the network for optimal feature extraction from this initial spatial information. In MFAS-Net, potential initial information is refined and transferred using the initial feature refinement and transfer block (IFRT-Block) approach. These initial features are transferred to the deep section of the network where, at the multi-level, they are combined with deep stage features. Feature accumulation at the multi-level improves network training and consequently enhances its segmentation performance.

Detailed fish assessment requires accurate segmentation. Both proposed networks perform semantic segmentation that classifies every pixel of the image into the fish and non-fish (background) classes. The proposed semantic segmentation methods enable precise soft computation and morphological assessment. The effectiveness of the proposed methods is validated on two publicly available datasets, namely DeepFish (Saleh et al., 2020) and semantic segmentation of underwater imagery (SUIM) (Islam et al., 2020). The contribution of the present study can be summarized as follows.

- Two novel networks were developed for robust fish segmentation in challenging underwater environments. EFS-Net and MFAS-Net are the base and final networks, respectively.
- EFS-Net uses a series of convolution layers at the start of the network for optimal low-level feature extraction. In MFAS-Net, initial low-level features are refined and transferred to a deep section of the network using IFRA-Block for improved performance. In addition, MFAS-Net accumulates initial and deep section features at multi-levels to further enhance accurate fish segmentation in underwater environments.
- Both networks, EFS-Net and MFAS-Net, were evaluated on two challenging underwater fish segmentation datasets (DeepFish and SUIM). The proposed approach outperformed state-of-the-art methods and exhibited promising performance with superior computational efficiency (it required only 3.57 million trainable parameters).
- The proposed models and algorithm are made publicly available[1].

The remaining paper is organized as follows. We discuss the existing methods in section 2. The proposed work and results are presented in sections 3 and 4, respectively. A detailed discussion is provided in section 5 and a brief conclusion of this work is given in section 6.

2 Related work

Globally, artificial intelligence has introduced revolutionary changes to traditional industries. Furthermore, computer vision has automated manual processes and provided robust solutions to reduce human efforts (Arsalan et al., 2022b). Machine vision provided versatile solutions in numerous fields, such as computer-aided diagnosis, autonomous driving, object tracking, intelligent agriculture, automatic fault detection, smart surveillance, remote sensing, smart decision support system, under-water imagery, human-machine interface,

robot-assisted surgery, and many more (Cui et al., 2022; Haider et al., 2022b; Mahmood et al., 2022a). Recently, semantic segmentation has become a widely accepted computer vision-based method. Semantic segmentation networks are usually trained with training data and tested on unseen data. In segmentation, every pixel of the image is classified into desired or undesired classes. Therefore, morphological assessments, measurements, region-based ratios, and visual monitoring can be conducted using segmentation algorithms. In a study (Arsalan et al., 2022d), a pool-less residual segmentation network (PLRS-Net) was used to segment the retinal vessels. This method transfers instant spatial information with residual connectivity. However, PLRS-Net requires a large number of parameters (7.3 million) to complete its training (Arsalan et al., 2022d). Underwater semantic segmentation is challenging because of the unfavorable visual conditions. In the work by Rahnemoonfar and Dobbs (2019), a deep feature-based approach; dense module, dilated convolution, and inception are used for the segmentation of underwater sonar imagery to detect potholes.

In Zhang et al. (2022), a dual pooling-aggregated attention network was employed for segmenting fish underwater. Position attention and channel attention modules were used for the aggregation of spatial context information to highlight the context dependencies for fish segmentation (Zhang et al., 2022). In their study, the proposed framework was evaluated using two datasets of underwater fish segmentation. The use of ResNet50 as the backbone can be defined as the limitation in the study by Zhang et al. (2022). Another study, Laradji et al. (2021), proposed a weakly supervised approach for underwater fish segmentation that uses a CNN, one section of which provides a pixel-wise score and the other an affinity matrix. Aggregation was performed for both outputs, and consequently, the refined pixel-wise output was generated (Laradji et al., 2021). The limitation of the study was the inability to detect several fish in the same image because it used point-level annotation (Laradji et al., 2021). In Ditria et al. (2020), a deep feature-based object detection approach was used to analyze the abundance of fish. This method presents three models for detecting species underwater and compares the results with the detection accuracy of human experts. A limitation to their study was using the ResNet50 model pre-trained on the ImageNet. Similarly, in Kandimalla et al. (2022), a deep learning method was presented for the detection and classification of fish. Herein, the widely used you only look once (YOLO) machine learning model was combined with Kalman filters to classify and track different species of fish. This method was evaluated with limited data; a single dataset was used for the detection and classification performance evaluation. In addition, no pixel-level detection was performed; therefore, no morphological analyses or species measurements were possible (Kandimalla et al., 2022).

Subsequently, semantic segmentation was performed using DeepLabv3+ for underwater pixel-wise scene detection in another study (Liu and Fang, 2020). This method employed a basic architecture (DeepLabv3+) and used XceptionNet as the backbone, which can be considered the limitation of this method (Liu and Fang, 2020). The Labao and Naval (2017) method, employs a fully convolutional residual network (ResNet-FCN) with 152-layers to segment fish underwater. This method can better deal with the illumination and background changes in underwater environments. However, the network used in this method is outdated and computationally expensive (Labao and Naval, 2017). The Abe et al. (2021) study refers to a framework that is designed for the identification and tracking of fish. In the Badrinarayanan et al. (2017) method, widely used SegNet was chosen for detecting fish at the pixel level (Abe et al., 2021). This method has the limitation of relying on SegNet, which has vanishing gradient problems and requires a large number of trainable parameters (Abe et al., 2021). A study by Islam et al. (2020) used a dataset of underwater objects with a deep residual network for segmentation. In this work, an encoder-decoder structure was presented with optional residual skip blocks (Islam et al., 2020). Each residual skip block consists of three convolutional layers in combination with the rectified linear unit (ReLU) and batch normalization (BN) layers (Islam et al., 2020). Although this work exhibited excellent segmentation performance with competitive computational efficiency, improvements can be made (Islam et al., 2020).

Existing methods suffered from many limitations that have driven the development of new models. Some of the existing methods use the backbone of different architectures or rely on pre-trained networks/preprocessing that can make the model dependent, complex, and less efficient. Similarly, few methods use point-level annotation therefore such a framework shows an inability for multiple fish detection. Most of the networks exhibit a degraded performance for underwater small and unclear objects. Existing methods that deliver competitive accuracies suffer from poor computational efficiency and require a large number of trainable parameters. Proposed methods are developed to address all these problems and to overcome the limitations of existing methods.

3 Materials and methods

3.1 Overview of the proposed methods

An overview of the proposed architecture is shown in Figure 1. Experimentation was performed for underwater fish segmentation task using two publicly available datasets. High-performance segmentation is challenging in an underwater environment because of the difficult imaging environment.

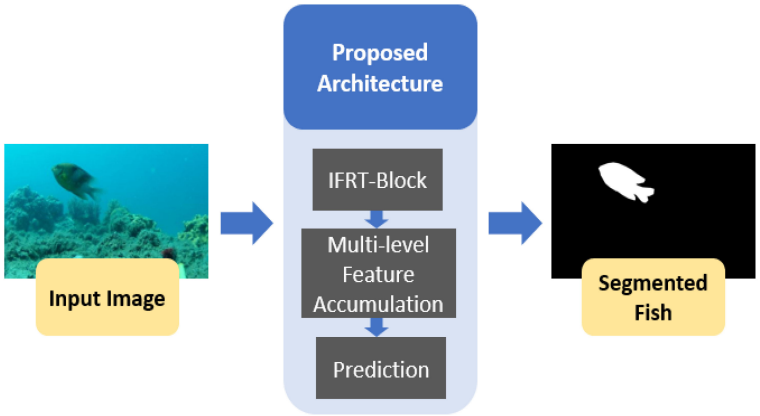


FIGURE 1
Overview of the proposed method.

Underwater imaging produces serious limitations in contrast and sharpness that create challenging conditions for segmentation. Underwater segmentation becomes even more challenging because of the sizes and shapes of fish as well as the effects of the background and illumination. Additionally, detecting small-sized fish under unfavorable background conditions is challenging. Most of the existing methods failed to provide better performance because of the small final feature map size, vanishing gradient problem, and unavailability of appropriate feature aggregation mechanisms. To overcome all these problems and deal with challenges, two networks were developed for fish segmentation underwater. Input images from both datasets are provided to the network for training and testing. To train the network fast, the training data split was resized using nearest neighbor interpolation and input to the network. Resizing training images is a common practice for the training of deep learning models. In our experiments, training images were resized to 400 × 400 pixels for efficient training of the network. However, images can be resized to any other size keeping the original image's size in view.

The initial layers of a CNN hold potential low-level information (Hosseinzadeh Kassani et al., 2022). The proposed method uses a large number of convolution layers in the early stage of the network to optimize extraction from this low-level information. These valuable initial features are further refined and transferred to the deeper stages of the network in the IFRT-Block. Subsequently, the transferred initial features are accumulated with the features from the deeper layers. For feature empowerment and effective learning, feature accumulation occurs at multiple levels in the network. Finally, a prediction mask is generated to detect pixels of the desired class. Unlike most of the existing methods, the proposed methods do not require pre-processing to achieve desirable results. In addition, the proposed methods are not computationally expensive and require a small number of trainable parameters to complete the network training.

3.2 Fish segmentation using EFS-net

Underwater image segmentation requires an efficient and effective architecture to obtain acceptable results. Fish imaging in a natural marine environment produces a wide variety of fish images with different sizes, shapes, and backgrounds. Therefore, accurate fish segmentation is challenging in the marine underwater environment. In this study, networks were designed to overcome these challenges. EFS-Net is the base network of this study and its network architecture is shown in Figure 2. In deep learning models, layers close to the image input layer have potential low-level initial information (Hosseinzadeh Kassani et al., 2022). Herein, a series of five convolution layers were deployed to extract valuable features from low-level information. This low-level information processing helps boost the segmentation performance.

Subsequently, the feature map size of the initial spatial information was reduced using a strided convolutional (strided-Conv) layer. In EFS-Net, the strided-Conv layer was used in place of despite using pooling layers. Pooling layers produce spatial loss that negatively affects the learning process (Abdeldaim et al., 2018). Moreover, an excessively small final feature map size degrades the detection accuracy, particularly for small objects. In underwater fish segmentation, several images contain small-sized fish. Therefore, in EFS-Net, the final feature map size was sufficiently large to detect the pixels of small-sized

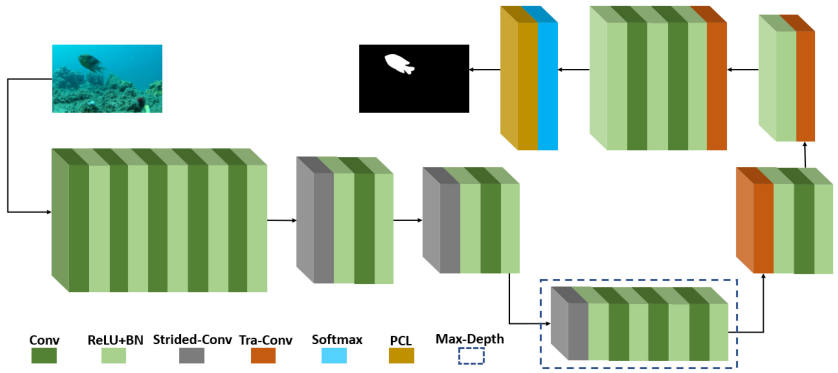


FIGURE 2
Proposed EFS-Net architecture.

fish. Only three strided-Conv layers with a stride value of 2 were used to reduce the feature map size. Every convolutional layer in the network was followed by ReLU and BN combinations for activations and normalization. In Figure 2, the maximum depth (Max-Depth) of the network is outlined by a dashed bounding box. The last strided-Conv layer further reduced the feature map size and applied maximum filters. Therefore, in Max-Depth, the final feature map size was minimal (50 × 50) with the maximum number of channels (256) in the network. The CNNs possess better semantic information in the Max-Depth area of the network (Kreso et al., 2017). Thus, to better learn the semantics of the image, more convolution layers (03) were used in the Max-Depth of the network.

The feature map size of spatial features in Max-Depth was increased using transposed convolution layers (Tra-Conv). In EFS-Net, three Tra-Conv layers were used to increase the feature map size back to the initial size. Note that EFS-Net avoided having an unpooling layer by using Tra-Conv layers to avoid any kind of spatial loss. Unlike pooling and unpooling layers, both strided and Tra-Conv layers are trainable layers; hence, they both contributed to the learning of the network (Kreso et al., 2017). The last Tra-Conv layer provided the feature to the next convolution layer for further feature empowerment. The last convolution layer applied two filters and, for the softmax function, provided two channels to match the number of classes (02). Finally, the dice pixel classification layer (PCL) classified each pixel into fish and non-fish (background) classes. Notably, more convolution layers were used in the initial stage and Max-Depth sections because these stages are crucial for effective learning. Despite showing promising performance, EFS-Net exhibited some performance degradation when segmenting indistinct small fish with similar backgrounds (Detailed quantitative and qualitative results are provided in subsections 4.4–4.7)

3.3 Fish segmentation using MFAS-net

MFAS-Net is the final network of this study. MFAS-Net uses the architectural base of EFS-Net and was developed to further increase the segmentation accuracies for underwater fish segmentation. As stated, EFS-Net exhibited certain performance limitations in segmenting obscure small-sized fish. To overcome this problem and enhance the overall segmentation performance, MFAS-Net introduced the IFRT-block with a multi-level feature accumulation scheme. MFAS-Net architectural design is presented in Figure 3. Layers close to the input image contain potential spatial features (Hosseinzadeh Kassani et al., 2022). In the IFRT-Block, the spatial features are further refined and transferred to the deeper layers of the network. IFRT block is based on three convolution layers with one feature accumulation (FA-1) junction. Initial low-level spatial features from four different points are concatenated in FA-1. The IFRT-Block transfers processed initial features to FA-4 using two convolution layers.

In FA-2, the downsampled spatial features are concatenated with the transferred features from the high stride (stride = 4) skip path. The first strided-Conv layer of this skip path uses a stride of 4 to reduce the feature map size and its output is provided to FA-2 through a convolution layer. This accumulation of features from different levels and scales allows diverse learning and helps improve the segmentation performance. Subsequently, features from the first high strided-Conv layer are input to the second strided-Conv layer where the feature map size is further reduced and its output is provided to FA-3. In FA-3, direct features from the second strided-Conv layer and features from a convolution layer are concatenated with the final downsampled features in Max-Depth. In CNNs, deeper layers are more effective for preserving image semantics (Kreso et al., 2017). FA-3 lies in Max-Depth; therefore, it has a vital impact on improving segmentation performance.

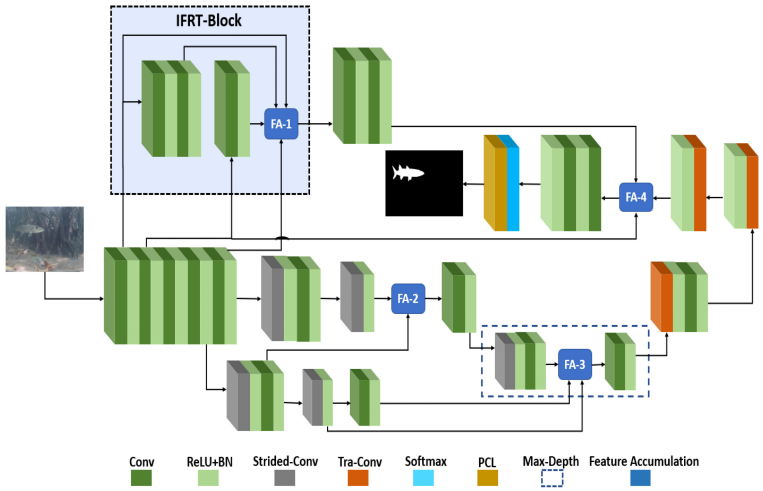


FIGURE 3
Proposed MFAS-Net architecture.

After the feature map is up-sampled back to the initial size, final feature accumulation (FA-4) is applied. In FA-4, the up-sampled accumulated features (F_{us}) are concatenated with the direct initial spatial features (F_{is}) and the initial refined features (F_{ir}) from the IRFT-Block. The feature accumulation process can be further explained using the diagram in Figure 4. After feature accumulation, FA (m) in the IFRT-Block transfers F_{ir} for final accumulation. F_{is} is the initial spatial feature that originated from the early stage of the network to be concatenated in FA-4 with identity mapping. At Max-Depth of the network, after feature accumulation FA (n), the accumulated features from Max-Depth are up-sampled. Additionally, F_{us} from the last Tra-Conv layer is provided to FA (p) for final feature accumulation as follows.

$$F_{fp} = F_{ir} \copyright F_{is} \copyright F_{us}, \tag{1}$$

where F_{fp} denotes the final feature for prediction and is generated by concatenating three features. The F_{fp} is provided to two convolution layers for final feature extraction before prediction and the feature change because of convolution is indicated by ∇. Finally, a prediction mask is generated by the pixel classification layer based on ∇F_{fp} .

$$\nabla F_{fp} = \nabla (F_{ir} \copyright F_{is} \copyright F_{us}), \tag{2}$$

This feature accumulation from different levels of the network empowers the feature and helps enhance the segmentation performance. In addition, the feature accumulation scheme enables the network to minimize the vanishing gradient problem and maintain a better pixel-wise prediction even for small-sized unclear fish images. The configuration details of layers of MFAS-Net are presented in Table 1

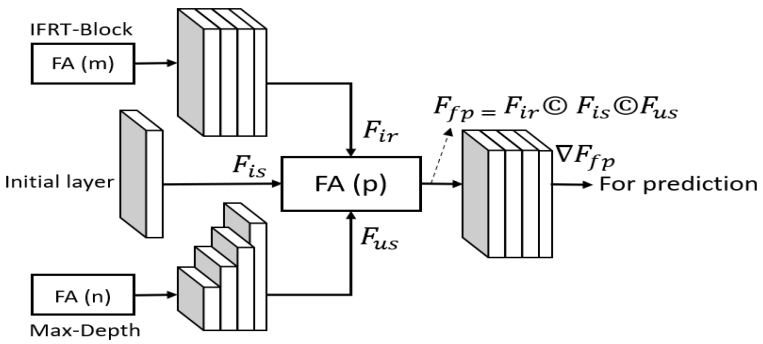


FIGURE 4
Schematic illustrating feature accumulation.

TABLE 1 Layers configurational details of MFAS-Net with the number of required trainable parameters (Resized feature map of 400 × 400 is used to show the configuration of the layers) (Str-Conv, Strided convolution layer; Tra-Conv ,Transposed convolution layer; IB, IFRT-Block; SP, Skip path; Conv, Convolution).

Name of layer	Size	No. of filters	Output features (width × height × number of channels)	Parameters
Conv 1 + ReLU 1	3 × 3 × 32	32	400 × 400 × 32	896
BN 1	–	–		64
Conv 2 + ReLU 2	3 × 3 × 32	32		9248
BN 2	–	–		64
Conv 3 + ReLU 3	3 × 3 × 32	32		9248
BN 3	–	–		64
Conv 4 + ReLU 4	3 × 3 × 32	32		9248
BN 4	–	–		64
Conv 5 + ReLU 5	3 × 3 × 32	32		9248
BN 5	–	–		64
Conv 1-IB + ReLU 1-IB	3 × 3 × 32	32		9248
BN 1-IB	–	–		64
Conv 2-IB + ReLU 2-IB	3 × 3 × 32	32		9248
BN 2-IB	–	–		64
Conv 3-IB + ReLU 3-IB	3 × 3 × 64	64	400 × 400 × 64	18496
BN 3-IB	–	–		128
Conv 1-SP + ReLU 1-SP	3 × 3 × 32	32	400 × 400 × 32	46112
BN 1-SP	–	–		64
Conv 2-SP + ReLU 2-SP	3 × 3 × 32	32		9248
BN 2-SP	–	–		64
Str-Conv 3-SP + ReLU 3-SP	3 × 3 × 64	64	100 × 100 × 64	18496
BN 3-SP	–	–		128
Conv 4-SP + ReLU 4-SP	3 × 3 × 128	128	100 × 100 × 128	73856
BN 4-SP	–	–		256
Str-Conv 5-SP + ReLU 5-SP	3 × 3 × 128	128	50 × 50 × 128	147584
BN 5-SP	–	–		256
Conv 6-SP + ReLU 6-SP	3 × 3 × 128	128		147584
BN 6-SP	–	–		256
Str-Conv 1 + ReLU 1-S	3 × 3 × 64	64	200 × 200 × 64	18496
BN 1-S	–	–		128
Conv 6 + ReLU 6	3 × 3 × 64	64		36928
BN 6	–	–		128
Str-Conv 2 + ReLU 2-S	3 × 3 × 128	128	100 × 100 × 128	73856
BN 2-S	–	–		256
Conv 7 + ReLU 7	3 × 3 × 128	128		295040
BN 7	–	–		256
Str-Conv 3 + ReLU 3-S	3 × 3 × 256	256	50 × 50 × 256	295168
BN 3-S	–	–		512
Conv 8 + ReLU 8	3 × 3 × 256	256		590080
BN 8	–	–		512
Conv 9 + ReLU 9	3 × 3 × 256	256		1179904
BN 9	–	–		512
Tra-Conv 1 + ReLU 1-T	3 × 3 × 128	128	100 × 100 × 128	295040
BN 1-T	–	–		256
Conv 10 + ReLU 10	3 × 3 × 128	128		147584
BN 10	–	–		256
Tra-Conv 2 + ReLU 2-T	3 × 3 × 64	64	200 × 200 × 64	73792

(Continued)

TABLE 1 Continued

Name of layer	Size	No. of filters	Output features (width × height × number of channels)	Parameters
BN 2-T	–	–		128
Tra-Conv 3 + ReLU 3-T	3 × 3 × 32	32	400 × 400 × 32	18464
BN 3-T	–	–		64
Conv 11 + ReLU 11	3 × 3 × 32	32		27680
BN 11	–	–		64
Conv 12 + ReLU 12	3 × 3 × 2	2	400 × 400 × 2	578
BN 12	–	–		4
		Total number of required trainable parameters		3,575,046

4 Results

4.1 Experimental data and environment

In this study, both proposed networks were evaluated using two underwater fish segmentation-related datasets, DeepFish (Saleh et al., 2020) and SUIM (Islam et al., 2020). DeepFish is a large-scale fish dataset containing a total of 40,000 images from different marine habitats. The purpose of developing the DeepFish dataset was to monitor the fish dynamics along with shape and size assessments. This dataset is divided into three categories: FishLoc (fish location), FishClf (fish classification), and FishSeg (fish segmentation). In FishLoc, point-level annotations are provided to specify the location of fish. FishClf includes the classification labels. Last, FishSeg has pixel-level expert annotations for the segmentation of fish. In this study, we focused on segmentation; therefore, only FishSeg was used in our experiments. The FishSeg data split contains 310 training, 124 validation, and 186 testing images. FishSeg is a challenging subset of DeepFish because it consists of samples with different shapes, sizes, illumination conditions, and backgrounds. Examples from DeepFish (FishSeg) are shown in Figure 5.

SUIM dataset contains a total of 1525 underwater images of multiple categories, including fish. SUIM provides a test split with separate annotations for each category. Therefore, in the experiments, we worked with fish and other vertebrate categories and used this data for fish segmentation. Sample images from the SUIM dataset are presented in Figure 6.

The fish segmentation experimental work was performed using a desktop computer Intel ® Core™ i7 CPU950@3.7 GHz with 32 GB of RAM and an NVIDIA GeForce GTX 1070 GPU with 8 GB of graphical memory (GeForce GTX 1070, accessed on January 25, 2022). MATLAB 2020b (MATLAB R2020b, accessed on January 25, 2022) was used for the architectural development, training, and testing of the networks.

4.2 Training proposed networks

Both networks, EFS-Net and MFAS-Net, were trained on DeepFish and SUIM datasets separately. The training images

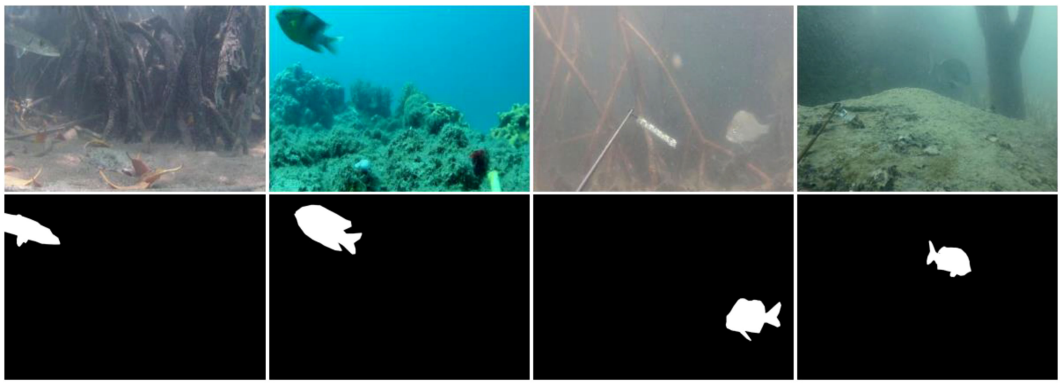


FIGURE 5
DeepFish (FishSeg) sample images with corresponding ground truth images. (The white pixels in the ground truth images belong to fish).

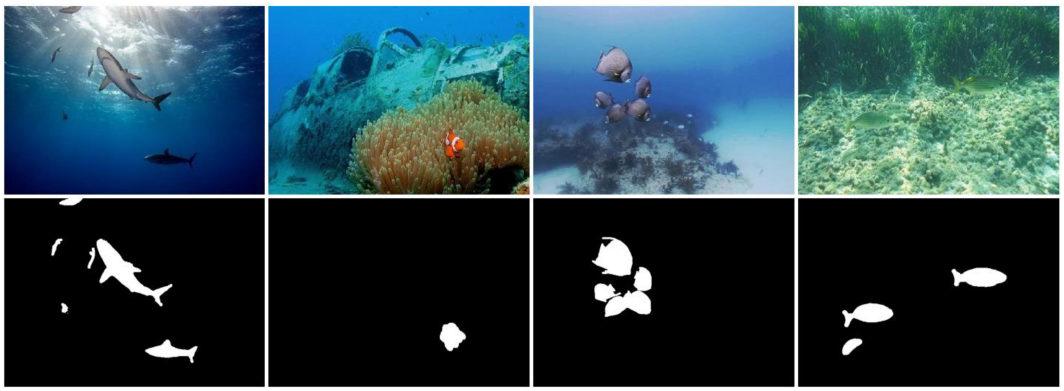


FIGURE 6
SUIM sample images with corresponding ground truth images. (The white pixels in the ground truth images belong to fish).

from both datasets were resized to 400 × 400 for time-efficient training. The image resizing used nearest-neighbor interpolation. The data splits of both datasets were defined by the dataset providers. Proposed methods are evaluated with the same data splits used by the previous research (Laradji et al., 2021) for a fair comparison. Overfitting was avoided using early stopping and data augmentation. The Adam optimizer was used because of its fast convergence and data handling capabilities (Kingma and Ba, 2014). The training loss and accuracy plots of MFAS-Net are presented in Figure 7, which shows high training accuracy with progressively decreasing loss.

Underwater imaging has numerous challenging aspects, such as visual limitations, logistic hurdles, and equipment costs; thus, only limited underwater data is available for experimentation. In addition, the annotation of images requires special resources. Augmentation, which transforms data through numerous operations, is widely used to overcome data limitations. To do so, we used different arithmetic and geometric operations including image flipping, translation, cropping, and random rotation for data augmentation. These geometric and arithmetic operations help in producing a wide variety of training images and contribute to effective learning of the network. MATLAB (MATLAB R2020b, accessed on January 25, 2022) functions and commands are used for the above-mentioned image operations. In addition, dice loss (L_d) (Drozdzal et al., 2018) was used in our experiments. L_d is among the top loss functions used in segmentation tasks to

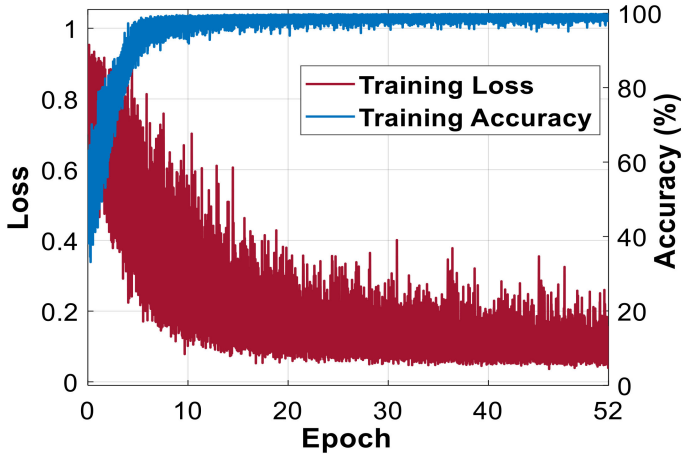


FIGURE 7
Training accuracy and loss plot.

guide the network for an effective learning throughout the training process. L_d is widely used in segmentation tasks because it covers the class imbalance problems, improves the segmentation performance, and minimizes the metric during network backpropagation (Drozdzal et al., 2018). Mathematically, L_d is expressed as:

$$L_d = -\frac{2\sum_i f_i \quad k_i + 1}{\sum_i f_i + \sum_i k_i + 1}, \tag{3}$$

where the ground truth label is represented by k_i; $k_i \in \{0,1\}$, whereas, f_i refers to the network's final output and $f_i \in \{0,1\}$. In equation (3), constant value in the denominator is included for smoothing (Drozdzal et al., 2018).

4.3 Testing of proposed networks

Both proposed models were tested on two publicly available datasets, DeepFish and SUIM. For evaluation, the trained networks were applied to the testing images. The networks generate a binary prediction mask that was compared with the ground truth image for final results. In both datasets, the ground truth images have binary pixel values. As shown in Figures 5 and 6, desired region (fish) is annotated in white (pixel value = 255), whereas non-fish (the background) is represented in black (pixel value = 0). Moreover, the network generates a binary prediction mask for pixel-wise comparison with the ground truth image. Mean intersection-over-union (mIoU) (Zhang et al., 2022) was used to evaluate the proposed methods. mIoU is widely used and accepted for the evaluation of segmentation tasks. The mathematical expression for mIoU is as follows

$$\text{mIoU} = \frac{tp}{tp + fp + fn}, \tag{4}$$

True positive (tp) pixels are those pixels that are correctly predicted as the desired class by the proposed network. The pixels that, according to ground truth, belong to the desired class but the proposed network incorrectly predicts as being of an undesired class, are categorized as false negative (fn) . If, according to the ground truth image, those pixels belonging to an undesired class but marked as being of the desired class are categorized as false positive (fp) .

4.4 Comparison of results between EFS-net and MFAS-net for fish segmentation on the DeepFish dataset (ablation studies)

EFS-Net and MFAS-Net were both evaluated on the DeepFish dataset for fish segmentation. The DeepFish dataset contains several challenging cases, such as small-sized, indiscernible, and multiple adjacent fish. Nonetheless, the proposed architectures exhibited promising performance. Table 2 presents numerical results that confirm the superior performance of MFAS-Net compared to EFS-Net. The IFRT-Block and multi-level feature accumulation enabled MFAS-Net's better segmentation performance.

The comparative qualitative segmentation results obtained by EFS-Net and MFAS-Net are shown in Figure 8 (rows 1-4) and show promising segmentation performance by both proposed architectures even in the challenging cases of segmenting small-sized and indiscernible fish. In addition, the visual results confirm the effectiveness of MFAS-Net compared with EFS-Net. MFAS-Net exhibits better performance because of the initial feature refinement and feature accumulation schemes. Segmented fish is detected based on different labels using the further processing of component labeling. In Figure 8 (row 5), the example of segmentation-based fish detection and counting result is shown. Despite the challenging case of small and unclear fish, MFAS-Net detects all the fish, and provides the accurate information of fish position and counting. On the other hand, EFS-Net could not detect small fish because of its blurred structure, and presented an inaccurate information of fish counting. Figure 9 shows that the poor visual segmentation results obtained by EFS-Net and MFAS-Net can be attributed to barely discernible fish having little contrast against the background. Nevertheless, compared to EFS-Net, MFAS-Net still exhibits a better segmentation performance.

4.5 Comparison of the segmentation results with state-of-the-art methods on the DeepFish dataset

To confirm the effectiveness of the proposed methods, the segmentation performances of both methods were compared with state-of-the-art methods. The state-of-the-art methods were fine-tuned with the two experimental datasets which

TABLE 2 Comparison between EFS-Net and MFAS-Net's numerical results on the DeepFish dataset.

Methods	Background IoU	Foreground IoU	mIoU
EFS-Net (Proposed base)	98.94	82.77	90.85
MFAS-Net (Proposed final)	99.15	84.86	92.005

(Results are reported in percentages).

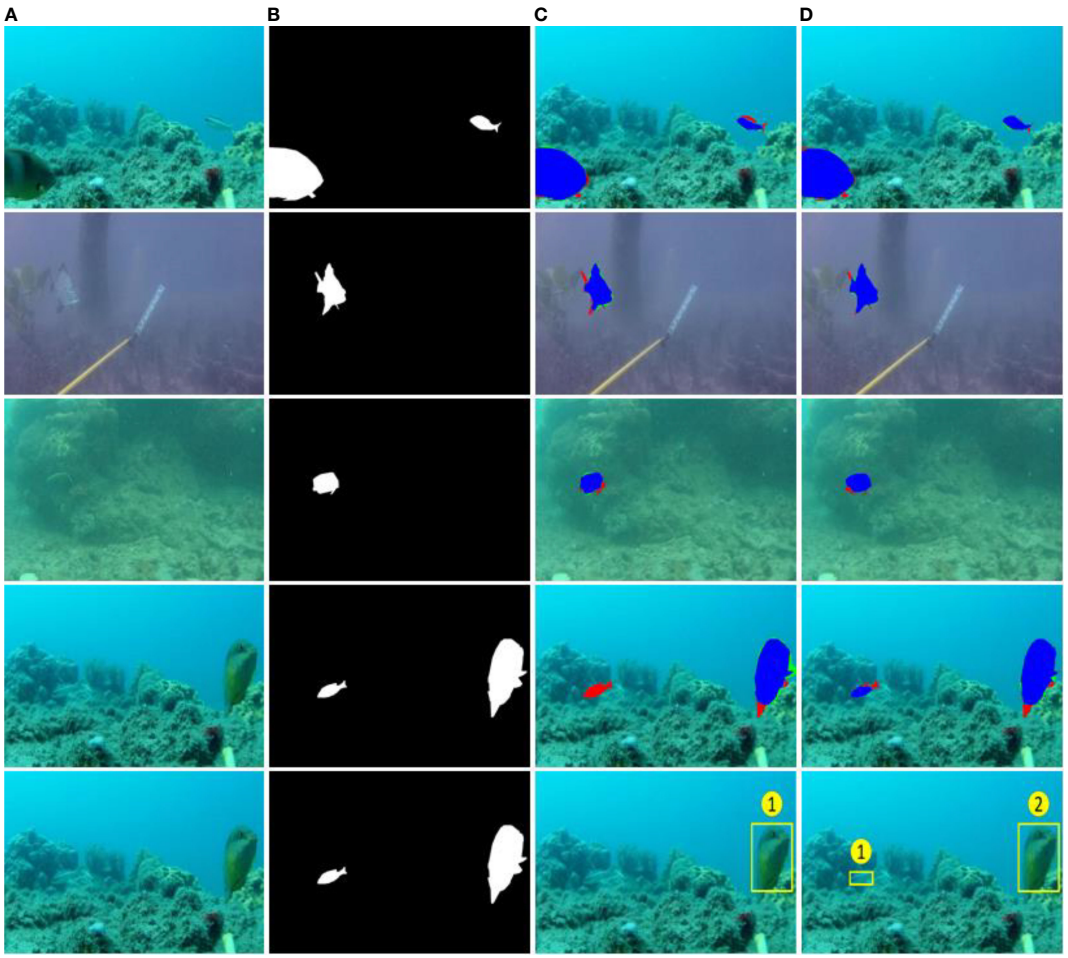


FIGURE 8
Good qualitative results by EFS-Net and MFAS-Net on the DeepFish dataset. Rows 1-4: Segmentation results; **(A)** original test image, **(B)** ground truth image, **(C)** segmented images achieved by applying EFS-Net, and **(D)** segmented images achieved by MFAS-Net (the fp and tp pixels are referred by green and blue colors, respectively. Pixel representation for fn is done with red color). Row 5: segmentation-based fish detection and counting.

were also used for the fine-tuning of our method for the fair comparisons. The comparative numerical results presented in Table 3 reveal that EFS-Net achieved a competitive performance, whereas MFAS-Net outperformed the state-of-the-art methods. These results were obtained without any pre-processing or applying biases during training. The results obtained by MFAS-Net validate the effectiveness of MFAS-Net architecture.

4.6 Comparison of results between EFS-net and MFAS-net for fish segmentation on SUIM dataset (ablation studies)

Both proposed networks were further tested on the SUIM dataset to validate the proposed methods' robustness for these tasks. The SUIM dataset is considerably different from the DeepFish dataset. Compared with the DeepFish dataset, fish in SUIM dataset differ in types, sizes, illumination effects, and backgrounds. However, both networks in general, and MFAS-Net in particular, achieve superior segmentation results. Comparative numerical results using EFS-Net and MFAS-Net presented in Table 4 further confirm better performance by MFAS-Net over EFS-Net. The better performance by MFAS-Net can be attributed to the accumulation of low-level initial features with the deep layers feature that help to improve underwater segmentation performance. In Figure 10 (rows 1-4), good segmentation visual results are presented to compare the segmentation performance of MFAS-Net with EFS-Net on the SUIM dataset. The visual results confirm that MFAS-Net produces better segmentation results than EFS-Net. In an underwater environment, segmentation is challenging because of the visual implications, fish sizes, background effects, and

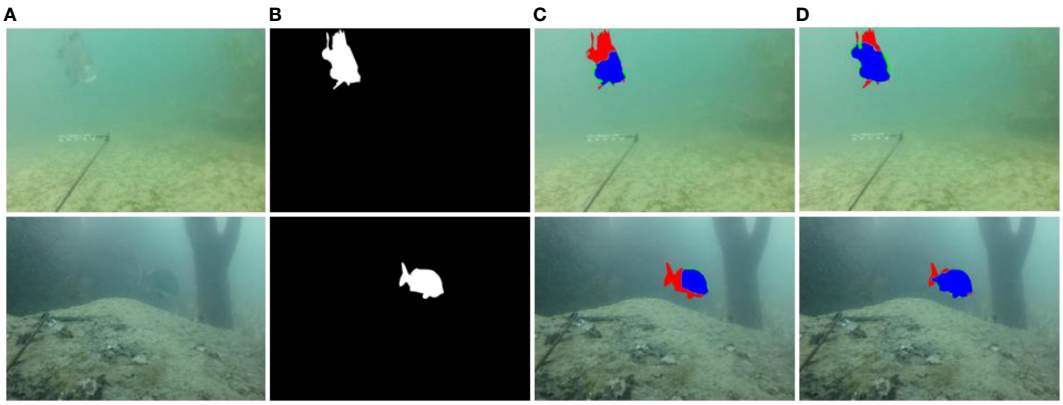


FIGURE 9
Poor segmentation sample qualitative results by EFS-Net and MFAS-Net on the DeepFish dataset **(A)** original test image, **(B)** ground truth image, **(C)** segmented images achieved by applying EFS-Net, and **(D)** segmented images achieved by MFAS-Net (the fp and tp pixels are referred by green and blue colors, respectively. Pixel representation for fn is done with red color).

indistinct object boundaries. However, MFAS-Net delivers a good segmentation performance using its effective architecture. Component labeling is applied to detect the fish in segmented images. The sample result for segmentation-based fish detection with the count is shown in Figure 10 (row 5). MFAS-Net accurately detects the challenging small-sized fish, and presents the correct information of fish counting. However, EFS-Net fails to detect the smaller fish positioned at the lower-left side of the image, and presents the false information of fish counting, consequently. Figure 11 exhibits poor segmentation caused by indistinct outlines. Nonetheless, MFAS-Net still delivers better performance than EFS-Net.

4.7 Comparison of segmentation results with state-of-the-art methods on SUIM dataset

In Table 5, the proposed methods' numerical results on the SUIM dataset are compared with those of the state-of-the-art

TABLE 3 Comparative results by the proposed methods with state-of-the-art methods on the DeepFish dataset. (Results are reported in percentages).

Methods	Background IoU	Foreground IoU	mIoU
SUIM-Net (Islam et al., 2020)	99.03	78.40	88.71
SegNet (Badrinarayanan et al., 2017)	98.89	68.94	83.91
DeepLab-v3 (Chen et al., 2017)	99.11	71.35	85.23
PSPNet (Zhao et al., 2017)	99.15	72.61	85.88
FCN (Long et al., 2015)	99.21	66.30	82.75
CCNet (Huang et al., 2019)	99.12	78.36	88.74
DANet (Fu et al., 2019)	99.02	78.87	88.95
GANet (Zhang et al., 2019)	99.16	79.34	89.25
DRANet (Fu et al., 2021)	99.33	79.42	89.37
SPNet (Hou et al., 2020)	99.21	80.64	89.92
SANet (Zhang and Yang, 2021)	99.27	80.71	89.99
HANet (Choi et al., 2020)	99.25	81.37	90.31
DGCNet (Zhang et al., 2020)	99.21	81.42	90.32
GFFNet (Li et al., 2020)	99.20	81.49	90.35
DPANet (Zhang et al., 2022)	99.31	82.86	91.08
A-LCFCN+PM (Laradji et al., 2021)	99.3	73.0	86.2
EFS-Net (Proposed base)	98.94	82.77	90.85
MFAS-Net (Proposed final)	99.15	84.86	92.005

TABLE 4 Comparison of EFS-Net and MFAS-Net's numerical results on the SUIM dataset.

Methods	mIoU
EFS-Net (Proposed base)	75.56
MFAS-Net (Proposed final)	76.42

(Results are reported in percentages).

methods. The state-of-the-art methods were fine-tuned with the two experimental datasets which were also used for the fine-tuning of our method for the fair comparisons. The results reveal that both the proposed methods performed better than the other methods. In particular, MFAS-Net achieved the highest mIoU score owing to its effective architecture. Additionally, these results were achieved without pre-processing or excessive computational overheads while requiring only a small number of trainable parameters (details are given in subsection 5.1)

5 Discussion

In an underwater environment, effectively segmenting objects is challenging because the image quality is degraded. Most of the underwater images contain obscure objects with indistinct boundaries. This is exacerbated with fish because they are usually moving and sometimes the contrast with the background is too low to distinguish them. In addition, the marine environment has several other animals and objects in the background, which can mislead the neural network. Thus, achieving good segmentation performance in such challenging

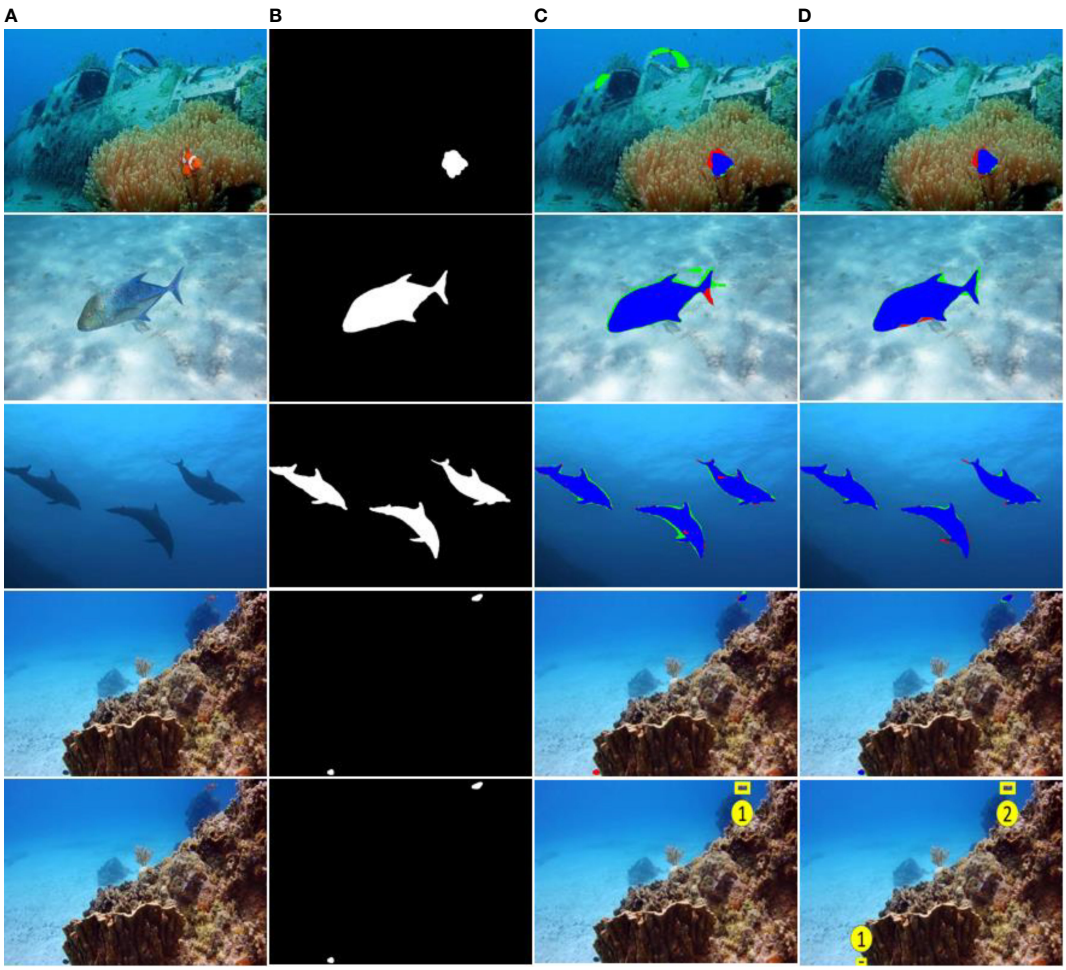


FIGURE 10
Good qualitative results by EFS-Net and MFAS-Net on the SUIM dataset. Rows 1-4: Segmentation results; **(A)** original test image, **(B)** ground truth image, **(C)** segmented images achieved by applying EFS-Net, and **(D)** segmented images achieved by MFAS-Net (the fp and tp pixels are referred by green and blue colors, respectively. Pixel representation for fn is done with red color). Row 5: segmentation-based fish detection and counting.

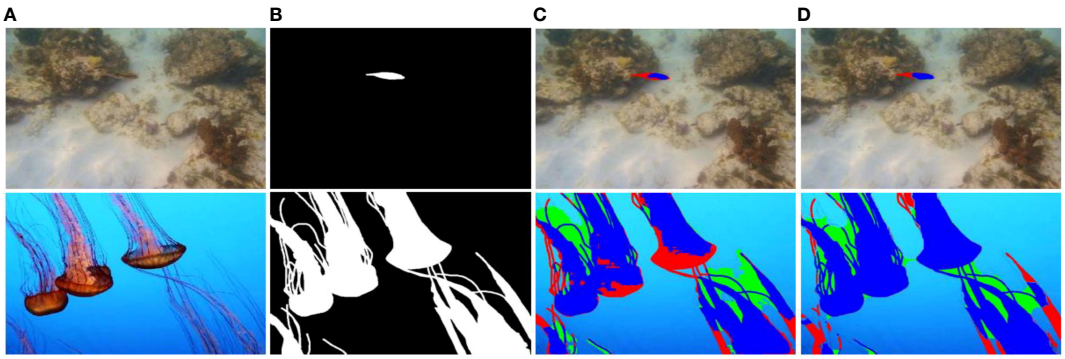


FIGURE 11
Poor segmentation visual results by EFS-Net and MFAS-Net on the SUIM dataset **(A)** original test image, **(B)** ground truth image, **(C)** segmented images achieved by applying EFS-Net, and **(D)** segmented images achieved by MFAS-Net (the fp and tp pixels are referred by green and blue colors, respectively. Pixel representation for fn is done with red color).

conditions is difficult. Despite these factors, the proposed networks demonstrated highly accurate segmentation However, as shown in the visual results, occasionally EFS-Net could not detect small-sized fish in murky conditions. Similarly, EFS-Net showed relatively a low performance in the case of fish with indistinct outlines. This is possibly because of the task difficulty, such as unclear objects and indistinct object outlines. Nevertheless, MFAS-Net overcomes these challenges by refining and transferring potential low-level information and finally accumulating the features at different levels of the network.

Proposed method can provide the pixel-wise segmentation result, and after the further processing of component labeling, the results of fish detection and counting can also be achieved. However, proposed method has the limitations for detection with the cases of overlapped fish in 'crowded' scenarios.

Our method focuses on the pixel-wise segmentation of fish in an underwater environment. Many latest researches have worked with the pixel-wise segmentation of fish because it can provide the detailed morphological measurements such as fish size and shape. In previous research (Laradji et al., 2021), the size and shape of fish are accounted as the main measure for fish habitat monitoring because this information can be used to assess the fish growth and can also be considered as the reference for feeding. Semantic segmentation offers pixel-wise prediction, therefore it can provide accurate information about the size, area, and shape of fish. However, this information cannot be accurately obtained by detection-based method because it can provide only the roughly detected box area including the object or instance. United nation (UN) Food and Agriculture findings state that 33% of important marine fish stocks are overfished, worldwide (FAO, 2018). Moreover, 11% of total fish are discarded because of undersized catching (Pe´rez Roda et al., 2019). Fish size assessment through segmentation can also help in avoiding the catching of undersized fish (Laradji et al., 2021). Similarly, another study (Zhang et al., 2022) reports that automatic measurement of fish data using semantic segmentation can be very helpful for the sustainable development of marine fisheries. Dataset provider of DeepFish (Saleh et al., 2020) refers that the segmentation labels are helpful to analyze the fish habitat by estimating the fish size and shape.

In short, the segmentation of fish has a significant impact on the morphological assessment of fish, and it can be very useful for fish monitoring and the sustainable development of fisheries. Due to these reasons, many state-of-the-art methods have researched about the pixel-wise semantic segmentation of fish as shown in Tables 3 and 5.

TABLE 5 Comparison between EFS-Net and MFAS-Net's numerical results on the SUIM dataset.

Methods	mIoU
SegNet (Badrinarayanan et al., 2017)	69.23
DeepLab-v3 (Chen et al., 2017)	62.72
LCFCN (Laradji et al., 2018)	59.8
PL-FCN (Bearman et al., 2016)	68.3
A-LCFCN+PM (Laradji et al., 2021)	74.9
EFS-Net (Proposed base)	75.56
MFAS-Net (Proposed final)	76.42

(Results are reported in percentages).

5.1 Comparison of computational efficiency

Computational requirements of the networks are equally important in a framework. Numerous deep learning-based architectures are capable of achieving a competitive result, but they require a large number of trainable parameters that render the framework computationally inefficient. The architecture proposed in this study exhibits outstanding results without

TABLE 6 Comparison between other methods and MFAS-Net's trainable parameters requirement.

Methods	The number of parameters
FCN8 CNN (Long et al., 2015)	69.74 M
PSPNet MobileNet (Zhao et al., 2017)	63.96 M
DeepLab-v3 (Chen et al., 2017)	41.25 M
SegNet (Badrinarayanan et al., 2017)	29.4 M
SUIM-Net (Islam et al., 2020)	3.86 M
MFAS-Net (Proposed final)	**3.57 M**

M, Million.

compromising computational efficiency. Proposed networks are designed in such a way that it provides a high segmentation performance requiring a small number of trainable parameters for their complete training. In a CNN, Max-Depth is the most computationally expensive part of the network because of the maximum number of channels operating in it. As shown in Figure 3, MFAS-Net uses less number of convolutional layers (only 2) in Max-Depth to reduce the parameters requirement. In addition, as presented in Table 1, MFAS-Net also uses a maximum number of 256 channels to contain the parameters. In Table 6, the required trainable parameters are compared, which, with its low computational overhead, confirm the superior performance of MFAS-Net. Requiring only 3.57 million (M) parameters for complete training, MFAS-Net is the lowest compared with the other methods. Modern research trends encourage robust models that can achieve higher accuracies with lower computational overheads. However, models have different architectural designs, performance limitations, and computational requirements. Proposed methods achieve superior performance without compromising computational efficiency because of their customized efficient network designs.

5.2 Analysis with class activation maps

In supervised learning, CNNs generally learn from training data and evaluation is performed on the unseen testing set. In CNNs, learning is a black box; however, this progressive learning can be visually interpreted using heat activation maps (Selvaraju et al., 2017). During learning, the main features considered by the CNN can be analyzed using activation maps. Figure 12 shows heat activation maps extracted from different stages of the architecture. This visual interpretation confirms that MFAS-Net progressively learns from fish images. Additionally, it confirms that MFAS-Net primarily considers the features of the desired class (fish) without biases.

6 Conclusion

Fish are vital to the maintenance of sustainable marine environments. Therefore, fish monitoring and assessment are essential for managing resources in the marine ecosystem. Conventionally, the manual methods employed for fish observation and assessment are time-consuming, labor-intensive, and error-prone. To meet this need, two novel architectures were developed for high-performance fish segmentation. In an underwater environment, achieving high segmentation accuracy is challenging because of the visual limitations underwater. However, both proposed methods exhibited promising results with superior computational efficiency. EFS-Net is a shallow architecture with a series of convolution layers at the early stage of the network for optimal low-level feature extraction. MFAS-Net is the final network and refines valuable initial features and transfers them to the deep stage of the network for feature accumulation. As shown in the ablation study, feature refinement and accumulation enable MFAS-Net to perform better segmentation even for the challenging barely discernible small-sized fish with indistinct outlines. Furthermore, multi-level feature accumulation improves the overall learning of the network and produces enhanced segmentation performance. Both proposed networks, EFS-Net and MFAS-Net, were evaluated on two publicly available databases. The proposed methods outperformed state-of-the-art methods with a small number of required trainable parameters (3.57 million). The fish detection and

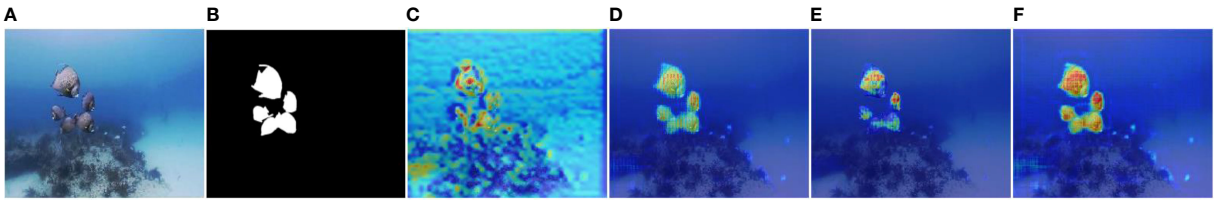


FIGURE 12
Visual depiction of progressive learning by MFAS-Net using activation maps taken from different stages of the architecture. **(A)** original image. **(B)** Ground truth image. Heat activation maps are obtained from the **(C)** ReLU 2, **(D)** ReLU 6-SP, **(E)** ReLU 8, and **(F)** ReLU 12 layers of the MFAS-Net, as given in Table 1.

counting based on segmentation and further processing of component labeling cannot adequately work when there are overlapped instances. This is a limit of the proposed method, i.e., it is not able to detect instances in 'crowded' scenarios. In future work, we would research the method to detect instances in 'crowded' scenarios. In addition, we would optimize the networks further, and perform the instance segmentation for fish detection.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://alzayats.github.io/DeepFish/ https://irvlab.cs.umn.edu/resources/suim-dataset.

Ethics statement

Ethical review and approval was not required for the animal study because in this research, we used two open databases of DeepFish and SUIM which can be freely used for academic purpose. Therefore, ethical review and approval was not required for this study.

Author contributions

AH designed the method with experiments and wrote the first draft of the manuscript supervised by KRP. MA, JC, and HS performed the data collection and verified the experimental results. KRP critically reviewed the initial manuscript and provided helpful input. All authors contributed to the article and approved the submitted version.

Funding

This research was supported in part by the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (MSIT) through the Basic Science Research Program (NRF-2021R1F1A1045587), in part by the NRF funded by the MSIT through the Basic Science Research Program (NRF-2022R1F1A1064291), and in part by the MSIT, Korea, under the ITRC (Information Technology Research Center) support program (IITP-2022-2020-0-01789) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdeldaim, A. M., Sahlol, A. T., Elhoseny, M., and Hassanien, A. E. (2018) Computer-aided acute lymphoblastic leukemia diagnosis system based on image analysis. In: *Advances in soft computing and machine learning in image processing studies in computational intelligence* (Accessed 28 March 2022).

Abe, S., Takagi, T., Torisawa, S., Abe, K., Habe, H., Iguchi, N., et al. (2021). Development of fish spatio-temporal identifying technology using segnet in aquaculture net cages. *Aquac. Eng.* 93, 102146. doi: 10.1016/j.aquaeng.2021.102146

Arsalan, M., Haider, A., Choi, J., and Park, K. R. (2022a). Detecting blastocyst components by artificial intelligence for human embryological analysis to improve success rate of in vitro fertilization. *J. Pers. Med.* 12, 124. doi: 10.3390/jpm12020124

Arsalan, M., Haider, A., Choi, J., and Park, K. R. (2022b). Diabetic and hypertensive retinopathy screening in fundus images using artificially intelligent shallow architectures. *J. Pers. Med.* 12, 7. doi: 10.3390/jpm12010007

Arsalan, M., Haider, A., Koo, J. H., and Park, K. R. (2022c). Segmenting retinal vessels using a shallow segmentation network to aid ophthalmic analysis. *Mathematics* 10, 1536. doi: 10.3390/math10091536

Arsalan, M., Haider, A., Won Lee, Y., and Ryoung Park, K. (2022d). Detecting retinal vasculature as a key biomarker for deep learning-based intelligent screening and analysis of diabetic and hypertensive retinopathy. *Expert Syst. Appl.* 200, 117009. doi: 10.1016/j.eswa.2022.117009

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 2481–2495. doi: 10.1109/TPAMI.2016.2644615

Bearman, A., Russakovsky, O., Ferrari, V., and Fei-Fei, L. (2016). "What's the point: Semantic segmentation with point supervision," in *Proceedings of the European conference on computer visio* (Amsterdam, Netherlands: Springer), 549–565.

Chen, L.-C., Papandreou, G., Schroff, F., and Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv*. 1–14. doi: 10.48550/arXiv.1706.05587

Choi, S., Kim, J. T., and Choo, J. (2020). "Cars can't fly up in the sky: Improving urban-scene segmentation *via* height-driven attention networks," in *Proceedings of the Computer Vision and Pattern Recognition*, Seattle, Online, USA, 2020 June 16-18. (New York, U.S: IEEE) 9373–9383.

Cui, Y., Chen, R., Chu, W., Chen, L., Tian, D., Li, Y., et al. (2022). Deep learning for image and point cloud fusion in autonomous driving: A review. *IEEE Trans. Intell. Transp. Syst.* 23, 722–739. doi: 10.1109/TITS.2020.3023541

Ditria, E. M., Lopez-Marcano, S., Sievers, M., Jinks, E. L., Brown, C. J., and Connolly, R. M. (2020). Automating the analysis of fish abundance using object detection: Optimizing animal ecology with deep learning. *Front. Mar. Sci.* 7. doi: 10.3389/fmars.2020.00429

Drozdzal, M., Chartrand, G., Vorontsov, E., Shakeri, M., Di Jorio, L., Tang, A., et al. (2018). Learning normalized inputs for iterative estimation in medical image segmentation. *Med. Image Anal.* 44, 1–13. doi: 10.1016/j.media.2017.11.005

FAO (2018). *The state of world fisheries and aquaculture. meeting the sustainable development goals* (Rome, Italy: Food and Agriculture Organization of the United Nations). Available at: http://www.fao.org/3/i9540en/I9540EN.pdf.

Fu, J., Liu, J., Jiang, J., Li, Y., Bao, Y., and Lu, H. (2021). Scene segmentation with dual relation-aware attention network. *IEEE Trans. Neural Netw. Learn. Syst.* 32, 2547–2560. doi: 10.1109/TNNLS.2020.3006524

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., et al. (2019). "Dual attention network for scene segmentation. in 3146–3154," in *Proceedings of the Computer Vision and Pattern Recognition*, California, USA. (New York, U.S: IEEE) 2019 June 17-20. 3146–3154.

Haider, A., Arsalan, M., Lee, M. B., Owais, M., Mahmood, T., Sultan, H., et al. (2022a). Artificial intelligence-based computer-aided diagnosis of glaucoma using retinal fundus images. *Expert Syst. Appl.* 207, 117968. doi: 10.1016/j.eswa.2022.117968

Haider, A., Arsalan, M., Lee, Y. W., and Park, K. R. (2022b). Deep features aggregation-based joint segmentation of cytoplasm and nuclei in white blood cells. *IEEE J. Biomed. Health Inform.*, 26 1–1. doi: 10.1109/JBHI.2022.3178765

Hosseinzadeh Kassani, S., Hosseinzadeh Kassani, P., Wesolowski, M. J., Schneider, K. A., and Deters, R. (2022). Deep transfer learning based model for colorectal cancer histopathology segmentation: A comparative study of deep pre-trained models. *Int. J. Med. Inf.* 159, 104669. doi: 10.1016/j.ijmedinf.2021.104669

Hou, Q., Zhang, L., Cheng, M.-M., and Feng, J. (2020). "Strip pooling: Rethinking spatial pooling for scene parsing," in *Proceedings of the Computer Vision and Pattern Recognition*, Seattle, Online, USA. (New York, U.S: IEEE) 2020 June 16-18. 4003–4012.

Huang, Z., Wang, X., Huang, L., Huang, C., Wei, Y., and Liu, W. (2019). "CCNet: Criss-cross attention for semantic segmentation," in *Proceedings of the International Conference on Computer Vision*, Seoul, Korea. (New York, U.S: IEEE) 2019 Oct-27 to Nov-2. 603–612.

Islam, M. J., Edge, C., Xiao, Y., Luo, P., Mehtaz, M., Morse, C., et al. (2020). "Semantic segmentation of underwater imagery: Dataset and benchmark," in *Proceeding of the IEEE International Conference on Intelligent Robots and Systems*, Las Vegas, NV, USA. (New York, U.S: IEEE) 2020 Oct-24 to Jan-24. 1769–1776.

Kandimalla, V., Richard, M., Smith, F., Quirion, J., Torgo, L., and Whidden, C. (2022). Automated detection, classification and counting of fish in fish passages with deep learning. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.823173

Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv*. 1–15. doi: 10.48550/arXiv.1412.6980

Kreso, I., Segvic, S., and Krapac, J. (2017). "Ladder-style DenseNets for semantic segmentation of Large natural images," in *Proceedings of the International Conference on Computer Vision*, Venice, Italy. (New York, U.S: IEEE) 2017 October 22-29. 238–245.

Labao, A. B., and Naval, P. C. (2017). "Weakly-labelled semantic segmentation of fish objects in underwater videos using a deep residual network," in *Proceedings of the Asian Conference on Intelligent Information and Database Systems*, Kanazawa, Japan: Springer. (New York, U.S: IEEE) 2017 April 3-5. 255–265.

Laradji, I. H., Rostamzadeh, N., Pinheiro, P. O., Vazquez, D., and Schmidt, M. (2018). "Where are the blobs: Counting by localization with point supervision," in *Proceedings of the European Conference on Computer Vision*, München, Germany. (New York, U.S: IEEE) 2018 September 10-13. 547–562.

Laradji, I., Saleh, A., Rodriguez, P., Nowrouzezahrai, D., Azghadi, M. R., and Vazquez, D. (2020). Affinity LCFCN: Learning to segment fish with weak supervision. *arXiv* 1–10, doi: 10.48550/arXiv.2011.03149

Laradji, I. H., Saleh, A., Rodriguez, P., Nowrouzezahrai, D., Azghadi, M. R., and Vazquez, D. (2021). Weakly supervised underwater fish segmentation using affinity LCFCN. *Sci. Rep.* 11, 17379. doi: 10.1038/s41598-021-96610-2

Liu, F., and Fang, M. (2020). Semantic segmentation of underwater images based on improved deeplab. *J. Mar. Sci. Eng.* 8, 188. doi: 10.3390/jmse8030188

Li, X., Zhao, H., Han, L., Tong, Y., Tan, S., and Yang, K. (2020). "Gated fully fusion for semantic segmentation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, New York, USA, (New York, U.S: IEEE) 2020 February 7-12. 11418–11425.

Long, J., Shelhamer, E., and Darrell, T. (2015). "Fully convolutional networks for semantic segmentation," in *Proceedings of the Computer Vision and Pattern Recognition*, Boston, USA, (New York, U.S: IEEE) 2015 June 7-12. 3431–3440.

Mahmood, T., Cho, S. W., and Park, K. R. (2022a). DSRD-net: Dual-stream residual dense network for semantic segmentation of instruments in robot-assisted surgery. *Expert Syst. Appl.* 202, 117420. doi: 10.1016/j.eswa.2022.117420

Mahmood, T., Kim, S. G., Koo, J. H., and Park, K. R. (2022b). Artificial intelligence-based tissue phenotyping in colorectal cancer histopathology using visual and semantic features aggregation. *Mathematics* 10, 1909. doi: 10.3390/math10111909

Owais, M., Baek, N. R., and Park, K. R. (2021). Domain-adaptive artificial intelligence-based model for personalized diagnosis of trivial lesions related to COVID-19 in chest computed tomography scans. *J. Pers. Med.* 11, 1008. doi: 10.3390/jpm11101008

Pe´rez Roda, M. A., Gilman, E., Huntington, T., Kennelly, S. J., Suuronen, P., Chaloupka, M., et al. (2019). *A third assessment of global marine fisheries discards* (Rome: FAO), 1–58.

Rahnemoonfar, M., and Dobbs, D. (2019). "Semantic segmentation of underwater sonar imagery with deep learning," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, Yokohama, Japan. (New York, U.S: IEEE) 2019 July-28 to Aug-2. 9455–9458.

Saleh, A., Laradji, I. H., Konovalov, D. A., Bradley, M., Vazquez, D., and Sheaves, M. (2020). A realistic fish-habitat dataset to evaluate algorithms for underwater visual analysis. *Sci. Rep.* 10, 14671. doi: 10.1038/s41598-020-71639-x

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). "Grad-CAM: Visual explanations from deep networks *via* gradient-based localization," in *Proceedings of the International Conference on Computer Vision*, Venice, Italy, 2017 October 22-29. (New York, U.S: IEEE) 618–626.

Sultan, H., Owais, M., Park, C., Mahmood, T., Haider, A., and Park, K. R. (2021). Artificial intelligence-based recognition of different types of shoulder implants in X-ray scans based on dense residual ensemble-network for personalized medicine. *J. Pers. Med.* 11, 482. doi: 10.3390/jpm11060482

Zhang, L., Li, X., Arnab, A., Yang, K., Tong, Y., and Torr, P. H. S. (2020). Dual graph convolutional network for semantic segmentation. *arXiv*. New York, U.S: IEEE. doi: 10.48550/arXiv.1909.06121

Zhang, P., Liu, W., Wang, H., Lei, Y., and Lu, H. (2019). Deep gated attention networks for Large-scale Street-level scene segmentation. *Pattern Recognit.* 88, 702–714. doi: 10.1016/j.patcog.2018.12.021

Zhang, W., Wu, C., and Bao, Z. (2022). DPANet: Dual pooling-aggregated attention network for fish segmentation. *IET Comput. Vis.* 16, 67–82:1–18. doi: 10.1049/cvi2.12065

Zhang, Q.-L., and Yang, Y.-B. (2021). "SA-Net: Shuffle attention for deep convolutional neural networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Toronto, Canada. Japan, 2021 June 6-11. New York, U.S: IEEE 2235–2239.

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). "Pyramid scene parsing network," in *Proceedings of the Computer Vision and Pattern Recognition*, Honolulu, USA, 2017 July 22-25. (New York, U.S: IEEE) 2881–2890.

frontiers | Frontiers in Marine Science

TYPE Brief Research Report
PUBLISHED 13 October 2022
DOI 10.3389/fmars.2022.1024339

Check for updates

OPEN ACCESS

EDITED BY
Rizwan Ali Naqvi,
Sejong University, South Korea

REVIEWED BY
Humaira Nisar,
Tunku Abdul Rahman University, Malaysia
Muhammad Arsalan,
Dongguk University Seoul, South Korea

*CORRESPONDENCE
Muhammad Tariq Mahmood
tariq@koreatech.ac.kr

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 21 August 2022
ACCEPTED 23 September 2022
PUBLISHED 13 October 2022

CITATION
Ali U and Mahmood MT (2022) Underwater image restoration through regularization of coherent structures. *Front. Mar. Sci.* 9:1024339. doi: 10.3389/fmars.2022.1024339

COPYRIGHT
© 2022 Ali and Mahmood. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Underwater image restoration through regularization of coherent structures

Usman Ali[1] and Muhammad Tariq Mahmood[2*]

[1]Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, South Korea, [2]Future Convergence Engineering, School of Computer Science and Engineering, Korea University of Technology and Education, Cheonan, South Korea

Restoration of underwater images plays a vital role in underwater target detection and recognition, underwater robots, underwater rescue, sea organism monitoring, marine geological survey, and real-time navigation. Mostly, physics-based optimization methods do not incorporate structural differences between the guidance and transmission maps (TMs) which affect the performance. In this paper, we propose a method for underwater image restoration by utilizing a robust regularization of coherent structures. The proposed method incorporates the potential structural differences between TM and the guidance map. The optimization of TM is modeled through a nonconvex energy function which consists of data and smoothness terms. The initial TM is taken as a data term whereas the smoothness term contains static and dynamic structural priors. Finally, the optimization problem is solved using majorize-minimize (MM) algorithm. The proposed method is tested on benchmark dataset and its performance is compared with the state-of-the-art methods. The results from the experiments indicate that the proposed regularization scheme adequately improves the TM, which results in high-quality restored images.

KEYWORDS

underwater images, image restoration, robust regularization, coherent structures, optimization problem

1 Introduction

Restoration of underwater images is a challenging task and it plays a vital role in underwater target detection and recognition, underwater robots, underwater rescue, sea organism monitoring, marine geological survey, and real-time navigation Jian et al. (2021). Underwater images suffer from strong light absorption, scattering, color distortion, and noise from the artificial light source. Image formation in the water is shown in Figure 1A. It can be observed that the image is formed through the three types of lights 1) the reflected light comes to the camera directly after striking the object, 2) the

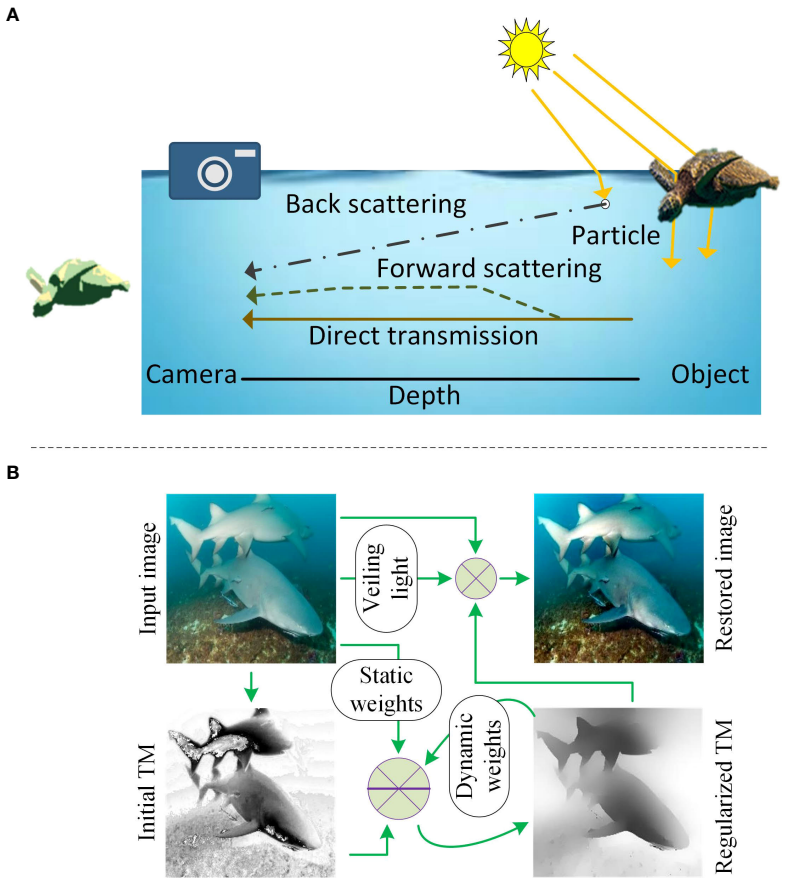


FIGURE 1
(A) Underwater image formation. Object signal is attenuated along the line of sight (direct transmission). Light scattered from the environment (e.g., particle) carries no information of the scene. Forward scattered light blurs the scene. **(B)** Proposed method. Firstly, following the physical scattering model, a transmission map (TM) and veiling light are computed from the input image. Then, the initial TM is improved by solving a robust energy function that utilizes a nonconvex regularizer and two types of regularization weights. Finally, an enhanced image is restored by using the regularized TM.

forward-scattered light that deviates from its way from the original direction after striking the object, 3) the back-scattering light comes to the camera after encountering particles Han et al. (2020); Islam et al. (2020).

Underwater image restoration is well studied in terms of enhancement, noise reduction, defogging/dehazing, segmentation, saliency detection, color constancy, and color correction Jian et al. (2021). Methods for underwater images can be divided into physics and machine/deep learning-based methods Islam et al. (2020). A comprehensive survey on deep learning-based underwater image enhancement algorithms is done in the study Anwar and Li (2020). This study suggested that the algorithms can further be divided into several categories, e.g., encoder–decoder models, modular designs, multi-branch designs, depth-guided networks, and dual generator GANs. In a study, generative adversarial networks (GANs)-based method is proposed to improve the quality of visual underwater scenes Fabbri et al. (2018). As mostly, deep learning-based models are computationally expensive, so a convolutional neural network (CNN)-based diver detection model is suggested to balance the trade-offs between robustness and efficiency Islam et al. (2019). In another work, the authors employed the cycle-consistent adversarial networks to generate synthetic underwater images, and then a deep residual framework-based model is developed for image enhancement Liu et al. (2019). The deep learning approach has become the state-of-the-art solution and it has provided reasonable performance.

Physics-based methods usually restore the image by computing a transmission map (TM). The initial TM is improved using various techniques including guided filtering, statistical models, and matting algorithms Lu et al. (2013). For instance, in a work, guided trigonometric bilateral filters are

applied for improving the TM and then a color correction algorithm is applied for enhancing the image visibility Lu et al. (2013). In another work, an underwater enhancement method is proposed that provided two versions of the restored image. This method is based on the minimum information loss principle and histogram distribution prior Li et al. (2016). In a recent work, the authors suggested methods for image restoration and color correction by taking into account the different optical water types. The revised model for image formation Akkaynak and Treibitz (2018) has been used and a depth map is also required as an input for their solutions Berman et al. (2021). Another recent work with improved results uses a locally adaptive color correction method using the minimum color loss principle and the maximum attenuation map-guided fusion strategy Zhang et al. (2022). In another study, an underwater normalized total variation (UNTV) model is suggested for underwater image dehazing and deblurring that uses sparse prior knowledge of blur kernel. The blur kennel is obtained by using an iterative reweighted least squares algorithm Xie et al. (2022). Mostly, physics-based optimization-based methods optimize TM by utilizing weights from some guidance map that depends on the input image. However, these methods do not incorporate structural differences between the guidance map and TM. Consequently, images recovered are of poor quality.

In this paper, we propose a method for underwater image restoration by utilizing a robust regularization of coherent structures in image and transmission map. The proposed method incorporates the potential structural differences between TM and the guidance map. The optimization of TM is modeled through a nonconvex energy function which consists of data and smoothness terms. The initial TM is taken as a data term whereas the smoothness term contains static and dynamic structural priors. Finally, the optimization problem is solved using majorize-minimize (MM) algorithm. The proposed method is tested on benchmark dataset and its performance is compared with the state-of-the-art methods. The results from the experiments indicate that the proposed regularization scheme adequately improves the TM, which results in high-quality restored images.

2 Proposed method

The proposed method for underwater image restoration can be divided into three steps as shown in Figure 1B. In the first step, the veiling light and initial TM are computed. In the second step, initial TM is regularized through the proposed nonconvex energy framework. During this step, static and dynamic weights are computed from the input image and iteratively regularized TM. In the last step, an image is restored using the underwater imaging model. These steps have been described in detail in the following sections.

2.1 Veiling light and initial transmission map

Based on the Koschmieder's law Koschmieder (1924), only a small portion of the reflected light reaches the observer and it causes poor visibility. The formation of images underwater has been described in Figure 1A. Usually, a linear interpolation-based model is used to describe the image formation in scattering media like water Han et al. (2020); Islam et al. (2020). Recently, a refined image formation model has been proposed in Akkaynak and Treibitz (2018). This revised model tries to explain the instabilities of current models however it also has certain limitations Li et al. (2021). Contrarily, the widely used physical scattering model Han et al. (2020); Berman et al. (2021) that describes the formation of images underwater is as follows,

$$I(x) = J(x)n(x) + V(1 - n(x)) \tag{1}$$

where $x = (x,y)$ denotes the pixel coordinates, $\boldsymbol{I}$ is the observed intensity (i.e., underwater image), $\boldsymbol{J}$ is the scene radiance, $\boldsymbol{V}$ is the global veiling light, and n is the medium transmission map (TM). When the medium is homogeneous, TM can be expressed as $n(\boldsymbol{x}) = e^{-\beta d(\boldsymbol{x})}$, where β is the medium extinction coefficient, and $d(\boldsymbol{x})$ is the depth. Actually, this β is dependent on the color channel, however, for simplicity, we have used it same for all three channels. The main goal of underwater image restoration methods is to recover $\boldsymbol{J}$, $\boldsymbol{V}$, and n from $\boldsymbol{I}$. Although the initial TM n can be obtained through any priors mentioned in the literature, in this work, veiling light V and initial TM n were computed using the haze-lines (HL) prior (Berman et al., 2020). The initial TM is computed per-pixel and is not spatially coherent. To improve this initial TM n, we propose to apply non-convex regularization. The improvement in TM ultimately leads to the restoration of images that are of better quality.

2.2 Model

This paper proposes optimizing the initial TM n by efficiently minimizing the following energy function

$$E(\hat{n}) = \sum_{x\in\Omega}(\hat{n}(x) - n(x))^2 + \lambda\sum_{x\in\Omega}\sum_{x'\in N_x} w_{x,x'}(s)\psi_\rho(\hat{n}(x')) \tag{2}$$

where $\hat{n}$ is the regularized (target) TM, Ω is the 2D spatial domain of TM, λ controls the smoothness level by adjusting the significance of two terms on the right-hand side, N_x is a 2D neighborhood window centered at $\boldsymbol{x}$. $w_{x,\acute{x}}(s)$ is the spatially varying weighting function computed from guidance s. The neighborhood coherence (smoothness) between pixels located at positions $\boldsymbol{x}$ and $\acute{x}$ is enforced adaptively using spatially varying

weights $w_{x,\acute{x}}$. To benefit from the advantages of guided filtering He et al. (2013); Shen et al. (2015), we have incorporated guidance signal s in our proposed framework. Specifically, the spatial regularization (smoothness) weights have been computed from s instead of n or $\hat{n}$. Gray-scale image computed from the input image $\boldsymbol{I}$ has been taken as the guidance s. The idea is that if $s(\boldsymbol{x})$ is considerably different from $s(\acute{x})$ (e.g., if $\boldsymbol{x}$ and $\acute{x}$ are across an edge in the gray-scale input image), then $\hat{n}(x)$ should have a little effect in the regularization of $\hat{n}(\acute{x})$. We define these weights using Gaussian distance in space and intensity as,

$$w_{x,\acute{x}}(s) = \exp\left(-\mu(x-\acute{x})^2\right)\exp\left(-\nu\left(s(x)-s(\acute{x})\right)^2\right), \quad (3)$$

where the first term is the spatial filter that would decrease the weight $w_{x,\acute{x}}$ if the distance between $\boldsymbol{x}$ and $\acute{x}$ is large, second is the intensity range filter that would decrease the weight $w_{x,\acute{x}}$ if the intensity difference between $s(\boldsymbol{x})$ and $s(\acute{x})$ is large. μ and ν are the positive parameters defined by the user. These parameters control the decay rate of the spatial and intensity range filter, respectively, and thus adjust the regularization (smoothness) bandwidth. The proposed robust regularizer is the parameterized squared hyperbolic tangent function defined as

$$\psi_\rho(\hat{n}(x)-\hat{n}(\acute{x})) = \tanh\left(\rho(\hat{n}(x)-\hat{n}(\acute{x}))^2\right) \quad (4)$$

where ρ djusts the skewness of this function. This function maps any $(\hat{n}(x)-\hat{n}(\acute{x}))$ to the range [0 1] and can compresses large values to approach to 1. While observing this function, it can be inferred that this function $\psi_\rho(j)$ penalizes large gradients of $\hat{n}$ less than L_2 or $L1$ regularizer function during filtering. This results in better preservation of high-frequency features (e.g., edges and corners). In other words, our function restrains the large deviations to be fused together.

2.3 Regularization

In the literature, mostly, a convex energy function is minimized for regularization. Contrarily, we have proposed and solved a non-convex energy function (Eq. 2). For such non-convex energy functions, the optimization is non-trivial. To solve this optimization problem, we have used a sophisticated technique of majorize minimization (MM) algorithm. This algorithm performs two steps several times. In the first (i.e., majorization) step, a convex surrogate function for the objective function is created. In the second (i.e., minimization) step, a local minimum is found for the surrogate function. These two steps are followed interchangeably several times until the algorithm converges. While the iteration number k is increased, the values of $E(\hat{n})$ corresponding to the set $\{\hat{n}^{(k)}\}$ decrease monotonically.

At the majorization step, $E^{(k)}$ for $\mathcal{E}$ is achieved by substituting the regularizer $\psi_\rho(j)$ with $\Psi^i_\rho(j)$ in (2) as follows

$$E^{(k)}(\hat{n}) = \sum_{x\in\Omega}[(\hat{n}(x)-n(x))^2 + \lambda\sum_{\acute{x}\in\mathcal{N}_x} w_{x,\acute{x}}(s)\Psi_\rho^{\left(\hat{n}^{(k)}(x)-\hat{n}^{(k)}(\acute{x})\right)}(\hat{n}(x)-\hat{n}(\acute{x}))], \quad (5)$$

where $\Psi^i_\rho(j)$ is a surrogate function for $\psi_\rho(j)$. That is, $\Psi^i_\rho(j)$ stays above the $\psi_\rho(j)$, and they touch each other only at $j=i$. The convex function $E^{(k)}$ is easy to be minimized by taking its first derivative with respect to $\hat{n}$. The normal equation of (5) is

$$0 = (\hat{n}(x)-n(x)) + \lambda\sum_{\acute{x}\in\mathcal{N}_x} w_{x,\acute{x}}(s)\eta^{(k)}_{x,\acute{x}}(\hat{n})(\hat{n}(x)-\hat{n}(\acute{x})), \quad (6)$$

where,

$$\eta^{(k)}_{x,\acute{x}}(\hat{n}) = \left[1-\left(\psi_\rho\left(\hat{n}^{(k)}(x)-\hat{n}^{(k)}(\acute{x})\right)\right)^2\right]\rho. \quad (7)$$

The output $\hat{n}$ is obtained through the vectorized form of (6) by iteratively solving the linear systems of the form, $\hat{n}^{(k)} = (I+\lambda L^{(k)})^{-1}\ n$, where $\boldsymbol{I}$ is an identity matrix, $\boldsymbol{n}$ and $\hat{n}$ denote the column vectors of n and $\hat{n}$, respectively, and $\boldsymbol{L}$ is a Laplacian matrix.

2.4 Recovered image

The goal of underwater image restoration is to recover the scene radiance $\boldsymbol{J}(\boldsymbol{x})$ from $\boldsymbol{I}(\boldsymbol{x})$ based on Eq. 1. Once the regularized TM $\hat{n}(x)$ is obtained, the scene radiance $\boldsymbol{J}(\boldsymbol{x})$ can be recovered by using,

$$J(x) = \frac{I(x)-V}{\hat{n}(x)} + V\ . \quad (8)$$

3 Results and discussion

Our proposed regularization-based scheme (Eq. 2) involves a number of parameters. We firstly describe what are the values of these parameters and how these values are determined. We have performed extensive experiments on a variety of underwater images and empirically found the optimal values for these parameters. Those optimal values are $\lambda = 200$, $\rho = 2.5$, $\mu = 1$, and $\nu = 200$. These same values have been used for all the images tested in this work. Now, we visually examine how our regularization scheme improves the initial TM. To do so, few underwater images have been taken from the EUVP dataset Islam et al. (2020). These images are named as jellyfish, shark, angel butterfly, coral leaf, snake eels, red snapper, mangrove, and yellow fish, and these images have been shown in the first row of Figure 4. Their initial TMs have been shown in the second row, where the color variation between black (dark) and white

(bright) corresponds to the variation in the TM values. It can be seen that the initial TMs have abrupt variations even for the neighboring similar depth regions. Moreover, in these initial TMs, the object boundaries have mingled with the background. If these inaccurate TMs are used for image restoration, their inaccuracies cause degradation in the quality of restored images. This is indicated by the images shown in the third row. These images are obtained through Eq. 8 by using the initial TMs, i.e., $\hat{n}(bfx) = n(x)$. It can be seen that these images suffer from poor visibility. Further, these images have several glitches like, detail loss, color shift, and dimmed light. The inaccuracies of initial TMs need to be addressed for accurate image restoration. We have improved these initial TMs through our non-convex regularization scheme. Our regularized TMs have been displayed in the fourth row. These TMs are considerably better than the initial TMs in a number of ways. Like, our regularized TMs have retained the structures as well as edges of the objects in the scene. On the one hand, these regularized TMs are adequately smooth which ensures the consistency of structures in the spatial domain. On the other hand, sharp structural edges are retained which agrees with the depth discontinuities in the scene. These attributes of the regularized TMs characterize the geometry of the scene. The images restored by using these regularized TMs in Eq. 8 are shown in the last row. It can be observed that these images have higher visibility as compared to the images in the third row. In these images, the details and the natural appearance of the scene are well-preserved. These restored images have rich color information with no artifacts like color saturation. In short, we can also compare the quality of restored images in the third and fifth rows of Figure 2 on the likert scale. Accordingly, each of the images in the third row is of 'poor' quality as compared to the corresponding image in the fifth row which is of 'good' quality. This comparison indicates the necessity of improving the initial TM, and that the proposed regularization scheme is effective in improving the initial TM.

Next, we quantitatively evaluated the advantage gained by regularizing the initial TM. To do so, we computed the perception-based image quality evaluator (PIQE) Venkatanath et al. (2015) for the images, where lower values of this metric indicate better perception quality of the images. PIQE values (before applying regularization, after applying regularization) for the images shown in Figure 2 are jellyfish (42.76,36.20), shark (53.36,38.44), angel butterfly (49.26,24.96), coral leaf (44.07,22.27), snake eels (44.08,34.34), red snapper (50.71,30.55), mangrove (46.89,28.60), and yellow fish (48.46, 23.87). It can be observed that PIQE values of images restored from the initial TMs are worse as compared to the values of images restored from the regularized TMs. This indicates that the regularization of initial TM results in the restoration of better quality images.

Finally, we compared the performance of the proposed method with several state-of-the-art methods. Among the compared methods, eight are learning-based: (i) image enhancement based on generative adversarial network with paired (IE-GAN), and (ii) unpaired training (IE-GAN-UP) Islam et al. (2020), (iii) GAN with cycle-consistency loss

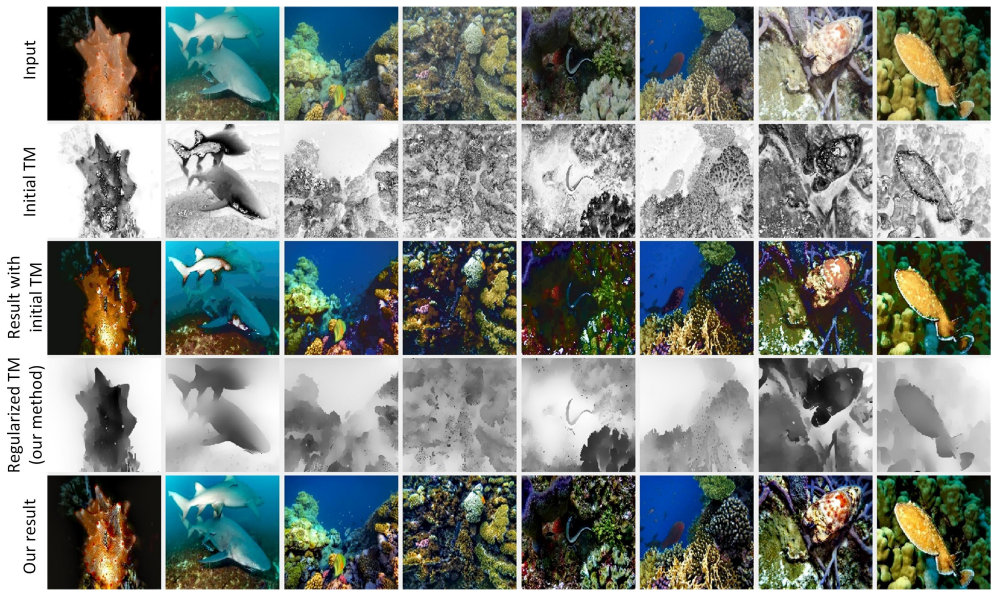


FIGURE 2
Improvement of initial TMs using the proposed regularization scheme. The images restored using the regularized TMs are of better quality as compared to the images restored using the initial TMs.

(CycleGAN) Zhu et al. (2017), (iv) underwater GAN with gradient penalty (UGAN-P) Fabbri et al. (2018), (v) Pix2Pix Isola et al. (2017), (vi) least-squared GAN (LS-GAN) Mao et al. (2017), (vii) GAN with residual blocks in the generator (Res-GAN) Li et al. (2017), and (viii) Wasserstain residual GAN (Res-WGAN) Arjovsky et al. (2017). Two physics-based methods: (i) multi-band fusion-based enhancement (Mband-En) Cho et al. (2018), and (ii) underwater color restoration based on haze-lines (Uw-HL) Berman et al. (2021), are also included for comparison. The output restored images of these methods have been shown in Figure 4 for few underwater images from the EUVP dataset Islam et al. (2020). It can be seen that the IE-GAN and IE-GAN-UP have increased the visibility to some extent but shifted the colors. CycleGAN and LS-GAN exhibit poor visibility, and loose the object boundaries and texture details. UGAP-N, Pix2Pix, Res-GAN, Res-WGAN, Mband-En, and Uw-HL over saturate the objects. Pix2Pix and LS-GAN often fail to improve global brightness as well. Mband-En and Uw-HL have shifted the colors by a large extent. On the other hand, our proposed method provides good quality results which are free from the above mentioned artifacts of the compared methods. Our proposed method adequately enhances the visibility without any color-shifting artifacts. Further, the output images of our method retain the fine details. We also used the quantitative measures to evaluate the quality of the images of Figure 3. Two quantitative measures naturalness image quality evaluator (NIQE) Mittal et al. (2012), and perception-based image quality evaluator (PIQE) Venkatanath et al. (2015) have been computed and shown in the Figure 4. Lower values of these measures reflect better perceptual quality of the image. The eleven bars for each image respectively correspond to the images restored by approaches IE-GAN, IE-GAN-UP Islam et al. (2020), CycleGAN Zhu et al.

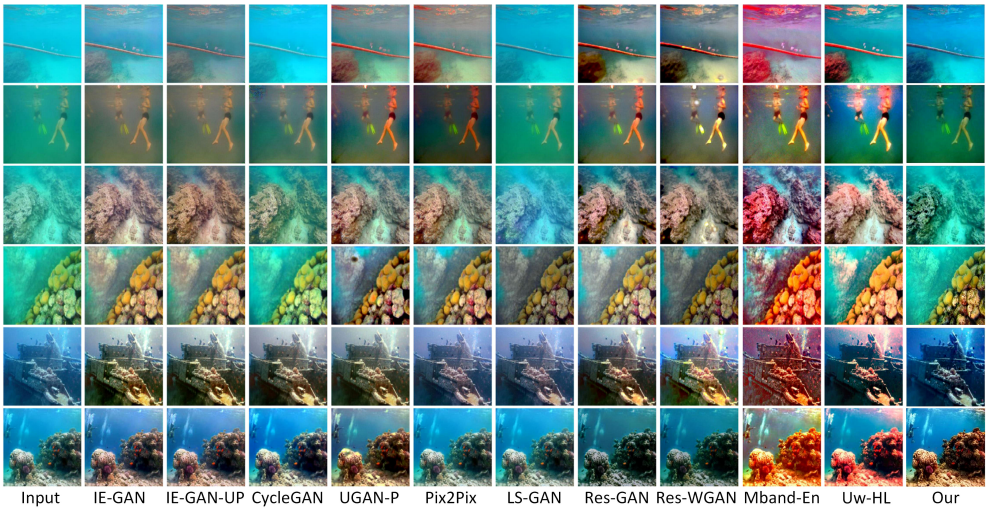


FIGURE 3
Comparison of our method with other approaches IE-GAN, IE-GAN-UP Islam et al. (2020), CycleGAN Zhu et al. (2017), UGAN-P Fabbri et al. (2018), Pix2Pix Isola et al. (2017), LS-GAN Mao et al. (2017), Res-GAN Li et al. (2017), Res-WGAN Arjovsky et al. (2017), Mband-En Cho et al. (2018), and Uw-HL Berman et al. (2021).

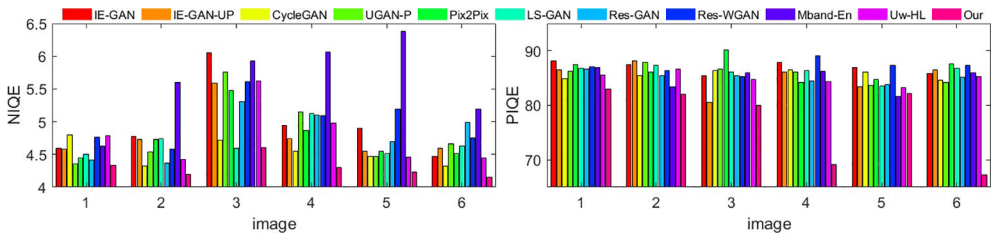


FIGURE 4
Quantitative measures NIQE Mittal et al. (2012) and PIQE Venkatanath et al. (2015) for the six images shown in Figure 3. The eleven bars for each image respectively correspond to the images restored by different compared approaches and our method.

(2017), UGAN-P Fabbri et al. (2018), Pix2Pix Isola et al. (2017), LS-GAN Mao et al. (2017), Res-GAN Li et al. (2017), Res-WGAN Arjovsky et al. (2017), Mband-En Cho et al. (2018), Uw-HL Berman et al. (2021), and our method. It can be seen that the proposed method attains the least values for both of these quantitative measures for almost all six images. These values suggest that the perception qualities of our restored images are better than the input images and the restored images by the compared methods. From this qualitative and quantitative comparison among the quality of restored images, it can be deduced that our proposed method outperforms the state-of-the-art methods.

4 Conclusion

In this paper, a robust regularization-based method has been proposed for the underwater image restoration. Usually, the TM suffers from several artifacts like, abrupt variations which are inconsistent with the scene, and object boundaries mingled with the background. These inaccuracies in TM lead to the image restoration of degraded quality. We have formulated a nonconvex energy function for the optimization of initial TM. As a result, the regularized TM is free from the artifacts; it is adequately smooth as well as retains the sharp boundaries. This improvement in TM results in image restoration of better quality. The experimental results demonstrated that the proposed method is remarkably effective for the restoration of underwater images and it outperforms the state-of-the-art methods.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://irvlab.cs.umn.edu/resources/euvp-dataset.

Author contributions

UA and MM contributed to conception and design of the study. UA wrote the code for simulations and collected the data. MM performed the analysis. UA and MM wrote the draft of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported in parts by the Creative Challenge Research Program (2021R1I1A1A01052521), BK-21 FOUR Program funded by the Ministry of Education, by the Basic Research Program (2022R1F1A1071452) and Basic Science Research Program (2022R1A4A3033571) funded by the Korea government (MSIT: Ministry of Science and ICT), through the National Research Foundation of Korea (NRF).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Akkaynak, D., and Treibitz, T. (2018). "A revised underwater image formation model," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6723–6732. doi: 10.1109/CVPR.2018.00703

Anwar, S., and Li, C. (2020). Diving deeper into underwater image enhancement: A survey. *Signal Processing: Image Communication* 89, 115978. doi: 10.1016/j.image.2020.115978

Arjovsky, M., Chintala, S., and Bottou, L. (2017). "Wasserstein generative adversarial networks," in *International conference on machine learning (PMLR)*. eds. D. Precup and T. Yee Whye. Sydney, NSW, Australia: PMLR 70, 214–223. Available at: https://proceedings.mlr.press/v70/arjovsky17a.html

Berman, D., Levy, D., Avidan, S., and Treibitz, T. (2021). "Underwater single image color restoration using haze-lines and a new quantitative dataset," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43 (8), 2822–2837. doi: 10.1109/TPAMI.2020.2977624

Berman, D., Treibitz, T., and Avidan, S. (2020). "Single image dehazing using haze-lines," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42 (3), 720–734. doi: 10.1109/TPAMI.2018.2882478

Cho, Y., Jeong, J., and Kim, A. (2018). "Model-assisted multiband fusion for single image enhancement and applications to robot vision," in *IEEE Robotics and Automation Letters* 3 (4), 2822–2829. doi: 10.1109/LRA.2018.2843127

Fabbri, C., Islam, M. J., and Sattar, J. (2018). "Enhancing underwater imagery using generative adversarial networks," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. 7159–7165. doi: 10.1109/ICRA.2018.8460552

Han, M., Lyu, Z., Qiu, T., and Xu, M. (2020). "A review on intelligence dehazing and color restoration for underwater images," in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 50 (5), 1820–1832. doi: 10.1109/TSMC.2017.2788902

He, K., Sun, J., and Tang, X. (2013). "Guided image filtering," in IEEE transactions on pattern analysis and machine intelligence, 35 (6), 1397–1409. doi: 10.1109/TPAMI.2012.213

Islam, M. J., Fulton, M., and Sattar, J. (2019). "Toward a generic diver-following algorithm: Balancing robustness and efficiency in deep visual detection," in *IEEE Robotics and Automation Letters*, 4 (1), 113–120 8543168. doi: 10.1109/LRA.2018.2882856

Islam, M. J., Xia, Y., and Sattar, J. (2020). "Fast underwater image enhancement for improved visual perception," in *IEEE Robotics and Automation Letters*, 5 (2), 3227–3234. doi: 10.1109/LRA.2020.2974710

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). "Image-to-image translation with conditional adversarial networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 5967–5976. doi: 10.1109/CVPR.2017.632

Jian, M., Liu, X., Luo, H., Lu, X., Yu, H., and Dong, J. (2021). Underwater image processing and analysis: A review. *Signal Processing: Image Communication* 91, 116088. doi: 10.1016/j.image.2020.116088

Koschmieder, H. (1924). Theorie der horizontalen Sichtweite. *Beitrage zur Physik der freien Atmosphare* 12, 33–53.

Li, C., Anwar, S., Hou, J., Cong, R., Guo, C., and Ren, W. (2021). Underwater image enhancement *via* medium transmission-guided multi-color space embedding," in *IEEE Transactions on Image Processing* 30, 4985–5000. doi: 10.1109/TIP.2021.3076367

Li, C.-Y., Guo, J.-C., Cong, R.-M., Pang, Y.-W., and Wang, B. (2016). "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," in *IEEE Transactions on Image Processing*, 25 (12), 5664–5677. doi: 10.1109/TIP.2016.2612882

Li, J., Liang, X., Wei, Y., Xu, T., Feng, J., and Yan, S. (2017). "Perceptual generative adversarial networks for small object detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 1951–1959. doi: 10.1109/CVPR.2017.211

Liu, P., Wang, G., Qi, H., Zhang, C., Zheng, H., and Yu, Z. (2019). "Underwater image enhancement with a deep residual framework," in *IEEE Access*. 7, 94614–94629. doi: 10.1109/ACCESS.2019.2928976

Lu, H., Li, Y., and Serikawa, S. (2013). "Underwater image enhancement using guided trigonometric bilateral filter and fast automatic color correction," in *2013 IEEE International Conference on Image Processing* 3412–3416. doi: 10.1109/ICIP.2013.6738704

Mao, X., Li, Q., Xie, H., Lau, R. Y., Wang, Z., and Smolley, S. P. (2017). "Least squares generative adversarial networks," Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2813–2821. doi: 10.1109/ICCV.2017.304

Mittal, A., Soundararajan, R., and Bovik, A. C. (2012). Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*. 20 (3), 209–212. doi: 10.1109/LSP.2012.2227726

Shen, X., Zhou, C., Xu, L., and Jia, J. (2015). "Mutual-structure for joint filtering," in *Proceedings of the IEEE International Conference on Computer Vision*. 3406–3414. doi: 10.1109/ICCV.2015.389

Venkatanath, N., Praneeth, D., Bh, M. C., Channappayya, S. S., and Medasani, S. S. (2015). "Blind image quality evaluation using perception based features," in *2015 Twenty First National Conference on Communications (NCC) (IEEE)*. 1–6. doi: 10.1109/NCC.2015.7084843

Xie, J., Hou, G., Wang, G., and Pan, Z. (2022). "A variational framework for underwater image dehazing and deblurring," in *IEEE Transactions on Circuits and Systems for Video Technology*, 32 (6), 3514–3526. doi: 10.1109/TCSVT.2021.3115791

Zhang, W., Zhuang, P., Sun, H.-H., Li, G., Kwong, S., and Li, C. (2022). "Underwater image enhancement *via* minimal color loss and locally adaptive contrast enhancement," in *IEEE Transactions on Image Processing* 31, 3997–4010. doi: 10.1109/TIP.2022.3177129

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*. 2242–2251. doi: 10.1109/ICCV.2017.244

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 11 November 2022
DOI 10.3389/fmars.2022.1047053

Check for updates

OPEN ACCESS

EDITED BY
Hong Song,
Zhejiang University, China

REVIEWED BY
Yaoming Zhuang,
Northeastern University, China
Kang Ryoung Park,
Dongguk University Seoul,
South Korea

*CORRESPONDENCE
Zhuang Zhou
jonahzhou199526@gmail.com

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 17 September 2022
ACCEPTED 17 October 2022
PUBLISHED 11 November 2022

CITATION
Lai Y, Zhou Z, Su B, Xuanyuan Z, Tang J, Yan J, Liang W and Chen J (2022) Single underwater image enhancement based on differential attenuation compensation.
Front. Mar. Sci. 9:1047053.
doi: 10.3389/fmars.2022.1047053

COPYRIGHT
© 2022 Lai, Zhou, Su, Xuanyuan, Tang, Yan, Liang and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Single underwater image enhancement based on differential attenuation compensation

Yunting Lai[1], Zhuang Zhou[1]*, Binghua Su[1], Zhe Xuanyuan[2], Jialin Tang[1], Jialin Yan[1], Wanxin Liang[1] and Jiongjiang Chen[1]

[1]Key Laboratory of Intelligent Detection in Complex Environment of Aerospace Land and Sea, Beijing Institute of Technology, Zhuhai, China, [2]Hong Kong Baptist University United International College, Beijing Normal University, Zhuhai, China

High quality underwater images and videos are important for exploitation tasks in the underwater environment, but the complexity of the underwater imaging environment makes the quality of the acquired underwater images generally low. To correct the chromatic aberration and enhance the sharpness of underwater images in order to improve the quality of underwater images, we based on the differential compensation proposed a Differential Attenuation Compensation (DAC) method. The underwater image is contrast stretched to improve the contrast of the image, as well as the underwater image is denoised, for the red channel with serious loss of detail information we choose the blue and green channels with more detail information to compensate for this, and finally the image is restored through the grayscale world to obtain more realistic colors. Our method is qualitatively and quantitatively compared with multiple state-of-the-art methods in the public underwater image dataset, underwater image enhancement benchmark (UIEB) and enhancing underwater visual perception (EUVP), demonstrating that the underwater images processed by our method better resolve the problems of chromatic aberration and blur, with more realistic color, detail and better underwater image quality evaluation indicators

KEYWORDS

underwater image, image enhancement, contrast stretching, differential attenuation compensation, machine vision

1 Introduction

The ocean is rich in material resources, but due to the complex underwater environment, the tasks related to ocean development are challenging. As a key technology for ocean development, underwater imaging technology can effectively assist related tasks of ocean development. Underwater imaging technology is an

effective way to obtain underwater images, and it is also used in various underwater scenarios, such as the laying of submarine cables (Ortiz et al., 2002), the survey of deep sea, and the observation of fish schools (Howell et al., 2021). Obtaining effective underwater environment information and target information is the premise of ocean development. However, due to the complex underwater imaging environment, the quality of underwater information collected by common sensors is low and difficult to use for other tasks.

In the marine environment, the complex imaging environment leads to serious degradation of underwater images collected by ordinary imaging equipment, and problems such as color distortion, contrast reduction, and image blurring occur. In the underwater imaging process, different wavelengths of light have different attenuation rates in underwater propagation (Clarke and James, 1939). In the visible wavelength band, red light has a long wavelength penetration is the weakest, most easily absorbed by water, its propagation distance underwater is about 4 meters; blue light and green light has a short wavelength, the absorption effect of water is small, the propagation distance is farther. The variability in the attenuation of different wavelengths of light in the underwater environment leads to a general problem of color distortion and an overall blue-green hue in underwater images. Meanwhile, there are a large number of suspended particles and plankton in the water, which scatter the imaging light and background light making the underwater images blurred (Wozniak and Dera, 2007).

Therefore, eliminating blurring and chromatic aberration of underwater images caused by differential attenuation and scattering through underwater image processing techniques is the key to obtaining high-quality underwater images.

We propose a pixel processing-based underwater image enhancement method DAC based on the differential attenuation of different wavelengths of light in the underwater environment. The quality of underwater images is enhanced by contrast stretching and color differential attenuation compensation of underwater images. The following contributions are made in this paper:

1. We decompose the underwater image to be processed under RGB color space based on the differential attenuation of different wavelengths of light in the underwater environment, further decompose each channel image into a base layer containing image structure information and a detail layer containing image texture information, and propose a method to compensate R channel detail information with G and B channel detail information.
2. Our proposed method is a new approach to underwater image processing that is not based on the underwater physical imaging model, but only on image pixels, and is capable of simultaneously removing image blur and correcting image chromatic aberration. And the paper contains a large number of experiments based on pixel information, such as verifying the associativity of the mean value of G and B channel with image hue.
3. We provide codes for this paper, every one can get codes from https://github.com/lailaiyun/Single-Underwater-Image-Enhancement-Based-on-Differential-Attenuation-Compensation.

2 Related work

The scattering effect of light by suspended particles and a large number of plankton in the water, as well as the absorption effect of water on different wavelengths of light cause problems such as color attenuation, low contrast and blurring of underwater images, which have serious impact on the exploitation of the ocean. The enhancement of underwater imaging technology can provide more useful information for the development of the ocean, which can be divided into two methods, one is the enhancement of hardware equipment, such as underwater imagers and LIDAR imaging systems (Phillips et al., 2019; Egorenko and Efremov, 2020). Unfortunately, for hardware equipment enhancements often require higher costs. The other approach is to process the acquired underwater information by algorithms (Sahu et al., 2014). Among them, there are also three methods for underwater image processing, physical model-based image enhancement, non-physical model-based image enhancement, and deep learning-based image enhancement.

Non-physical model-based image enhancement is used to enhance the quality of underwater images by directly processing the pixel values of the underwater images. The classical non-physical model algorithm for histogram equalization (Kaur et al., 2011) uses contrast stretching (Abdullah-Al-Wadud et al., 2007) to adjust the histogram of an image by distributing the concentrated pixel points in the histogram evenly throughout the Gray World Algorithm, redistributing the pixel values of the image, and increasing the gap between the gray levels of the histogram, thus achieving contrast stretching. In the deep ocean, where light propagation is obstructed, artificial light sources are often used for image acquisition, but using histogram equalization for underwater images that are unevenly lit and contain areas that are too dark or too bright can overstretch the contrast of the processed image. In this regard, adaptive histogram equalization (Pizer et al., 1987) is proposed to solve the problem of global equalization of the image, which divides the image into small blocks of equal size and performs local histogram stretching for each small block of image to solve the problem of image overstretching, but this proposed method introduces the problem of noise being amplified, which affects the peak signal-to-noise ratio of the image. In this regard,

adaptive histogram equalization is proposed to solve the problem of global equalization of the image, which divides the image into small blocks of equal size and performs local histogram stretching for each small block of image to solve the problem of image overstretching, but this proposed method introduces the problem of noise being amplified, which affects the peak signal-to-noise ratio of the image. Later, many improvements and fusion methods were proposed by researchers, Ancuti et al. (Ancuti et al., 2012; Ancuti et al., 2017) proposed a multi-scale fusion strategy approach combining contrast enhancement and color correction, which largely increased the color of underwater images. Besides, frequency domain method (Yang et al., 2021) is also a non-physical model image processing method. This method converts the pixel point and position information in space to other processing space, and filters the high frequency information and low frequency information of the image, such as high pass filtering and low pass filtering. Then it is back-converted to the spatial domain to obtain the enhanced underwater image. Huang (Huang et al., 2018) processes the image in RGB and CIE-Lab color space, and stretches the histogram of the image to enhance the quality of image. Kashif (Iqbal et al., 2010) proposed an unsupervised colour correction method (UCM) for underwater image enhancement. UCM stretches the image on RGB and HSI colour space to archive color correction.

The underwater imaging model proposed by Jaffe-McGlamery (McGlamery, 1980) recovers the degradation process of underwater images in the form of a physical model and is the basis for underwater image enhancement based on the physical model. This underwater imaging model suggests that the light energy of an underwater image consists of three components as long as: forward scattering (scattering of light by the underwater scene), backward scattering (scattering of light by impurities in the water), and directly transmitted light energy (Lu et al., 2016). Image enhancement algorithms based on physical models are often used to enhance the quality of underwater images by reducing the forward and backward scattering of images (Li et al., 2018). Based on the underwater imaging model, HE et al. (He et al., 2010) proposed a classical de-fogging algorithm Dark Channel Prior (DCP), which is able to estimate the depth map of the underwater image scene and achieve image clarity. Paulo et al. (Drews et al., 2013) proposed Underwater Dark Channel Prior (UDCP) by combining the properties of color decay of underwater images in order to repair underwater degraded images and obtain underwater images that are clearer and contain realistic colors. In addition to this, Akkaynak et al. (Akkaynak and Treibitz, 2018) (Akkaynak and Treibitz, 2019) proposed a method for de-watering underwater images and videos based on an underwater image imaging model, which realistically restores the colors of underwater images, making them closer to the colors as well as the true form of land images. Wei (Song et al., 2018) proposes underwater light appreciation prior to estimate the parameters of the model, ambient background light and transmission.

In recent years, the development of deep learning has led to outstanding performance in various fields, and it has gained wide attention in underwater image processing. Generative adversarial network models based on game ideas are also often used in underwater image enhancement (Engin et al., 2018), but in deep learning, pairs of datasets (containing degraded images and corresponding high-quality images) are usually required to train the network models. Cycle Generative Adversarial Networks (Cycle-GAN) (Zhu et al., 2017) was proposed to solve the problem of no paired dataset. Based on the paired dataset generated by Cycle-GAN, Wang et al. (Wang et al., 2019) proposed an underwater color image enhancement algorithm Underwater Generative Adversarial Networks (UWGAN) based on generative and adversarial network to make blurred, color-biased underwater images clearer and more colorful. Islam et al. combined both supervised and unsupervised learning methods to propose a multimodal objective function-basedfully-convolutional conditional Generative Adversarial Networks based model for real-time underwater image enhancement, and refer to as FUnIE-GAN (Islam et al., 2020), which optimizes the loss function of generative adversarial networks and provides enhancing underwater visual perception (EUVP) datasets that can be used for both one-way and two-way training. Even though Cycle-GAN is able to train the network by synthesising underwater image datasets, there is still a gap between the synthesised underwater image datasets and the real underwater images, and the network model is not always well trained. Therefore, the robustness and generalization of deep learning-based underwater image enhancement algorithms still have more room for improvement compared to traditional algorithms based on physical and non-physical models. Ankita (Naik et al., 2021) proposes a shallow neural network architecture, Shallow-UWnet which maintains performance and has fewer parameters than the state-of-art models. Chen (Chen et al., 2021) proposed a new a new underwater image enhancement method based on deep learning and image formation model, we refer to as Image Formation. This method works well, but lacks interpretability.

3 Differential attenuation compensation

The complex imaging environment in the ocean causes severe degradation of the underwater images acquired by underwater imaging systems. In the aqueous medium, water molecules and various substances contained in the water absorb light, and the absorption has the property of increasing with decreasing wavelength. The absorption of light energy by the water body makes the underwater images suffer from low contrast, color shift

and distortion. In addition, impurities in the water also cause blurring and lack of ground realism in the underwater images. As shown in Figure 1, we propose the Differential Attenuation Compensation (DAC) method for the problems of underwater images in the above paper, which uses a three-step strategy of contrast enhancement, image decomposition and R channel attenuation compensation to process underwater images without relying on the underwater imaging system and thus improve the quality of underwater images.

3.1 Contrast enhancement

The underwater environment contains a large number of suspended particles, and because of the scattering effect, the underwater image will be blurred, like a shroud of fog, and the contrast is low. And in environments such as the deep sea and other natural light can't be illuminated, the use of artificial light sources and other converging light illumination, the target scene is not uniformly illuminated, the images collected in this environment, some areas are brighter, some are darker. Therefore, we use the Contrast Limited Adaptive Histogram Equalization (Reza, 2004) (kumar Rai et al., 2012) method to stretch each channel of the image.

$$I_{CE}^{C} = CLAHE(I^{C}) \tag{1}$$

$C \in \{R,G,B\}$ denotes the red, green, and blue channels of the image, I^{C} denotes the C hannel image of the original underwater image, $CLAHE()$ enotes the image processed with the Contrast Limited Adaptive Histogram Equalization method, and I_{CE}^{C} denotes the underwater image after contrast enhancement. Contrast Limited Adaptive Histogram Equalization is an improvement to the adaptive histogram equalization mentioned in the related work above.

As shown in Figure 2, the algorithm uses a pre-set threshold to limit the maximum magnitude of the histogram, crops the histogram after computing the mapping function, and then distributes the cropped pixel values uniformly to each gray level of the image, suppressing the noise of the adaptive histogram equalization. Also, bilinear interpolation is used to stitch the divided image blocks in the adaptive histogram equalization to remove the boundaries between image blocks.

We divide the underwater image into several equal-sized image blocks evenly and calculate the histogram and the corresponding mapping function in each image block separately, mapping the pixel points located at the boundaries with the mapping function of the adjacent image block. For non-boundary pixel points, the mapping values of the four adjacent image blocks for that pixel value are calculated separately for linear interpolation.

As shown in Figure 3, it is the image before and after the CLAHE method stretched and its channel images and the corresponding histogram. The histogram distribution of each channel before and after stretching is similar, but in the underwater image after stretching, the R channel has more pixels distributed in the large pixel value range, and the B channel has more pixels distributed in the small pixel value range. The stretched underwater image visibly partially removes the blue hue or green hue of the original underwater image.

3.2 Image decomposition

Image decomposition refers to the decomposition of an image into two parts, structure and texture. The structure part is the larger scale base object in the image, we refer to as *BL* and the texture part is the smaller scale detail object, we refer to as *DL* Before image decomposition, the image needs to be pre-processed with noise removal to prevent noise from being

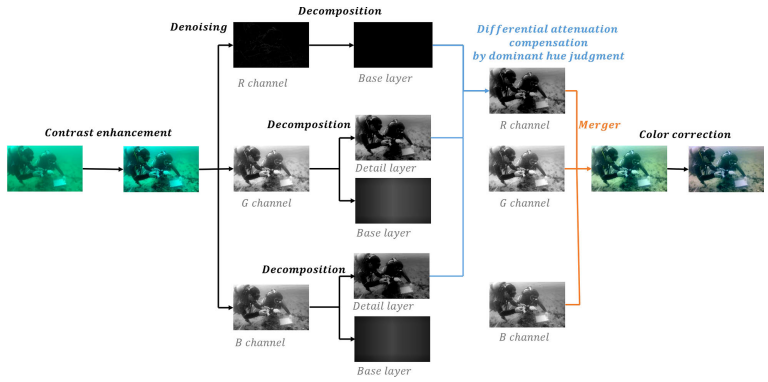


FIGURE 1
Algorithm flow chart. Firstly, the image is contrast stretched, the base layer as well as the detail layer of the image is extracted, then the color attenuation is compensated, and finally the color of the image is restored by the Gray World.

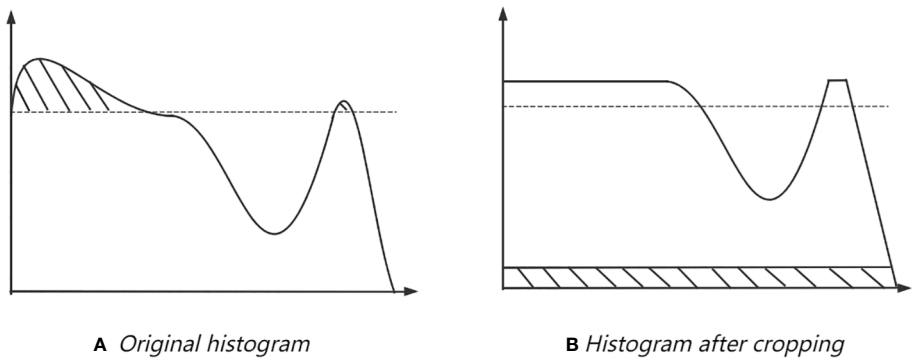


FIGURE 2
Contrast limiting principle. The cropped pixels are evenly divided into gray levels. **(A)** is the original histogram, **(B)** is the histogram after cropping, the cropped pixels are uniformly filled at each gray level.

considered as detail information and affecting the decomposition result of the image.

$$I_{CE}^{C} = BL^{C} + DL^{C} \tag{2}$$

Where $C \in \{R,G,B\}$ equation 2 shows that the C channel of contrast- enhanced image I_{CE}^{C} is viewed as an accumulation of two parts, the base layer BL^{C} and the detail layer DL^{C}

Due to the impact of plankton and suspended particles on imaging in the underwater environment, the captured underwater images are superimposed with too much interference and severe noise. Then we decompose the image into two parts, the base layer and the detail layer, using the image mean value as the threshold. As shown in equation 3 we perform mean filtering on the C channel of the contrast-boosted image and consider the filtered image as the base layer of the channel.

$$BL^{C} = I_{CE}^{C} * Z \tag{3}$$

Where Z denotes the mean filter and * denotes the convolutional transport. The selection of filter template size is adaptively adjusted according to the image resolution. Figure 4 shows the effect of using a 4×4 filter template, a 4×4 filter template, the minimum size of the image (filter size=minm, n) and the maximum size of the image (filter size=maxm, n) as filter templates for filtering the underwater image, respectively. The template with too small size will make the underwater image too smooth and produce red edges. In this paper, we choose the largest size of the image as the filtering template, which can have a better smoothing effect on the detail area of the underwater image and retain the brightness information of the underwater image, which is convenient for detail extraction. As mentioned above we consider the image to be composed of a base layer and a detail layer, so the detail layer for this channel is:

$$DL^{C} = I_{CE}^{C} - I_{CE}^{C} * Z \tag{4}$$

As shown in Figures 4A, B are the images processed with 4×4 filter template and 4×4 filter template respectively, in which the divers and objects and the edges of the images have obvious red lines, which make the images distorted. The size of the images used in our experiments is mostly around 800 × 500, much larger than 4 and 40. 4 (C) is the processing result of the image when the size of filter template is *min {m, n}* and there is obvious red distortion at the edges of the image, which makes

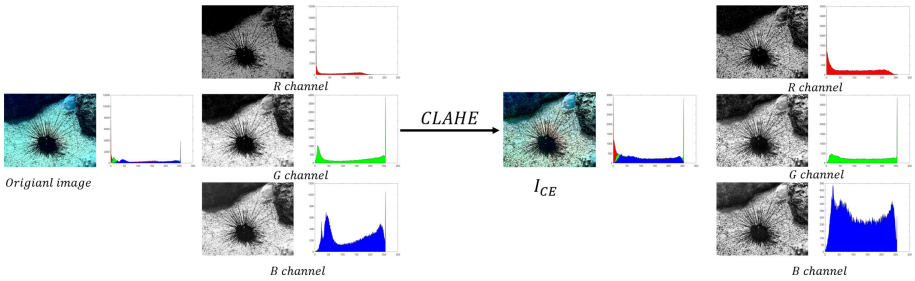


FIGURE 3
Comparison before and after contrast stretching.



A 4×4 *filter template* **B** 40×40 *filter template* **C** min *of the image size* **D** max *of the image size*

FIGURE 4
Comparison of the results of processing images with different filters. **(A–D)** is 4 × 4 filter templates, 40 × 40 filter templates, min of the image size and max of the image size respectively.

the image quality significantly degraded. We have gone through a lot of experiments to change the size of filter template, and found that when the size of filter template is $max\ \{m, n\}$ the filtered image results are the best, as shown in 4 (D). In the figure we only show the results of our individual experiments, but we have actually tried many size of filter templates.

3.3 R channel attenuation compensation

Compared to red light, blue and green light decay relatively slowly in water and travel the farthest distance. Therefore, we believe that the R channel information of the image is more lost, while the G channel and B channel information is more retained. We observed several datasets of underwater images that have been used more by researchers and found that most of the underwater images appear blue and green, with most of the images that appear blue being images taken in the deep sea, while the green images were taken in relatively close waters. We also found in our field research that the seawater close to the coastline is generally very turbid due to the current and sediment, normally appearing yellow, gradually appearing green away from the coast, and gradually appearing blue as it continues to move away from the coast. For the underwater image with green hue, we believe that the G channel detail is the most complete information retained by attenuation, and for the underwater image with blue hue, we believe that the B channel detail is the most complete information retained by attenuation. This is because under natural lighting conditions, red light is generally depleted by attenuation at about $4m$ underwater.

As shown in Figure 3, the images before and after stretching by the CLAHE method and their respective channel images as well as the corresponding histograms are shown. Underwater images in blue hue, the pixel values of the R channel are mostly distributed around pixel values equal to 0. The corresponding R channel images have many black dots, and during image processing, when the pixel value of a pixel is very close to 0, the computer may assume that the pixel value of the pixel is 0. Therefore, the image information corresponding to this part may be lost during the processing, and the corresponding position in the image will become a black dot. In addition, when the pixel value is small, it is difficult for the human eye to distinguish the details in a darker image.

As shown in Figure 7 and Figure 8, we found that the pixel values of the image R channel in the blue or green hue images are relatively small after extensive statistical experiments, and the vast majority of the R channel pixels in Figure 7 and Figure 8 are below 50. Therefore, we believe that the loss of detail information in the R channel of the image is serious, and we need to compensate for the loss of detail information in the R channel with the G channel or B channel.

After a large number of image tests, it was found that the blue underwater images had the largest B channel mean and the green underwater images had the largest G channel mean. The pixel values of the R channel of the blue-green hue underwater images are all relatively small.



FIGURE 5
The green hue underwater images.

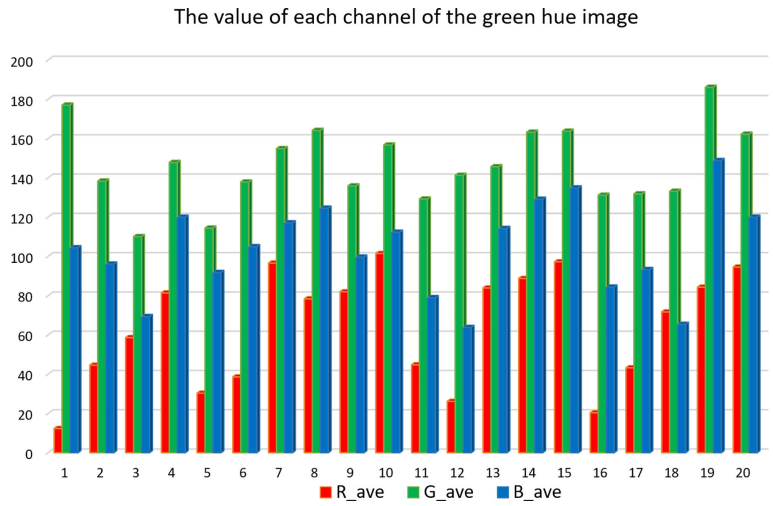


FIGURE 6
Comparison of mean values of green hue underwater images in Figure 5 for each channel.



FIGURE 7
The blue hue underwater images.

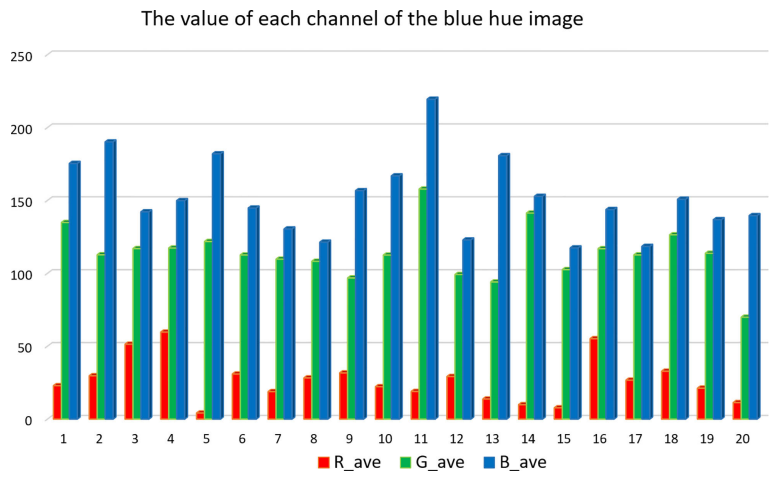


FIGURE 8
Comparison of mean values of blue hue underwater images in Figure 6 for each channel.

Among the blue underwater images and the green underwater images, as shown in Figure 5 and Figure 6, we selected 20 representative images for analysis respectively, and the analysis results are shown in Figure 7 and Figure 8. We therefore compare the mean values of the G channel and the B channel of the image to determine whether to compensate the detail information of the R channel with the G channel or the B channel.

$$R_{cp} = \begin{cases} BL^R + DL^B \cdot \text{if} \cdot B_{ave} > G_{ave} \\ BL^R + DL^G \cdot \text{if} G_{ave} > B_{ave} \end{cases} \quad (5)$$

Where G_{ave}, B_{ave} are the mean values of G channel and B channel of the image, respectively. $R_{cp}$is the R channel after compensating the detail information, the above equation shows that when G_{ave}> B_{ave} the image as a whole is green, the R channel is compensated by the detail layer of the image G channel, and when B_{ave}> G_{ave} the image as a whole is blue, the R channel is compensated by the detail layer of the image B channel. Finally, R_{cp} I_{CE}^G I_{CE}^B are combined as the result of underwater image after R channel attenuation compensation.

As shown in Figure 7 and Figure 8, for the underwater image in green hue, the mean value of the G channel is the largest among the mean values of the RGB channels, and for the underwater image in blue hue, the mean value of the B channel is the largest among the mean values of the RGB channels. The mean value of the R channel is the smallest in both the blue hue underwater image and the green hue underwater image, and even in the vast majority of cases the mean value of the R channel is less than 50. In DAC, we first determine whether the image is dominated by green or blue, and then compensate for the lost detail information of the R channel with the detail information of the dominant color. In DAC, we first determine whether the image is dominated by green or blue, and then compensate for the lost detail information of the R channel with the detail information of the dominant color.

We think that the underwater image after compensating the R channel partially eliminates the effect of attenuation on the underwater image, but there is still some deviation between the underwater image and the real color of the target scene, so we introduce the Gray World Algorithm (Fu et al., 2017) to eliminate the effect of different wavelengths of light attenuation on the image and restore the real color of the target scene. In our experiments, we found that most of the results would show an overall red color of the image, so we used the Gray World Algorithm with the following restrictions on the red channel.

$$k_R = \alpha \frac{Gray}{R_{ave}} \quad (6)$$

$$k_C = \frac{Gray}{C_{ave}}, C \in G, B \quad (7)$$

$$k_C = \frac{Gray}{C_{ave}}, C \in G, B \quad (8)$$

$$I_{\text{out}}^C = k_C I_{CE}^C, C \in G, B \quad (9)$$

Where, $Gray$ is the mean of the R, G, B channel mean. α is a weight coefficient to control the color recovery of R channel, and the value range is [0,1]. If α is too small, it will make the image red compensation insufficient, resulting in the loss of some color information and the overall blue-green mixed color of the image. If the α is too large, it will make the image overcompensate and appear reddish overall, especially the background part of the image will appear pink or purple. In this paper, we choose α = 0.8 or better performance. R_{ave} G_{ave} B_{ave} are the mean values of R_{cp} I_{EC}^G I_{EC}^B respectively. k_R k_G and k_B are scale parameters of R, G, B channel respectively. I_{out}^C denotes the C channel of the output image.

4 Experiment

4.1 Analysis for underwater image enhancement benchmark dataset

In the experiments of this paper, the UIEB (Li et al., 2019) underwater image enhancement benchmark dataset is used, which consists of 950 real-world underwater images covering the diversity of underwater environments and contains a variety of underwater scenes, such as deep-sea fish, coral reefs, submarine cables, underwater antiquities, etc. Most of the underwater images in the dataset show blue-green color, which satisfies the original intention of our proposed algorithm. In this paper, we conduct experiments based on the UIEB underwater image dataset, and to verify the effectiveness of our algorithm, we compare it with other underwater image enhancement methods for quantitative and qualitative analysis.

In this paper, two types of image quality evaluation metrics are used to demonstrate the effectiveness of our algorithm. They are the quality evaluation metrics for non-reference images and the quality evaluation metrics for reference images. We used the underwater color image quality evaluation metric (UCIQE) (Yang and Sowmya, 2015) as an image quality evaluation metric for non-reference images. UCIQE is a linear combination of color intensity, saturation and contrast and is an evaluation of uneven chromatic aberrations, blurring and low contrast in underwater images. The smaller value of MSE indicates that the processed image is closer to Ground Truth, and the larger values of PSNR, SSIM, and UCIQE indicate better image quality. The CLAHE algorithm is able to stretch the contrast of underwater images in a limited way, which increases the contrast and sharpness of the image, improving the quality of underwater images. However, the range of application is limited and can result in overexposure or

underexposure of underwater images that are too bright or too dark, as well as loss of detail in the image.

As shown in Figure 9, a histogram calculation of the original image reveals that there are few pixel points in the red channel of the unprocessed underwater image. In order to compensate for the lost detail information in the R channel, we want to stretch the image by CLAHE to make the pixel value of the R channel image larger, as shown in Figure 9B, the image after CLAHE stretching has more pixels in the large pixel value range for each RGB channel compared to the original image, especially the peak of the histogram of the R channel is obviously shifted to the right, and the pixel value of the R channel has significantly increased. However, due to the less information in the R channel, the R channel information of the stretched image is still less, and most of the pixel values are still distributed in the range of [0,35], so there is no obvious red color in the CLAHE processed image, and the image still shows a blue-green hue. While our method compensates for the red channel detail information of the underwater image, the histogram of the processed image shows a peak around the pixel value equal to 135, with significantly more pixels distributed in the large pixel value range than in the original image and the CLAHE-processed image, as shown in Figure 9C. The results of our method clearly eliminate the blue-green hue of the original and CLAHE processed images.

As shown in Figure 10, we compare our algorithm with CLAHE. From the comparison, it can be seen that our algorithm clearly eliminates the blue-green hue of the original image and has a significant enhancement of the details of the image. Figures 10A, C show that our method clearly eliminates the yellow-green hue in both the original image and the CLAHE processing results. Because there is no hue interference, the detailed information on the bill held in the diver's hand is clearer, and the information on the packaging of the item in Figure 10C is clearer. As shown in Figures 10B, D, our method clearly eliminates the blue hue in the Original image and CLAHE processing results, so the fish outline and the details of the rocks in the background are clearer, and our algorithm eliminates the overstretched red shadows in CLAHE as shown in Figure 10D, restoring the true shape of the image. Figure 10E, F showed that the information of the objects in the boat was clearer because the strong green hue were eliminated, so the outline of the objects in the image was clearer, and there was no exposure of the upper outline of the objects after the CLAHE method processing. Figure 10G showed that each fish in the school of fish in the image was clearer because the green hue was eliminated.

Gray World Algorithm makes the mean values of the three channels of the underwater image converge to the same gray value, which can eliminate the influence of ambient light on the image and restore the true color of the image. Figure 11 shows the comparison of the effect between the algorithm in this paper and Gray World Algorithm. As can be seen from the figure, the color balance in Gray World Algorithm makes the underwater images have serious chromatic aberration and the overall color of underwater images is reddish. As shown in Figure 11: Figures 11A, B The effect images processed by Gray World Algorithm have too much red information, which makes the image lose its real color. Besides, according to Figure 11C it can be found that the underwater image processed by Gray World Algorithm has lost details. Therefore, we compensated the image

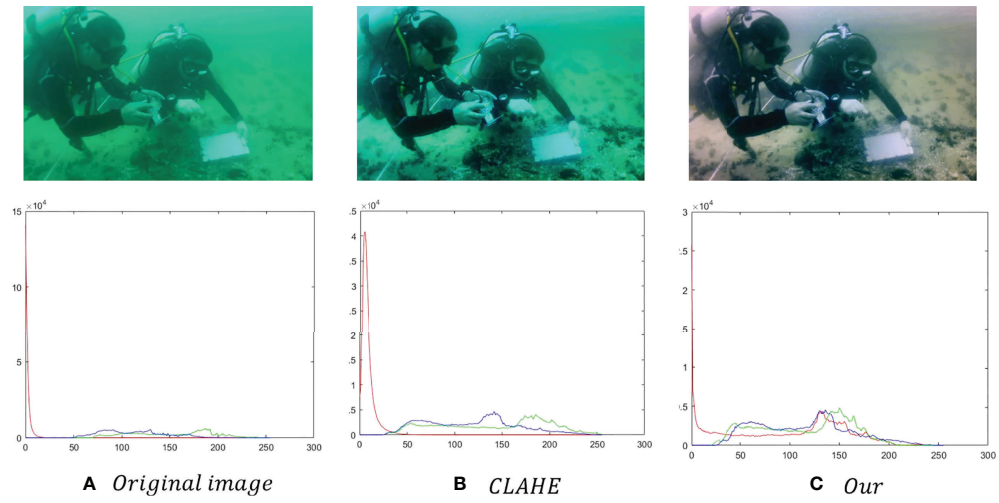


FIGURE 9
Image Contrast Stretching. **(A)** is the histogram of the original image, **(B)** is the histogram of the image stretched by the CLAHE algorithm, **(C)** is the histogram of the image stretched by our DAC.

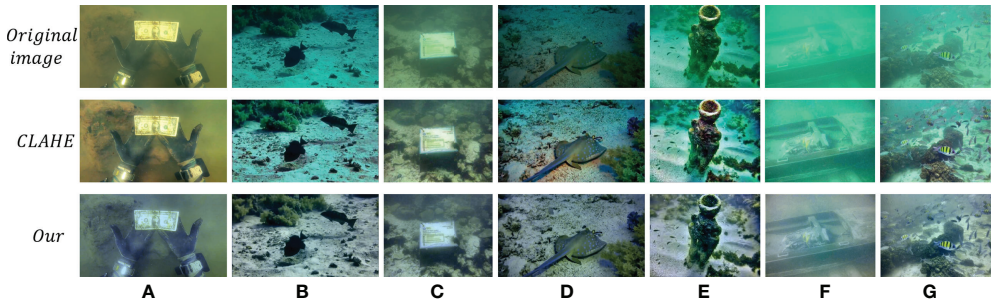


FIGURE 10
Effect of comparison with CLAHE. **(A–H)** are underwater images in various scenes processed using the CLAHE algorithm and our DAC algorithm, respectively.

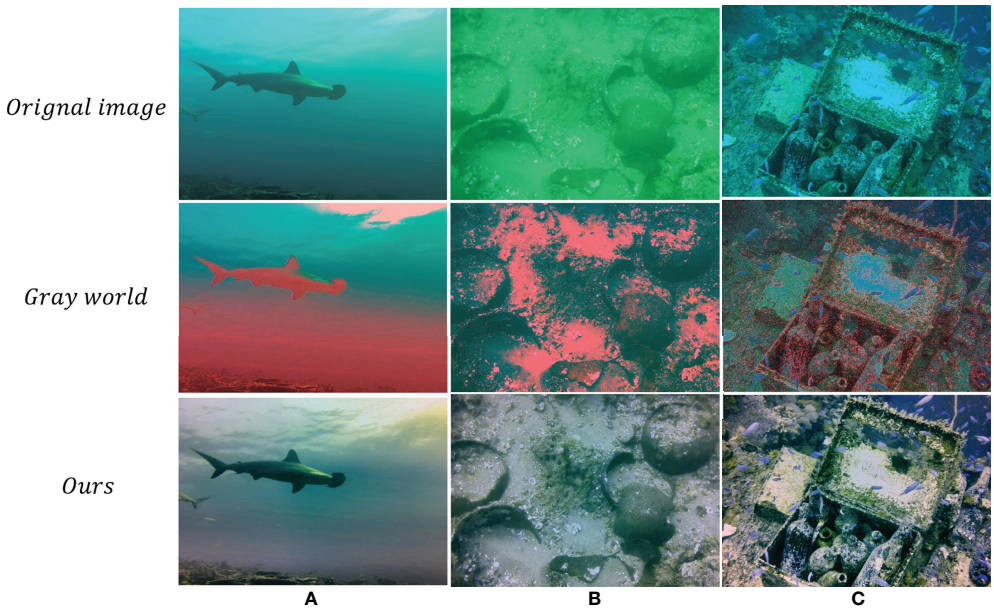


FIGURE 11
The result of comparing with Gray World Algorithm. **(A–G)** are underwater images in various scenes processed using the Gray World algorithm and our DAC algorithm, respectively.

TABLE 1 The quantitative analysis of comparing with Gray World Algorithm.

Figure 11	Method	MSE ↓	PSNR ↑	SSIM ↑
(A)	Gray world	168.66	19.55	0.931
(A)	Our	**116.86**	**22.76**	**0.948**
(B)	Gray world	109.24	18.63	0.763
(B)	Our	**98.53**	**21.09**	**0.813**
(C)	Gray world	157.614	**18.632**	0.451
(C)	Our	**100.788**	17.451	**0.763**

The bold values are the best values in the comparison result or our results. ↓ indicates that the smaller the value, the better the image quality, ↑ indicates that the larger the value, the better.

for color attenuation and obtained images with more uniform colors and richer details.

In addition, as shown in Table 1, the bold values are the better values in the quantitative comparison. we also quantitatively analyze our method with Gray World Algorithm and we can find that our algorithm has a greater advantage over Gray World Algorithm in the three metrics of MSE, PSNR and SSIM (Hore and Ziou, 2010).

To demonstrate that our algorithm has a significant improvement on underwater images, we performed a qualitative analysis to make a comparison with UCM (Iqbal et al., 2010), UDCP, ULAP (Song et al., 2018), CLAHE, Gray World Algorithm, Image Information (Chen et al., 2021) and FUnIE-GAN based on generative adversarial networks, and a comparison of the experimental results is shown in the following Figure 12.

In addition, we conduct the Mean Opinion Score (MOS) test as subjective test. We find 25 volunteers on university campuses to evaluate the images in Figure 12. The full score is 5 points. 25 volunteers are made up of teachers and students. The mean score of 25 volunteers is shown in the Table 2. The bold values are the best value in the comparison. The enhanced images by our DAC method get the max score in each comparison. This also means that the enhanced images by our DAC have obvious advantages in the subjective test and more in line with human visual aesthetics. Everyone can find the table of result scores by subjective test in https://github.com/lailaiyun/Single-Underwater-Image-Enhancement-Based-on-Differential-Attenuation-Compensation.

As shown in Figure 12, by comparing with other algorithms, we can see that our algorithm has a significant enhancement effect on the image. The underwater image we processed attenuates the color shift of the image, reduces the blue and green hue of the image, and restores the true color of the image, making it visually closer to the real sense of land. In addition, we also enhance the image clarity and make the image more detailed. UCM stretches the image in RGB and HSV color space to enhance the quality of the image. FUnIE-GAN and Image formation all claim to improve the contrast of the image. However, through experimental comparison, it can be found that our DAC method is significantly better than their processing results in improving the contrast and details of underwater images. Therefore, our method has obvious advantages compared with the general method of improving image quality by stretching.

Table 3 shows the results of the quantitative analysis of our method with other methods for MSE, PSNR, SSIM, and UCIQE metrics. The bold values are our results in comparison. The analysis of the above table shows that the results of our method are only larger than those of the Image Formation method in the

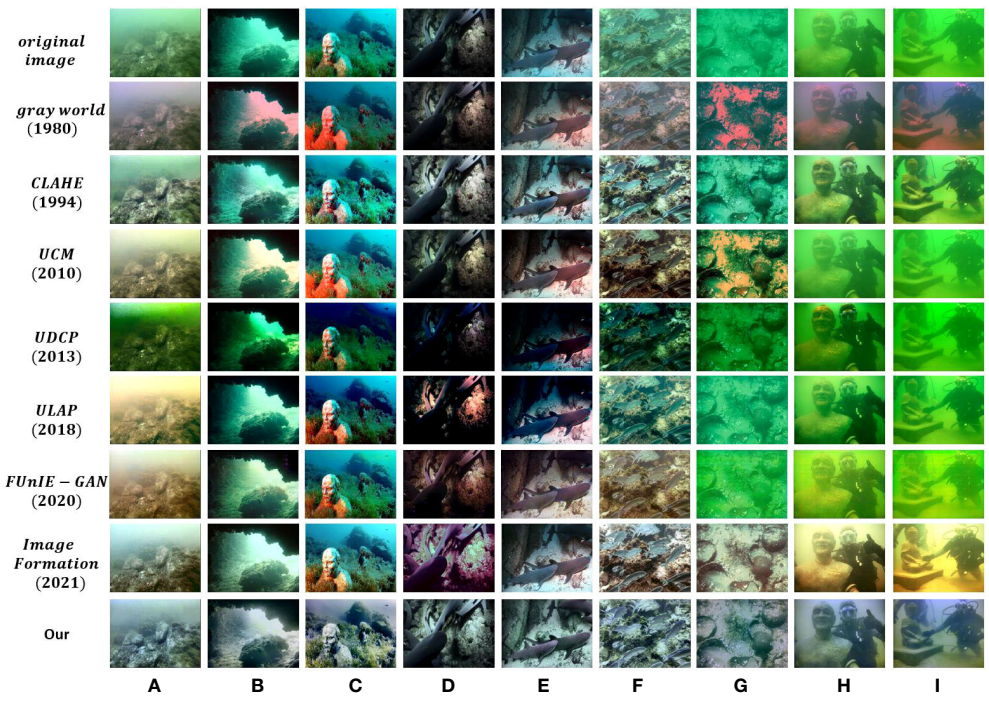


FIGURE 12
The DAC compares with classical and state-of-the-art methods on the UIEB dataset. **(A–I)** is the comparison results of original image with UCM, UDCP, ULAP, CLAHE, gray world, FUnIE-GAN, Image Information and our DAC in each underwater scene.

TABLE 2 MOS of images in Figure 12.

MOS ↑	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)
UCM	3	2.48	2.28	2.8	2.84	2.84	1.8	1.72	1.84
UDCP	1.76	2	1.8	1.8	1.92	2.12	2.08	2.2	1.96
ULAP	2.44	2.76	2.92	1.96	2.48	2.76	2.44	2.28	2.08
CLAHE	3.68	3.44	3.4	3.6	3.72	3.72	3.16	3.12	2.96
Gray World	2.24	1.8	2.08	2.32	2.96	2.2	1.52	1.8	1.96
FunIE-GAN	2.4	2.88	2.68	2.08	2.8	2.68	1.84	2.12	2.04
Image Formation	3.36	2.8	3.16	2.36	3.4	3.72	3.72	2.64	2.8
Our	**4.08**	**4.08**	**4.04**	**4.28**	**4.72**	**4.48**	**4.04**	**4**	**4.2**

The bold values are the best values in the comparison result or our results. ↑ indicates that the larger the value, the better.

MSE comparison. The results of our method are only smaller than the Image Formation method in the PSNR comparison, but the results of our method are much higher than the Image Formation method in both the SSIM and UCIQE comparisons. And our method achieved the maximum value in the comparison of SSIM and only scored less than the UCM method in the comparison of UCIQE. Overall, it seems that our method still has a clear advantage over other methods in quantitative analysis.

We also compared the details of the images for each method. The details of the stone statue and the box were compared as shown in Figures 13, 14. In Figure 13 it can be found that other algorithms have some exposure on the stone image. Our algorithm corrects the problem of uneven illumination of the stone image, enhancing the dark areas and weakening the bright areas of the image. And it can be found in Figure 14 that the image processed by the algorithm of this paper has more details and the text part is clearer, which effectively proves the improvement of the image details by the algorithm of this paper.

4.2 Analysis for enhancing underwater visual perception dataset

In addition to the UIEB dataset, the experiments in this paper also use the synthetic underwater image dataset EUVP provided by (Islam et al., 2020) in FUnIE-GAN. This dataset consists of unpaired data and paired data. We select the Underwater Dark in the paired data as dataset of our experiment, and randomly selected 8 underwater images for qualitative and quantitative analysis.

In addition, we also conduct the Mean Opinion Score (MOS) test as subjective test. We find 25 volunteers on university campuses to evaluate the images in Figure 15. The full score is 5 points. 25 volunteers are made up of teachers and students. The mean score of 25 volunteers is shown in the Table 4. The bold values are the best values in the MOS test. The enhanced images by DAC get the max score in each comparison. This also means that the enhanced images by our DAC have obvious advantages in the subjective test and more in line with human visual aesthetics. Everyone can find the table of result scores by subjective test in https://github.com/lailaiyun/Single-Underwater-Image-Enhancement-Based-on-Differential-Attenuation-Compensation.

As shown in Figure 15, the algorithms in this paper are shown the comparison of the image processing results with the traditional defogging algorithm CLAHE, DCP, the proposed UDCP based on the DCP method, the relative global histogram stretching(RGHS) method proposed by Huang (Huang et al., 2018), Underwater Light Appreciation Prior (ULAP), and Compressed Model for Underwater Image Enhancement (UWnet) proposed by Ankita (Naik et al., 2021).

The results of the CLAHE, DCP, UDCP, and RGHS methods removed the blurring of the images, but did not

TABLE 3 The quantitative analysis of comparing with other methods on UIEB dataset. Each value is the mean of the processing results of each method in UIEB dataset.

Method	MSE ↓	PSNR ↑	SSIM ↑	UCIQE↑
UCM	109.5064	20.9337	0.857	0.6408
UDCP	224.3492	11.537	0.5564	0.5816
ULAP	118.7066	19.5166	0.8368	0.6025
CLAHE	100.1878	19.9017	0.8884	0.6247
Gray World	114.1646	19.3598	0.8109	0.5617
Image Formation	63.4474	24.4169	0.8293	0.5133
Our	**92.8167**	**20.9417**	**0.8885**	**0.6164**

The bold values are the best values in the comparison result or our results. ↓ indicates that the smaller the value, the better the image quality, ↑ indicates that the larger the value, the better.

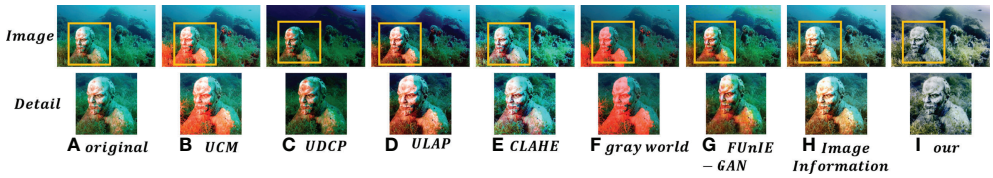


FIGURE 13
Comparison of stone statue detail. **(A–I)** are the detail information comparison of the original image with UCM, UDCP, ULAP, CLAHE, gray world, FUnIE-GAN, Image Information and our DAC processed stone image, respectively.

eliminate the blue-green hue of the underwater images. The resultant image details after Image Formation method and UWnet method processing are not as clear as the details of the image after our DAC method processing, as shown in the comparison of Figure 15D, E), the sea urchin spine after DAC method processing has more detail information.

As shown in Table 5, the data reflects the comparison of each method. , the bold values are our results in the comparison. Our DAC results achieve good values in MSE and SSIM, and outperform most other methods, this means our DAC results are similar to ground truth. From our experiences, larger PSNR value do not seem to correlate with better visual quality of images in many cases. Therefore, although the PSNR value of the DAC result is not large, the visual effect is indeed better than most methods, the detailed analysis about are shown in subjective tests. Our DAC result get the lowest mean value UCIQE in comparing. This is because UCIQE is a color related metric, while the EUVP dataset is a synthetic dataset, and the corresponding ground truth colors are richer, but the color of the underwater environment in the real world is dull, as shown in the first row of Figure 16. The color of the image processed by our DAC method is more consistent with the actual situation.

5 Application

On the one hand, underwater image enhancement tasks can provide high quality underwater images and videos that conform to human visual habits, and on the other hand, they also serve as a basis for other underwater development tasks by enhancing the quality of underwater images and videos to improve the robustness and accuracy of tasks such as underwater target detection. In recent years, many researchers have proposed some new research ideas by combining the underwater image enhancement task with the object detection task. Yeh et al. (Yeh et al., 2021) proposed a light-weight deep neural network (LDN), the network contains color conversion network and Object detection network, which quantitatively proved that correcting the color of underwater images can improve the accuracy of underwater target detection. Liu et al. (Liu et al., 2022) solved the problem of low contrast and loss of color in underwater images by a self-adaptive global histogram method and introduced the convolutional block attention module (CBAM) in YOLO v5 to adapt the network to target detection in underwater environments. Zhao et al. (Zhao et al., 2021) designed a new composite backbone network (CBresnet) and an enhanced path aggregation network (EPANet) by improving the residual network (ResNet) to form a novel composite fish detection framework. The method demonstrates a strong detection capability for underwater environmental targets.

Although we do not propose new methods applicable to underwater target detection based on underwater image enhancement in this paper, we have actually been working on target detection tasks in underwater environments for many years, and we have shown our latest progress in our other work. As shown in the Figure 17 and Table 6 we briefly demonstrate

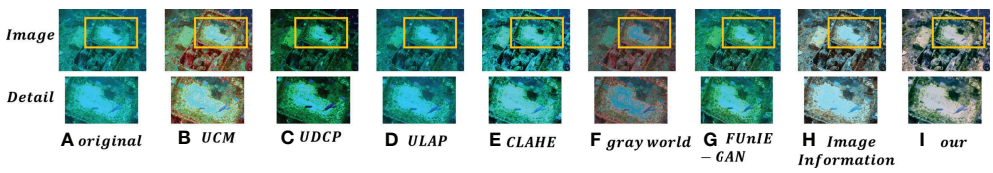


FIGURE 14
Comparison of text on the box. **(A–I)** are the textual information comparison of the original image with UCM, UDCP, ULAP, CLAHE, gray world, FUnIE-GAN, Image Information and our DAC processed box image, respectively.

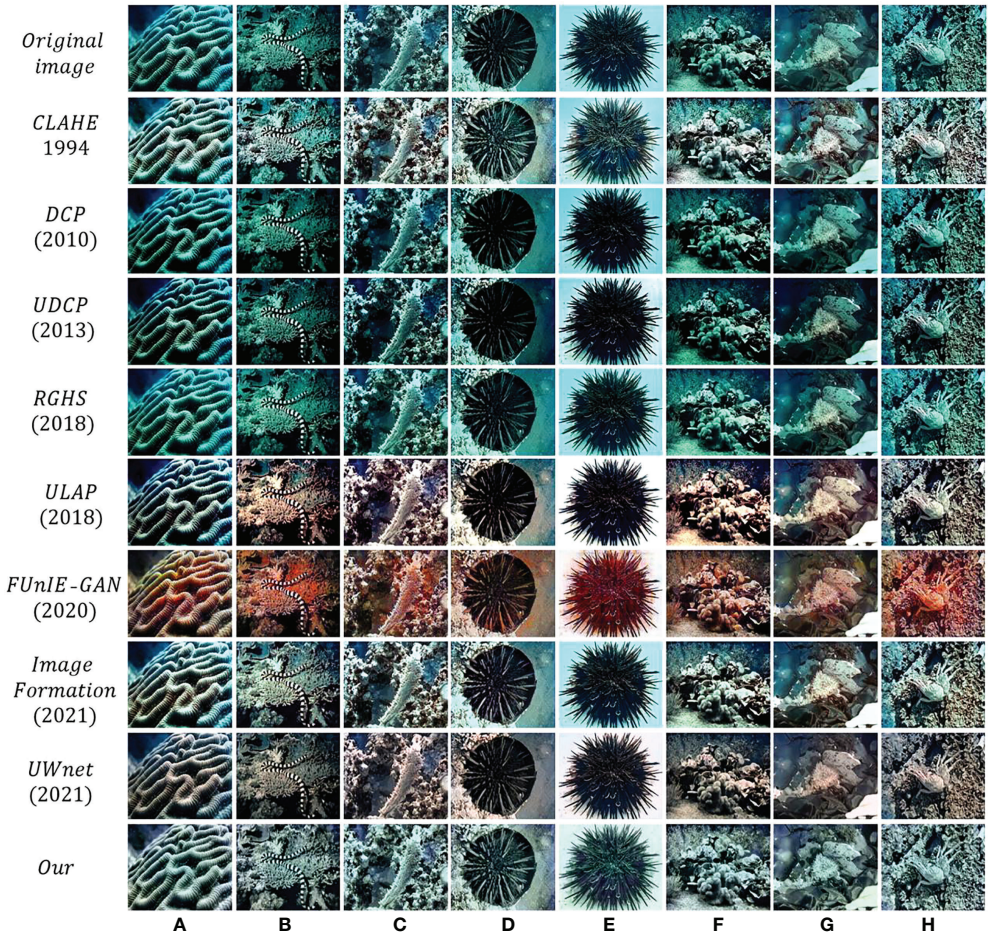


FIGURE 15
The DAC compares with classical and state-of-the-art method on the EUVP dataset. **(A–H)** is the comparison results of original image with CLAHE, DCP, UDCP, RGHS, ULAP, FUnIE-GAN, Image Information, UWnet and our DAC in each underwater scene.

the performance improvement of the DAC method mentioned in this paper for the underwater target detection task.

We have enhanced the underwater images provided by 2022 China Underwater robot professional contest, Everyone can get it from: http://www.urpc.org.cn/index.html. It can be found that the target information of the processed image detection becomes clear. We applied the processed image to YOLO v7 (Wang et al., 2022) target detection, and it can be found that more targets can

TABLE 4 MOS of images in Figure 15.

MOS	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)
CLAHE	3.28	3.44	3.32	3.4	3.48	3.2	3.48	3.52
DCP	2.76	2.64	2.64	2.84	2.76	2.56	2.72	2.76
UDCP	2.88	2.72	2.48	2.56	2.8	2.28	2.28	2.84
RGHS	2.88	2.76	2.84	2.84	2.84	2.72	2.48	2.76
ULAP	3.28	2.96	2.92	3	3.44	2.56	2.6	3.36
FunIE-GAN	2.24	2.12	1.96	2.12	1.92	2.08	1.84	1.72
Image Formation	3.2	3.2	3.12	3.24	3.08	3.2	2.84	3.52
UW net	3.16	3.16	2.72	3.04	3.64	3	2.64	3.36
Our	**3.56**	**3.72**	**3.48**	**4.2**	**4.08**	**3.6**	**3.6**	**3.68**

The bold values are the best values in the comparison result or our results.

TABLE 5 The quantitative analysis of comparing with other methods on EUVP dataset. Each value is the mean of the processing results of each method in EUVP dataset.

Method	MSE ↓	PSNR ↑	SSIM ↑	UCIQE ↑
CLAHE	82.2157	19.5938	0.8324	0.5861
DCP	120.48	20.7352	0.8848	0.5781
UDCP	179.685	15.4622	0.746	0.595
RGHS	64.5001	25.7784	0.9282	0.5903
ULAP	87.8695	22.4938	0.85347	0.5887
FUnIE-GAN	81.2569	23.3168	0.7747	0.5838
Image Formation	68.716	23.5785	0.909	0.6002
UWnet	80.0802	25.7124	0.8482	0.5411
Our	**87.8287**	**19.3929**	**0.8363**	**0.5321**

The bold values are the best values in the comparison result or our results. ↓ indicates that the smaller the value, the better the image quality, ↑ indicates that the larger the value, the better.

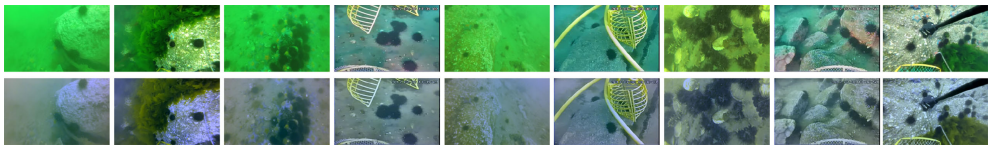


FIGURE 16
Original underwater images provided by 2022 China Underwater robot professional contest, and DAC results of these images.

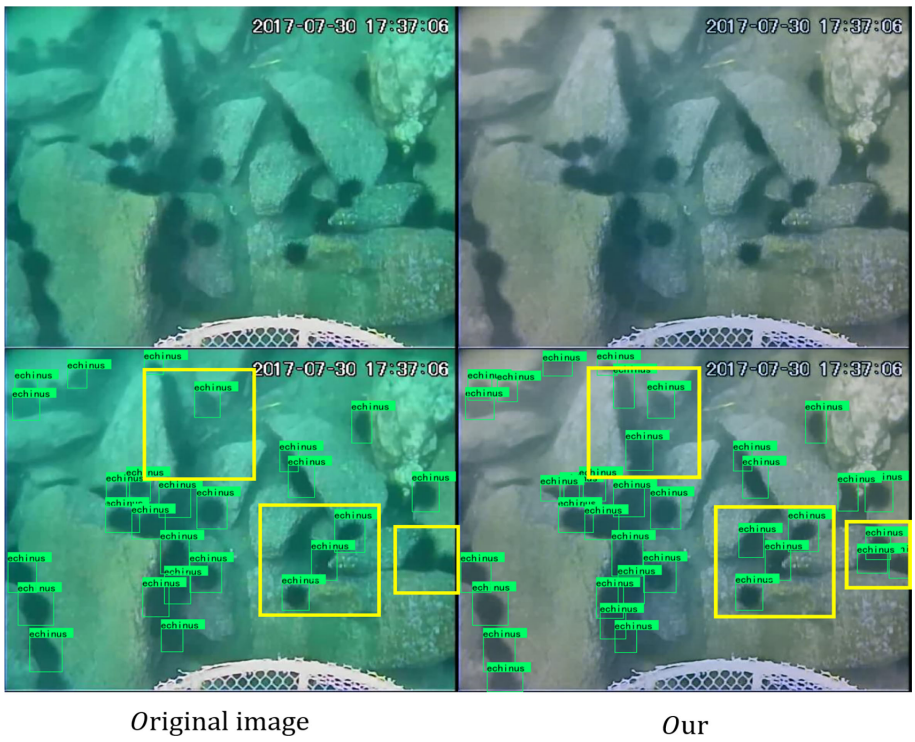


FIGURE 17
Comparison of YOLO v7 target detection results.

TABLE 6 All mAPs for original underwater images dataset and enhanced underwater images dataset by CLAHE, DCP, UDCP, RGHS, ULAP, Image Formation and DAC method, are got at the 100th epoch.

Dataset	mAP(%) ↑
Original Image	74.32
Enhanced Image by CLAHE	73.91
Enhanced Image by DCP	72.70
Enhanced Image by UDCP	75.26
Enhanced Image by RGHS	73.68
Enhanced Image by ULAP	75.79
Enhanced Image by Image Formation	77.02
Enhanced Image by DAC	77.49

↑ indicates that the larger the value, the better.

be detected in the enhanced image, as shown in Figure 17. Application experiments related to YOLO v7 in the environment with Intel(R) Core (TM) i7-12700KF@3.61 GHz CPU, 16GB RAM, NVIDIA GeForce RTX 3080 Ti graphics card, Windows 10 Professional, Python version 3.8, CUDA version 11.6, and Adam gradient descent optimizer. PyTorch version 1.12.0, CUDA version 11.6. Gradient descent optimizer is Adam. Learning rate update during training is step. Maximum learning rate is 0.001. Frozen training batch size is 8, unfrozen training batch size is 4. Momentum is 0.937.

As shown in Figure 16, the first row are original images, the second row are DAC results of these images. These original images provided by 2022 China Underwater robot professional contest, these images are all collected from real underwater scenes, we can clearly see that the colors of the images are relatively monotonous. This reflects the fact that the underwater image in the real world has dull color. Therefore, parameters related to color will not get high value. This may explain the lower values of color related UCIQE parameter in our quantitative analysis in Table 5 on the synthetic underwater dataset EUVP in Section 4.2.

As shown in the Figure 17, the top left image in the figure shows the original unprocessed image, the bottom left image shows the detection result of the unprocessed image, the top right image shows the underwater image after the DAC method, and the bottom right image shows the detection result of the processed underwater image. As the yellow box in the figure shows, the processed underwater image is able to identify more sea urchins.

In the comparison experiment, we process the original images provided by 2022 China Underwater robot professional contest with CLAHE, DCP, UDCP, RGHS, ULAP, Image Formation and our DAC method respectively, and do not change the label, and then train the YOLO v7 model. The model can realize underwater target detection after trained. Because the output images size of the FUnIE-GAN and UWnet methods changes, resulting in a mismatch between the image and the label, we do not compare the improvement of YOLO v7 with the images processed by the FUnIE-GAN and UWnet methods. At the 100th epoch, we get the original underwater image dataset and the model mAPs of the datasets processed by each method. As shown in Table 6, dataset enhanced by CLAHE, DCP, RGHS, the corresponding mAP is lower than that of the Original Image, which may be related to the distortion of some images processed by these methods. The mAP values corresponding to the datasets enhanced by UDCP, ULAP, Image Formation, and DAC methods have been significantly improved, and the mAP value corresponding to the dataset processed by the DAC method is the largest, get the increase of 3.17%. This proves that the enhanced underwater images by our DAC method can significantly improve the performance of related underwater object detection and outperform other methods.

After demonstrating the performance improvement of our DAC method on the task of underwater image object detection. In order to verify that the underwater images processed by our algorithm improve the efficiency of the vision task, it is shown

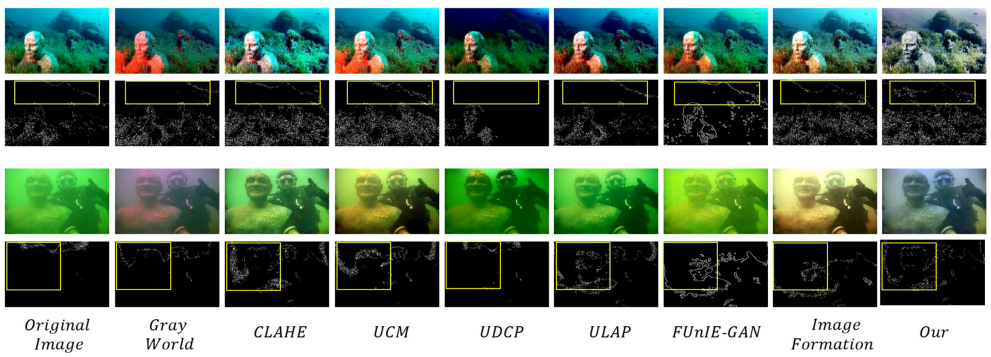


FIGURE 18
Canny operator edge detection results.

that we also applied our processed images to the edge detection of the Canny operator (Canny, 1986), as shown in Figure 18, there is the comparison that the Canny edge detection results of the original image and the image enhanced by each method.

In Figure 18, the first and third rows are the original image and the result images after processing by each method, and the second and fourth rows are the results of Canny edge detection corresponding to each image. In the comparison of the second row, the outline of the background stone in the image enhanced by our method is clearer, and the outline of the seagrass in the image is also clear. In the comparison in the fourth row, the outlines of the stone in the images processed by our method are clearer and the background noise is less. This proves that our DAC method improves the quality of underwater images to a certain extent and can improve the performance of underwater exploitation related tasks.

6 Conclusion

In this paper, we propose a pixel processing-based underwater image enhancement method DAC, which decomposes the information of each channel of the underwater image in RGB color space based on the characteristic of differential attenuation of different wavelengths of light in the underwater environment, compensates the R channel detail information of the image, eliminates the image blur and corrects the image chromatic aberration, and obtains the underwater image closer to the real color. In this paper, we propose a pixel processing-based underwater image enhancement method DAC, which decomposes the information of each channel of the underwater image in RGB color space based on the characteristic of differential attenuation of different wavelengths of light in the underwater environment, compensates the R channel detail information of the image, eliminates the image blur and corrects the image chromatic aberration, and obtains the underwater image closer to the real color. To verify the effectiveness of the DAC algorithm, we demonstrate the superiority of our algorithm by qualitative and quantitative analysis in the experimental section and the application section.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: UIEB:https://li-chongyi.github.io/proj_benchmark.html. EUVP:https://irvlab.cs.umn.edu/resources/euvp-dataset.

Ethics statement

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

YL completes the main work of this paper. ZZ, BS and ZX guide YL in the research of this work. JT, WL, JC and JY help complete the part of experiment. All authors are involved in revising the manuscript, proofreading, and approving the submit.

Funding

Our work was supported in part by the Key Platform and Scientific Research Project for General Universities, Guangdong, 2022, grant 2022ZDZX4061, in part by the Science and Technology Program of Social Development, Zhuhai, 2022, grant 2220004000195, in part by the Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, BNU-HKBU United International Collegeproject code 2022B1212010006 and in part by Guangdong Higher Education Upgrading Plan (2021-2025) with UIC research grant R0400001-22 and R201902.

Acknowledgments

Throughout the writing of this thesis. I have received a great deal of support and assistance. I would like to express my gratitude to all those who helped me during the writing of this thesis. My deepest gratitude goes first and foremost to my supervisor, Zhuang Zhou, for his constant encouragement and guidance. He has walked me though all the stages of the writing of this thesis. Without his consistent and illuminating instruction, this thesis could not have reached its present form. Second, I would like to express my heartfelt gratitude to professor Binghua Su, who provided me a good platform to complete my thesis. Finally, I am indebted to my parents, other authors of this paper, and my friends for their continuous support and encouragement.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdullah-Al-Wadud, M., Kabir, M. H., Dewan, M. A. A., and Chae, O. (2007). A dynamic histogram equalization for image contrast enhancement. *IEEE Trans. Consum. Electron.* 53, 593–600. doi: 10.1109/TCE.2007.381734

Akkaynak, D., and Treibitz, T. (2018). "A revised underwater image formation model," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. Vol. 19. (Salt Lake City, UT, USA: IEEE), 6723–6732. doi: 10.1109/CVPR.2018.00703

Akkaynak, D., and Treibitz, T. (2019). Sea-Thru: A method for removing water from underwater images. *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.* 20, 1682–1691. doi: 10.1109/CVPR.2019.00178

Ancuti, C. O., Ancuti, C., De Vleeschouwer, C., and Bekaert, P. (2017). Color balance and fusion for underwater image enhancement. *IEEE Trans. image Process.* 27, 379–393. doi: 10.1109/TIP.2017.2759252

Ancuti, C., Ancuti, C. O., Haber, T., and Bekaert, P. (2012). "Enhancing underwater images and videos by fusion," in *2012 IEEE conference on computer vision and pattern recognition*, Vol. 12. (IEEE), 81–88.

Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 679–698. doi: 10.1109/TPAMI.1986.4767851

Chen, X., Zhang, P., Quan, L., Yi, C., and Lu, C. (2021). Underwater image enhancement based on deep learning and image formation model. *arXiv* 33, 2101.00991. doi: 10.48550/arXiv.2101.00991

Clarke, G. L., and James, H. R. (1939). Laboratory analysis of the selective absorption of light by sea water. *JOSA* 29, 43–55 3. doi: 10.1364/JOSA.29.000043

Drews, P., Nascimento, E., Moraes, F., Botelho, S., and Campos, M. (2013). "Transmission estimation in underwater single images," in *Proceedings of the IEEE international conference on computer vision workshops*, Vol. 825–830. (Sydney, NSW, Australia: IEEE), 18. doi: 10.1109/ICCVW.2013.113

Egorenko, M. P., and Efremov, V. S. (2020). "Mirror-lens camera system for underwater drones," in *26th international symposium on atmospheric and ocean optics, atmospheric physics*, (SPIE) 11560, 601–605.6.

Engin, D., Genç, A., and Kemal Ekenel, H. (2018). "Cycle-dehaze: Enhanced cyclegan for single image dehazing," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, Vol. 825–833. (Lake City, UT, USA: IEEE), 21. doi: 10.1109/CVPRW.2018.00127

Fu, X., Fan, Z., Ling, M., Huang, Y., and Ding, X. (2017). "Two-step approach for single underwater image enhancement," in *2017 international symposium on intelligent signal processing and communication systems (ISPACS)*, vol. 789–794. (IEEE), 27.

He, K., Sun, J., and Tang, X. (2010). Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 2341–2353. doi: 10.1109/TPAMI.2010.168

Hore, A., and Ziou, D. (2010). "Image quality metrics: Psnr vs. ssim," in *2010 20th international conference on pattern recognition*, Vol. 2366–2369. (IEEE), 30.

Howell, K. L., Hilário, A., Allcock, A. L., Bailey, D., Baker, M., Clark, M. R., et al. (2021). A decade to study deep-sea life. *Nat. Ecol. Evol.* 5, 265–267. doi: 10.1038/s41559-020-01352-5

Huang, D., Wang, Y., Song, W., Sequeira, J., and Mavromatis, S. (2018). "Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition," in *International conference on multimedia modeling*. (Springer), 453–465.

Iqbal, K., Odetayo, M., James, A., Salam, R. A., and Talib, A. Z. H. (2010). "Enhancing the low quality images using unsupervised colour correction method," in *In 2010 IEEE International Conference on Systems, Man and Cybernetics*, Vol. 31. (IEEE), 1703–1709.

Islam, M. J., Xia, Y., and Sattar, J. (2020). Fast underwater image enhancement for improved visual perception. *IEEE Robotics Automation Lett.* 5, 3227–3234 24. doi: 10.1109/LRA.2020.2974710

Kaur, M., Kaur, J., and Kaur, J. (2011). Survey of contrast enhancement techniques based on histogram equalization. *Int. J. Advanced Comput. Sci. Appl.* 2. doi: 10.14569/IJACSA.2011.020721

kumar Rai, R., Gour, P., and Singh, B. (2012). Underwater image segmentation using clahe enhancement and thresholding. *Int. J. Emerg. Technol. Advanced Eng.* 2, 118–123.26.

Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., et al. (2019). An underwater image enhancement benchmark dataset and beyond. *IEEE Trans. Image Process.* 29, 4376–4389 28. doi: 10.1109/TIP.2019.2955241

Li, Y., Lu, H., Li, K.-C., Kim, H., and Serikawa, S. (2018). Non-uniform de-scattering and de-blurring of underwater images. *Mobile Networks Appl.* 23, 352–362 16. doi: 10.1007/s11036-017-0933-7

Liu, Z., Zhuang, Y., Jia, P., Wu, C., Xu, H., and Liu, Z. (2022). A novel underwater image enhancement algorithm and an improved underwater biological detection pipeline. *J. Mar. Sci. Eng.* 10, 1204 36. doi: 10.3390/jmse10091204

Lu, H., Li, Y., Xu, X., He, L., Li, Y., Dansereau, D., et al. (2016). "Underwater image descattering and quality assessment," in *Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP 2016) (Institute of Electrical and Electronics Engineers Inc.)*, Vol.1998–2002. (Phoenix, AZ, USA: IEEE), 15. doi: 10.1109/ICIP.2016.7532708

McGlamery, B. (1980). A computer model for underwater camera systems. *In Ocean Optics VI (SPIE)* 208, 221–231 14. doi: 10.1117/12.958279

Naik, A., Swarnakar, A., and Mittal, K. (2021). "Shallow-uwnet: Compressed model for underwater image enhancement (student abstract)," in *Proceedings of the AAAI conference on artificial Intelligence*, Vol. 35. (Salt Lake City, UT, USA), 15853–15854.

Ortiz, A., Simó, M., and Oliver, G. (2002). A vision system for an underwater cable tracker. *Mach. Vision Appl.* 13, 129–140 1. doi: 10.1007/s001380100065

Phillips, B. T., Licht, S., Haiat, K. S., Bonney, J., Allder, J., Chaloux, N., et al. (2019). Deepi: A miniaturized, robust, and economical camera and computer system for deep-sea exploration. *Deep Sea Res. Part I: Oceanographic Res. Papers* 153, 1031365. doi: 10.1016/j.dsr.2019.103136

Pizer, S. M., Amburn, E. P., Austin, J. D., Cromartie, R., Geselowitz, A., Greer, T., et al. (1987). Adaptive histogram equalization and its variations. *Comput. vision graphics image Process.* 39, 355–368 10. doi: 10.1016/S0734-189X(87)80186-X

Reza, A. M. (2004). Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *J. VLSI Signal Process. Syst. signal image video Technol.* 38, 35–44 25. doi: 10.1023/B:VLSI.0000028532.53893.82

Sahu, P., Gupta, N., and Sharma, N. (2014). A survey on underwater image enhancement techniques. *Int. J. Comput. Appl.* 87, 7. doi: 10.5120/15268-3743

Song, W., Wang, Y., Huang, D., and Tjondronegoro, D. (2018). "A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration," in *Pacific rim conference on multimedia*, vol. 32. (Springer), 678–688.

Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M. (2022). Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* 38. doi: 10.48550/arXiv.2207.02696

Wang, N., Zhou, Y., Han, F., Zhu, H., and Yao, J. (2019). Uwgan: underwater gan for real-world underwater color restoration and dehazing. *arXiv* 23, 1912.10269. doi: 10.48550/arXiv.1912.10269

Wozniak, B., and Dera, J. (2007). *Light absorption by suspended particulate matter (SPM) in sea water* Vol. 4 (Springer).

Yang, M., and Sowmya, A. (2015). An underwater color image quality evaluation metric. *IEEE Trans. Image Process.* 24, 6062–6071 29. doi: 10.1109/TIP.2015.2491020

Yang, N., Zhong, Q., Li, K., Cong, R., Zhao, Y., and Kwong, S. (2021).). a reference-free underwater image quality assessment metric in frequency domain. *Signal Process.: Image Commun.* 94, 116218 13. doi: 10.1016/j.image.2021.116218

Yeh, C.-H., Lin, C.-H., Kang, L.-W., Huang, C.-H., Lin, M.-H., Chang, C.-Y., et al. (2021). Lightweight deep neural network for joint learning of underwater object detection and color conversion. *IEEE Trans. Neural Networks Learn. Syst.* 35. doi: 10.1109/TNNLS.2021.3072414

Zhao, Z., Liu, Y., Sun, X., Liu, J., Yang, X., and Zhou, C. (2021). Composited fishnet: Fish detection and species recognition from low-quality underwater videos. *IEEE Trans. Image Process.* 30, 4719–4734 37. doi: 10.1109/TIP.2021.3074738

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *In Proceedings of the IEEE international conference on computer vision*, Vol. 22. (Salt Lake City, UT, USA), 2223–2232.

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 15 November 2022
DOI 10.3389/fmars.2022.1030113

Check for updates

OPEN ACCESS

EDITED BY
Hong Song,
Zhejiang University, China

REVIEWED BY
Inam Ullah,
Shandong Jianzhu University, China
Shafqat Khan,
University of Technology Malaysia, Malaysia

*CORRESPONDENCE
Salman Jan
salman.jan@unikl.edu.my

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 28 August 2022
ACCEPTED 25 October 2022
PUBLISHED 15 November 2022

CITATION
Ahmad I, Rahman T, Khan I, Jan S, Musa S and Uddin MI (2022) RACE-SM: Reliability and adaptive cooperation for efficient UWSNs using sink mobility.
Front. Mar. Sci. 9:1030113.
doi: 10.3389/fmars.2022.1030113

COPYRIGHT
© 2022 Ahmad, Rahman, Khan, Jan, Musa and Uddin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

RACE-SM: Reliability and adaptive cooperation for efficient UWSNs using sink mobility

Irfan Ahmad[1], Taj Rahman[1], Inayat Khan[2], Salman Jan[3,4]*, Shahrulniza Musa[3] and M. Irfan Uddin[5]

[1]Department of Physical & Numerical Science, Qurtuba University of Science & Information Technology, Peshawar, Pakistan, [2]Department of Computer Science, University of Engineering & Technology, Mardan, Pakistan, [3]Malaysian Institute of Information Technology, Universiti Kuala Lumpur, Kuala Lumpur, Malaysia, [4]Department of Computer Science, Bacha Khan University, Charsadda, Pakistan, [5]Institute of Computing, Kohat University of Science and Technology, Kohat, Pakistan

Underwater Wireless Sensor Networks (UWSNs) are the most crucial method for exploring the hidden resources under the water. It enables many underwater applications, such as military, commercial, disaster prevention, ocean sampling, and other emergencies. Data transmission through a single relay node creates a hotspot, which will minimize the network lifetime and reduce the network reliability. Therefore, the cooperative technique is essential for transferring data between the source and the destination. This research proposes an improved version of Reliability and Adaptive Cooperation for Efficient (RACE), a well-known cooperative routing protocol for UWSNs known as RACE-SM. RACE-SM solved the single relay node issues by using the sink mobility scheme. All sensor nodes transfer data directly to the sink node if the sink node is in the communication range. Otherwise, sensor nodes use the cooperative combining strategies scheme to send the data from the source to the destination or sink node. The performance of the proposed method is then compared with the current protocols. The simulation results show that the RACE-SM outperforms in average up to 40.60%, 59%, 278%, and 77% than current protocols in terms of alive nodes, energy consumption, packet delivery ratio (PDR), and end-to-end delay respectively.

KEYWORDS

UWSNs, energy efficient routing, routing protocols, sink mobility, cooperative routing

1 Background

Around 75% of the Earth's surface is covered by water. There are multiple applications for underwater sensor networks (UWSNs), such as oil and gas, various types of mines, sea surveillance, etc. (Nayyar et al., 2018; Ahmad et al., 2022). Approximately 10% of the total ocean volume has been explored throughout human history (Nayyar et al., 2019; Ahmad et al., 2021).

In UWSNs, many sensor nodes are available in the network (Erol et al., 2007; Raj Priyadarshini and Sivakumar, 2020). These nodes are responsible for sensing the data and forwarding it to the sink node or base station. The data packets received at the sink node then proceed to multiple applications. The accuracy of data packets received at the sink node depends on error-free communication. Radio waves are primarily not used for communication in UWSNs environments because they are easily affected by water (Teekaraman et al., 2019). Instead of radio waves, UWSNs use acoustic waves for communication due to their best performance. There is multiple uniqueness between the UWSNs and underwater acoustic networks in terms of scalability, flexibility, and localization (Nayyar and Balas, 2019; Anwar et al., 2021).

One of the significant concerns in UWSNs is reducing energy consumption. The node battery replacement is challenging due to the UWSN's nature. However, the researchers are focusing on designing innovative routing schemes to have error-free communication and a better network lifetime. When designing the routing protocol, it's essential to identify the minimum power of the node battery, transmission loss, and packet delivery ratio (Teekaraman et al., 2019; John et al., 2020; Zeb et al., 2022). The sink mobility technique is the best way to reduce the energy consumption in UWSNs. In this method, sink nodes move around the predefined area and collect the data from the sensor nodes.

Network reliability is another of the most critical challenges for underwater communication. The cooperative routing technique is the most reliable way to achieve network reliability. Cooperative routing routes data through a multihop system to ensure network reliability (Shah et al., 2018; Yahya et al., 2019; Mhemed et al., 2021; Zeb et al., 2022). The cooperative strategy also helps decrease the channel effects and provides reliable data transmission to the sink node.

1.1 Motivation

There are many problems that UWSNs are faced with them now a days such as replacement of sensor node battery, propagation delay, limited bandwidth, and minimizing energy consumption. and charging the UWSNs deployed sensors are very difficult in real environments. Data forwarding through multiple nodes from source to destination is better to solve the network reliability issues. Still, the latency will be high due to various nodes in the data processing. The existing Reliability and Adaptive Cooperation for Efficient UWSNs (RACE) uses a single hop method for data transmission. Using a single node for data transmission creates the hotspot. Therefore, the forwarder node drains energy quickly due to the continuous transmission of the data packets toward the sink, which will cause to minimize the network lifetime (Ahmed et al., 2016b).

Due to the aforementioned limitations, this research article introduces an enhanced version of RACE called Reliability and Adaptive Cooperation for Efficient UWSNs using sink mobility (RACE-SM). The proposed method transfers data to the sink node in two ways. First, when the sensor node detects the sink node, the communication range will transfer the data directly to the sink node instead of the multi-hopping method. Second, data are transmitted through the relay nodes. When the sink node is not in the communication range, the sensor nodes transfer the data to relay nodes and then to the sink node.

1.2 Objectives

- To minimize the energy consumption, the RACE-SM scheme is developed for UWSNs in which data are transferred to the mobile sink directly rather than *via* a multi-hopping technique. The deployment area is partitioned into four portions for better performance and to avoid collision between the sink nodes. Sink nodes move around the specified area and collect the data from the sensor nodes.
- Consider the underwater channel modeling to reduce path, channel, and transmission losses, reducing the end-to-end delay.
- Single relay routing techniques reduce the reliability of the network. RACE-SM also uses the cooperative routing technique to improve the network's reliability.
- The proposed routing scheme performance is evaluated through the MATLAB simulation tool, which shows the improvement in energy consumption, End–to–end delay, PDR, and number of alive nodes.

1.3 Paper organization

The rest of the article is prepared as follows: literature review discussed in section 2, section 3 discusses the proposed RACE-SM routing protocol, section 4 discusses the simulation results of the compared and proposed routing scheme, and finally, section 5 presents the conclusion of the work.

2 Literature review

Energy Efficient Regional Base Cooperative Routing (EERBCR) protocol with the sink mobility technique has been proposed in (Gul et al., 2021). The authors' main focus in this work is moving the sink node to the predefined path to gather the data from the sensor nodes. The basic concept in this research is to keep all the sensor devices in sleeping mode. When they receive the Hello packet from the sink nodes, it activates itself. When the sink moves from the specific region again, it will broadcast another message to inform the sensor nodes living in the area. The proposed protocol outperforms DEADS, DBR, and EEDBR routing protocols based on the experiment results. As network reliability and transmission loss are the key factors in data transmission, the authors do not discuss them in the mentioned protocol.

The solution for the problem of the hole coverage in Underwater Acoustic Sensor Networks (UWASN) has been proposed in (Raj Priyadarshini and Sivakumar, 2020). The stated problem directly affects the currently active node threshold value limit, which causes node destruction. The authors improved the coverage and data transmission by introducing the simple value of those already known parameters. Moreover, due to the movement of water motions for some of the parameters like currents, winds, and waves, the topology will be changing. This paper is also limited to improving the network lifetime while transmission loss and end-to-end delay are not considered.

A new multi-layer routing protocol for discovering the best route and enhancing network lifetime, energy utilization, and End–to–end delay has been proposed in (Gomathi and Martin, 2018). The authors used the efficient discovery method to find the shortest path for data transmission. The simulation results show that the proposed routing protocol has performed better in all the above metrics but not good in network throughput.

The data retransmission to the failure node has been improved with the E2R2P routing protocol in (Anuradha and Srivatsa, 2019)—the essential work of the research on reducing energy consumption in UWSNs. The authors have enhanced data retransmission to those nodes that are already dead or unreachable due to any problem. As in the stated work, the computation has decreased. Therefore, the network lifetime and performance have been extended.

The energy consumption and path loss issue during the data transmission has been achieved with SPARCO (Stochastic Performance Analysis with Reliability and Cooperation) (Ahmed et al., 2016a). This protocol uses the cooperative routing technique. All the nodes are placed with a single Omni-directional antenna for signal broad-casting. The simulation results show that the mentioned routing protocol performed better than other routing protocols regarding transmission, path loss, and energy consumption.

The energy consumption due to the unnecessary hello packets has been controlled with a localization-free, energy-efficient, and link reliable (E2LR) routing protocol (Tariq et al., 2021). Furthermore, the E2LR routing protocol uses a composite metric for the forwarder hop selection to reduce the E2E delay and packet loss. The experiment results show that the stated routing protocol performed better energy usage during the data forwarding and information sharing phase than R-ERP2R and H2-DAB routing protocols. Moreover, the result also shows that the E2LR routing protocol performed better in terms of End–to–end delay, packet delivery ratio, and network lifetime than compared protocols.

The joint optimization framework of forward and hold mechanisms, sink mobility, data aggregation, and adoptive depth threshold (dth) for improving network lifetime, reducing energy consumption, increasing throughput, and reducing the propagation delay has been designed in (Latif et al., 2020). The authors performed the experiments in a 3-D underwater sensor network environment for better results displaying. The simulation results show that the proposed technique performs better in packet drop, network lifetime, and throughput than both iAMCTD and Mobicast routing schemes. Moreover, the simulation results show that the proposed routing scheme provides better performance for real-time delay intolerant applications over the existing algorithms.

A reliable and energy-efficient routing scheme named Energy Balanced Efficient and Reliable Routing (EBER) has been proposed (Wadud et al., 2019). The author focuses on the forwarding nodes and nodes' energy to reduce energy conservation. Moreover, the authors deployed two sink nodes to decrease network latency in areas with high traffic burdens. The simulation results show that the WDFAD-DBR performed better in PDR, energy usage, and lesser duplicate packets than the EBER 2 routing protocol.

An energy-balanced routing strategy for underwater sensor networks has been proposed in (Bouabdallah et al., 2017). The main concept of this study is to study and overcome the problem of the holes and overcome energy consumption. The authors also worked on the transmission load redistribution for nodes during the communication between the source and destination nodes. The network lifetime is improved by solving the energy holes problem based on a result.

An efficient energy-gathering routing protocol for UWSNs has been proposed in (Banaeizadeh and Toroghi Haghighat, 2020). The authors have solved three issues in different steps in the stated routing protocol. Firstly, the energy usage issue is solved by combining the sensor nodes into other node groups managed by the sender node. The mentioned node will be responsible for gathering the data from their neighbors in one hope of communication. Secondly, the proposed MAC routing protocol has improved the packet loss and the collision in the way that the normal sensor nodes will send their data to the forwarder node on specific committed slots. Finally, the graph structure has decreased the data collision of the overall network. The simulation results, which are taken from the NS2

simulation, show that EEDG performed better in throughput, end-to-end delay, and energy consumption.

Energy and reliable routing protocols by the name Cooperative Effective Energy and Reliable Delivery (CoEERD) and Effective Energy and Reliable Delivery (EERD) have been proposed in (Ullah et al., 2019) to solve the reliability and adverse channel issues which cause packet corruption. In the first proposed routing protocol (EERD), the forwarder transfers the packets through a reliable path with less energy usage. Still, this routing protocol uses a single route for all the data transmitting, which interferes with the reliability of data packets because of the unpredictable environment. Therefore, the authors introduced the second routing protocol called CoEERD, which will use the cooperative routing technique for data transferring between the source and destination nodes. The relay node concept has been adopted for data forwarding from the sensor nodes to the sink nodes in the cooperative method. The simulation results show that the EERD and Co-EERD performed better in delay, reliable packet transfer, and energy usage than ODBR and CoDBR.

The fuzzy logic technique (FLT) improves (Bu et al., 2018) performance by selecting the best forwarder nodes based on energy information and node position factors. Interference between nodes can be reduced using this method. The use of remaining energy ratios (RERs) and the interference of the fuzzy logic techniques minimizes the energy required in the design, saving resources. There are many nodes in the network, and only one is chosen as a destination because it has the lowest position in terms of distance and the maximum amount of energy. The system introduces a slight lag, but it saves energy in the long run. While this increases throughput, it compromises network security by increasing the likelihood of data packet loss.

The authors in (Pappas et al., 2018) use two data collectors to facilitate random access to the IoT wireless network. When transmitting data packets *via* random access, nodes and aggregators are used. Data packets are slotted for transmission by the aggregators, who cooperate at the network level. Additionally, they allow for network scalability by providing throughput. As a result, they may have a finite waiting time. In addition, they aim to make IoT data collection easier. Aggregators add complexity to the system.

The researchers in (Javaid et al., 2017b) divide the network into three equal parts. An effective way to save energy and improve the ratio of transmitted packets is to divide the network. The method considers the sink's horizontal and vertical movement, which aids in reducing packet losses. Nodes in the network use collaboration to forward data using the cooperative approach—these aids in achieving the highest possible RODP. The information is amplified and decoded by the RLN before being sent to the DSN. The BER is verified at each stop along the way when going from one node to another. The packet is successfully sent to its intended destination if the threshold value exceeds the BER value. The nodes are selected based on the cost function, and the network is built. Choosing a destination node depends on the lowest depth, SNR, and maximum residual energy. The protocol has a high RODP and extends the life of batteries at the same time.

In (Javaid et al., 2017a), the authors show two new ways to use less energy and prevent void zones from forming. The first strategy relies heavily on mobility sink nodes to reduce node energy usage. You'll use routing in conjunction with the first method in the second method. By routing packets *via* many paths, the cooperative system improves network reliability. While packet loss improves in the second system, transmission delays grow as RODP increases, all at the cost of total energy use. The first method saves energy and ensures that the nodes stay up and running for a long time.

In (Ullah et al., 2019), the authors presented new wireless network routing algorithms based on multiple cognitive access. The methods used a cognitive multiple access technique to bypass traditional cooperation. Two new routing strategies have been introduced to help minimize the consumption of resources. As a result, the proposed protocols outperform more traditional forms of cooperation. The goals of the proposed solutions are to increase network throughput and keep the network stable. However, it has a significant energy cost and prolongs the response time.

A multi-hop transmission system is proposed in (Balaji et al., 2019). Data packets are transmitted from one device to another before being sent to the sink's final destination. Fuzzy logic uses the cluster head to select the data packets going from the sender node to the sink. Three criteria are used in this approach: trust factor, current energy, and distance from the sink node. If there are more cluster heads between the sink and the cluster head, the cluster head uses a fuzzy logic technique to select the best cluster head for reaching the data. The optimal forwarding node must be chosen to elect the neighbor node closest to the sink. The high trust factor and sink node distance identify the best forwarding node. As a result of the protocol, the network's overall lifetime and overhead are raised.

In (Liu et al., 2018), the authors suggest a protocol (RECRP) that does not require additional hardware to define the location. The optimal minimum-max technique dynamically controls trans-mission power and channel frequency parameters. It achieves energy efficiency through the use of two-hop forwarding capabilities. The protocol also has the advantage of preventing communication voids. This protocol uses less energy per node and reduces the end-to-end delay while maintaining a higher PDR than other methods.

The authors in (Mhemed et al., 2021) focused on reducing energy consumption, increasing network lifetime, and enhancing the packet delivery ratio. The authors used the void avoidance technique to achieve the mentioned task. The task has been accomplished, as shown in the simulation result, but

reliability is the essential issue in networking and is not considered in the proposed protocol. To reduce energy consumption and maximize throughput, the authors (Yahya et al., 2019) used a single broadcasting antenna to transfer data between the source and distention. This paper also focuses on the single relay node selection for data transferring, which caused early death of the relay node due to the burden of sharing the other sensor nodes' data with the sink node.

To address the high energy consumption and throughput, the authors of (Ahmed et al., 2016b) focused solely on the cooperative technique. Still, the limitation is that using the relay node for data transmission creates a hotspot, and the relay node will die quickly. This research work motivation is from work presented in (Ahmed et al., 2016a; Yahya et al., 2019; Mhemed et al., 2021), leading to issues such as using a single relay node end-to-end delay in achieving the network's reliability. We introduce the RACE-SM routing protocol to increase the network's lifetime and reliability and reduce transmission loss using the sink mobility technique.

3 Proposed protocol design scheme

The proposed RACE-SM routing protocol is briefly explained in this section. The RACE-SM is a cooperative routing scheme that reduces power consumption and keeps the network for a long-time.

3.1 Network configuration and path initialization

A 1000m dimensions network with the 3D platform was developed. The simulation area is divided into four equal squares labeled Upper right square (URS), Upper left square (ULS), Bottom right (B.R.), and Bottom left (B.L.). Sink nodes (S.N.) move in a three-cornered path to gather data from the sensor nodes in each portion. Randomly installed nodes sense the characteristics and transform them into packets. So that they can be used in consequent processing steps, the data is sent to the Sink node. Each of the nodes has direct communication with the sink nodes. Acoustic waves communicate between nodes, and each node is equipped with an acoustic modem.

The position selection method for the autonomous underwater vehicle in (Cheng, 2005) is used to gather the sink node's location information. The mobile sinks, which are GPS-enabled, are installed at the top of the water. The speed and time arrival of the acoustic waves are used to calculate the distance from these mobile sinks. We use the close distance to the sink nodes on the surface to get the coordinates, which tells us where the sink node is now located. The path initialization for sink nodes is designed to cover the whole network and select more nodes. SN1, SN2, SN3, and SN4 initiate their movements in their respective regions from the corner, mid-boundary, center, and center of the B.L as shows in Figure 1.

Due to interference and collisions, S.N. movement starts from a different location instead of the same point, guaranteeing that all the sink nodes must not come to the center simultaneously. The middle or center of the network is defined as below:

$$M(\ q_m\ ,\ w_m\ ) = \left(\frac{q_i + q_f}{2}\right), \left(\frac{w_i\ + w_f}{2}\right) \tag{1}$$

where$(\ q_m\ ,\ w_m\ )$, $(\ q_m\ ,\ w_m\ )$ and $(\ q_i\ ,\ w_i\ )$ consequently, the final, center, and initial points of the network.

The S.N. can move in three directions: horizontally, vertically, and diagonally. All S.N.s begin motion from their reference point and keep moving through the network along with a specified position.

The diagonal movements of all S.N. are mathematically indicated by shifting the predefined rom the current position (q_c, w_c) to $(\ q_n\ ,\ w_n\ )$ that is the new location:

$$SN1(\ q_n\ ,\ w_n\ ) = (q_c - o,\ w_c - o) \tag{2}$$

$$SN2(\ q_n\ ,\ w_n\ ) = (q_c + o,\ w_c - o) \tag{3}$$

$$SN3(q_n, w_n) = (q_c + o, w_c + o) \tag{4}$$

$$SN4(q_n, w_n) = (q_c - w_c + o) \tag{5}$$

To avoid collisions, when the S.N. reaches the network's angle, it starts to move along the network's border in either a vertical or horizontal direction. SN2 and SN4 move vertically, and SN1 and SN2 move horizontally. The Next coordinates for every S.N. are as follows:

$$SN1(q_n, w_n) = (q_c + o, w_c) \tag{6}$$

$$SN2(q_n, w_n) = (q_c, w_c + o) \tag{7}$$

$$SN3(q_n, w_n) = (q_c - o, w_c) \tag{8}$$

$$SN4(q_n, w_n) = q_c, w_c - o) \tag{9}$$

To move vertically, S.N.s must be in the middle of a network, and their horizontal motions must be reversed to do so:

$$SN2(q_n, w_n) = (q_c, w_c + o) \tag{10}$$

$$SN2(q_n, w_n) = (q_c - o, w_c) \tag{11}$$

$$SN3(q_n, w_n) = (q_c, w_c - o) \tag{12}$$

$$SN4(q_n, w_n) = (q_c + o, w_c) \tag{13}$$

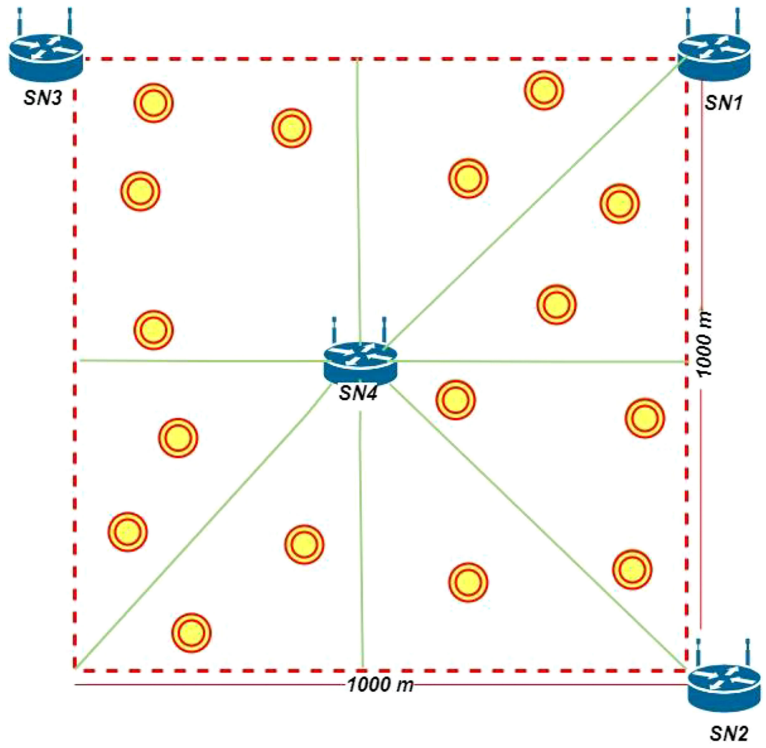


FIGURE 1
Proposed Network Model.

3.2 Sink nodes limitation to move around in the networks

S.N.s are initially deployed at a specific location on the network, and afterward, they start to move along the specified route. As a result, how can one restrict an S.N.'s mobility in a network to follow a predetermined path? Is there any S.N. that doesn't leave the network? How does an S.N. know where it is in the network when it changes directions? Since all S.N.s move in the same region, their memory includes information about the region's coordinates to change their movement direction. Those are the corner, mid-boundary, and center of the network. It then compares its current coordinates with the stored ones when it steps forward.

The movement direction is changed if it equals one of the stored coordinates. Other than that, it continues to move on the same path. If fixed quantity o is added or subtracted, the S.N.s' movement direction is defined. Suppose SN1 starts its movement from a network corner and moves horizontally. While on the next step, SN1 checks its current position with the saved coordinates. Whereas, when the current coordinates do not match any previously saved ones, they will continue to move in the exact directions.

To increase the x coordinate, the z-coordinate is increased to a permanent value while the coordinate persists the same. By adding the permanent value o to q, the SN1's horizontal movement is noted and stops the SN1's in the network. Also, when the q and w coordinates are equal to the middle point, only the w coordinate increases. Whenever the coordinate's points are similar to the center coordinates, they move to the diagonal path. All S.N.s in the network must follow this same path.

The x and y values increment pushes the S.N. to follow the determined path to track it. A change in S.N. y and x coordinates determines a diagonal movement, while a difference only in the x or y coordinate specifies horizontal or vertical mobility for S.N. It is done by comparing the S.N.'s current position with the network's center, corner, and mid-boundary coordinates, saved in its memory. When the S.N.'s current coordinates match the last numbers, it switches its route.

The currents in the water can cause the S.N. to deviate from its proposed route. When the S.N. moves forward, it calculates an error e, which keeps it on track. It is compared to the surface sinks' estimated position value when an S.N. moves forward from its current location of (q_c, w_c) to the next position (q_n, w_n) to confirm the S.N.'s actual position and identify drifts. As soon as an S.N. moves forward, it compares its new position (q_n, w_n)

including its original position (q_c, w_c) and discovers the error as follows e:

$$e = |q_c - q_n| - 1 \tag{14}$$

$$e = |w_c - w_n| - 1 \tag{15}$$

e should be equal to 0 if the ocean current does not interfere with the S.N.'s mobility. For example, if e has a negative or positive value, it will be added or subtracted, whether negative or positive. It's calculated as follows in the case of a positive error:

$$SN(\ q_n\ ,\ w_n\ ) = (q_c \pm o - e,\ \ w_c\ \pm o - e\ ) \tag{16}$$

It's calculated as follows in the case of a negative error:

$$SN(q_n, w_n) = (q_c \pm o + e, w_c \pm o + e\ ) \tag{17}$$

For example, S.N. is at position (500m, 500m). It then moves to the following location (501m, 501m). There is no interference from the water currents in this case and, therefore, no error. A negative or positive value for e specifies that the S.N. has different from its path. It can be traced directly by subtracting or adding the error value when discovering the following position when the water currents disturb its function.

The reliability of the UWSNs channel is introduced to ignore the loss in transferring the data at the time of the outage channel. Our proposed work also emphasized the reliability of the link to achieve it. Multihop channels are formed by bringing together many links, creating a chain of sensor nodes that transmit data from one source to another and finally to the final destination. If all data is successfully sent, it will be regarded as a successful E2E transmission. E2E reliability R is used to express an occurrence likelihood of a case. Hence, R is written as follows (Ahmed et al., 2015):

$$\mathrm{R} = 1 - \left\{ 2\pi \mathrm{H}\ \mathrm{x} \exp\left(\frac{2^{\left(\frac{\mathrm{R}}{\mathrm{B}}\right)} + \mathrm{SL} + \mathrm{p(d,f)}}{20}\right) \right\} \tag{18}$$

where R is the function based on two nodes' distance which is a point to point link, state of the channel and depth of the water, and the(d, f) describe the distance and frequency, respectively, S.L. is the source level, p is the probability, and E2E channel total reliability could be determined from the equation (18).

$$\mathrm{R} = 1 - \left\{ \sum_{\mathrm{i}-1}^{\mathrm{n}} \left(2\pi \mathrm{Hi}\ \mathrm{x} \exp\left(\frac{2^{\left(\frac{\mathrm{R}}{\mathrm{B}}\right)} + \mathrm{SL} + \mathrm{p(d,f)i}}{20}\right)\right) \right\} \tag{19}$$

3.3 Data exchange, link establishment and data forwarding

A Sink Node and a node establish a link before exchanging the data; every S.N. with a 150-meter transmission range broadcasts the info packet to establish the connection. As a part of this information packet, the S.N. and node I.D.s are included. The information packet received by a node indicates the presence of the S.N. within the node's 150 m coverage range. The reception of an info packet indicates that a data exchange link has been established between S.N. and a node. If a node receives the information packet, the data are transferred with the S.N. any data received to S.N. is then transferred directly to the next S.N. or the surface sink node.

In contrast, SN3 and SN4 forward data to SN2 and SN1. It will be an additional load on other sensor nodes when transferring data through the multihop transmission to the surface. Data is directly received by S.N.s, which reduces the burden on nodes by transferring data directly from nodes to the surface. The S.N. transmits info packets to conserve the node's energy because the network lifetime is reduced when a node sends an info packet. To send data packets to the sink node, the nodes that send the data must first look for the closest mobile Sink in its communication range, and if it finds one, it will send a packet of data directly to that mobile Sink.

Otherwise, packets are sent through a relay node to reach their destination (Shah et al., 2016). Two-step communication is performed between the Relay and S.N. First, the distance of neighbors and the source node in the transmission range is calculated with S.N.

Then, in the second phase the distance between all neighbors is compared, the nearest neighbor with the shortest distance is chosen as a gateway. So, the packets are sent to the S.N. in this manner. The S.N. transmission range is computed after a predetermined amount of time. Figures 2, 3 show both types of data forwarding, respectively.

3.4 Cooperation phase

Figure 4 shows a two-phase transmit system that lets the source node and the relay node communicate without interfering. Phase one and phase two make together with the entire working process. R represents the best relay node, S represents the source, and D represents the destination.

S transmits data to D and R simultaneously in the first stage, whereas R transmits the data received to n the other stage. The overall distance between the sender and sink node is $d_1 + d_2$, as indicated in Figure 3. At R and D, data from phase 1 transmission can be written as follows (Ahmed et al., 2015):

$$\mathrm{Y_{S.R.}} = \mathrm{h_{S.R.}x_S} + \mathrm{n_{SR}} \tag{20}$$

$$Y_{SD} = h_{SD}x_S + n_{SD} \tag{21}$$

where h_{SD} and h_{SR} are the co-efficient of the channel, respectively, and x_S is the symbol of sender information between the source to destination and source to relay,

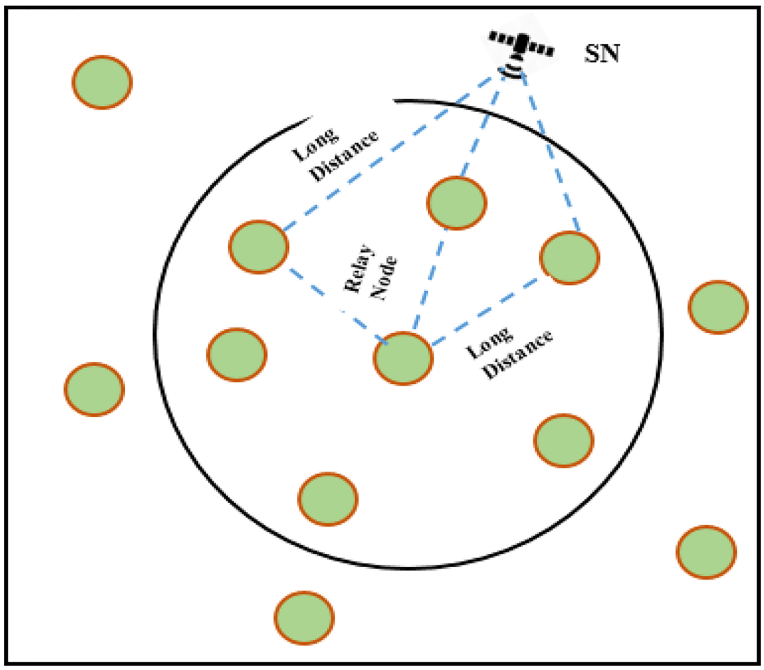


FIGURE 2
Packet transmission trough Relay node.

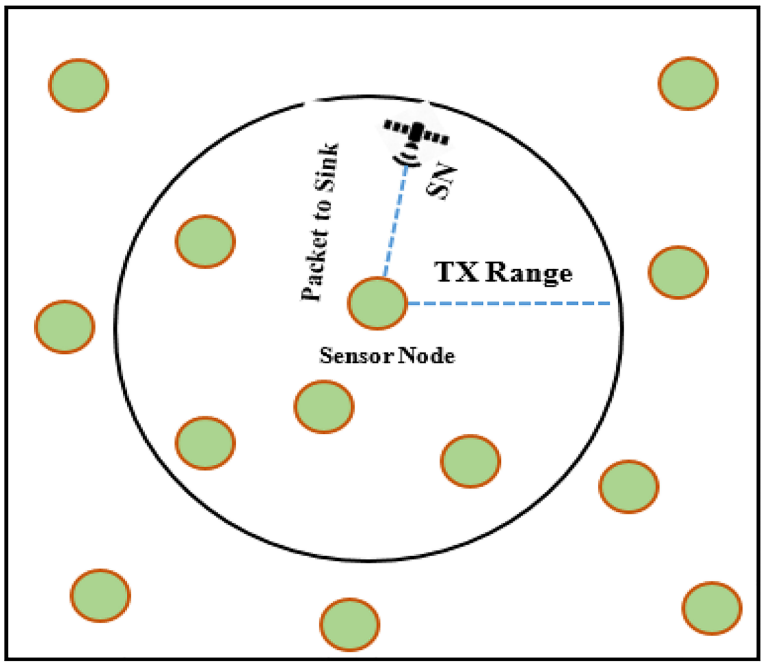


FIGURE 3
Direct Packet transmission SN.

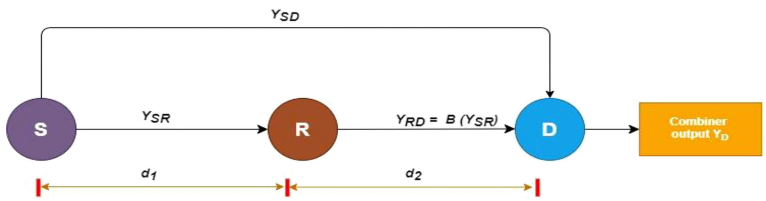


FIGURE 4
Three Sensor Node System M.

respectively. In the second processing stage, some received data from stage 1 Rb was retransmitted to D destination. Therefore the received data from the second stage can be represented as (Ahmed et al., 2015).

$$Y_{RD} = h_{RD}x_S\ f\ (Y_{SR}) + n_{RD} \quad (22)$$

The source node receives data from the relay node using f as the function (y_{SR}). There are three noise components in the links between source-relay (n_{SR}), source-destination (n_{SD}), and source-relay-destination (n_{RD}). The flowchart in Figure 5 and Algorithm 1 explain the proposed RACE-SM routing protocol.

1. **Start**
2. Network Configuration;
3. Divided into four regions;
4. N = Total number of Nodes;
5. SN_i (q,w) = GPS points of SN_i;
6. ID_{SNi} = SN ID;
7. RE = Residual energy of a node
8. SN = Transmission range of SN
9. Neighbor [j] = All neighbors of node i
10. **While** *Packet reached to SN == false* **do**
11. SN_i(q, w) = SN_i(q± o, w± o) /Diagonal movement
12. Or SN_i (q, w) = SN_i(q± o, w) //Horizontal movement
13. Or SN_i (q, w) = SN_i(q, w± o) //Vertical movement
14. Send an info packet;
15. **if** *Info packet == received*
16. **elseif** *SN in range = = true* **then**
17. Forward data;
18. **if** RE[j]> 0 and Distance [j] < Distance [j +1] **then**
19. *Neighbor* (j) selects as a *forwarder*
20. **else**
21. Neighbor (j+1) selects as a forwarder
22. **else**
23. wait for *SN*
24. *end*
25. ***End***

ALGORITHM 1. DATA TRANSMISSION USING SINK MOBILITY

3.5 Relay node selection and routing phase

If the channel is effective, a source sensor node (*S*) can identify which of its n neighboring sensor nodes is most suited to transfer data to the Sink node. Source and relay sensor nodes are selected by comparing their respective weights.

Using the sensor node with the greater value of W_i, the transmitted data is sent. Directional transmission is possible if the source's residual energy is greater than to the relay node's remaining energy. Otherwise, the communication should go through the intermediate nodes.

If E_{re} (R) < E_{re} (S), then use direct transfer (23)
else E_{re} (R) ≥ E_{re} (S), then use relay node

The relay node uses the Amplify and Forward (AF) technique to transfer the data, which applies an amplification factor before sending the signal to the destination node from the source node.

3.6 Combining strategy

The destination sensor node D uses the SNRC (signal-to-noise ratio combining) technique to combine the signals from the source S and the relay R. The SNRC combines signals at the receiver. Each array element's SNR is weighted equally. Equal ratio combining (ERC) doesn't work better since it doesn't consider small-scale variations (low SNR) while combining. SNRC, on the other hand, does. The formula for calculating SNRC is:

$$Y_d = x_1 y_{sd} + x_2 y_{rd} \quad (24)$$

where X_2 shows the weight of the data transmission on the relay route, Yd represents the signal of output that is combined at the receiver node D, and X1 specifies the weight of the direct data transfer route. The Algorithm 2 shows the data transmission using relay nodes.

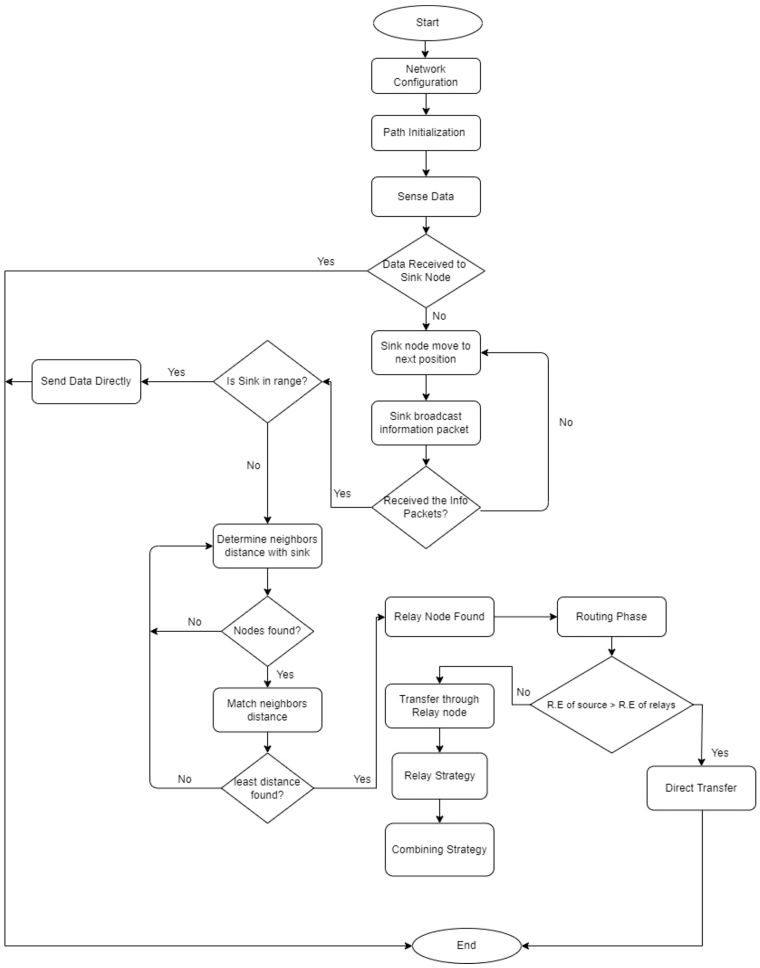


FIGURE 5
Flow chart of the proposed RACE-SM routing protocol.

Start
SRN: Source node
R.N.: Relay node
R. E: Residual energy
S.N.: Sink node
Find Relay node
Check the weight value
If $BER \leq 0.5$ **then**
Find RN
RN found = **true**
Else $BER > 0.5$
Find RN = **False**
Then repeat step No 3
If $R.E\ of\ source \geq R.E\ of\ Relay$
Forward Data = **true**
Packet reached to SN = **true**
Else *Use multi-hoping RN path*
Packet reached to SN = **do**
Packet accepted = **true**
end if
end if
End

ALGORITHM 2. DATA TRANSFERRING THROUGH THE RELAY NODE

4 Results and performance evaluation of RACE-SM and RACE

The following section defines the simulation scheme of the RACE-SM and also the comparison of the RACE-SM and RACE routing protocols. Simulation parameters are defined in Table 1. Additionally, this section defines the primary performance metrics for all compared protocols.

TABLE 1 Simulation parameters.

Parameters	Value
Simulation Deployment Width	1000 m
Simulation Deployment Depth	1000 m
Simulation Deployment Breadth	1000 m
No of sensor Nodes	225
No of Sink Nodes	10
Transmission Range	220 m

4.1 Performance metrics

Performance metrics for all compared protocols are defined as follows:

- **Residual Energy:** It described the distinction between the startup nodes' power and the power of the nodes utilized during the operation.
- **Network Lifetime:** The overall time spent by running the network is referred to as network lifetime.
- **Throughput:** The total number of efficiently transmitted packets at the sink is called throughput.
- **Path-Loss:** Path-loss is a unit of measurement used to describe the difference between the transmitted and received power of a transmitter and a receiver (dB).

4.2 Results, discussions and analysis of RACE-SM with existing protocols

This section compares and evaluates the proposed RACE-SM and existing routing schemes. Sensor nodes are deployed randomly for all simulation techniques. This research work is modeled in a 3D environment with a height of 1000m x 1000m x 1000m with ten (10) multiple mobile sink nodes on the water surface and 225 sensor nodes in the functional area deployed randomly. The sensor node sends the data to the nearest node, which transfers it to another neighbor node.

There is efficient cooperation between the sensor nodes that carry the same physical parameters, most importantly depth and weight threshold, with their adjacent nodes to stay informed about the network's changing conditions. After predefined intervals, nodes calculate their distances to their neighbors. Sensors communicate with the higher layer *via* neighboring sensors until the information receives at the mobile sink node. The sink node is in charge of the cooperating sensors' depth thresholds and adaptive mobility.

4.2.1 Total energy consumption

Figure 6 shows the energy consumption of the proposed scheme and other existing routing methods. The proposed scheme's energy consumption is the lowest compared to the current routing protocols. The proposed RACE-SM uses the sink

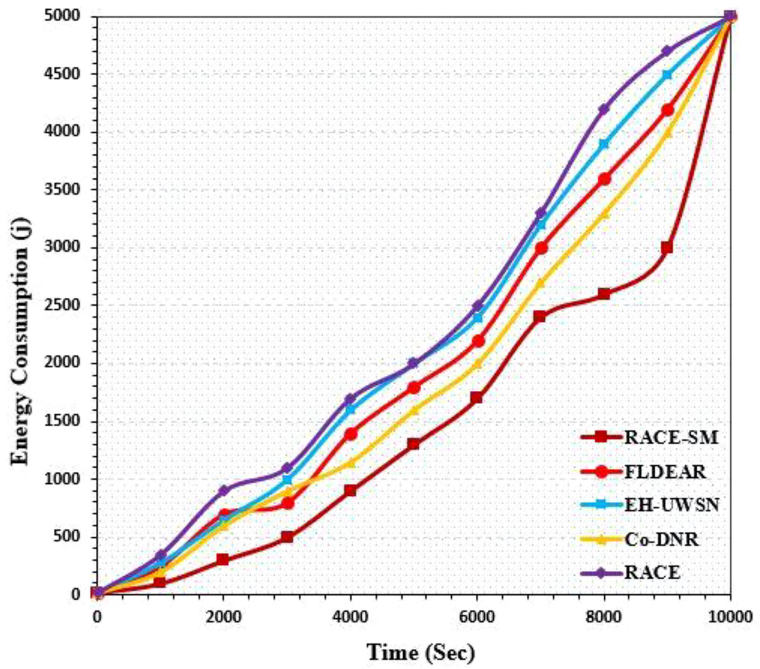


FIGURE 6
Energy Consumption.

mobility technique for data collection from the sensor nodes and direct data transmission between the sender and the sink node. Furthermore, direct data transmission between the source and sink nodes minimizes energy consumption because only one sensor node is part of the data transmission process. RACE uses relay nodes to send and receive data, which uses more energy. Co-DNR, EH-UWSN, and FLDEAR, on the other hand, send data to the sink node through multiple hops. When transferring data through various nodes, higher energy consumption occurs.

The energy consumption of Co-DNR is lower than the other three protocols, and Co-DNR uses at most two nodes for data transmission, while FLDEAR and EH-UWSN use multiple nodes to transfer data from the source to the destination. FLDEAR has lower energy consumption than the EH-UWSN, and it's due to controlling the collision of sensor nodes. FLDEAR uses the priority number for each sensor node, and data transmission is done through this number, preventing data retransmission to the sink node.

The energy consumption of both RACE-SM and RACE routing protocols shows in Figure 7. As the graph illustrates, RACE-SM uses energy more efficiently than RACE because efficient data forwarding with the support of load balancing and adjacent nodes is eventually achieved. In RACE, using a single relay node was a big issue that created hotspots that would use more energy than other sensor nodes and die early (Table 2).

4.2.2 Packet delivery ratio

The PDR of the proposed and existing routing schemes shows in Figure 7. As illustrated in the figure, the PDR of the RACE-SM is much higher than Co-DNR, EH-UWSN, FLDEAR, and RACE. The main reason behind achieving a higher PDR has reduced BER. The other main reason is using the function's parameters, such as the lowest distance and residual energy, for selecting the relay node for transferring from source to destination. Using the minimum number of nodes between the source and destination consumes minimum power and time to transmit the data packets.

Co-DNR, FLDEAR, and EH-UWSN have the same PDR at the start but gradually decrease after the 4000 rounds. FLDEAR and Co-DNR use multiple nodes for transferring the data. Therefore, the value decreases while EH-UWSN uses two relay nodes with a destination. Thus, the value is better than the other two schemes. The RACE is also a cooperative routing scheme but using a single relay node with a destination consumes much energy, and the nodes drain the power quickly. Therefore, the PDR is much lower as compared to other routing schemes. From the start to the end of the simulation, RACE-SM has a higher PDR than existing routing schemes (Table 3).

4.2.3 Total alive nodes

The total number of alive nodes in the proposed and existing schemes is plotted in Figure 8. The results show that

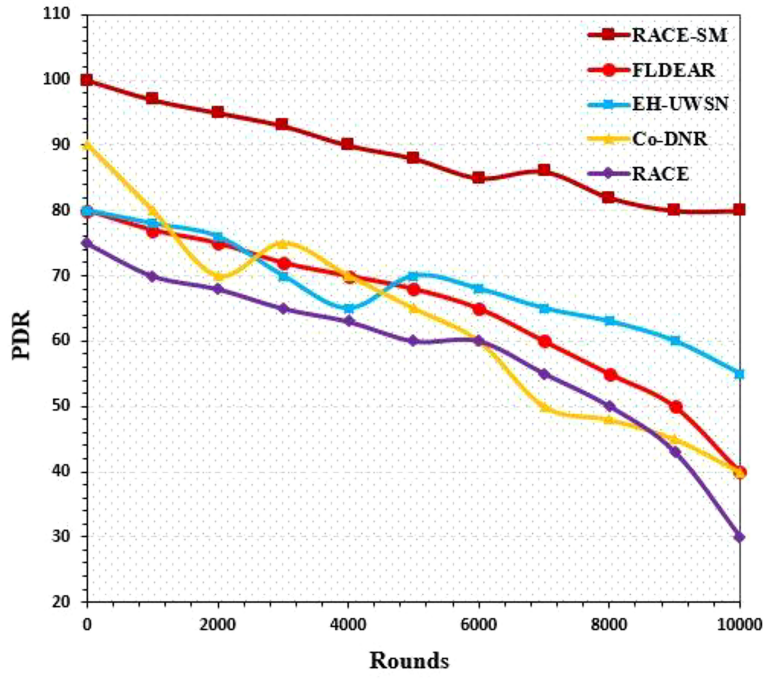


FIGURE 7
Packet Delivery Ratio (PDR).

TABLE 2 Protocols addressing energy consumption.

Time	RACE-SM	FLDEAR	EH-UWSN	Co-DNR	RACE
0	20	20	20	20	20
1000	100	250	280	200	350
2000	300	700	650	600	900
3000	500	800	1000	900	1100
4000	900	1400	1600	1150	1700
5000	1300	1800	2000	1600	2000
6000	1700	2200	2400	2000	2500
7000	2400	3000	3200	2700	3300
8000	2600	3600	3900	3300	4200
9000	3000	4200	4500	4000	4700
10000	5000	5000	5000	5000	5000

TABLE 3 Protocols addressing PDR.

Rounds	RACE-SM	FLDEAR	EH-UWSN	Co-DNR	RACE
0	100	80	80	90	75
1000	97	77	78	80	70
2000	95	75	76	70	68
3000	93	72	70	75	65
4000	90	70	65	70	63
5000	88	68	70	65	60
6000	85	65	68	60	60
7000	86	60	65	50	55
8000	82	55	63	48	50
9000	80	50	60	45	43
10000	80	40	55	40	30

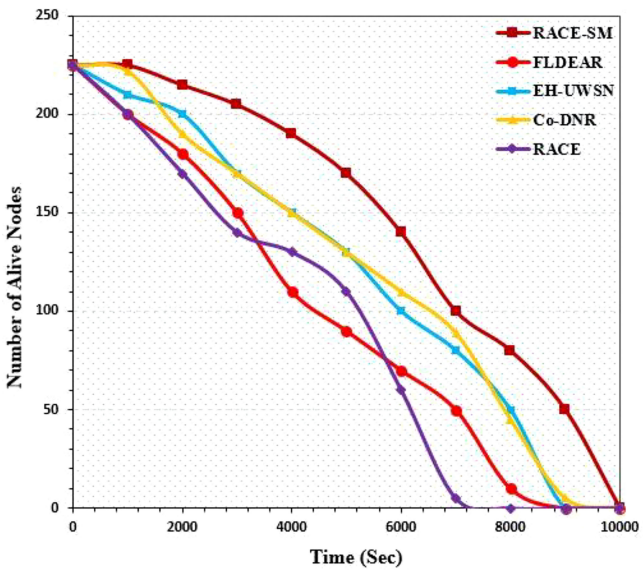


FIGURE 8
Total Number of alive Nodes.

the number of active nodes in RACE-SM is higher than in the other three strategies. The main reason for this is using the minimum energy consumption. Additionally, selecting the best forwarder node for data transmission is also affected because it will use minimum energy, and the data will be transferred at the lowest cost. Therefore, the number of alive nodes will be high. The Co-DNR is the second one with the maximum number of active nodes because it uses the minor sink nodes to transfer the data packets from source to destination. The total number of active nodes in RACE-SM at round 6000 is 145, and Co-DNR & EH-UWSN have 110 and 100 active nodes, respectively. The total number of active nodes in RACE at 6000 rounds is 110. Still, after the 6000 rounds, nodes die significantly quicker than RACE-SM, and it's due to RACE using a single forwarder which will create a hotspot getting burden on a single node and drain energy quickly (Table 4).

4.2.4 End-to-end delay of the Network

The results of the end-to-end delay of the proposed and other three routing schemes are shown in Figure 9. As illustrated in the figure, RACE-SM has the lowest latency of all other existing techniques. The reason behind this is sink mobility. Data transferring are done directly between the

TABLE 4 Protocols addressing Live Nodes.

Time	RACE-SM	FLDEAR	EH-UWSN	Co-DNR	RACE
0	225	225	225	225	225
1000	225	200	210	222	200
2000	215	180	200	190	170
3000	205	150	170	170	140
4000	190	110	150	150	130
5000	170	90	130	130	110
6000	140	70	100	110	60
7000	100	50	80	89	5
8000	80	10	50	45	0
9000	50	0	0	5	0
10000	0	0	0	0	0

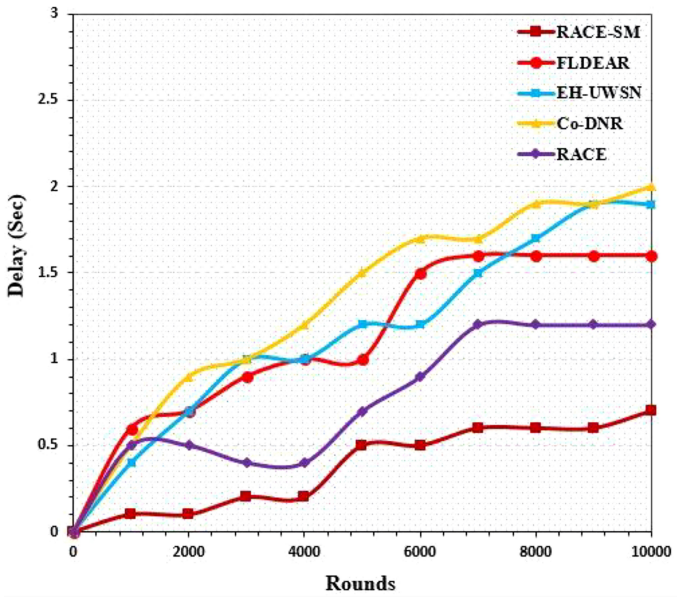


FIGURE 9
Average end-to-end delay.

TABLE 5 Protocols addressing End-to-End Delay.

Rounds	RACE-SM	FLDEAR	EH-UWSN	Co-DNR	RACE
0	0	0	0	0	0
1000	0.1	0.6	0.4	0.5	0.5
2000	0.1	0.7	0.7	0.9	0.5
3000	0.2	0.9	1	1	0.4
4000	0.2	1	1	1.2	0.4
5000	0.5	1	1.2	1.5	0.7
6000	0.5	1.5	1.2	1.7	0.9
7000	0.6	1.6	1.5	1.7	1.2
8000	0.6	1.6	1.7	1.9	1.2
9000	0.6	1.6	1.9	1.9	1.2
10000	0.7	1.6	1.9	2	1.2

source and destination. RACE-SM uses the sink mobility method in which sensor nodes communicate the data directly to the sink when the sink is in the communication range; therefore, the number of nodes is reduced between the source and destination.

The latency of the RACE is lower than Co-DNR, FLDEAR, and EH-UWSN. It's due to using a single relay node for data transmission between the source and destination. Co-DNR, FLDEAR, and EH-UWSN are approximately the same in latency due to using multiple nodes for data transmission between the source and destination (Table 5).

5 Conclusion

This research article contains the design of the RACE-SM algorithm for UWSNs. RACE-SM minimizes the energy consumption, the reliability of the network, the PDR, and the number of active sensor nodes. For efficient data collection, the network is divided into multiple portions. In RACE-SM, to gather data correctly, the sink node moves around the network, and when it senses the data, it collects it from the nodes. The sensor node sends the data directly to the sink node when the sink is in communication range, which will help to improve the PDR, reduce energy consumption, and improve network lifetime. Otherwise, data will be transferred through single-hop and multihop transmission methods to enhance the network reliability and reduce packet loss. The utilization of the cooperative technique makes the distance between neighboring nodes lower among neighboring nodes to achieve reliable data transmission between the source and destination. Also, the cooperative scheme achieves load balancing and improves network lifetime. Compared to the current routing schemes, the proposed method performed better up to 51.50%, 28.50%, 26.40%, and 56% compared to FLDEAR, EH-UWSN, Co-DNR, and RACE respectively in terms of live nodes. In terms of energy consumption, the proposed method is better up to 51.50%, 67.30%, 36.50%, and 79.50% compared to FLDEAR, EH-UWSN, Co-DNR, and RACE respectively. The RACE-SM archives higher PDR of 264%, 226%, 283%, and 337% over FLDEAR, EH-UWSN, Co-DNR, and RACE respectively. In terms of end-to-end delay, the RACE-SM performed better up to 80%, 84%, 102%, and 41% over FLDEAR, EH-UWSN, Co-DNR, and RACE respectively. The overall analysis shows that on average with existing routing schemes the proposed method outperforms in in terms of using less energy, having less delay, and being more stable. In future we aim to improve in other parameters such as hop count and want to develop a routing protocol for real time environments.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This work is supported by Malaysian Institute of Information Technology, Universiti Kuala Lumpur, Malaysia.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ahmad, I., Rahman, T., Zeb, A., Khan, I., Othman, M. T., and Hamam, H. (2022). Cooperative energy-efficient routing protocol for underwater wireless sensor networks. *Sensors* 22 (18), 6945. doi: 10.3390/s22186945

Ahmad, I., Rahman, T., Zeb, A., Khan, I., Ullah, I., Hamam, H., et al. (2021). Analysis of security attacks and taxonomy in underwater wireless sensor networks. *Wireless Commun. Mobile Computing* 2021, 1–15. doi: 10.1155/2021/1444024

Ahmed, S., Javaid, N., Ahmad, A., Ahmed, I., Durrani, M. Y., Ali, A., et al. (2016a). SPARCO: Stochastic performance analysis with reliability and cooperation for underwater wireless sensor networks. *J. Sensors* 2016, 1–17. doi: 10.1155/2016/7604163

Ahmed, S., Javaid, N., Shahid, S., Hadi, F., and Azeem, M. Q. (2016b). "RACE: Reliability and adaptive cooperation for efficient underwater sensor networks," in *2016 International Conference on Open Source Systems & Technologies (ICOSST)*. 122–128 (IEEE).

Ahmed, S., Javaid, JavaidN., Khan, F. A., Durrani, M. Y., Ali, A., Shaukat, A., et al. (2015). Co-UWSN: Cooperative energy-efficient protocol for underwater WSNs. *Int. J. Distributed Sensor Networks* 11, 891410. doi: 10.1155/2015/891410

Anuradha, D., and Srivatsa, S. K. (2019). Energy effectual reconfigurable routing protocol (E2R2P) for cluster based underwater wireless sensor networks. *J. Ambient Intell. Humanized Computing* 12, 1–8. doi: 10.1007/s12652-019-01413-z

Anwar, K., Rahman, T., Zeb, A., Khan, I., Zareei, M., and Vargas-Rosales, C. (2021). Rm-adr: Resource management adaptive data rate for mobile application in lorawan. *Sensors* 21 (23), 7980. doi: 10.3390/s21237980

Balaji, S., Golden Julie, E., and Harold Robinson, Y. (2019). Development of fuzzy based energy efficient cluster routing protocol to increase the lifetime of wireless sensor networks. *Mobile Networks Appl.* 24, 394–406. doi: 10.1007/s11036-017-0913-y

Banaeizadeh, F., and Toroghi Haghighat, A. (2020). An energy-efficient data gathering scheme in underwater wireless sensor networks using a mobile sink. *Int. J. Inf. Technol.* 12, 513–522. doi: 10.1007/s41870-020-00445-5

Bouabdallah, F., Zidi, C., and Boutaba, R. (2017). Joint routing and energy management in underwater acoustic sensor networks. *IEEE Trans. Network Service Manage.* 14, 456–471. doi: 10.1109/TNSM.2017.2679482

Bu, R., Wang, S., and Wang, H. (2018). Fuzzy logic vector–based forwarding routing protocol for underwater acoustic sensor networks. *Trans. Emerging Telecommunications Technol.* 29, e3252. doi: 10.1002/ett.3252

Cheng, W.-H. (2005). Study mobile underwater positioning system with expendable and multi-functional bathythermographs. *Ocean Eng.* 32, 499–512. doi: 10.1016/j.oceaneng.2003.12.012

Erol, M., Vieira, L. F., and Gerla, M. (2007). "Localization with Dive'N'Rise (DNR) beacons for underwater acoustic sensor networks," in *Proceedings of the Second Workshop on Underwater Networks* Montréal: Québec, Canada. 97–100.

Gomathi, R. M., and Martin Leo Manickam, J. (2018). Energy efficient shortest path routing protocol for underwater acoustic wireless sensor network. *Wireless Pers. Commun.* 98, 843–856. doi: 10.1007/s11277-017-4897-5

Gul, H., Ullah, G., Khan, M., and Khan, Y. (2021). EERBCR: Energy-efficient regional based cooperative routing protocol for underwater sensor networks with sink mobility. *J. Ambient Intell. Humanized Computing* 2021, 1–13. doi: 10.1007/s12652-020-02781-7

Javaid, N., Ejaz, M., Abdul, W., Alamri, A., Almogren, A., Niaz, I. A., et al. (2017a). Cooperative position aware mobility pattern of AUVs for avoiding void zones in underwater WSNs. *Sensors* 17, 580. doi: 10.3390/s17030580

Javaid, N., Hussain, S., Ahmad, A., Imran, M., Khan, A., and Guizani, M. (2017b). Region based cooperative routing in underwater wireless sensor networks. *J. Network Comput. Appl.* 92, 31–41. doi: 10.1016/j.jnca.2017.01.013

John, S., Menon, V. G., and Nayyar, A. (2020). "Simulation-based performance analysis of location-based opportunistic routing protocols in underwater sensor networks having communication voids," in *Data management, analytics, and innovation* (Singapore: Springer), 697–711.

Latif, K., Javaid, N., Ullah, I., Kaleem, Z., Abbas Malik, Z., and Nguyen, L. D. (2020). DIEER: delay-intolerant energy-efficient routing with sink mobility in underwater wireless sensor networks. *Sensors* 20, 3467. doi: 10.3390/s20123467

Liu, J., Yu, M., Wang, X., Liu, Y., Wei, X., and Cui, J. (2018). RECRP: An underwater reliable energy-efficient cross-layer routing protocol. *Sensors* 18, 4148. doi: 10.3390/s18124148

Mhemed, R., Comeau, F., Phillips, W., and Aslam, N. (2021). Void avoidance opportunistic routing protocol for underwater wireless sensor networks. *Sensors* 21, 1942. doi: 10.3390/s21061942

Nayyar, A., Ba, C. H., Cong Duc, N. P., and Binh, H. D. (2018). Advances in Intelligent Systems and Computing Advances in Intelligent Systems and Computing, Spinger in *International Conference on Industrial Networks and Intelligent Systems*, Cham. 191–207 (Springer).

Nayyar, A., and Balas, V. E. (2019). "Analysis of simulation tools for underwater sensor networks (UWSNs)," in *International conference on innovative computing and communications 2019* (Singapore: Springer), 165–180.

Nayyar, A., Puri, V., and Le, D. N. (2019). "Comprehensive analysis of routing protocols surrounding underwater sensor networks (UWSNs)," in *Data management, analytics, and innovation 2019* (Singapore: Springer), (pp. 435–450).

Pappas, N., Dimitriou, I., and Chen, Z. (2018). "Network-level cooperation in random access IoT networks with aggregators," in *2018 30th International Teletraffic Congress (ITC 30)* Vienna, Austria. 245–253 (IEEE).

Raj Priyadarshini, R., and Sivakumar, N. (2020). Enhancing coverage and connectivity using energy prediction method in underwater acoustic WSN. *Journal of Ambient Intelligence and Humanized Computing* 11, 2751–2760. doi: 10.1007/s12652-019-01334-x

Shah, S., Khan, A., Ali, I., Ko, K.-M., and Mahmood, H. (2018). Localization free energy efficient and cooperative routing protocols for underwater wireless sensor networks. *Symmetry* 10, 498. doi: 10.3390/sym10100498

Shah, P. M., Ullah, I., Khan, T., Hussain, M. S., Khan, Z. A., Qasim, U., et al. (2016). "MobiSink: Cooperative routing protocol for underwater sensor networks with sink mobility," in *2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA)*. 189–197 (IEEE).

Tariq, M., Ayaz, M., Subhan, F., and Abbas, M. Z. (2021). Energy efficient and link reliable routing (E2LR) scheme for underwater sensor networks. *Peer-to-Peer Netw. Appl.* 14, 1870–1888. doi: 10.1007/s12083-021-01105-6

Teekaraman, Y., Sthapit, P., Choe, M., and Kim, K. (2019). Energy analysis on localization free routing protocols in UWSNs. *Int. J. Comput. Intell. Syst.* 12, 1526–1536. doi: 10.2991/ijcis.d.191203.001

Ullah, U., Khan, A., Zareei, M., Ali, I., Khattak, H. A., and Din, I. U. (2019). Energy-effective cooperative and reliable delivery routing protocols for underwater wireless sensor networks. *Energies* 12, 2630. doi: 10.3390/en12132630

Wadud, Z., Ismail, M., Qazi, A. B., Khan, F. A., Derhab, A., Ahmad, I., et al. (2019). An energy balanced efficient and reliable routing protocol for underwater wireless sensor networks. *IEEE Access* 7, 175980–175999. doi: 10.1109/ACCESS.2019.2955208

Yahya, A., Islam, S. U., Zahid, M., Ahmed, G., Raza, M., Pervaiz, H., et al. (2019). Cooperative routing for energy efficient underwater wireless sensor networks. *IEEE Access* 7, 141888–141899. doi: 10.1109/ACCESS.2019.2941422

Zeb, A., Wakeel, S., Rahman, T., Khan, I., Uddin, M. I., and Niazi, B. (2022). Energy-efficient cluster formation in IoT-enabled wireless body area network. *Comput. Intell. Neurosci.* 2022, 1–11. doi: 10.1155/2022/2558590

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 17 November 2022
DOI 10.3389/fmars.2022.1056300

Check for updates

OPEN ACCESS

EDITED BY
Hong Song,
Zhejiang University, China

REVIEWED BY
Narayanamoorthi R,
SRM Institute of Science and Technology, India
Siquan Yu,
Shenyang Institute of Automation (CAS), China
Helton Maia,
Federal University of Rio Grande do Norte, Brazil
Chao Qi,
Nanjing Agricultural University, China

*CORRESPONDENCE
Zhuang Zhou
jonahzhou199526@gmail.com

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 28 September 2022
ACCEPTED 25 October 2022
PUBLISHED 17 November 2022

CITATION
Yan J, Zhou Z, Zhou D, Su B, Xuanyuan Z, Tang J, Lai Y, Chen J and Liang W (2022) Underwater object detection algorithm based on attention mechanism and cross-stage partial fast spatial pyramidal pooling. *Front. Mar. Sci.* 9:1056300. doi: 10.3389/fmars.2022.1056300

COPYRIGHT
© 2022 Yan, Zhou, Zhou, Su, Xuanyuan, Tang, Lai, Chen and Liang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Underwater object detection algorithm based on attention mechanism and cross-stage partial fast spatial pyramidal pooling

Jinghui Yan[1], Zhuang Zhou[1]*, Dujuan Zhou[1,2], Binghua Su[1], Zhe Xuanyuan[3], Jialin Tang[1], Yunting Lai[1], Jiongjiang Chen[1,2] and Wanxin Liang[1]

[1]Key Laboratory of Intelligent Detection in Complex Environment of Aerospace Land and Sea, Beijing Institute of Technology, Zhuhai, China, [2]Faculty of Innovation Engineering, Macau University of Science and Technology, Macau, China, [3]Faculty of Science and Technology, Hong Kong Baptist University United International College, Beijing Normal University, Zhuhai, China

For the routine target detection algorithm in the underwater complex environment to obtain the image of the existence of blurred images, complex background and other phenomena, leading to difficulties in model feature extraction, target miss detection and other problems. Meanwhile, an improved YOLOv7 model is proposed in order to improve the accuracy and real-time performance of the underwater target detection model. The improved model is based on the single-stage target detection model YOLOv7, incorporating the CBAM attention mechanism in the model, so that the feature information of the detection target is weighted and enhanced in the spatial dimension and the channel dimension, capturing the local relevance of feature information, making the model more focused on target feature information, improved detection accuracy, and using the SPPFCSPC module, reducing the computational effort of the model while keeping the model perceptual field unchanged, improved inference speed of the model. After a large number of comparison experiments and ablation experiments, it is proved that our proposed ACFP-YOLO algorithm model has higher detection accuracy compared with Efficientdet, Faster-RCNN, SSD, YOLOv3, YOLOv4, YOLOv5 models and the latest YOLOv7 model, and is more accurate for target detection tasks in complex underwater environments advantages.

KEYWORDS

Underwater Object detection, ACFP-YOLO, YOLOv7, attention, SPPFCSPC

Introduction

Underwater target detection refers to the localization and identification of a specific target in an underwater scene. The technology is widely used in underwater cable laying, oil exploration, salvage and rescue, marine fish detection, undersea aquaculture, underwater navigation, smart fishery farming, underwater target striking and other

fields (Lin and Zhao, 2020) (YU, 2020) (Klausner and Azimi-Sadjadi, 2019). Although target detection algorithms for ground targets are relatively well established, the detection of underwater targets in the underwater environment still faces many challenges. The main reason is that the underwater medium, underwater light conditions, underwater submarine environment, etc. are more complex than the surface environment (Qiang et al., 2020) (Lei et al., 2022). Due to the differential attenuation of different wavelengths of light in water, scattering of light by plankton and suspended particles in water (Wei et al., 2021), making the target in underwater images and videos blurred and with severe color cast, it seriously affects the features of the target and creates serious obstacles for feature learning and recognition understanding of underwater targets. Therefore, underwater target detection continues to face a huge challenge (Jiang and Wang, 2020).

In order to improve the detection of fuzzy underwater targets and small underwater targets in the underwater environment, while maintaining the efficiency of the algorithm and ensuring good detection of the model. In this paper, based on the framework of YOLOv7 (Wang et al., 2022) we introduce the Convolutional Block Attention Module(CBAM) (Woo et al., 2018) attention mechanism and Cross-Stage Partial Fast Spatial Pyramid Pooling(SPPFCSPC) module, and propose a detection model that is suitable for target detection in underwater environment and has stronger feature extraction ability and better detection speed in underwater scenes.

Section 2 of this paper introduces the development process of underwater target detection and the current research problems in this field. Section 3 introduces the overall architecture of our proposed ACFP fusion model in detail, and explains the theory of CBAM attention mechanism and SPPFCSPC module. Section 4 introduces our experimental environment, experimental parameters, datasets and evaluation indicators. Section 5 is the result part. We conduct qualitative and quantitative analysis through the compatibility comparison of different attention mechanisms, ablation experiments, and comparison experiments with mainstream algorithms. Section 6 is a summary of the entire article, illustrating the advantages of our method for high-accuracy and real-time underwater object detection scenarios, as well as future research directions.

The contribution of this paper is shown below:

1. In this paper, we propose a target detection network model for underwater environment based on the improvement of YOLOv7. In which we fuse the CBAM attention mechanism and the SPPFCSPC module in the YOLOv7 model, this fusion idea effectively improves the detection accuracy of the model for underwater fuzzy targets and small targets, and provides an effective solution for the underwater target detection task.
2. On the URPC dataset, the mAP value of ACFP-YOLO is 80.62%, and the detection speed FPS value is 64.21. On the underwater garbage detection dataset, the mAP value is 74.92%, and the detection speed FPS value is 65.56. On both datasets, ACFP-YOLO achieves the highest detection accuracy and has better inference speed.
3. After a large number of comparative experiments and ablation experiments, we quantitatively and qualitatively verify that our model is superior to classical and state-of-the-art methods in the task of underwater environmental target detection from different perspectives. The detailed experiments provide a detailed idea and an important basis for other researchers to refer to our work.

Related work

In the early 21st century some researchers started to study underwater target detection algorithms, and with the development of artificial intelligence, new branches of research on underwater target detection algorithms have emerged. For the present the field is divided into two branches, one is the traditional algorithm based on the detailed feature analysis of the image, and the other is the neural network model algorithm based on the training of a large number of underwater images, by extracting features, analyzing the image features, extracting the target information and obtaining the final detection results (Tinghui et al., 2022).

Traditional target detection algorithms mainly digitize images and then apply mathematical theoretical knowledge to analyze and model them. However, the same target from different angles presents a different shape on the picture, and there are great difficulties in mathematical modeling of it, and the established model is difficult to be put into the application of realistic underwater scenes. QIU et al. (2019) proposed an underwater motion target detection algorithm based on surface feature ripples, this algorithm in photoelectric polarization imaging mode for underwater target detection, which became a mature masterpiece of traditional algorithms in underwater target detection. But, the traditional algorithm for modeling the features of a certain class or a certain target in underwater targets in different water quality environments is a major limitation, and the robustness for detecting underwater images with different complex backgrounds under different lighting is poor.

With the development of deep learning in recent years, target detection algorithms based on convolutional neural networks have been developed (Girshick et al., 2015) and have achieved better results in recent years, and this type of algorithm

has significant advantages in various detection tasks. Villon et al. (2016) did a study on the performance of traditional and deep learning algorithms, which used a two-stage extraction of HOG features and the use of an SVM classifier to compare detection on a coral reef fish dataset, and the results showed that the deep learning algorithm has more advantages. The current mainstream methods can be divided into two-stage target detection algorithms and one-stage target detection algorithms. Among them, the two-stage target detection algorithm performs the detection task in two stages, generating the suggestion frame first in the first stage and then making predictions in the second stage. This type of algorithm has high detection accuracy but is slow, and its representative algorithms are the Faster R-CNN (Ren et al., 2015) series. The one-stage target detection algorithm treats the target detection task as a single regression problem, and although it is slightly lower in accuracy than the two-stage algorithm, it has a faster detection speed, represented by the YOLO family of algorithms, such as YOLO (Redmon et al., 2016), YOLO9000 (Redmon and Farhadi, 2017), YOLOv3 (Redmon and Farhadi, 2018), YOLOv4 (Bochkovskiy et al., 2020), YOLOv5 (Jocher, 2020)and SSD (Liu et al., 2016).

Chen et al. (2020) proposed a new sample-weighted super network (SWIPENET) and a robust training paradigm–curriculum Multi Class Adaboost (CMA), for underwater images with small blurred samples. Shi et al. (2021) proposed an improved Faster-RCNN based underwater detection algorithm to improve the detection accuracy of Faster-RCNN in underwater scenes for problems such as low quality of underwater images, overlapping or occluded targets, and different sizes or shapes. Zeng et al. (2021) proposed a Faster R-CNN-AON network for the complex and variable underwater environment with limited acquired sample images, and introduced an adversarial network to improve the overall detection performance of the model. However, the optimization algorithm based on two-stage algorithm still has the problems of low efficiency and poor real-time performance. To address this problem above, Tinghui et al. (2022) proposed an improved YOLOv5 underwater target detection network model, the model has unique performance in underwater target detection and maintains with real-time and accuracy. Lei et al. (2022) used Swin Transformer as the base backbone of YOLOv5 for underwater image blurring, making the network suitable for underwater images with blurred targets.

Attention mechanism and cross-stage partial fast spatial pyramid pooling (ACFP)

We present our underwater target detection model in detail in Section 3. We based on the YOLOv7 framework, incorporating the CBAM attention mechanism, Using the channel attention mechanism and spatial attention mechanism, the channel weight of the detection target is increased, while the perceptual field of the target to the original image is expanded, allowing the model to pay more attention to the feature information of the detection target. We improve the original SPPCSPC module to SPPFCSPC module, which reduces the computation of the model and improves the inference speed of the model while keeping the perceptual field unchanged.

Network model for underwater target detection

Before feeding the network with images, we first perform a distortion-free affine transformation of the original input image to a 640×640 size image, which is then used as the input to the model. The overall network framework model after our fusion is shown in Figure 1, where the modules marked with red boxes are the parts of the model fusion. The input images are first passed through the backbone network for feature extraction. In order to retain multi-scale information, the backbone network provides a variety of different scales and outputs the multi-scale feature maps to the neck network as the input of the neck network. After the neck network, the fusion of feature maps containing shallow fine-grained information and deep semantic information is combined, thus enhancing the expressive power of the network and assigning the multi-scale learning task to multiple detection networks of different sizes. Finally, the feature information is integrated and transformed into detection prediction output.

Backbone network is used for feature extraction of images, such as texture, color and shape of images. It can provide multiple scales, multiple combinations of sense field sizes and center steps, thus meeting the requirements of different scales and categories. The extraction process of backbone network is shown in Figure 1A, firstly, it goes through 4 CBS modules for convolution, normalization and activation, and then it is after the E-ELAN module and MP module to extract features alternately, leading to the output of the last 3 E-ELAN modules as the input of neck. Among them, the E-ELAN module is composed of multiple convolutional layers, and the MP module is composed of MaxPool and CBS modules as shown in Figure 2C, D.

In order to enable the model to learn diverse information and improve the performance of target detection, the role of neck network is to disperse the multi-scale output learning provided by backbone network to multiple feature maps and fuse the learned multi-scale information together, which improves the perceptual wildness of the model while effectively separating the most important contextual features and avoiding the image distortion problem to some extent. As in Figure 1B, the neck network is a PAFPN structure, consisting of a modified

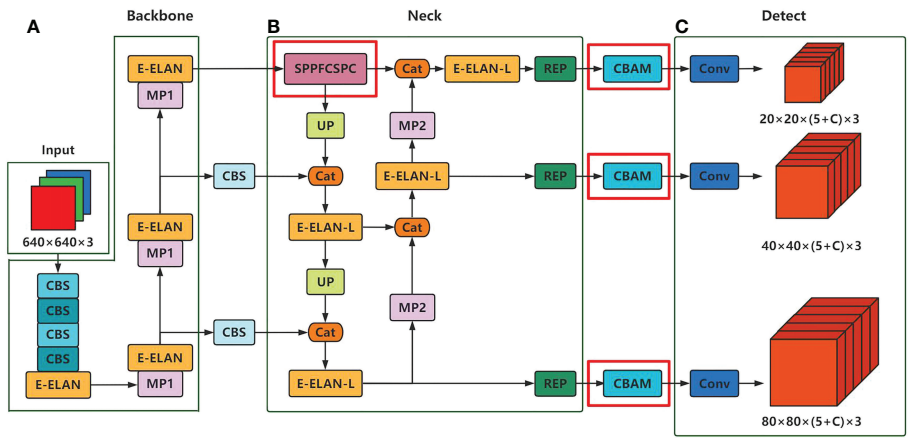


FIGURE 1
Overall architecture diagram of the model. Where **(A)** denotes the backbone network for feature extraction, **(B)** denotes the neck network for feature fusion, **(C)** denotes the detect network used to obtain the model prediction results, and *C* denotes the number of categories in the dataset.

FPN (Lin et al., 2017) and PANet (Liu et al., 2018) structure, for extracting features and fusing them. The PAFPN structure is basically the same as that of YOLOV5, except that the PAFPN structure of YOLOv7 uses the E-ELAN-L module for feature extraction and fusion, and the MP module for down sampling, this makes this structure more capable of strengthening features than previous PAFPN structures. Finally, after two feature extractions of the input image by backbone network and neck network, the 1×1 convolution is used to integrate the feature information to convert it into the final prediction information, as shown in Figure 1C, to obtain the prediction results of the model.

SPPFCSPC for ACFP

Spatial pyramid pooling is more effective than simply using maximum pooling to increase the received range of backbone features, significantly separating the most important contextual features, and this structure outputs fixed-size feature vectors after multi-scale feature extraction to increase the perceptual field of the network.

The SPPCSPC structure and the SPPFCSPC structure are shown in Figure 3. The SPPCSPC structure in YOLOv7 uses three independent pooling layers with different sizes of pooling kernels to compute a spatial pyramid pooling structure. The

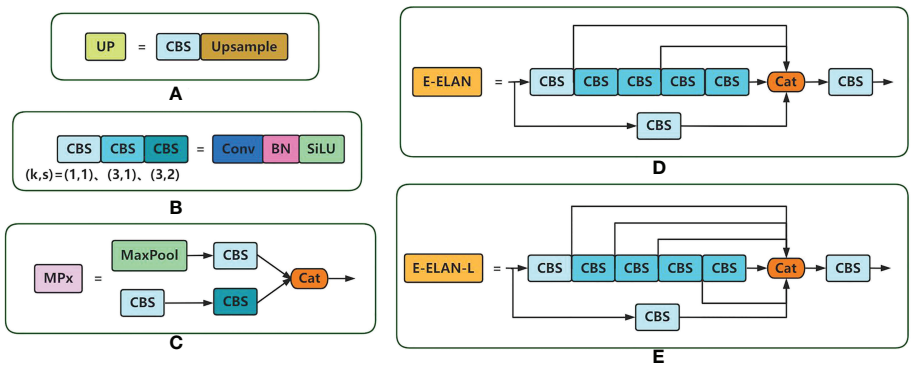


FIGURE 2
Structure diagram of the model part of the module. Where **(A)** denotes the upsampling module, **(B)** denotes the composition of different convolution modules, where *k* denotes the size of the convolution kernel and *s* denotes the convolution step size, **(C)** denotes the basic structure of the MP module, **(D)** denotes the basic structure of the E-ELAN module, and **(E)** denotes the basic structure of the E-ELAN-L module.

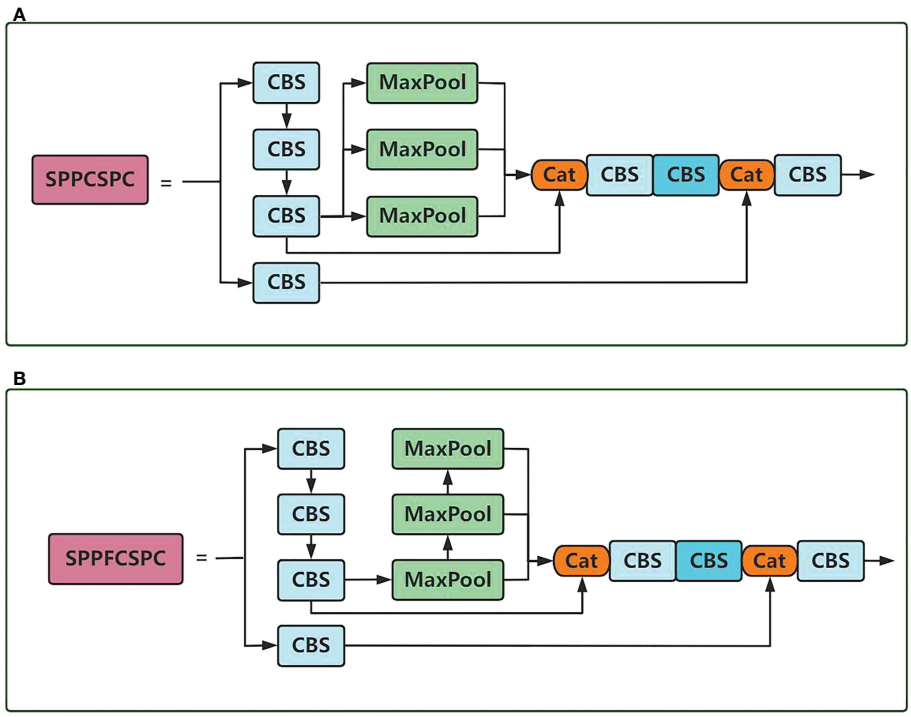


FIGURE 3
(A) indicates the SPPCSPC structure diagram, **(B)** indicates the SPPFCSPC structure diagram.

relevant pooling part of the equation is shown in equation (1), but the three pooling have the same input, and the results of the larger pooling kernel can be calculated on the computational results of the smaller output results of the pooling kernel, reducing the computational effort without changing the perceptual field of the module.

$$\mathbf{R}(\mathbf{F}) = \mathbf{MaxPool}_{k=5}^{p=2}(\mathbf{F}) \circledast \mathbf{MaxPool}_{k=9}^{p=4}(\mathbf{F}) \circledast \mathbf{MaxPool}_{k=13}^{p=6}(\mathbf{F}) \tag{1}$$

Where R denotes the output result, $\circledast$ denotes tensor stitching, and F denotes the input feature layer.

The SPPFCSPC structure is optimized for the SPPCSPC structure, and the pooling part is calculated as shown in equations (2), (3), (4), and (5), linking three separate pooling uses less computation on the output results of the pooling layer of the smaller pooling kernel, yielding the pooling layer results of the larger pooling kernel, gaining speedup while keeping the perceptual field constant.

$$R_1(\mathbf{F}) = \mathbf{MaxPool}_{k=5}^{p=2}(\mathbf{F}) \tag{2}$$

$$R_2(R_1) = \mathbf{MaxPool}_{k=5}^{p=2}(R_1) \tag{3}$$

$$R_3(R_2) = \mathbf{MaxPool}_{k=5}^{p=2}(R_2) \tag{4}$$

$$\mathbf{R_4} = \mathbf{R_1} \circledast \mathbf{R_2} \circledast \mathbf{R_3} \tag{5}$$

Where R_1 denotes the pooling layer result for the minimum pooling kernel, R_2 denotes the pooling layer result for the medium pooling kernel, R_3 denotes the pooling layer result for the maximum pooling kernel, and R_4 denotes the final output result, $\circledast$ denotes tensor stitching.

Attention for ACFP

CBAM is an attention mechanism module that incorporates two dimensions of feature channel information and feature space information. As shown in Figure 4, CBAM processes the incoming feature layers by the channel attention mechanism and the spatial attention mechanism, respectively, and automatically obtains the importance level for each feature channel and feature space by learning, and uses the obtained importance level to enhance features and suppress features that are not important for the current task. The overall equation of CBAM is summarized as shown in equation (6)(7):

$$\mathbf{F}' = \mathbf{M_c}(\mathbf{F}) \otimes \mathbf{F} \tag{6}$$

$$\mathbf{F}'' = \mathbf{M_s}(\mathbf{F}') \otimes \mathbf{F}' \tag{7}$$

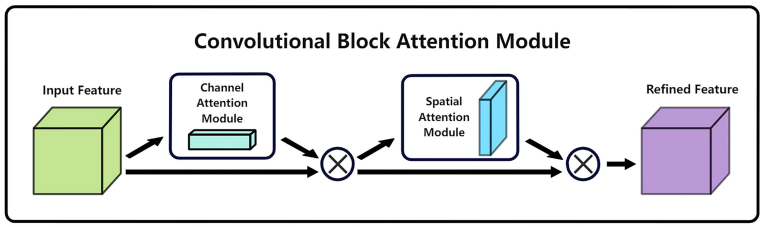


FIGURE 4
Overall structure of CBAM attention mechanism.

The working process of CBAM is to first multiply the input feature layer F through the channel attention mechanism and the obtained M_c with the input feature layer F to obtain the output of strengthening and suppression on the channel F', and then use F' as the input of the spatial attention mechanism, the obtained M_s is multiplied with F' to obtain the final output F'' of reinforcement and suppression in the channel content and spatial location.

The module used for the channel attention mechanism is shown in Figure 5A, which consists of MaxPool, AvgPool, and Shared MLP. The related equation is shown in equation (8):

$$\begin{aligned}\mathbf{M_c}(\boldsymbol{F}) &= \sum(\mathbf{MLP}(\mathbf{AvgPool}(\mathbf{F})) + \mathbf{MLP}(\mathbf{MaxPool}(\mathbf{F}))) \\ &= \sum\left(W_1(\mathbf{W}_0(\mathbf{F}^c_{\mathbf{avg}})) + W_1(\mathbf{W_0}(\mathbf{F}^c_{\mathbf{max}}))\right)\end{aligned} \tag{8}$$

Where σ denotes the sigmoid function and W_0 W_1 denote the two shared fully connected layers that make up the *MLP*.

For the input single feature layer $F(H\times W\times C)$, MaxPool and AvgPool are performed in the $H\times W$ dimension respectively to compress the feature layer to $1\times1\times C$. The one-dimensional parameters after MaxPool compression retain the feature texture of the original feature layer and converge the important information to distinguish the object features. The one-dimensional parameters compressed by AvgPool retain the global visual information of $H\times W$ before compression and have a larger perceptual area. After that, the results of MaxPool and AvgPool are fed into Shared *MLP* network for processing, and then the two processed results are summed to obtain the feature map channel weights.

The module used for the spatial attention mechanism is shown in Figure 5B, which consists of MaxPool, AvgPool and conv layer. The related equation is as in (9):

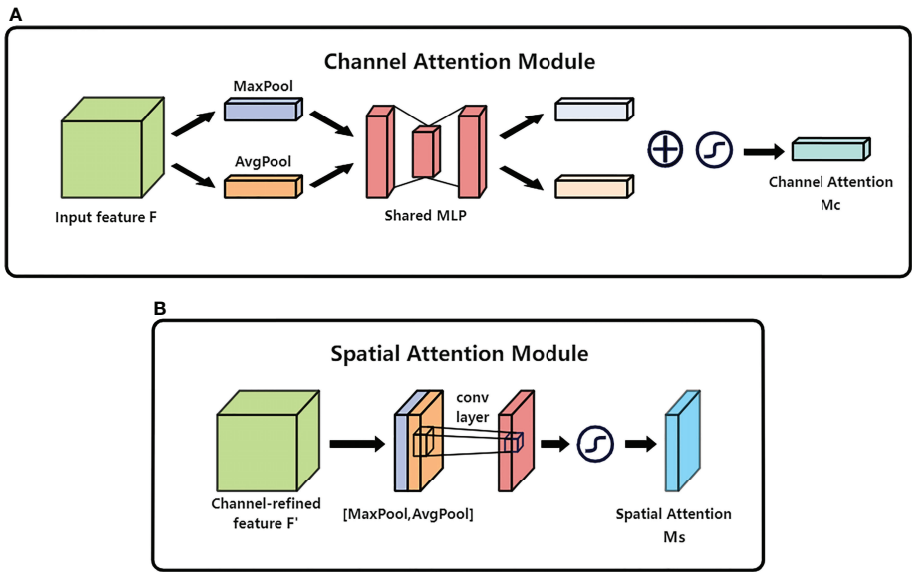


FIGURE 5
(A) denotes the specific structure of the channel attention mechanism branch and **(B)** denotes the specific structure of the spatial attention mechanism branch.

$$\begin{aligned}\mathbf{M_s}(\mathbf{F}) &= \sum(\mathbf{f}^{7\times7}([\mathbf{AvgPool}(\mathbf{F});\mathbf{MaxPool}(\mathbf{F})])) \\ &= \sum\left(\mathbf{f}^{7\times7}([\mathbf{F}^s_{avg};\mathbf{F}^s_{\max}])\right)\end{aligned} \tag{9}$$

Where σ denotes the sigmoid function and $f^{7\times7}$denotes the convolution operation with a convolution kernel size of 7 × 7.

For a single feature layer $F(H\times W\times C)$ input in, MaxPool and AvgPool are performed in the channel dimension respectively to compress the feature layer to $H\times W\times1$. The compressed feature layer focuses the effective information of the region in the space and is used to extract the efficient information region along the channel, after which the results of both are concat and then convolutional dimensionality reduction is performed to obtain the feature map space weights, thus capturing the local relevance of the feature information.

Loss function

All the experiments in this paper use training without the auxiliary training head, therefore, we only describe the loss function when training without the auxiliary training head in the following, and the overall loss calculation formula is shown in equation (10):

$$\mathcal{L}_{v7}(\mathrm{t_p},\mathrm{t_{gt}}) = \sum_{\mathrm{m=0}}^{\mathrm{M}} \left[\alpha_{\mathrm{m}}^{\mathrm{bal}}\alpha_{\mathrm{box}}\sum_{\mathrm{i=0}}^{\mathrm{S^2}}\sum_{\mathrm{j=0}}^{\mathrm{B}}I_{\mathrm{kij}}^{\mathrm{obj}}\mathcal{L}_{\mathrm{CIoU}} + \alpha_{\mathrm{obj}}\sum_{\mathrm{i=0}}^{\mathrm{S^2}}\sum_{\mathrm{j=0}}^{\mathrm{B}}\mathrm{I}_{\mathrm{kij}}^{\mathrm{obj}}\mathcal{L}_{\mathrm{obj}} + \alpha_{\mathrm{cls}}\sum_{\mathrm{i=0}}^{\mathrm{s^2}}\sum_{\mathrm{j=0}}^{\mathrm{B}}I_{\mathrm{kij}}^{\mathrm{obj}}\mathcal{L}_{\mathrm{cls}}\right] \tag{10}$$

Where M denotes the output feature layer, S^2 denotes the cell, B denotes the number of anchors on each cell. a_{box}, a_{obj}, a_{cls} denote the weights of the corresponding terms, and the values taken in the experiment are a_{box}=0.05, a_{obj}=1.0, a_{cls}=0.03. I_{kij}^{obj} is the control function, which indicates whether the mth output feature map, the i-th cell, the j-th anchor box is a positive sample, if it is a positive sample, it is 1, and vice versa, it is 0. t_{p},t_{gt} is the prediction vector and ground-truth vector. a_m^{bal} is used to balance the weights of the output feature map of each scale, and the values are [4.0,1.0,0.4], which correspond to 80×80, 40×40, 20 ×20 for the output feature maps.

The localization coordinate loss expressed by the Bounding Box boundary regression loss function, using CIoU loss, calculates the localization loss of positive samples only, as shown in (11)(12):

$$\begin{aligned}\mathcal{L}_{\mathrm{CIoU}}(\mathrm{b},b_{\mathbf{gt}}) &= 1-\mathrm{CIOU} \\ &= 1-\left(\mathrm{IOU}-\frac{d_o^2}{d_c^2}-\frac{v^2}{1-\mathrm{IOU}+\mathrm{u}}\right)\end{aligned} \tag{11}$$

$$\mathbf{u} = \frac{\mathbf{4}}{\pi^2}\left(\mathbf{arctan}\ \frac{b_{gt-w}}{b_{gt-h}} - \mathbf{arctan}\ \frac{b_W}{b_h}\right)^2 \tag{12}$$

Where b,b_{gt} denotes the prediction frame vector b_{x},b_{y},b_{w},b_{h} and the ground-truth vector $b_{\mathrm{gt\text{-}x}}$, $b_{\mathrm{gt\text{-}y}}$,$b_{\mathrm{gt\text{-}w}}$, $b_{\mathrm{gt\text{-}h}}$.IOU denotes the intersection ratio between the prediction frame and the ground-truth. d_o denotes the Euclidean distance between the prediction frame and the center point of the ground-truth. d_c denotes the diagonal distance between the prediction frame and the smallest outer rectangle of the ground-truth v denotes the impact factor measuring the aspect ratio.

The target confidence loss function adopts BCE loss, which only calculates the objective loss of the samples obtained from positive sample matching, and the specific formula (13):

$$\mathcal{L}_{\mathrm{obj}}(\mathrm{p_o},\mathrm{p_{iou}}) = \mathrm{BCE}^{\mathrm{sig}}_{\mathrm{obj;}}(\mathrm{p_o},\mathrm{p_{iou}};\mathrm{w_{obj}}) \tag{13}$$

Where p_o denotes the target confidence score in the prediction frame, p_{iou} denotes the prediction frame and the IOU value of the ground-truth corresponding to it.

The classification loss function, using BCE loss, calculates the classification loss of positive samples only, and the specific calculation formula is as (14):

$$\mathcal{L}_{\mathrm{cls}}(c_{\mathrm{p}},c_{\mathrm{gt}}) = \mathrm{BCE}^{\mathrm{sig}}_{\mathrm{cls}}(c_{\mathrm{p}},c_{\mathrm{gt}};w_{\mathrm{cls}}) \tag{14}$$

Where c_p denotes the probability of the target category in the prediction frame, c_{gt} denotes the probability of the category of the ground-truth to which the prediction frame corresponds.

Experiments

Experimental environment and hyperparameter settings

All experimental data in this paper are measured in the same environment. The hardware environment uses Intel(R) Core (TM) i7-12700KF@3.61 GHz CPU, 16GB RAM, NVIDIA GeForce RTX 3080 Ti graphics card. The system environment is Windows 10 Professional Edition. Python version 3.8, PyTorch version 1.12.0, CUDA version 11.6.

The relevant parameters in the experiment are shown in the Table 1. The gradient descent optimizer used to update the convolution kernel parameters is Adam, and the optimizer Momentum is 0.937, the learning rate update method during the training process is step, the maximum learning rate is 0.001, the frozen training batch size is 8. The epoch of freezing training is 50, the batch size of unfreezing training is 4, the epoch of unfreezing training is 50, all experiments only load the pre-training weights of the backbone network part, and other parts

TABLE 1 Experiment-related hyperparameter settings.

Hyperparameter	Freeze_train	Epoch	Batch_size	Max_learning_rate	Optimizer	Momentum	Lr_decay
Value	True	1-50	8	0.001	Adam	0.937	Step
	False	51-100	4				

are trained from scratch, and the total training epoch is 100. The frozen training model only trains other parts except the backbone network, and the entire network model is trained when unfreezing training.

Dataset

There are 2 datasets used in the experiments in this paper. The main experiments are performed on the URPC (Lab, 2018) dataset, and the auxiliary verification experiments are performed on the underwater garbage detection dataset (Fulton et al., 2019). For the URPC dataset, in order to enable the model to learn more features, more pictures with the same category are added, and the category of waterweeds is added. In order to make the category distribution of the training set and test set more reasonable, we replace Some pictures of the test set. The dataset consists of 4571 images, including 3771 training images and 800 testing images, covering 5 target categories: scallop, holothurian, starfish, echinus, and waterweeds. The underwater garbage detection data set has a total of 7337 pictures, including 6206 training pictures, 1461 test pictures, and 13 categories marked, namely timestamp, Paper, Wood, Bio, Metal, Rov, Plastic, Unknown, Papper, Platstic, Rubber, Cloth and Fishing. The pictures of the dataset were taken in the real marine environment, and the pictures have problems such as color distortion, low contrast, blurred feature information, etc., and there are occlusions, dense targets, and uneven distribution of the number of targets in different categories, which gives underwater problems. Object detection brings great challenges.

Evaluation indicators

There are seven main indicators used in this study to test the performance of the model. Precision(P) represents the proportion of the positive class that the model considers to be a positive class, and the calculation formula is in Equation 15. Recall(R) represents the proportion of the positive class divided by the model to the total positive class, and the calculation formula is in Equation 16. Average Precision (AP) means that each class is composed of Precision and Recall taking different thresholds The area under the curve, the larger the value, the better the recognition accuracy of the class, the formula for calculation is in Equation 17. The mean Average Precision (mAP) represents the average AP of all classes, and the larger the value, the better the model The better the accuracy of identifying the target, the calculation formula is in Equation 18. Frame Per Second (FPS) (Liu et al., 2009) represents the number of frames processed by the model per second, reflecting the speed of the model inference, the larger the value, the faster the inference speed of the model, and the better the model performance. Billions of floating point operations per second (GFLOPS) is the number of computations required by the model and measures the complexity of the model. Number of parameters(params) is the sum of the parameters in the model and is used to evaluate the model size.

$$\mathbf{P(Precision)} = \frac{\mathbf{TP}}{\mathbf{TP + FN}} \tag{15}$$

$$\mathbf{R(Recall)} = \frac{\mathbf{TP}}{\mathbf{TP + FN}} \tag{16}$$

$$\mathbf{AP} = \int_0^1 \mathbf{P(r)dr} \tag{17}$$

$$\mathbf{mAP} = \frac{\mathbf{1}}{\mathbf{N}} \sum_{\mathbf{n=1}}^{\mathbf{N}} \mathbf{A}P_{\mathbf{n}} \tag{18}$$

Where *TP* represents the number of positive samples predicted by the model correctly, and *FP* represents the number of positive samples predicted by the model that are actually negative samples. *FN* represents the number of positive samples predicted by the model to be negative. *P* represents the precision of this class, *r* represents the recall of this class, *N* represents the number of all classes, and AP_n represents the average precision of class *n*.

Results

Compatibility of attention mechanisms

SENet (Hu et al., 2018) is a typical implementation method of channel attention mechanism, which focuses on obtaining the enhanced weights of the input feature layer on the channel, but ignores the weight information of the target spatial position. ECA (Wang et al., 2020) is also an implementation form of the channel attention mechanism, which obtains the enhancement weight of each feature layer by obtaining cross-channel information. Although it has better cross-channel information, it also ignores the spatial information of the target. The CA (Hou

et al., 2021) attention mechanism embeds the location information into the channel attention, decomposes the channel attention into two feature encoding processes, aggregates the features along two spatial directions respectively, and obtains the weight of the fusion channel information and spatial information. CBAM combines the channel attention mechanism and the spatial attention mechanism to deal with the channel weight and the spatial weight respectively, that is, it pays attention to both the channel information and the spatial information. Different attention mechanisms focus on different information directions, and different models have different compatibility with different attention mechanisms.

We chose to use the CBAM attention mechanism in our improved model, and to verify the compatibility of the CBAM attention mechanism with the model, we compared it with the models without fused attention mechanism, fused SENet attention mechanism, fused ECA attention mechanism, and fused CA attention mechanism in separate experiments.

A visualization method in deep learning was used in the experiments for qualitative analysis, which is the Gradient Weighted Class Activation Mapping (Grad-CAM) (Zhou et al., 2016), used to show the differences in the regions of interest for the different attention mechanisms introduced by the model, reflecting the degree of influence of different regions on the results. In this case, the feature importance increases sequentially from blue to red light.

As shown in Figure 6, compared with the visualization results of other attention mechanisms, the overall coverage area of the heat map of the CBAM attention mechanism is larger, indicating that the model focuses on a larger learning area at locations with targets, and the overall feature extraction of the targets is more adequate, which is beneficial to the detection of small targets, and the red area also becomes larger, indicating that the effective target feature information is enhanced and the model is more focus on the target information that should have been focused on. From the experimental results, it can be seen that the introduction of CBAM attention mechanism makes the model pay more attention to the feature information of the target to be recognized, and suppresses the effect of target features that are not obvious due to the complex underwater background, and shows better results compared with other attention mechanisms.

The experiments were quantitatively analyzed with mAP assessment criteria. We changed only the attention mechanism module, and then measured the mAP values of each model, and compared the mAP values of different models to assess the compatibility of different attention mechanisms with the models, and the data of the comparison experiments are shown in Table 2.

The experimental data show that the model incorporating the CBAM attention mechanism has higher detection accuracy compared to the models incorporating the SE attention mechanism, incorporating the ECA attention mechanism, incorporating the CA attention mechanism, and not incorporating the attention mechanism. The detection accuracy of the model with fused SE attention mechanism and ECA attention mechanism decreased by 0.75% and 0.95%,

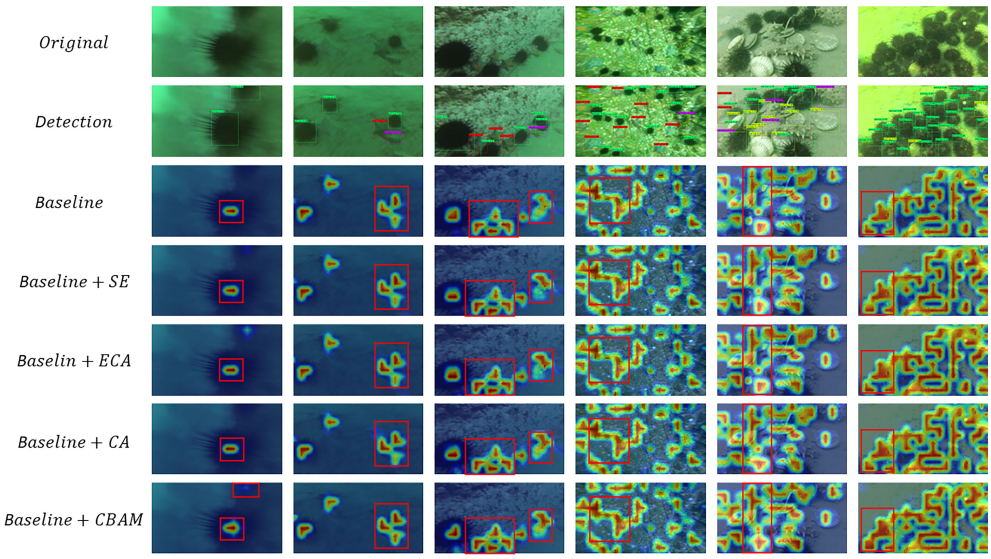


FIGURE 6
Heatmap of different attention mechanisms. Original shows the data set image. Detection shows the detection results of Baseline+CBAM, and the Baseline shows the heat map of YOLOv7. Baseline+XX shows the heat map of YOLOv7 integrated into the XX attention mechanism (XX is SE, ECA, CA, CBAM).

TABLE 2 mAP measurements for different attention mechanisms.

Model	Attention	Input shape	mAP(%)
YOLOv7	–	640×640	78.98
YOLOv7	SENet	640×640	78.23
YOLOv7	ECA	640×640	78.03
YOLOv7	CA	640×640	79.39
YOLOv7	CBAM	640×640	**80.67**

The bold value is the best value in the comparison.

respectively. The detection accuracy of the model with fused CA attention mechanism improved by 0.41%. The detection accuracy of the model with fused CBAM attention mechanism improved by 1.69%. Compared with the original YOLOv7 model, which is more adaptable to the underwater scenario and has better model compatibility.

Ablation experiments

As described in Section3, we introduced the CBAM attention mechanism in the model, as well as the improved SPPCSPC module. To verify the effectiveness of the improved YOLOV7 model for underwater target detection, we controlled a variable by the control variable method and quantitatively analyzed the experimental results. In the experiments, we measured the mAP and FPS values of each model and compared them by metrics to verify the importance of the improved module for the model. Three models were designed for comparison with the improved model in this experiment, where experiment 1 represents the original YOLOV7 model, experiment 2 incorporates the CBAM attention mechanism based on experiment 1, experiment 3 replaces the SPPCSPC module with the SPPFCSPC module based on experiment 1, and experiment 4 is the improved model. The experimental data are shown in Table 3.

Comparing the data from Exp.1 and Exp.2, the model with the introduction of the CBAM attention mechanism improves the average detection accuracy (mAP) by 1.69% and slightly reduces the model inference speed, indicating that the CBAM attention mechanism uses channel attention to establish the correlation between channels, thus suppressing the non-essential feature information, while using the spatial attention mechanism to extract the spatial location of the target more effectively. Through the parallel action of both, the model pays more attention to the feature information of the detection target, thus improving the quality of the feature mapping and significantly improving the overall accuracy of the model, but the CBAM attention mechanism increases the complexity of the model and reduces the inference speed of the network; comparing the data of Exp.1 and Exp.3, the model inference speed (FPS) is improved by 0.85%, indicating that replacing the SPPCSPC module is replaced with the SPPFCSPC module, the model inference speed is improved while keeping the perceptual field unchanged; comparing the data of Exp.1 with Exp.4, the average detection accuracy (mAP) on the model is improved by 1.64% and the inference speed is slightly reduced, indicating that the YOLOv7 model that incorporates the CBAM attention mechanism and replaces the SPPCSPC module sacrifices a small portion of speed in exchange for higher detection accuracy, balancing the one-sided performance degradation brought by using either one alone, and making the overall performance of the model more superior.

Comparison with mainstream algorithms

Our proposed ACFP-YOLO algorithm has good feature extraction ability in complex underwater scenes, and has a faster detection speed, and has better performance in underwater target detection. In order to verify the superiority of the ACFP-YOLO algorithm in this paper in underwater detection, we compare the algorithm in this paper with Efficientdet (Tan et al., 2020), Faster-RCNN (F-RCNN), SSD, YOLOv3, YOLOv4, YOLOv5, YOLOv7 target detection mainstream algorithms, in On the same data set URPC, the same training method is used for network model training, and the superiority of different algorithm models is compared through qualitative analysis and quantitative analysis.

In the experiment, we qualitatively analyze the performance of the algorithm through the detection renderings of different models, and we select the model with better detection effect for analysis. Figure 7 shows Faster-RCNN-ResNet50, YOLOv3, YOLOv4, YOLOv5-l, YOLOv7 detection renderings. From the intuitive renderings, it can be concluded that the detection effect of F-RCNN is better than that of YOLOv3 and YOLOv4, and is comparable to that of YOLOv5, but the detected target

TABLE 3 The impact of the fusion of different modules of the model on the metrics.

Index	CBAM	SPPFCSPC	Input shape	mAP (%)	FPS
Exp.1	×	×	640x640	78.98	65.94
Exp.2	✓	×	640x640	80.67	62.29
Exp.3	×	✓	640x640	78.45	66.50
Exp.4	✓	✓	640x640	80.62	64.21

"✓" indicates that the module is incorporated in the model, and "x" indicates that the module is not incorporated in the model.

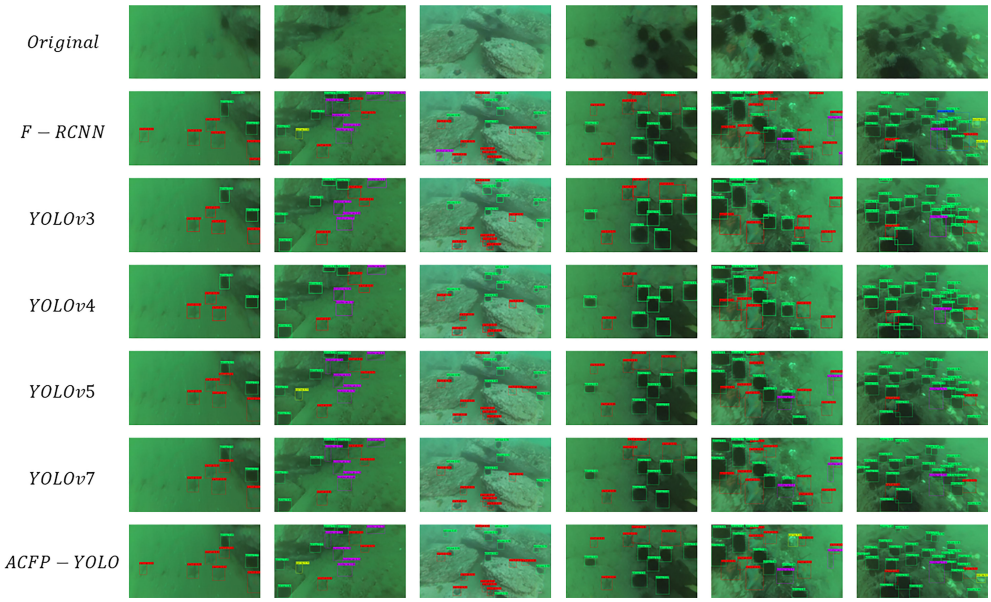


FIGURE 7
The first row *Original* represents the original image of the dataset, the second row represents the F-RCNN detection image, the third row represents the YOLOv3 detection image, the fourth row represents the YOLOv4 detection image, the fifth row represents the YOLOv5 detection image, and the sixth row represents the YOLOv7 detection image Figure, the seventh row represents the ACFP-YOLO detection map.

probability is generally lower than that of YOLOv5, and there is a target misjudgment. The ACFP-YOLO algorithm in this paper has better target recognition effect than YOLOv3, YOLOv4 and YOLOv5, and has fewer misjudgment and recognition compared with F-RCNN.

In order to better observe and compare the detection effect of the ACFP-YOLO algorithm, the ACFP-YOLO algorithm is compared with YOLOv7, and YOLOv7 has the best detection effect among other algorithms. Figure 8 shows the detection results of the ACFP-YOLO algorithm and the YOLOv7 algorithm. The target marked by the yellow box in the figure has blurred edges and distorted colors, which makes it difficult to identify the features similar to the background. The ACFP-YOLO algorithm is not obvious for such target features, the target edge is blurred, the detection effect of small targets is better, and the detection ability of edge feature information, overlapping and blurred targets is stronger. From the perspective of detection effect, the ACFP-YOLO algorithm has better feature extraction ability in complex underwater scenes, and improves the detection ability of small targets and targets with indistinct edge features. At the same time, the fusion of the CBAM attention mechanism enhances the spatial feature information of the model for small targets, improves the model's detection ability for small targets, and improves the detection ability of underwater targets at various scales.

In the experiment, we carried out quantitative comparative analysis of each model by measuring the mAP value, params value, GFLOPS value and FPS value of each model. Efficientdet used the D1 model for the experiment, and Faster-RCNN used the VGG backbone network and the ResNet50 backbone network for the experiment. All model comparison experimental measurement results are shown in Table 4.

Observing the experimental data, from the perspective of detection accuracy, the mAP value of the ACFP-YOLO algorithm is 80.62%, which is much higher than other mainstream target detection algorithms at present. The experimental data show that the ACFP-YOLO algorithm has more advantages in the underwater target detection task. From the unilateral point of view of detection speed, compared with models of the same scale, the improved model maintains a medium-to-high level of detection speed and has good real-time performance. Compared with the mainstream two-stage target detection algorithm Faster-RCNN (ResNet50), the improved YOLOv7 algorithm is 18.0% higher in accuracy and 152.97% faster in speed. The model size has increased by 9.245M. Compared with YOLOv5-l, the most widely used one-stage target detection algorithm in industry, mAP has increased by 3.45%, the detection speed has decreased by 1.29%, and the model size has decreased by 9.092M. Compared with the YOLOv7 algorithm, the improved ACFP-YOLO algorithm mAP increased by 1.64%, the detection speed decreased by 2.62%, and the model size increased by 0.345M. It shows that the improved ACFP-YOLO algorithm in this paper is an algorithm with high detection accuracy. While improving the

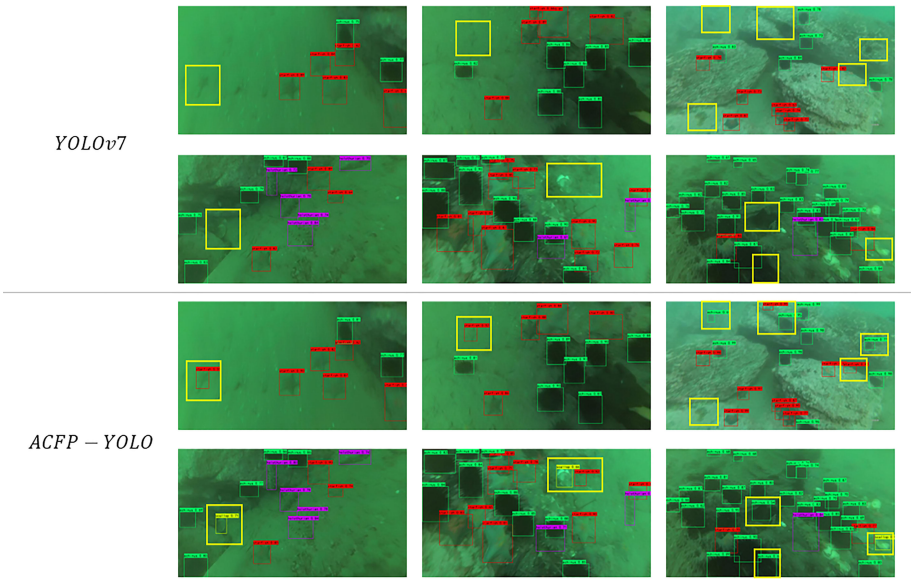


FIGURE 8
The detection effect of YOLOv7 and the detection effect of ACFP-YOLO.

detection accuracy of the model, it loses a small part of the detection speed.

In order to prove the superiority of the ACFP-YOLO algorithm in underwater scene performance, we use the same method to conduct a comparative experiment again on the underwater garbage dataset. We chose the detection effect of YOLOv7 and ACFP-YOLO, which has the best detection effect among other mainstream algorithms, for comparison. The Figure 9 shows the detection effect of ACFP-YOLO and YOLOv7 on the underwater garbage dataset. The original data set images are obtained at different water quality, different depths, and using different cameras. There are situations where the target features are not obvious, the features are attenuated, and the image background is complex. For this kind of picture, YOLOv7 shows a good detection effect, but ACFP-YOLO shows a better effect, we marked a part of the target objects with black boxes in Figure 9. Some of them were considered to be backgrounds by YOLOv7 because of the inconspicuous target features and small targets, and they were not detected. However, ACFP-YOLO has better detection ability for targets in this situation, and there is also a situation where the targets are occluded by each other. ACFP-YOLO can still detect it, while the detection effect of YOLOv7 is not as good as that of ACFP-YOLO. Through qualitative analysis, it can be seen that

TABLE 4 Performance metric values of mainstream target detection algorithms on URPC dataset.

Model	Input shape	mAP (%)	FPS	GFLOPS (G)	params (M)
SSD	416×416	57.73	100.93	116.159	24.146
YOLOv3	416x416	65.40	93.08	65.626	61.545
YOLOv4-tiny	416x416	52.19	215.96	6.84	5.883
YOLOv4	416x416	61.96	68.91	59.982	63.959
ACFP-YOLO	416x416	78.22	67.35	44.439	37.216
Efficientdet-D1	640x640	51.23	32.47	11.544	6.558
Faster-RCNN-VGG	600x600	59.41	36.51	369.817	136.771
Faster-RCNN-ResNet	600x600	62.62	25.40	940.972	28.316
YOLOv5-s	640x640	72.47	101.13	16.511	7.074
YOLOv5-l	640x640	77.17	65.05	114.627	46.653
YOLOv7-tiny	640x640	72.73	132.18	13.215	6.025
YOLOv7	640x640	78.98	65.94	105.182	37.216
ACFP-YOLO	640x640	80.62	64.21	105.191	37.561

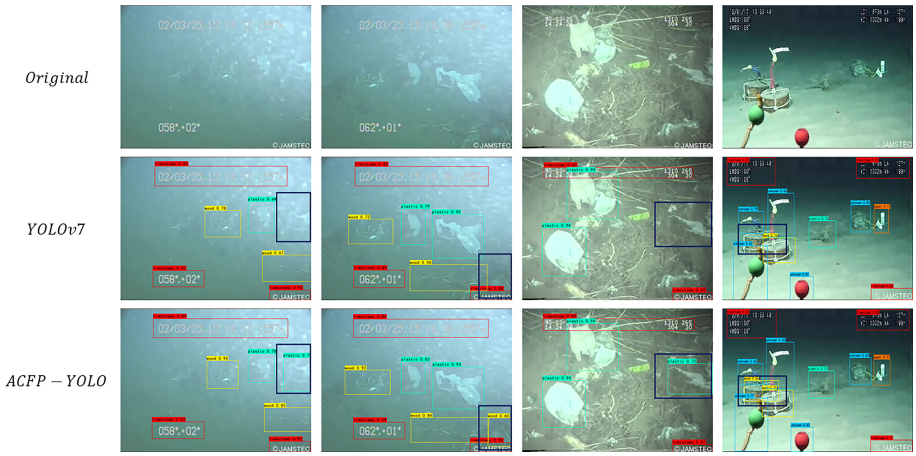


FIGURE 9
The first line Original is the original image of the dataset, the second line is the detection effect map of YOLOv7, and the third line is the detection effect map of ACFP-YOLO.

the ACFP-YOLO algorithm has better feature extraction ability in complex underwater scenes with different water quality and different depths, and has better detection ability for different color attenuation and occluded objects.

In the experiment of underwater garbage detection dataset, we selected consistent models for comparison, measured mAP value, GFLOPS value, params value and FPS of different models, and quantitatively analyzed the performance of each model through different indicators. All the comparative experimental measurement results in the experiment are shown in Table 5.

According to the data analysis in Table 5, from the perspective of detection accuracy, the mAP value of ACFP-YOLO algorithm on the underwater garbage dataset is 74.92%, which is still higher than other current mainstream target detection algorithms, the superior performance reflected in the underwater scene. The mAP value of ACFP-YOLO is 1.07% higher than that of YOLOv7, 2.60% higher than that of YOLOv5-l, and 4.38% higher than that of Faster-RCNN (ResNet). In terms of detection speed, the FPS value of the ACFP-YOLO algorithm is 65.56. Compared with models of the same scale, the detection speed remains at an upper-middle level and maintains good real-time performance. In terms of overall performance, the ACFP-YOLO algorithm improves mAP while losing a small part of the detection speed. It has better ability to extract features in complex underwater scenes. Our model is more suitable for efficient and accurate, and real-time requirements. Underwater missions.

TABLE 5 Performance Index Values of Mainstream Object Detection Algorithms on Underwater Garbage Detection Datasets.

Model	Input shape	mAP (%)	FPS	GFLOPS (G)	params (M)
SSD	461x416	60.21	143.85	61.951	25.216
YOLOv3	416x416	62.68	93.96	65.684	61.588
YOLOv4-tiny	416x416	56.03	217.79	6.853	5.902
YOLOv4	416x416	62.33	71.82	60.04	64.002
ACFP-YOLO	416x416	63.97	67.40	44.498	37.259
Efficientdet-D1	640x640	43.69	33.65	11.652	6.564
Faster-RCNN-VGG	600x600	69.89	43.81	370.014	136.935
Faster-RCNN-ResNet	600x600	70.54	26.89	941.071	28.398
YOLOv5-s	640x640	71.40	112.7	16.58	7.096
YOLOv5-l	640x640	72.32	67.30	114.765	46.696
YOLOv7-tiny	640x640	64.47	140.53	13.284	6.046
YOLOv7	640x640	73.85	66.83	105.32	37.259
ACFP-YOLO	640x640	74.92	65.56	105.328	37.604

In summary, the ACFP-YOLO algorithm, which integrates the CBAM attention mechanism and the SPPFCSPC module, achieves the highest detection accuracy compared with other mainstream algorithms in engineering applications, while maintaining a moderate level of detection and reasoning speed. Accuracy has a significant advantage in real-time underwater tasks.

Conclusion

In this paper, the ACFP-YOLO target detection model is proposed to address the problems of blurring and color deviation of images under the underwater map that make the extraction of object feature information difficult due to poor image quality. The model introduces the CBAM attention mechanism to enhance the extracted features in channel and spatial dimensions, which reduces the information loss in the feature extraction process and improves the overall feature extraction capability of the network, making the YOLOv7 model incorporating the attention mechanism have higher detection accuracy in underwater target scenes. The replacement of the SPPFCSPC module links the original three independent pooling layers together, reducing the model computation and obtaining faster model inference while keeping the perceptual field unchanged. By fusing the above two parts on YOLOv7, the improved ACFP-YOLO model has better performance in underwater target detection, and to a certain extent, solves the difficulties caused by the overlapping targets and complex background of underwater scenes to underwater target detection.

Our Future Work: Artificial intelligence has developed very rapidly, with numerous achievements in language translation, anomaly detection, target detection, and semantic segmentation, but few applications in intelligent exploitation of marine resources and underwater operations. In our future work, we will continue to study network models for underwater target recognition to improve the accuracy and speed of target recognition, and expand and enrich the dataset so that the models can be applied to more underwater scenarios with different conditions, and promote the application of AI in special underwater scenarios.

Data availability statement

Publicly available datasets were analyzed in this study, provided by 2022 China Underwater robot professional contest. This data can be found here: http://www.urpc.org.cn/index.html. Some experiments used the underwater garbage dataset, which can be found here: https://www.godac.jamstec.go.jp/dsdebris/e/index.html.

Author contributions

JY completed the main work of this paper. ZZ, ZX and BS guided JY in the research of this work. All authors contributed to the article and approved the submitted version.

Funding

This research was funded the Key Platform and Scientific Research Project for General Universities, Guangdong, 2022 (grant number 2022ZDZX4061), and the Science and Technology Program of Social Development, Zhuhai, 2022 (grant number 2220004000195). Our work was supported in part by the Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, BNU-HKBU United International College, project code 2022B1212010006 and in part by Guangdong Higher Education Upgrading Plan (2021-2025) with UIC research grant R0400001-22 and R201902.

Acknowledgments

Throughout the process of conducting the experiment and writing the paper. I received great support and help, and received many valuable comments. I would like to express my sincere gratitude to all those who have helped me throughout the process. First of all, I would like to thank my supervisor, Zhou Zhuang, who gave me valuable advice on academic matters. This thesis would not have been possible without his patient guidance, insightful criticism and professional instruction. Secondly, I would like to thank Prof. Bing-Hua Su, who provided a good platform for me to complete my thesis writing. I would like to thank the professors for providing me with valuable advice. Last but not least, I would like to thank my parents, classmates and friends for their constant support and encouragement.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv*, 10934. doi: 10.48550/arXiv.2004.10934

Chen, L., Zhou, F., Wang, S., Dong, J., Li, N., Ma, H., et al. (2020). Swipenet: Object detection in noisy underwater images. *arXiv*, 10006. doi: 10.48550/arXiv.2010.10006

Fulton, M., Hong, J., Islam, M. J., and Sattar, J. (2019). "Robotic detection of marine litter using deep visual detection models," in *2019 international conference on robotics and automation (ICRA)* (New York: IEEE), 5752–5758.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2015). Region-based convolutional networks for accurate object detection and segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 38, 142–158. doi: 10.1109/TPAMI.2015.2437384

Hou, Q., Zhou, D., and Feng, J. (2021). "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (New York: IEEE), 13713–13722.

Hu, J., Shen, L., and Sun, G. (2018). *Squeeze-and-excitation networks. computer vision and pattern recognition* (New York: IEEE).

Jiang, Z., and Wang, R. (2020). "Underwater object detection based on improved single shot multibox detector," in *2020 3rd International Conference on Algorithms, Computing and Artificial Intelligence*. (NY: ACM), 1–7.

Jocher, G. (2020) *Yolov5*. Available at: https://github.com/ultralytics/yolov5.

Klausner, N. H., and Azimi-Sadjadi, M. R. (2019). Performance prediction and estimation for underwater target detection using multichannel sonar. *IEEE J. Ocean. Eng.* 45, 534–546. doi: 10.1109/JOE.2018.2881527

Lab, P. (2018) *Underwater robot professional contest*. Available at: http://www.urpc.org.cn/index.html.

Lei, F., Tang, F., and Li, S. (2022). Underwater target detection algorithm based on improved yolov5. *J. Mar. Sci. Eng.* 10, 310. doi: 10.3390/jmse10030310

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. NY: IEEE, 2117–2125.

Lin, S., and Zhao, Y. (2020). Review on key technologies of target exploration in underwater optical images. *Laser Optoelectron. Prog.* 57, 060002. doi: 10.3788/LOP57.060002

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). "Ssd: Single shot multibox detector," in *European Conference on computer vision* (Berlin: Springer), 21–37.

Liu, T.-Y., and Tie, YL. (2009). Learning to rank for information retrieval. *Found. Trends® Inf. Retr.* 3, 225–331. doi: 10.1561/1500000016

Liu, S., Qi, L., Qin, H., Shi, J., and Jia, J. (2018). "Path aggregation network for instance segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (NY: IEEE), 8759–8768.

Qiang, W., He, Y., Guo, Y., Li, B., and He, L. (2020). Exploring underwater target detection algorithm based on improved ssd. *Xibei Gongye Daxue Xuebao/J. Northwest. Polytech. Univ.* 38, 747–754. doi: 10.1051/jnwpu/20203840747

Qiu, S., Xu, M., Jin, W., Yang, J., and Guo, H. (2019). Radon transform detection method for underwater moving target based on water surface characteristic wave. *Acta Optica Sin.* 39, 25–37. doi: 10.3788/AOS201939.1001003

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (NY: IEEE), 779–788.

Redmon, J., and Farhadi, A. (2017). "Yolo9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (NY: IEEE), 7263–7271.

Redmon, J., and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv*, 02767. doi: 10.48550/arXiv.1804.02767

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* 28, 91–99. doi: 10.48550/arXiv.1506.01497

Shi, P., Xu, X., Ni, J., Xin, Y., Huang, W., and Han, S. (2021). Underwater biological detection algorithm based on improved faster-rcnn. *Water* 13, 2420. doi: 10.3390/w13172420

Tan, M., Pang, R., and Le, Q. V. (2020). "Efficientdet: Scalable and efficient object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (NY: IEEE), 10781–10790.

Tinghui, H., Xinyu, G., Chunde, H., and Yueping, H. (2022). Research on underwater target detection algorithm based on fattention-yolov5. *Microelectron. Comput.* 39, 60–68. doi: 10.19304/J.ISSN1000-7180.2021.1261

Villon, S., Chaumont, M., Subsol, G., Villéger, S., Claverie, T., and Mouillot, D. (2016). "Coral reef fish detection and recognition in underwater videos by supervised machine learning: Comparison between deep learning and hog+ svm methods," in *Advanced Concepts for Intelligent Vision Systems- 17th International Conference, (ACIVS) 2016*. (Italy: Lecce), 160–171.

Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M. (2022). Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv*, 02696. doi: 10.48550/arXiv.2207.02696

Wang, Q., Wu, B., Zhu, P., Li, P., and Hu, Q. (2020). "Eca-net: Efficient channel attention for deep convolutional neural networks," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. NY: IEEE.

Wei, X., Yu, L., Tian, S., Feng, P., and Ning, X. (2021). Underwater target detection with an attention mechanism and improved scale. *Multim. Tools Appl.* 80, 33747–33761. doi: 10.1007/s11042-021-11230-2

Woo, S., Park, J., Lee, J.-Y., and Kweon, S. (2018). Cbam: convolutional block attention module. *Eur. conf Comput. Vision* 10, 978–973. doi: 10.1007/978-3-030-01234-2_1

YU, H. (2020). Research progresson object detection and tracking techniques utilization in aquaculture: a review. *J. Dalian Ocean Univ.* 35, 793–804. doi: 10.16535/j.cnki.dlhyxb.2020-263

Zeng, L., Sun, B., and Zhu, D. (2021). Underwater target detection based on faster r-cnn and adversarial occlusion network. *Eng. Appl. Artif. Intell.* 100, 104190. doi: 10.1016/j.engappai.2021.104190

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). "Learning deep features for discriminative localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (NY: IEEE), 2921–2929.

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 28 November 2022
DOI 10.3389/fmars.2022.1031549

Check for updates

OPEN ACCESS

EDITED BY
Ran Liao,
Tsinghua University, China

REVIEWED BY
Tatiana Novikova,
École Polytechnique, France
Lina Zhou,
Hong Kong Polytechnic University,
Hong Kong SAR, China

*CORRESPONDENCE
Xiaobo Li
lixiaobo@tju.edu.cn

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 30 August 2022
ACCEPTED 10 November 2022
PUBLISHED 28 November 2022

CITATION
Hu H, Huang Y, Li X, Jiang L, Che L, Liu T and Zhai J (2022) UCRNet: Underwater color image restoration *via* a polarization-guided convolutional neural network.
Front. Mar. Sci. 9:1031549.
doi: 10.3389/fmars.2022.1031549

COPYRIGHT
© 2022 Hu, Huang, Li, Jiang, Che, Liu and Zhai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

UCRNet: Underwater color image restoration *via* a polarization-guided convolutional neural network

Haofeng Hu[1,2], Yizhao Huang[1], Xiaobo Li[2]*, Liubing Jiang[3], Li Che[3], Tiegen Liu[1] and Jingsheng Zhai[2]

[1]Key Laboratory of Opto-Electronics Information Technology, School of Precision Instrument and Opto-electronics Engineering, Ministry of Education, Tianjin University, Tianjin, China, [2]School of Marine Science and Technology, Tianjin University, Tianjin, China, [3]School of Information and Communication, Guilin University of Electronic Technology, Guilin, China

Underwater images always suffer from low contrast and color distortion due to the wavelength-dependent scattering and absorption effects caused by particles existing in turbid water, especially in high turbidity conditions. Based on the polarization properties of the backscattering light, polarimetric methods can estimate the intensity level of the backscattering and the transmittance of the media. Accordingly, they can separate the target signal from the undesired ones to achieve high-quality imaging. In addition, learning-based polarimetric methods are effective for gray-model image restoration, but the learning-based polarimetric technique for color image restoration has yet to be considered. In this paper, we propose a 3- dimensional convolutional neural network, which maintains the correlation of polarization information among different polarization channel images as well as embodies polarization constraints, for underwater color image restoration. The experimental results verify that the proposed solution improves the image quality (i.e., the image contrast, details, and color) and outperforms other existing methods, especially when the turbidity of scattering media is high. The proposed solution can be readily applied to practical applications and potentially realize the clear vision in other scattering media, including biomedical imaging and remote sensing.

KEYWORDS

Polarization, polarimetric imaging, scattering media, imaging recovery, physical imaging

Introduction

High-quality imaging under turbid water or sea is significant in marine biology, archaeology, and military exploration (Li et al., 2018; Li et al., 2020; Hu et al., 2020; Li et al., 2022), as such applications significantly depend on the imaging systems' ability to obtain targeted object signals. However, the targeted object signal is scattered and

absorbed by the existing particles, resulting in severe degradation of the signal received by the detector. The backscattered light also veils the image and reduces the image contrast. Besides, for underwater color imaging, the levels of absorption and scattering by the existing particles for three channels (red, green, and blue, RGB) are different, which makes the image color distorted (Li et al., 2018; Li et al., 2019).

Various underwater image restoration methods have been developed to enhance the quality, including contrast and color. Examples of such methods include the histogram stretching (HS) method (Seeram, 2019) based on the image enhancement, color-line (Fattal, 2014) and dark channel prior (DCP) (He et al., 2010) based on physical models. We first recall the basic physical model of imaging through scattering media. The signal received by the detector can be divided into two parts (Liang et al., 2021). One part is the unpolarized light $D_c(x,y)$ obtained from the reflected light of the targeted object in the water after scattering by the existing particles, which can be written as the product of the reflected signal $L_c(x,y)$ and the transmission map $t_c(x,y)$, i.e., as shown in Eq. (1).

$$D_c(x,y) = L_c(x,y) \cdot t_c(x,y) \;\; c \in \{R,G,B\} \tag{1}$$

The transmission map tends to include abundant detailed information about the targeted object. The second part is the partially polarized backscattered light $B_c(x,y)$ caused by scattering particles, which can be expressed as:

$$B_c(x,y) = A_{c\infty} \cdot (1 - t_c(x,y)) \; c \in \{R,G,B\} \tag{2}$$

where $A_{c\infty}$ named air light or backscattered light denotes the ambient light scattered into the detector extending to infinity. The sum signal $I_c(x,y)$ received by the detector can be given by:

$$I_c(x,y) = D_c(x,y) + B_c(x,y) \; c \in \{R,G,B\} \tag{3}$$

The reflected signal of the targeted object can be obtained by stripping the backscattered light, thus achieving the goal of de-scattering in visual. Combining Eqs. (1) - (3), the reflected signal of the targeted object could be restored by Eq. (4):

$$L_c(x,y) = \frac{I_c(x,y) - A_{c\infty}}{t_c(x,y)} + A_{c\infty} \; c \in \{R,G,B\} \tag{4}$$

Based on the expression in Eq. (4), various methods have been proposed to recover underwater images by applying high-performance detectors (Skinner and Johnson-Roberson, 2017) or developing new processing algorithms. In 2003, Schechner (Schechner et al., 2003) et al. first introduced the polarization information, i.e., the degree of polarization (DoP), to solve the underwater image enhancement problem and achieved better results because the backscattered light is partially polarized, and the targeted object signal is unpolarized (Schechner et al., 2003; Schechner and Karpel, 2004; Liang et al., 2015; Hu et al., 2018; Wei et al., 2021; Qi et al., 2022). It should be noted that Schechner's model is significantly different from such polarimetric imaging methods based on the orthogonal state contrast (OSC) (Shao et al., 2006; Novikova et al., 2009; Novikova et al., 2010) or polarization difference (PD) (Nothdurft and Yao, 2006) images. This is because the OSC or PD-based methods only consider the polarization difference of targets while the practical physical degradation model, i.e., the expression in Eq. (3) is ignored. Liang (Liang et al., 2015) et al. developed Schechner's basic model and proposed a new polarization angle (AoP)-based polarimetric method to handle the image degeneration in scattering media, including foggy, hazy, and turbid environments. In 2021, Qi (Hu et al., 2021) designed a particular polarimetric method based on the typical polarization difference model and addressed the underwater color imaging issue well. This method solves the problem of HS for processing colors, ensuring the accuracy of mutually orthogonal polarization information recovery. Liu (Liu et al., 2019) et al. introduced the absorption coefficients of the water body particles and established the Lambertian body model algorithm based on a fundamental physical model. This method overcomes the problem that the recovered image has color distortion when using a fundamental physical model without considering the absorption coefficient of water body particles.

Recently, the deep learning method has developed rapidly and has been considered a successful way to outperform the traditional intensity-based ones and boost performance in polarimetric imaging techniques, including denoising, demosaicing, and de-scattering tasks (Sun et al., 2021; Liu et al., 2022; Ding et al., 2022). In contrast to the physical model, deep learning methods don't require complex physical models and prior knowledge due to powerful fitting capability. Hu (Hu et al., 2022) et al. proposed a well-designed neural network and proved that the combination of polarization model and neural network was beneficial to improving the image quality even in a high turbidity environment. It can effectively remove the scattering light and obviously be more robust than other traditional methods; but this method cannot deal with color images. A medium scale of color polarization image datasets from natural conditions through passive polarization imaging was built by Ding (Ding et al., 2022) et al. They constructed generative feature-fusion adversarial networks to extract different polarization angles features and obtain better results on both laboratory simulated and real natural datasets. Since passive polarization imaging relies on ambient light for illumination, the method is not suitable for the task of strong scattering environment de-scattering. Therefore, an effective way to enhance the image contrast and reduce the color distortion of underwater images is important for the current demands.

In this paper, we propose a 3- dimensional convolutional kernel-based polarization-guided network, which maintains the correlation of polarization information between images of different polarization angles and reflects polarization constraints for underwater color image recovery. The

superiority of the proposed method is demonstrated by comparing it with representative methods for digital image processing and deep learning. Furthermore, the proposed method can be applied to the de-scattering task on real polarization city foggy environments, which verifies its effectiveness and advantage. The remaining parts of the article are arranged as follows: In Section 2, we first introduce the applied method and methodology, including the designed network structure and training details; in Section 3, we present the experimental setup and perform the imaging experiments; Finally, we conclude this paper and draw future works in Section 4.

Methods and methodology

Network structure

In this section, we first introduce the structure of the proposed UCRNet. The network is based on the customized 3D convolution kernel and a two-step feature extraction strategy. As its significant advantage in addressing multiple channels and improving recovery accuracy, 3D convolution kernels have been successfully applied to various vision tasks (Varol et al., 2017; Carreira and Zisserman, 2017). For example, one of the most representative works is the two-stream inflated 3D Convolutional network (i.e., I3D) (Carreira and Zisserman, 2017). The 3D convolution kernel is inflated from the traditional 2D kernel and its dimension is transformed from $N\times N$ to $N\times N\times n$, where n endows the network with an additional temporal dimension, making the neural network better understand the correlation between video frames; therefore, it improves the accuracy of video action understanding. Inspired by this work, we applied the 3D convolution kernel for the task of polarimetric imaging through scattered media, i.e., under turbid water. Specially, the 3D convolution kernels were used to extract the relationship between channel pixels and different polarization angles (i.e., 0, 45 and 90 in this work). Besides, we design a two-step strategy to balance the network's depth and training time; the first part of the network structure uses 2D convolution kernels to extract and address former features (e.g., the color, edge, and shape) and the second part uses 3D convolution kernels for the deeper features (e.g., the global style and abstract features). This strategy ensures that the network structure is not particularly deep and saves training time. Figure 1 proposes the network structure of UCRNet.

The left and right side of Figure 1A shows the 3D RGB pixels spatial distributions of the input, output and label image, respectively. This figure represents the gray value of each pixel in a three-dimensional coordinate system with red, green, and blue as the coordinate axis. Each point's color denotes its true color in the related images. Besides, we also draw the projection on the R (red)-B (blue) and B (blue)-G (green) plane. In fact, 1) the more color the clearer the images, as the input low-quality images always perform a single color, e.g., the gray. 2) the more scattered the pixels clearer the images as the inputs are compressed into a narrow pixel space. Besides, we may observe that the restored image's point distribution of the 3D

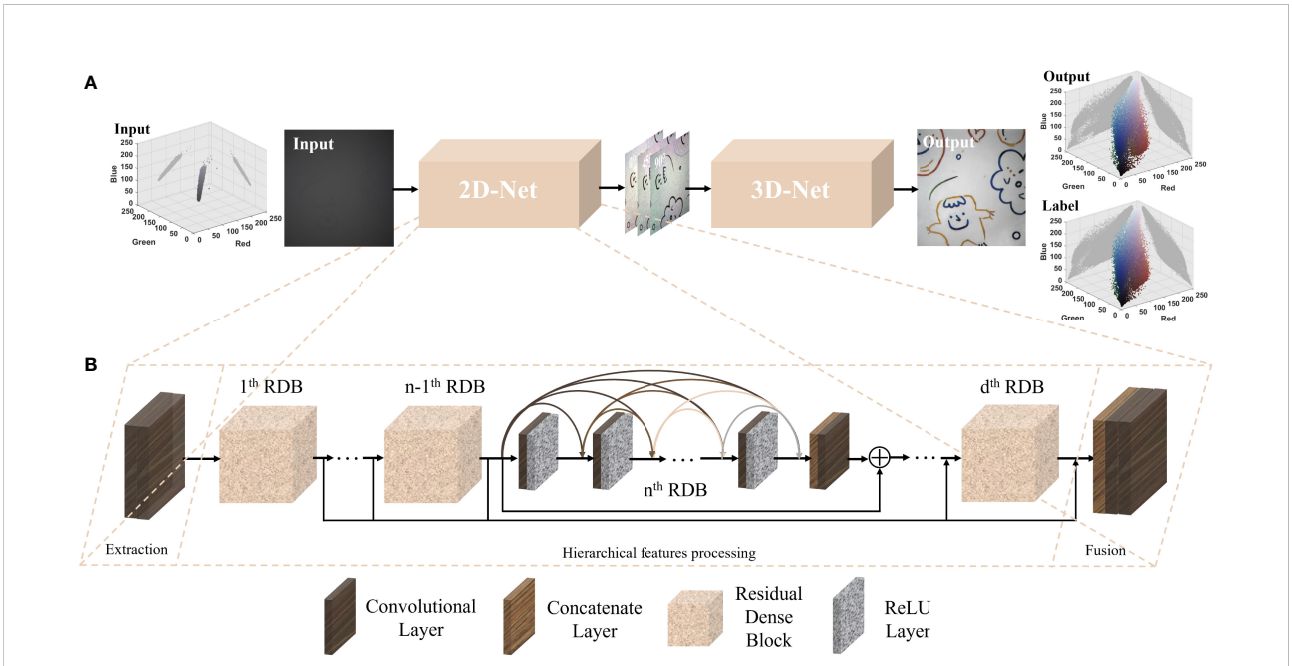


FIGURE 1
Network structure of UCRNet. **(A)** Overall model frame and 3D-RGB color models for input and output **(B)** Details of network structure.

RGB pixels is more scattered and is close to that of the label image. From Figure 1B (i.e., in the proposed network), we first use two convolutional layers to extract shallow features from the input image (i.e., the raw images captured by the DoFP polarization camera), which is then sent into the followed eight residual dense blocks (Zhang et al., 2018) (RDBs) to extract hierarchical features. After a global features fusion operation including one concatenate layer and two convolutional layers. All the above steps use 2D convolution kernels and are combined as the first step (or the first sub-network) of the UCRNet, named 2D-Net. It can be seen in Figure 1A that the output features of the 2D-Net are images with different polarization angles. Subsequently, the output features are thrown into the following sub-network (i.e., 3D-Net), in which all the kernels are the 3D type. The output of the whole network is intensity image S_0, and all the other structures and parameter designs are similar to the former 2D-Net. The detailed settings of the 3D-Net are listed in Table 1, where P and P_0 are channels of features that should be determined before training. P is also called the growth rate of the residual dense blocks, and $P \times k$ is the input channel of the k^{th} Conv-ReLU layer. Besides, the number of residual dense blocks in the 3D-Net is 4, and C is the total number of the Conv-ReLU layer in a single residual dense block. represents the hierarchical features from the preceding residual dense blocks, and F denotes the fusion operation. Eq. (5) can express the output image.

$$I_{out} = F(f_n, f_{n-1}, \cdots f_1, f_0) \tag{5}$$

Loss function

To guide the training, the designed loss functions include two types, i.e., the polarization loss and perceptual loss. Eq. (6) gives the polarization loss.

$$L_{out} = \frac{1}{N}\sum_i \| y_i^c(\theta) - \hat{y}_i^c(\theta) \|_1 \tag{6}$$

where L_{out} denotes the polarization loss for the final output and N means the scale of the training dataset. $y_i^c(\theta)$ indicates the i^{th} reconstructed polarization image corresponded to the polarization angle (i.e., θ) of 0, 45, or 90 degrees $\hat{y}_i^c(\theta)$ is as the i^{th} corresponding label image. The perceptual loss (Johnson et al., 2016) could be expressed as:

$$L_{per} = \frac{1}{CWH} \| \phi_j(I_{S_0}^c) - \phi_j(\hat{I}_{S_0}^c) \|_2^2 \tag{7}$$

where C, W and H are the images' channel, width, and height, respectively. ϕ_j is the j^{th} layer of the pre-trained VGG-16 network on the ImageNet dataset (He et al., 2015). $I_{S_0}^c$ and $\hat{I}_{S_0}^c$ respectively represent the output S_0 and label S_0 images, calculated as $y^c(0^\circ)+y^c(90^\circ)$. We compute the perceptual loss at layer relu1_2 of the VGG-16 pretrained network. By comparing the features of reconstructed and labeled images, we could take full advantage of the image feature extraction accuracy from networks pre-trained on enormous-scale datasets. Finally, L_{sum} in the following Eq. (8) denotes the summation of loss functions; λ_1 andλ_2 are the weights of polarization loss in Eq. (6) and perceptual loss in Eq. (7).

$$L_{sum} = \lambda_1 L_{out} + \lambda_2 L_{per} \tag{8}$$

where $\lambda_1 = L_{out}/L_{out}+L_{per}$ and $\lambda_2 = L_{per}/L_{out}+L_{per}$.

Training model details

The dataset includes 130 paired images collected and pre-processed under scattering media of full cream milk; 100 pairs were used as the training dataset, and the rest were divided into the validation dataset (15 pairs) and the test dataset (15 pairs). To expand the scale of the dataset and improve the training performance, the dataset is cropped to 64×64 pixels in the stride of 32 pixels in horizontal and vertical directions. As such, the scale of the training dataset reaches more than one hundred thousand images. Before the training, all the input images were normalized to [0,1], and each layer's weights were initialized using the initialization method in ref (He et al., 2015). We use Pytorch as the deep learning framework on the Nvidia RTX2080Ti GPU, and the initial learning rate is e^{-4} and decays by a factor of 0.6 every ten epochs. We trained the

TABLE 1 Parameters of the 3D-Net.

Part	Layer name	Input channel	Output channel	Kernel size
Extraction	Conv.	3	P_0	(3×3×3)
	Conv.	P_0	P	(3×3×3)
Hierarchical features processing	Conv.	$P \times k$	P	(3×3×3)
	ReLU.	/	/	/
	Conv.	$P \times (C+1)$	P	(1×1×1)
Fusion	Conv.	$4P$	P_0	(1×1×1)
	Conv.	P_0	P_0	(3×3×3)
	Conv.	P_0	3	(3×3×3)

network for 60 epochs using the Adam optimizer with a mini-batch size of 32.

Experiment

Dataset acquisition

Figure 2 presents the experimental setup for polarimetric imaging under turbid water, and all the image pairs in the dataset are captured by this setup. In this setup, the polarization state generator (PSG) is to produce linearly polarized light *via* a white broad-spectrum LED light source. The targeted object is placed in a water tank made of polymethyl methacrylate (PMMA) in size of 65 × 26 × 26 cm^3. The imaging device is a chromatic division of focal plane (DoFP) polarization camera (Lucid-PHX050S) with a pixel size of 2048 × 2448 × 3. As shown in Figure 2, each macro super-pixel of the chromatic DoFP polarization camera consists of four micro super-pixels, i.e., two for green (G), one for red (R), and one for blue channels (B); each micro color super-pixel further consists of four pixels, each of which has a micro wire-grid polarizer orientated at 0°, 45°, 90°, and 135°, and the related intensity images are denoted as $I_0^c(x,y)$, $I_{45}^c(x,y)$, $I_{90}^c(x,y)$ and $I_{135}^c(x,y)$, respectively, where c denotes the three color channels, i.e., R, G, and B. In practice, the circularly polarized light could perform better for the imaging through scattering media (Lewis et al., 1999; Xu and Alfano, 2005); yet the related optical system is more complex, and the data collection may become time-consuming. Based on that, we choose the chromatic DoFP polarization camera for the data acquisition.

For the imaging experiments, we first added clear water to the tank and recorded the related image as the label for training. Subsequently, as the scattering properties of milk were confirmed to be relatively close to those of seawater by the long-term experiment (Dubreuil et al., 2013), we added the full cream milk into the clear water to generate the scattering media. Then, we record the images with degraded polarization information in the scattering environment after the milk is uniformly diffused. The scattering coefficient of the medium is influenced by the concentration of fat microspheres and casein molecules in the milk solution, which is equal to 3.00 mc (cm^{-1}) for full cream milk, where mc represents the milk concentration in water (Piederrière et al., 2005). Notably, as the value of mc increases, the scattering coefficient becomes higher, and in this case, the polarization information of the object is further degraded. To ensure the reasonability of the experiments and the validity following verifications, we build the dataset in varying milk concentrations.

Visual comparisons

In this section, we perform imaging experiments to verify the effectiveness of the proposed network. Figure 3 first presents the raw images of two example scenes in dense turbid water. From Figure 3A, we may observe that most details are annihilated by the backscattered light, from which the color information cannot be distinguished. Besides, Figure 3B presents the related ground truths (GTs). The restored images by the proposed method are proposed in Figure 3C. From the results, we can see that all the details are well-addressed and the color information is significantly similar to GTs.

To further verify the proposed method, we also present the restored images by some representative works, i.e., a fusion algorithm (Ancuti et al., 2012) proposed by Cosmin Ancuti et al., MLLE (Zhang et al., 2022), and RDN-Net (Hu et al., 2020). The Fusion method fuses several images using different enhanced method through Laplace's pyramid to complement each other's strengths. The MLLE method uses the grayscale world assumption as a criterion to achieve the effect of color compensation through several iterations. It is worth pointing out

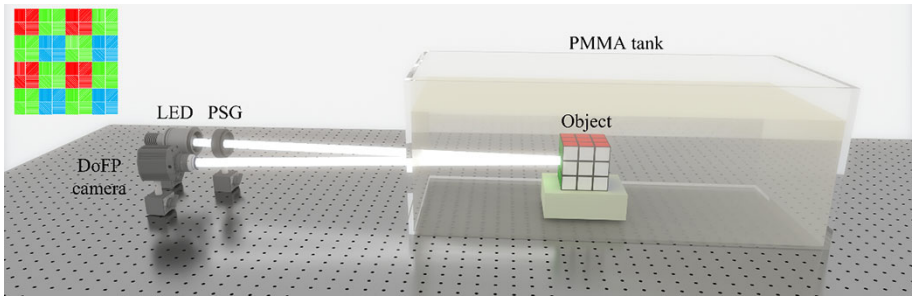


FIGURE 2
Simulation diagram of the experimental setup for dataset acquisition. The scattering medium shown in the simulation diagram is full cream milk. The RGB-polarization pattern array of chromatic DoFP polarization camera by Lucid vision is at the top-left corner.

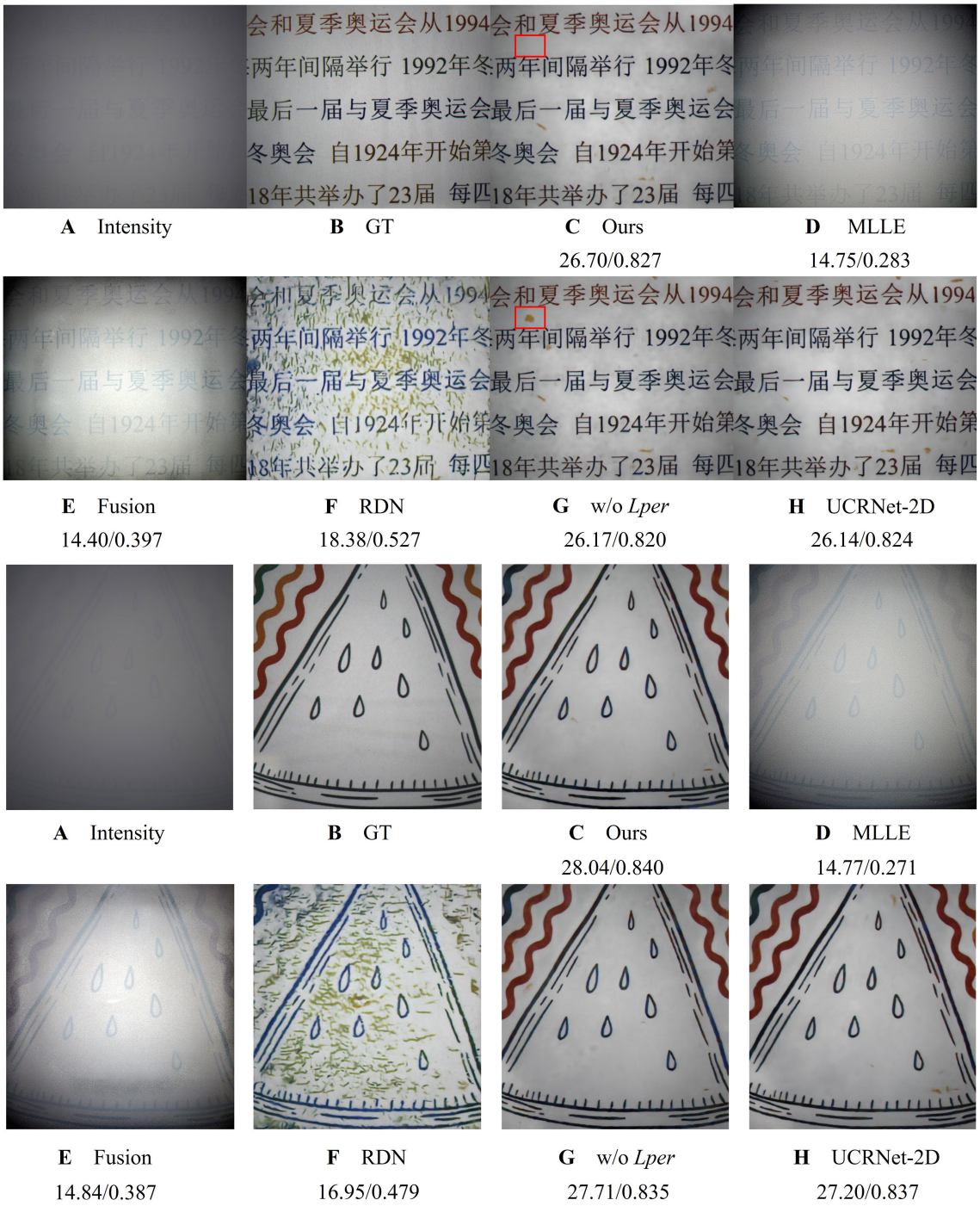


FIGURE 3
Results of the visual comparison. Two evaluation metrics below the image represent PSNR (dB) and SSIM respectively. **(A)** Intensity; **(B)** GT; Results by **(C)** Proposed method, **(D)** MLLE **(E)** Fusion, **(F)** RDN, **(G)** w/o L_{per}, and **(H)** UCRNet-2D. (1) and (2) are represented two samples in the test set.

that the Fusion method is influenced by the active light source, and the halo affects the recovery image qualification, as shown in Figure 3E. The common problem of these digital image processing-based recovery methods is that the backscattered light scattered by the scattering medium is not well suppressed. As a result, the images seem still "hazy". As shown in Figure 3F, this method greatly improves the image contrast but suffers from severe color distortion. In addition, we use two representative metrics, i.e., the peak of signal to noise ratio (PSNR) and structural similarity (SSIM), to quantificationally evaluate the final performance. The comparison values (PSNR/SSIM) are listed under the corresponding images in Figure 3. The value of the method based on digital image processing is significantly lower than that of the proposed method.

To verify that adding perceptual loss can improve the quality of the restored image, we remove the perceptual loss and the rest of the training details remain unchanged. Without (w/o) perceptual loss in Figure 3G, the result obtained has some yellow artifacts (framed by the red rectangle) in different positions of the first image. In addition, we can also observe the effect of over-enhancement (a broad edge along the pattern lines) on the boundaries of the second image. For the UCRNet-2D method, we change the 3D-Net of the proposed method to a 2D-Net. The parameters of a 3-D convolution kernel are approximately three times that of a 2-D convolution kernel. To ensure the consistency of the overall network parameters, twelve dense residual blocks are used in the UCRNet-2D method. The recovered images are close to the proposed method, but with a certain degree of artifacts. The superiority of the proposed method of recovery considering polarization guidance is demonstrated by comparison with UCRNet-2D method. Besides, Figure 4 shows an additional example to compare the performance in avoiding artifacts for the UCRNet-2D and -3D methods. From this figure, we may find that, compared with the UCRNet-3D's result, there are significant artifacts in that of UCRNet-2D, as shown in the red rectangles.

Furthermore, we also calculate the PSNR, SSIM, ZNCC (Zero-normalized cross-correlation) and UICM (underwater image colorfulness measure) average values on the test set, as shown in the Table 2. In terms of the average PSNR value, the proposed method shows a better performance than other



FIGURE 4
Results of the visual comparison. **(A)** Intensity image **(B)** Proposed method and **(C)** Result obtained by UCRNet-2D.

TABLE 2 Quantitative comparisons of different methods.

Methods	PSNR↑	SSIM↑	ZNCC↑	UICM↑
MLLE (Zhang et al., 2022)	13.679	0.274	0.402	**18.969**
Fusion (Ancuti et al., 2012)	12.890	0.361	0.429	17.062
RDN-Net (Hu et al., 2020)	15.463	0.423	0.622	9.673
w/o L_{per}	24.564	0.758	0.890	16.273
UCRNet-2D	24.783	0.762	0.893	16.475
Ours	**24.901**	**0.764**	**0.894**	16.654

Values in bold for each evaluation metric represent numerically best-performing data.

representative methods in Figure 3. This conclusion agrees with the results of the visual comparisons in Figures 3D–F, proving the superiority of the proposed method. Compared with w/o L_{per} and ours, the value is improved by 0.35dB, which further demonstrates the necessity of well-designed loss in the proposed method. As for the UCRNet-2D, the average value of the proposed method is increased by 0.12dB. Consequently, we can conclude that the polarization-guided network is significant in improving the network performance. ZNCC is for evaluating the cross-correlation between two images and the higher the value, the more correlated the two images are. Our proposed method performs best on the ZNCC evaluation metric. UICM is an evaluation metric that measures the deviation of the different color channels. Since the backscattered light of the images processed by MLLE and Fusion methods is not completely removed, the UICM evaluation metric is falsely high. The value of UICM also confirm our method features the better restored color information.

We also perform the proposed method on different turbidity and the results are shown in Figure 5. The corresponding PSNR/SSIM values are also presented under the restored results. From the Figure 5, we may observe that the proposed method can well restore the hazy images in both low and high turbidity levels, verifying the universality of our method.

Moreover, we extend the proposed method to other scattering media-real foggy environments. The dataset we used is Polarized Foggy Cityscapes-DBF dataset-Zurich in ref (Zhou et al., 2021). This dataset provides depth and foggy maps corresponding to different scattering coefficients, semantic segmentation maps, and intensity images of urban streets. Using the depth map and the corresponding scattering coefficient, the scattering length distribution map that can simulate the effect of haze gradually deepening from near to far can be obtained. The polarization characteristics are related to the object's material, taking advantage of semantic segmentation images to assign equal degrees of polarization value to objects with the same semantic information is more suitable for the natural environment. Paired images of different polarization directions are obtained using Malus law. Figure 6 presents the related extension results, which includes the comparison of intensity, GT, our restoration images as well as the calculated PSNR/SSIM values. From Figure 6, we may conclude that the trained model proposed in Section 2.1 is fine-tuned in the Zurich. The image details are well observed in the restored image, such as the bus stop (in the orange rectangle) nearby the trees in Figure 6A and the distant views (in the red rectangle) in Figure 6B.

Furthermore, Table 3. compares the imaging performance between our method and the selected the typical polarimetric methods, i.e., polarization difference imaging models proposed by Schechner (Skinner and Johnson-Roberson, 2017), and three non-polarimetric methods, including the (1) dark channel prior-based method (DCP) (He et al., 2010), (2) fusion method (Ancuti et al., 2012), and (3) an open dehazing platform developed by Baidu (Li et al., 2017). From the Table 3, one may find that our method achieved the highest values in most metrics, including PSNR, SSIM, ZNCC and NIQE. These compared methods are not capable of dealing with unevenly distributed haze environment images. Compared with the raw image and Baidu's results, our method has effectively improved the image's PSNR values by 10.9dB and 10.8dB, respectively. In conclusion, both subjective visual effects and objective evaluation metrics prove that this method could be extended to city foggy scattering environments. We must note that the recovery results contain more noise than GTs; this is because the

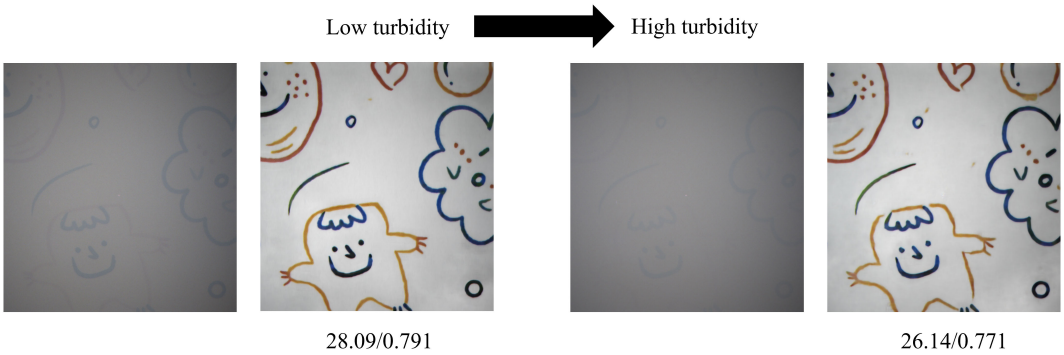


FIGURE 5
Results on low turbidity and high turbidity. Two evaluation metrics below the image represent PSNR (dB) and SSIM respectively.

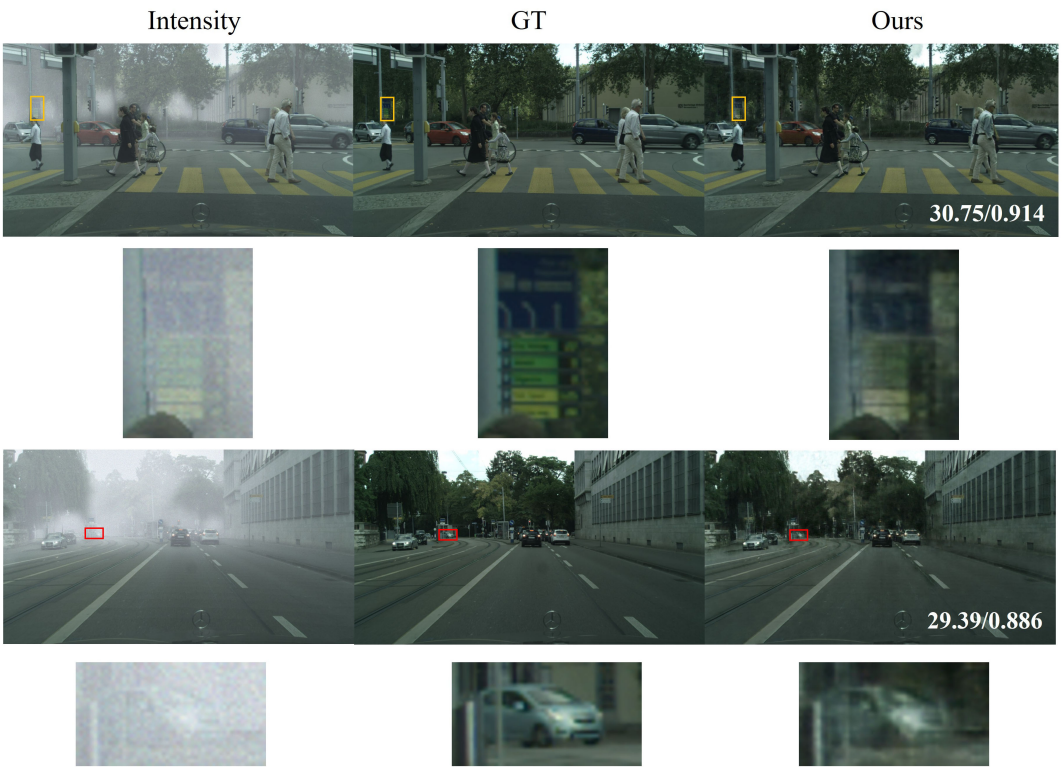


FIGURE 6
Visual de-scattering effects on Polarized Foggy Cityscapes-DBF dataset-Zurich. The numbers on the bottom-right corner of each image refer to its PSNR (dB) and SSIM. The bottom of each image represents the corresponding area in the rectangle frame. (1) and (2) are represented as two samples in the test set.

original intensity images are noisy. Besides, the scattering density of the foggy dataset is also lower than that in the turbid water we designed. The two problems can be further addressed by choosing a high-scattering but high-quality dataset, which is a promising work in future.

Conclusion

To our knowledge, this is the first work to introduce polarization information into a network to achieve the de-scattering effect on color imaging. We successfully constructed a polarization-guided network, which further improved the recovery effect of the network. We also show that the quality of restored images can be enhanced by introducing a well-designed loss function and the capability of the proposed method in different turbidity. Comprehensive experimental results confirm that the proposed method outperforms other representative methods, including learning-based methods and digital image processing-based methods. Moreover, we demonstrate that the technique can be extended to the field of image dehazing in natural environments. Our solution may

TABLE 3 Average evaluation metrics on the test set.

Methods	PSNR↑	SSIM↑	ZNCC↑	UICM↑	NIQE↓	BRISQUE↓
Intensity	16.56	0.715	0.605	0.002	4.964	18.836
DCP (He et al., 2010)	11.62	0.314	0.552	**0.004**	6.058	41.001
Baidu (Li et al., 2017)	16.67	0.661	0.743	0.003	5.087	**16.018**
Schechner (Skinner and Johnson-Roberson, 2017)	16.56	0.675	0.579	0.002	5.368	18.588
Fusion (Ancuti et al., 2012)	16.39	0.625	0.640	0.003	5.606	21.498
Ours	**27.49**	**0.899**	**0.967**	0.001	**2.846**	26.885

Values in bold for each evaluation metric represent numerically best-performing data.

find important applications in object detection and clear vision under strong scattering conditions (e.g., dense fog, deep-sea, and biological field) by adjusting the designs of loss functions and network structures according to the special applications.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

Conceptualization, XL, YH, and HH. Funding acquisition, HH. Methodology, XL, YH, and HH. Resources, XL and YH. Supervision, XL and HH. Validation, YH and XL. Visualization, XL and YH. Writing – original draft, HH, YH, and XL. Writing – review & editing, XL, YH, LJ, LC, TL, and JZ. All authors contributed to the article and approved the submitted version.

Funding

Guangxi Special Fund Project for Innovation-Driven Development (No. GuikeAA21077008) and National Natural Science Foundation of China (62075161, 62205243).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ancuti, C., Ancuti, C. O., Haber, T., and Bekaert, P. (2012). "Enhancing underwater images and videos by fusion," in *Proc. IEEE Conf. Comput. Vision Pattern Recognition*. (Washington, DC,: IEEE), 81–88.

Carreira, J., and Zisserman, A. (2017). "Quo vadis, action recognition? a new model and the kinetics dataset," in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 6299–6308.

Ding, X., Wang, Y., and Fu, X. (2022). Multi-polarization fusion generative adversarial networks for clear underwater imaging. *Optics Lasers Eng.* 152, 106971. doi: 10.1016/j.optlaseng.2022.106971

Dubreuil, M., Delrot, P., Leonard, I., Alfalou, A., Brosseau, C., and Dogariu, A. (2013). Exploring underwater target detection by imaging polarimetry and correlation techniques. *Appl. Optics* 52 (5), 997–1005. doi: 10.1364/AO.52.000997

Fattal, R. (2014). Dehazing using color-lines. *ACM Trans. Graphics (TOG)* 34 (1), 1–14. doi: 10.1145/2651362

He, K., Sun, J., and Tang, X. (2010). Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (12), 2341–2353. doi: 10.1109/TPAMI.2010.168

He, K., Zhang, X., Ren, S., and Sun, J. (2015). "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*. 1026–1034.

Hu, H., Han, Y., Li, X., Jiang, L., Che, L., Liu, T., et al. (2022). Physics-informed neural network for polarimetric underwater imaging. *Optics Express* 30 (13), 22512–22522. doi: 10.1364/OE.461074

Hu, H., Lin, Y., Li, X., Qi, P., and Liu, T. (2020). IPLNet: a neural network for intensity-polarization imaging in low light. *Optics Lett.* 45 (22), 6162–6165. doi: 10.1364/OL.409673

Hu, H., Qi, P., Li, X., Cheng, Z., and Liu, T. (2021). Underwater imaging enhancement based on a polarization filter and histogram attenuation prior. *J. Phys. D: Appl. Phys.* 54 (17), 175102. doi: 10.1088/1361-6463/abdc93

Hu, H., Zhang, Y., Li, X., Lin, Y., Cheng, Z., and Liu, T. (2020). "Polarimetric underwater image recovery via deep learning," in *Optics and lasers in engineering*, vol. 133. (Berkely, CA: Optics and Lasers in Engineering (Elsevier)), 106152.

Hu, H., Zhao, L., Li, X., Wang, H., and Liu, T. (2018). Underwater image recovery under the nonuniform optical field based on polarimetric imaging. *IEEE Photonics J.* 10 (1), 1–9. doi: 10.1109/JPHOT.2018.2791517

Johnson, J., Alahi, A., and Fei-Fei, L. (2016). "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on computer vision* New York, NY, USA: European conference on computer vision (Springer), pp.694–pp.711.

Lewis, G. D., Jordan, D. L., and Roberts, P. J. (1999). Backscattering target detection in a turbid medium by polarization discrimination. *Appl. Optics* 38 (18), 3937–3944. doi: 10.1364/AO.38.003937

Liang, J., Ren, L., Ju, H., Zhang, W., and Qu, E. (2015). Polarimetric dehazing method for dense haze removal based on distribution analysis of angle of polarization. *Optics Express* 23 (20), 26146–26157. doi: 10.1364/OE.23.026146

Liang, J., Ren, L., and Liang, R. (2021). Low-pass filtering based polarimetric dehazing method for dense haze removal. *Optics Express* 29 (18), 28178–28189. doi: 10.1364/OE.427629

Li, C., Guo, J., and Guo, C. (2018). Emerging from water: Underwater image color correction based on weakly supervised color transfer. *IEEE Signal Process. Lett.* 25 (3), 323–327. doi: 10.1109/LSP.2018.2792050

Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., et al. (2019). An underwater image enhancement benchmark dataset and beyond. *IEEE Trans. Image Process.* 29, 4376–4389. doi: 10.1109/TIP.2019.2955241

Li, X., Hu, H., Zhao, L., Wang, H., Yu, Y., Wu, L., et al. (2018). Polarimetric image recovery method combining histogram stretching for underwater imaging. *Sci. Rep.* 8 (1), 1–10. doi: 10.1038/s41598-018-30566-8

Li, X., Li, H., Lin, Y., Guo, J., Yang, J., Yue, H., et al. (2020). Learning-based denoising for polarimetric images. *Optics Express* 28 (11), 16309–16321. doi: 10.1364/OE.391017

Li, B., Peng, X., Wang, Z., Xu, J., and Feng, D. (2017). "Aod-net: All-in-one dehazing network," in *Proceedings of the IEEE international conference on computer vision*. 4770–4778.

Liu, F., Wei, Y., Han, P., Yang, K., Bai, L., and Shao, X. (2019). Polarization-based exploration for clear underwater vision in natural illumination. *Optics Express* 27 (3), 3629–3641. doi: 10.1364/OE.27.003629

Liu, H., Zhang, Y., Cheng, Z., Zhai, J., and Hu, H. (2022). Attention-based neural network for polarimetric image denoising. *Optics Lett.* 47 (11), 2726–2729. doi: 10.1364/OL.458514

Li, X., Xu, J., Zhang, L., Hu, H., and Chen, S. C. (2022). Underwater image restoration *via* stokes decomposition. *Optics Lett.* 47 (11), 2854–2857. doi: 10.1364/OL.457964

Nothdurft, R. E., and Yao, G. (2006). Effects of turbid media optical properties on object visibility in subsurface polarization imaging. *Appl. optics* 45 (22), 5532–5541. doi: 10.1364/AO.45.005532

Novikova, T., Bénière, A., Goudail, F., and De Martino, A. (2009). Sources of possible artefacts in the contrast evaluation for the backscattering polarimetric images of different targets in turbid medium. *Optics Express* 17 (26), 23851–23860. doi: 10.1364/OE.17.023851

Novikova, T., Bénière, A., Goudail, F., and De Martino, A. (2010). "Contrast evaluation of the polarimetric images of different targets in turbid medium: Possible sources of systematic errors," in *Polarization: Measurement, analysis, and remote sensing IX*, vol. 7672. (Bellingham, WA, USA: SPIE), 189–197.

Piederrière, Y., Boulvert, F., Cariou, J., Le Jeune, B., Guern, Y., and Le Brun, G. (2005). Backscattered speckle size as a function of polarization: Influence of particle-size and-concentration. *Optics Express* 13 (13), 5030–5039. doi: 10.1364/OPEX.13.005030

Qi, P., Li, X., Han, Y., Zhang, L., Xu, J., Cheng, Z., et al. (2022). U2R-pGAN: Unpaired underwater-image recovery with polarimetric generative adversarial network. *Optics Lasers Eng.* 157, 107112. doi: 10.1016/j.optlaseng.2022.107112

Schechner, Y. Y., and Karpel, N. (2004). "Clear underwater vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. I–I.

Schechner, Y. Y., Narasimhan, S. G., and Nayar, S. K. (2003). Polarization-based vision through haze. *Appl. Optics* 42 (3), 511–525. doi: 10.1364/AO.42.000511

Seeram, E. (2019). "Digital image processing concepts," in *Digital radiography* (Singapore: Springer), (pp.21–(pp39).

Shao, H., He, Y., Shao, Y., and Ma, H. (2006). "Contrast enhancement subsurface optical imaging with different incident polarization states," in *Fourth international conference on photonics and imaging in biology and medicine*, vol. 6047. (Bellingham, WA, USA: SPIE), 226–231.

Skinner, K. A., and Johnson-Roberson, M. (2017). "Underwater image dehazing with a light field camera," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 62–69.

Sun, Y., Zhang, J., and Liang, R. (2021). Color polarization demosaicking by a convolutional neural network. *Optics Lett.* 46 (17), 4338–4341. doi: 10.1364/OL.431919

Varol, G., Laptev, I., and Schmid, C. (2017). Long-term temporal convolutions for action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 40 (6), 1510–1517. doi: 10.1109/TPAMI.2017.2712608

Wei, Y., Han, P., Liu, F., and Shao, X. (2021). Enhancement of underwater vision by fully exploiting the polarization information from the stokes vector. *Optics Express* 29 (14), 22275–22287. doi: 10.1364/OE.433072

Xu, M., and Alfano, R. R. (2005). Circular polarization memory of light. *Phys. Rev. E* 72 (6), 065601. doi: 10.1103/PhysRevE.72.065601

Zhang, Y., Tian, Y., Kong, Y., Zhong, B., and Fu, Y. (2018). Residual dense network for image super-resolution. *In Proc. IEEE Conf. Comput. Vision Pattern Recognition* 133, 2472–2481. doi: 10.1109/CVPR.2018.00262

Zhang, W., Zhuang, P., Sun, H. H., Li, G., Kwong, S., and Li, C. (2022). Underwater image enhancement *via* minimal color loss and locally adaptive contrast enhancement. *IEEE Trans. Image Process.* 31, 3997–4010. doi: 10.1109/TIP.2022.3177129

Zhou, C., Teng, M., Han, Y., Xu, C., and Shi, B. (2021). "Learning to dehaze with polarization," in *Advances in neural information processing systems* (San Diego, CA: Advances in Neural Information Processing Systems (NeurLPS)), vol. 34, 11487–11500.

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 29 November 2022
DOI 10.3389/fmars.2022.1032287

Check for updates

OPEN ACCESS

EDITED BY
Hong Song,
Zhejiang University, China

REVIEWED BY
Amjad Ali Amjad,
Peking University, China
Peiyi Zhu,
Changshu Institute of Technology, China

*CORRESPONDENCE
Jialin Tang
01068@bitzh.edu.cn

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 30 August 2022
ACCEPTED 28 October 2022
PUBLISHED 29 November 2022

CITATION
Chen J, Tang J, Lin S, Liang W, Su B, Yan J, Zhou D, Wang L, Lai Y and Yang B (2022) RMP-Net: A structural reparameterization and subpixel super-resolution-based marine scene segmentation network. *Front. Mar. Sci.* 9:1032287. doi: 10.3389/fmars.2022.1032287

COPYRIGHT
© 2022 Chen, Tang, Lin, Liang, Su, Yan, Zhou, Wang, Lai and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

RMP-Net: A structural reparameterization and subpixel super-resolution-based marine scene segmentation network

Jiongjiang Chen[1], Jialin Tang[1*], Shounan Lin[1], Wanxin Liang[1], Binghua Su[1], Jinghui Yan[1], Dujuan Zhou[1,2], Lili Wang[1], Yunting Lai[1] and Benxi Yang[1]

[1]Key Laboratory of Intelligent Detection in Complex Environment of Aerospace Land and Sea, Beijing Institute of Technology, Zhuhai, China, [2]Faculty of Innovation Engineering, Macau University of Science and Technology, Macau, China

Ocean exploration has always been an important strategic direction for the joint efforts of all mankind. Many countries in the world today are developing their own underwater autonomous explorers to better explore the seabed. Vision, as the core technology of autonomous underwater explorers, has a great impact on the efficiency of exploration. Different from traditional tasks, the lack of ambient light on the seabed makes the visual system more demanding. In addition, the complex terrain on the seabed and various creatures with different shapes and colors also make exploration tasks more difficult. In order to effectively solve the above problems, we combined the traditional models to modify the structure and proposed an algorithm for the super-resolution fusion of enhanced extraction features to perform semantic segmentation of seabed scenes. By using a structurally reparameterized backbone network to better extract target features in complex environments, and using subpixel super-resolution to combine multiscale feature semantic information, we can achieve superior ocean scene segmentation performance. In this study, multiclass segmentation and two-class segmentation tests were performed on the public datasets SUIM and DeepFish, respectively. The test results show that the mIoU and mPA indicators of our proposed method on SUIM reach 84.52% and 92.33%mPA, respectively. The mIoU and mPA on DeepFish reach 95.26% and 97.38%, respectively, and the proposed model achieves SOTA compared with state-of-the-art methods. The proposed model and code are exposed *via* Github[1].

KEYWORDS

submarine exploration, underwater scene, RMP-Net, structural re-parameterization, multiscale fusion

1 https://github.com/QingWind6/RMP-Net.

1 Introduction

As we all know, the area of the ocean is about 360 million square kilometers, accounting for about 71% of the total surface area of the earth, but the degree of human exploration and development of the ocean is less than 5%. However, the ocean is an important source of food, energy, and minerals in the world, such as rich marine life, oil, natural gas, coal, and various rare metal resources. In addition, about 40% of the world's population live in coastal areas, and three quarters of the world's largest cities are located in coastal areas (Rayner et al., 2019).

In the process of using autonomous underwater explorers to explore the seabed, the detection and perception accuracy of the vision system is relatively high. For example, for challenging problems such as insufficient ambient light, related research studies have used convolutional neural networks to detect and identify marine organisms (Li et al., 2020; Xu et al., 2021; Zhang et al., 2021; Kraft et al., 2022). However, using target detection to achieve target object positioning is still not accurate enough. In order to effectively solve the above problems, we take semantic segmentation as an important research direction and use semantic segmentation to perform pixel-level segmentation of target objects so that pixel-level segmentation can be achieved. We then locate the target object.

Most of the previous studies are dealing with single-class or less-classification tasks, i.e., segmentation only of the foreground and the background (Saleh et al., 2020; Zhang et al., 2021). For example, Zhang et al. (2021) proposed a dual-pool aggregated attention network called DPANet, through which the pool-aggregated location attention module and the pool-aggregated channel attention module process the feature maps to perform pixel-level segmentation of marine organisms. The above research performs semantic segmentation of fish; that is, only the fish and the ocean background need to be segmented. For the multi-objective semantic segmentation task of the ocean, Islam et al. (2020) proposed a fully convolutional encoder–decoder model called SUIM-Net, which has two versions. One is to combine the residual structure and skip connection. The residual skip block (RSB) is formed, which can solve the problem of gradient explosion and gradient disappearance in relatively deep network training (He et al., 2015a). This improves the performance of the model by adding RSB connections to the decoder. The other version is to use the VGG convolutional neural network as the backbone network of SUIM-Net (Simonyan and Zisserman, 2014), which has the advantage of improving the segmentation performance of the network. Finally, the authors of this study chose SUIM-Net (VGG) as their final model, which shows that segmentation performance is still the key research direction of ocean image segmentation. In addition, the authors of this study published a large dataset they produced for the semantic segmentation of underwater images (Islam et al., 2020).

The above-mentioned previous studies used more traditional and classical neural networks, and most of them were binary classification tasks, that is, background and target objects. However, in an actual marine exploration environment, various creatures will appear, and using only a few classification tasks cannot meet practical engineering needs. In order to effectively improve the operational efficiency of autonomous underwater explorers, we propose a novel structure-reparameterization and subpixel convolutional super-resolution network (RMP-Net), which uses a structure-reparameterized backbone feature extraction network. It enables the network to have better feature extraction performance and certain real-time performance (Ding et al., 2021), and the subpixel convolution super-resolution module can better help the performance of the backbone network to better migrate to downstream tasks in feature recovery (Shi et al., 2016; Bousmalis et al., 2016). In the process of feature fusion, more pieces of contour and shallow semantic information of the target object are retained, so that the network model can realize the multitarget semantic segmentation task of the marine environment. In this study, we use the SUIM dataset (Islam et al., 2020) and the DeepFish dataset (Saleh et al., 2020) for experimental testing to better evaluate the comprehensive performance of the model through these two challenging datasets.

The contributions of this study can be summarized as follows:

1. This study proposes RMP-Net for the multiclassification semantic segmentation of marine scenes to meet the needs of autonomous underwater explorers for multitarget recognition and segmentation in seabed operations. It is expected to improve the performance of underwater operations' efficiency.
2. RMP-Net uses subpixel convolution to perform semantic information super-resolution recovery and fusion of the features learned by the backbone feature extraction network, which can efficiently fuse the multiscale feature layers obtained by the backbone network to further enhance the model's ability to operate under the sea. This shows the accuracy of the multiclass object segmentation in the environment.
3. RMP-Net is evaluated using both SUIM and DeepFish datasets and outperforms state-of-the-art methods. At the same time, the RMP-Net proposed in this study and the entire framework source code can be downloaded from Github.

The rest of the paper is organized as follows. We discuss the method proposed in this study in Section 2, present the experimental results in Section 3, discuss the results in Section 4, and summarize the results in Section 5.

2 Materials and methods

2.1 Structural reparameterization

Early convolutional neural networks achieved better results by stacking convolutional layers continuously, such as VGG and AlexNet (Krizhevsky et al., 2012; Simonyan and Zisserman, 2014). In recent years, in order to improve network performance, researchers have also developed many complex structures, such as introducing multibranch structures and using different convolution methods (Szegedy et al., 2014). Although these schemes can improve model performance, they will also bring other problems such as increasing memory consumption and affecting model inference speed (Zhang et al., 2017); therefore, VGG and ResNet are still widely used today (Simonyan and Zisserman, 2014; He et al., 2015a).

In order to effectively improve the accuracy of the plain structure of VGG, RepVGG is proposed (Ding et al., 2021). RepVGG introduces a multibranch structure based on VGG, which can greatly improve the model performance during training. In addition, RepVGG uses multibranch fusion technology to merge the convolution kernels BN in the multibranch structure during inference, which greatly improves the speed of the model in the inference phase.

Figure 1 shows a RepVGG Block structure, which uses a multibranch structure during training and a single-way structure during the inference phase above. Assuming that F(X) represents a 3x3 convolution and G(X) represents a 1x1 convolution, then

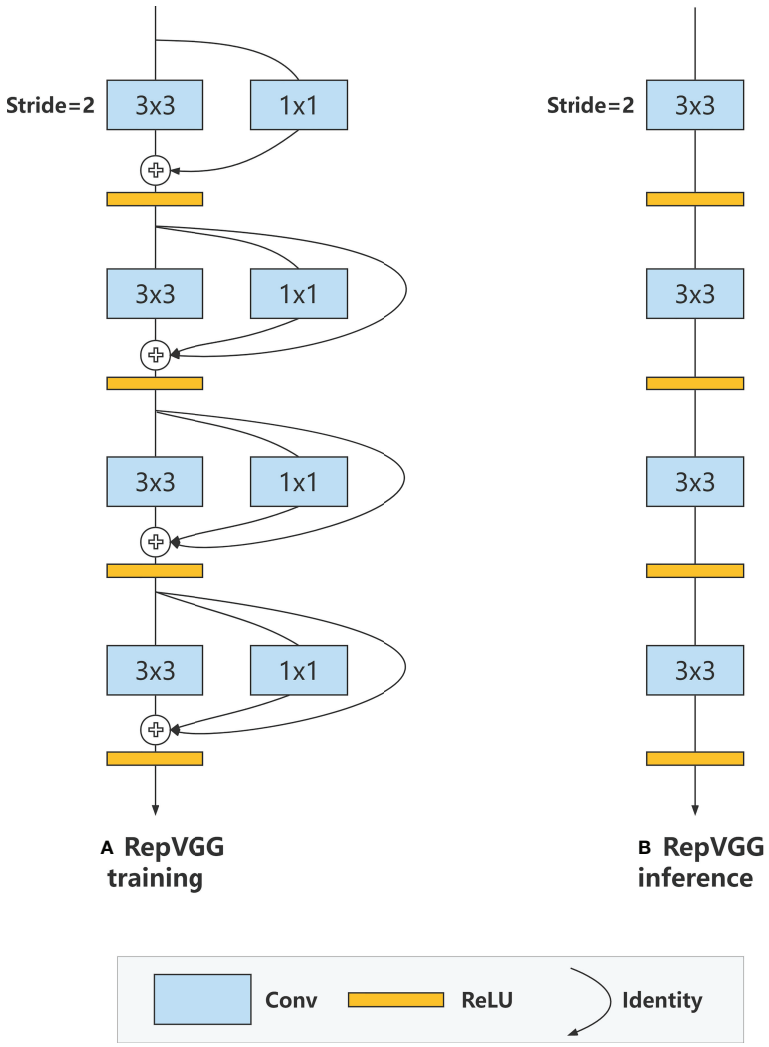


FIGURE 1
RepVGG Block structure, using a multibranch structure during training and a single-way structure during the inference phase.

the calculation made in a block can be expressed as Equation 1:

$$Out = F(X) + G(X) + X \tag{1}$$

In the process of structural reparameterization, we will fuse the convolutional and pooling layers, where it is assumed that W is the convolutional kernel weight, i is the ordinal number of the convolutional kernel, μ denotes the mean of the BN layer, and the variance of the BN layer is denoted by σ; γ_i and β_i denote the scale factor and the offset factor of the BN layer, respectively; and W' and b' denote the convolutional weight and bias after fusion, respectively. Then the weight of the ith convolutional kernel after fusion can be expressed as follows:

$$W'_{i,:,:,:} = \frac{\gamma_i}{\sigma_i} W_{i,:,:,:} \tag{2}$$

The weight of the ith BN layer after fusion can then be expressed as Equation 3:

$$b'_i = -\frac{\mu_i \gamma_i}{\sigma_i} + \beta_i \tag{3}$$

Figure 2 shows RepVGG, which performs the process of structural reparameterization, i.e., multibranch structural fusion above. For the 1x1 convolutional branch in the structure, we can first replace it with a 3x3 convolutional kernel and then move the values in the 1x1 convolutional kernel to the center of the 3x3 convolutional kernel, and fill the rest with zeros. For the identity branch, we can also equivalently convert it to a 3x3 convolution kernel, which gives us three 3x3 convolution kernels.

Finally, we only need to fuse and merge multiple convolutional kernels, i.e., to ßuck BN. Assuming that the weights of all branches are W, the input and output matrices are M, and the bias is B, we can represent the fused 3x3 convolution as Equation 4:

$$bn(M*W, \mu, \sigma, \gamma, \beta)_{:,i,:,:} = (M*W')_{:,i,:,:} + b'_i \tag{4}$$

We use RepVGG as the U-Net backbone feature extraction network, and the output of the five stages is effectively used as multiscale fusion features, which can increase the model's ability to extract features at different scales while reducing the number of computational parameters in model inference, and can effectively extract features and fuse them to improve model segmentation accuracy.

2.2 Subpixel convolution

In the process of camera imaging, the resulting image is actually discretized; where two adjacent pixels are connected macroscopically, in reality, there are countless tiny things between them microscopically. These are called subpixels. Subpixels actually exist, but due to the physical limitations of the acquisition device's sensor, the data can only be represented by approximation. To maximize the use of information in the image itself to improve super-resolution, Shi et al. (2016) proposed subpixel convolution.

Figure 3 shows a high-resolution image obtained by upsampling a low-resolution image using subpixel convolution. By super-resolution, upsampling a high-dimensional low-resolution feature map to obtain a low-dimensional high-resolution image can effectively recover the detailed information in the feature map. The principle of subpixel convolution is to learn the convolution of the corresponding number of channels in the penultimate layer of the model, (r^2, c), where c is the number of channels in the final output, and r is the upsampling multiplier to be performed. For example, if a 9-channel 3x3 feature map is upsampled by a factor of 3, the final size is a single-channel 9x9 feature map. The above pixel alignment operation can be described by the following principle:

$$PS(T)_{x,y,c} = T_{[x/r],[y/r],c \cdot r \cdot \bmod (y,r) + c \cdot \bmod (x,r)} \tag{5}$$

From the above equation, we can see that in subpixel convolution, consecutive c channels are taken from the number of channels of the feature map as a whole, and then the final multichannel upsampled map is obtained by rearranging the pixels. For a feature vector of size HxW channel number r^2, we can upsample it with subpixel convolution, assuming that H' is the output feature vector length, W' is the output feature vector width, and r' is the output feature vector channel number. Then the final output scale can be expressed by the following equation:

$$H' = rH, \quad W' = rW, \quad r' = 1 \tag{6}$$

We connect the subpixel convolution to the decoder in U-Net, pass the fused features at each scale to the subpixel convolution for super-resolution upsampling, set different subpixel convolution upsampling multipliers for each layer (Sun et al., 2020), and upsample the output of each layer into a feature vector of size 256x256. The advantage of this is that the edge information in the deep semantic information of low resolution and high dimension can be recovered and preserved as much as possible, whereas the consistent scale makes it convenient to fuse and sum the feature vectors.

2.3 Structure of RMP-Net model

The RMP-Net network is based on the improved U-Net model (Ronneberger et al., 2015) [19], and B2 is selected as the backbone feature extraction network after comparing the structure of multiple versions of RepVGG. The five stages of RepVGG-B2 have [1,4,6,14,1] layers and [64,128,256,512] widths, and the scaling hyperparameters are 2.5 and 5, respectively. Figure 4 shows the

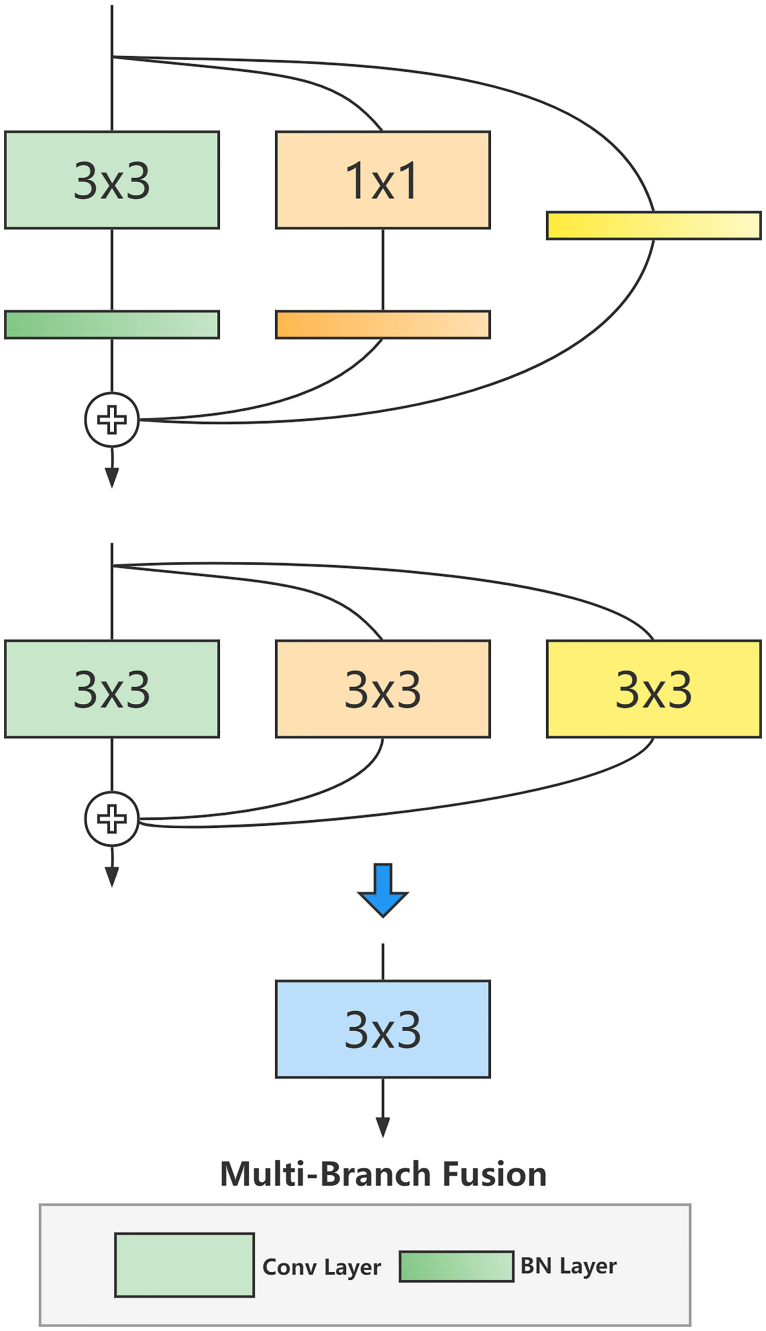


FIGURE 2
RepVGG performs the process of structural reparameterization, i.e., multibranch structural fusion.

overall structure of the RMP-Net network, where the backbone RepVGG is composed using a RepVGG block and uses subpixel convolution to upsample multiscale effective features for semantic information recovery below. The RMP-Net is mainly composed of four parts: input, backbone, multiscale fusion, and output prediction.

The first part of the data input is actually a pre-processing operation of the data, which contains the enhancement of the dataset, such as the rotation, cropping, and enlargement operations corresponding to the original image and labels, and the 1.5 times expansion of the dataset, which can effectively improve the robustness and accuracy of the model through data

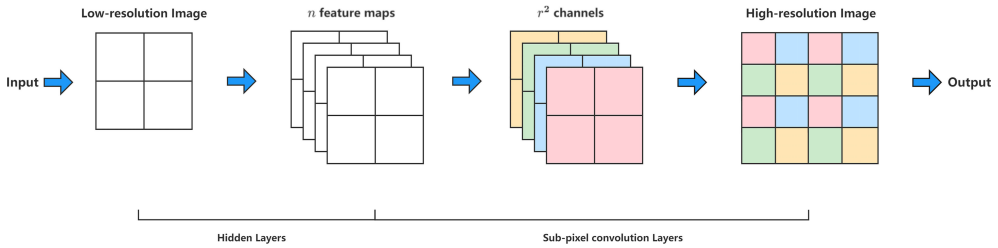


FIGURE 3
The high-resolution image is obtained by upsampling the low-resolution image using subpixel convolution. By super-resolution, upsampling a high-dimensional low-resolution feature map to obtain a low-dimensional high-resolution image can effectively recover the detailed information in the feature map.

enhancement (Cubuk et al., 2018). Here we will unify the input images to a size of 512x512.

The backbone network uses the RepVGG network, which mainly consists of multiple stages with different numbers of RepVGG blocks, and the RepVGG blocks are composed of VGGs with multibranch structure; the multibranch structure is used to obtain high performance during training (Szegedy et al., 2014), and the inference stage is turned into a single-way structure by the multibranch fusion algorithm to speed up the model inference. This allows the model to have high performance and faster inference speed at the same time. We will select five stages of RepVGG as the effective multiscale feature vectors for the subsequent enhanced feature fusion operation, which are [64,256,256], [160,128,128], [320,64,64], [640,32,32], and [2560,16,16] in order of scale, which can be referred in Table 1.

In the multiscale fusion part, the feature vectors extracted from the backbone network at different scales are first super-resolved by subpixel convolutional scale normalization, where we will process the five feature vectors output from the backbone network and upsample them by subpixel convolutional scale unification so that we can get five feature vectors with scales of [10,256,256], [8,256,256], [16,256,256], [32,256,256], and [64,256,256]. After summing the above feature vectors, we can obtain a feature vector map with the scale [130,256,256], which is the final fused feature vector map we need. Table 2 shows the scale of different layers.

The final prediction part classifies the output of the model by performing a Softmax operation on the predicted values of each pixel to obtain the multicategorization probability of the model for each pixel. The maximum probability value is obtained by argmax to determine the class of the pixel. After classifying a pixel, it is

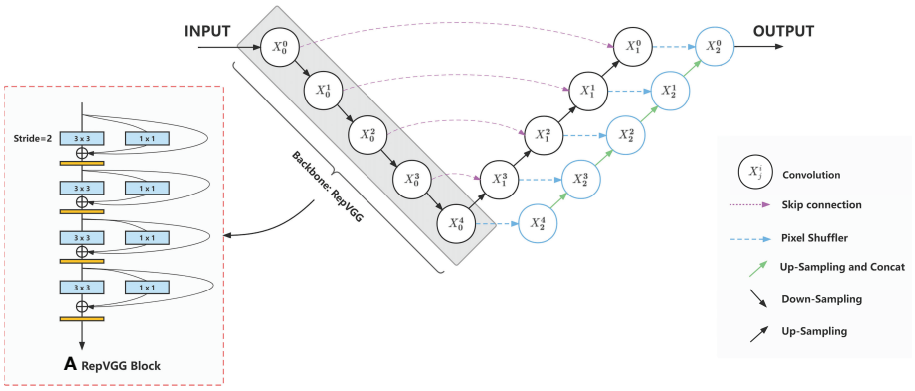


FIGURE 4
The overall structure of the RMP-Net network; the backbone RepVGG is composed using a RepVGG Block and uses subpixel convolution to upsample multiscale effective features for semantic information recovery.

TABLE 1 The scale of different effective feature layers of the backbone.

Name of Layer	Shape	Channels
Feat-Layer 1	256 x 256	64
Feat-Layer 2	128 x 128	160
Feat-Layer 3	64 x 64	160
Feat-Layer 4	32 x 32	640
Feat-Layer 5	16 x 16	2560

simply assigned a different color and mapped to the location of the corresponding pixel to complete the segmentation.

2.4 Loss function

The loss function used in the training model in this paper is cross-entropy loss (Rubinstein and Kroese, 2004; Ma et al., 2021), which is used when classifying pixel points using Softmax. It can be used to measure the difference between two probability distributions, and thus the difference between the distribution learned by the model and the true distribution. When using cross-entropy loss for multiclassification tasks, the mathematical relation can be expressed by the following equation:

$$L = \frac{1}{N}\sum_{i} L_i = -\frac{1}{N}\sum_{i}\sum_{c=1}^{M} y_{ic} \log\left(p_{ic}\right) \tag{7}$$

where M denotes the number of categories, and y_{ic} denotes the symbolic function (if the true category of sample i is equal to c, take 1; otherwise, take 0). p_{ic} then denotes the prediction probability that sample i belongs to category c. Different tasks may require the use of different loss functions, and the loss function will, to some extent, affect the final effect of the model after training. When using the cross-entropy loss function, the learning speed will be faster when the model is poor and slower when the model is good (Gonzalez and Miikkulainen, 2019).

Because the semantic segmentation we study this time is a multiclassification task, here we directly use cross-entropy loss as the loss function of the neural network to optimize the model.

TABLE 2 Model hyperparameters at different training stages.

Name of Layer	Shape	Channels
Sup-pixel-Up 1	256 x 256	10
Sup-pixel-Up 2	256 x 256	8
Sup-pixel-Up 3	256 x 256	16
Sup-pixel-Up 4	256 x 256	32
Sup-pixel-Up 5	256 x 256	64
Combined-Layer	256 x 256	130

2.5 Model training

The computer hardware and software environment for model training and the main performance indicators are shown as follows:

(1) CPU: Intel Xeon Silver 4110; (2) memory: 128G DDR4; (3) graphics card: 4 x Nvidia GeForce Titan Xp; (4) operating system: Ubuntu 18.04; (5) development environment: Python 3.8; PyTorch 1.11.0.

The network model training in this paper uses ImageNet-based pretraining weights as the initial parameters of the backbone network RepVGG for migration learning training (Deng et al., 2009). The parts other than the backbone network will be initialized with Kaiming to accelerate the model convergence speed (He et al. (2015b)). The whole training process will be divided into two parts: freeze trunk training and unfreeze trunk training. Between them, 0–50 is for freezing training, and 50–100 is for thawing training. The initial learning rate for frozen training is set to 5e-3, and the batch size is set to 16. The initial learning rate for unfreezing training is set to 5e-5, and the batch size is set to 8. The training optimizer used is the Adam optimizer (Kingma and Ba, 2014), and its parameters and learning rate are optimized by the StepLR adjuster; the step size is 1, and the gamma is 0.96. The training parameters are referred to in Table 3. Freeze training only loads the weights of the backbone network because the preweights used are obtained by training the ImageNet dataset. Such a classification network already has very good feature extraction capabilities, so the weights of the backbone network remain unchanged during the training process (Pan and Yang, 2010). Only the weight parameters of the decoder part are changed. Unfreezing training releases the weight of the entire network, so the weight parameters of the entire network will change during the training process, but the parameters of the backbone network are generally only fine-tuned and will not change too much.

Transfer learning can transfer the ability of the network used for one task to another task, which greatly improves the speed and effect of model training and enables the network to have relatively good results in different tasks and therefore in preweighting. The loading of is essential for 99% of model training.

3 Results

3.1 Results on SUIM

3.1.1 Qualitative analysis

The SUIM dataset focuses on the segmentation of ocean scenes [11]. It includes a total of 1,525 training images and 110 verification images. The entire dataset contains a total of eight

TABLE 3 Model hyperparameters at different training stages.

Hyperparameters	Epoch	Learning Rate	Batch Size	Freeze Train	Optimizer	LR Decay	Momentum
Values	1–50	5e-3	16	True	Adam	Step	0.9
	51–100	5e-5	8	False			

categories of target objects. The codes corresponding to the specific categories are shown in Table 4.

In Figure 5, the visualization results of the comparison test of each model network on the SUIM dataset are shown. We use images of each category from the test set as the test input of the model to more intuitively compare the multiclass segmentation performance of the model.

It can be seen from the segmentation comparison diagram in Figure 5 that the segmentation effect of the original U-Net is relatively poor, and there will be many misjudgments of target categories. In addition, the recovery effect of target segmentation integrity and edge information is also very broad. The U-Net model that replaces the VGG backbone with ResNet50 has relatively good prediction results, and there are much fewer misjudgments of the target category. Most of the pixels can be classified correctly, and most of the target area can be predicted. The prediction results of PSPNet are still very good (Zhao et al., 2017). Compared with the previous two, it can be seen from the fourth picture that the pixel accuracy of the model is much higher, and the segmentation effect is very good. The performance of the DeepLab model is comprehensively between the U-Net of the PSPNet and ResNet50 versions (Chen et al., 2017). The overall prediction effect of SUIM-Net is also good, but there will be some misclassifications.

The last column is the prediction result of RMP-Net proposed in this paper. It can be seen intuitively that its effect is better than that of the other models. There are a few cases of misjudgment of the target category in the six test images. The prediction results of the first picture show that RMP-Net has a high degree of restoration for the target contour feature recovery. From the results of the fourth picture, it can also be seen that the model has a better recovery effect on the edge contour feature details of the human hand. Experiments show that the test results of the RMP-Net model are better than those of the other models.

TABLE 4 Classes of objects annotated with pixels in the dataset and the corresponding codes.

Object category	Code
Background/water body	BW
Human divers	HD
Aquatic plants and seagrass	AS
Robots/instruments	RO
Reefs and invertebrates	RI
Fish and vertebrates	FV
Wrecks or ruins	WR
Seafloor and rocks	SR

3.1.2 Quantitative analysis

The evaluation metrics use two metrics commonly utilized in semantic segmentation tasks: mIoU and mPA, which can better reflect the comprehensive performance of the model.

The mIoU is the weighted average of the IoU of each class. The IoU of each class can be calculated by the confusion matrix, and the value on the diagonal of the confusion matrix is the intersection of the class; each row plus each column minus the value on the diagonal is the sum of the class.

After getting the IOU of each class, we only need to obtain the average to get the mIoU. Let N be the number of categories and Sum be the summation; then mIoU is given as follows:

$$\mathrm{mIoU} = \frac{Sum(IoU_i)}{N} \tag{8}$$

The mPA is the proportion of the number of pixels correctly classified for each class calculated separately, i.e., the CPA and then averaged cumulatively. Assuming that P is the accuracy of each category pixel, the mPA is given as follows:

$$mPA = \frac{Sum\ (P)}{N} \tag{9}$$

The SUIM dataset is used to test the segmentation accuracy of the model, and the corresponding verification models are RMP-Net, UNet (ResNet50), UNet (VGG16), PSPNet, DeepLabV3, DeepLabV3, and SUIM-Net. As mentioned above, the evaluation indicators mIoU and mPA are used to evaluate the model, and the final evaluation results are shown in Figure 5 and Table 5, 6.

Observing the two data comparison tables, the RMP-Net proposed in this paper is the best among all models in terms of comprehensive indicators. In the segmentation test on the SUIM dataset, mIoU reached 84.52%, and mPA reached 92.33%. Compared with the baseline, mIoU and mPA are improved by

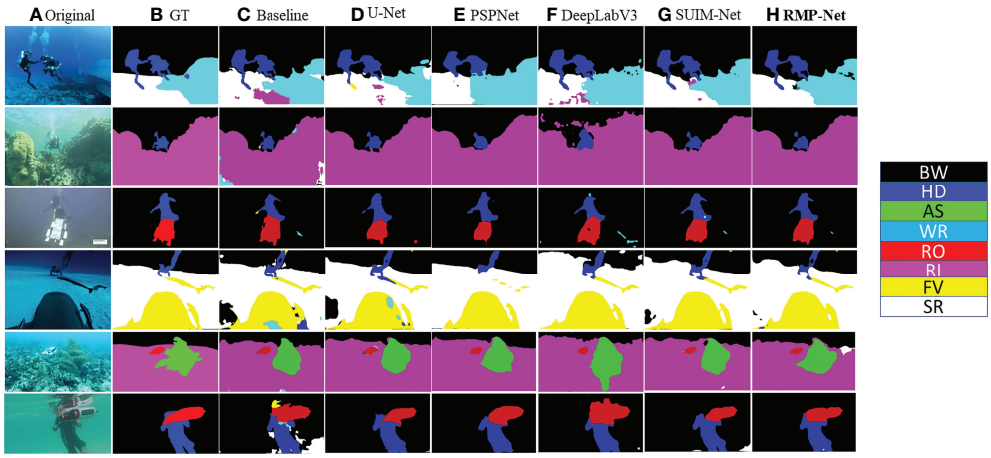


FIGURE 5
Qualitative comparative experiments on state-of-the-art segmentation models. **(A)** Input image. **(B)** Label. **(C)** Baseline. **(D)** U-Net (ResNet50). **(E)** PSPNet. **(F)** DeepLabV3. **(G)** SUIM-Net. **(H)** RMP-Net.

about 9.25% and 9.33%, respectively. We can see that among the eight categories in the dataset, the model has the highest classification accuracy for humans and coral reefs and invertebrates, reaching 96.31% and 92.93%, respectively. Compared with the SUIM-Net model, our proposed model achieves about 3.37% and 4.62% higher mIoU and mPA, respectively. The experimental results show that the effect of RMP-Net in this seabed image segmentation task has reached the expectation and can meet the needs of ocean multitarget segmentation. ()

In Figure 6, data visualization of the performance of different segmentation models on the dataset can be seen. It can be seen that RMP-Net has achieved the best results in multiple classifications and achieved SOTA in comparison with the most advanced methods, which further reflects the effectiveness of using RMP-Net in the task of ocean scene segmentation from the data.

3.2. Results on DeepFish

3.2.1. Qualitative analysis

In order to further verify the robustness and effectiveness of the proposed model, we selected the binary segmentation dataset DeepFish to test the model (Saleh et al., 2020). The DeepFish dataset focuses on fish in the marine environment and can be used in the fields of target detection, classification, and segmentation. In this experiment, we use the semantic segmentation part of the dataset as the experimental data. The segmentation dataset has a total of 620 images, and the corresponding label of the picture only contains two categories: fish and background. We will divide the training and verification set given by an official for model training and model verification. Figure 7 shows the final prediction effect of the model.

TABLE 5 Comparing mIoU metrics between different models on the SUIM dataset, where the metrics with the highest rankings are shown in bold.

Method	Backbone	mIoU(%)	IoU per category(%)					
			BG	RO	FV	HD	RI	WR
PSPNet	**MobileNet**	77.8	82.1	69.7	79.1	80.2	79.4	80.4
PSPNet	**ResNet50**	81.2	84.5	72.	81.2	83.4	85.2	83.9
DeepLabV3	**MobileNet**	78.7	81.2	70.9	79.0	79.3	83.2	81.0
DeepLabV3	**Xception**	80.7	82.9	70.5	80.9	82.6	85.2	85.2
U-Net	**VGG16**	77.3	80.2	69.2	77.3	77.	82.2	80.6
U-Net	**ResNet50**	79.0	83.	71.	79.6	78.3	84.5	81.3
SUIM-Net	**VGG**	81.7	86.5	70.3	**82.22**	84.1	86.6	85.4
RMP-Net	**RepVGG**	**84.52**	**89.53**	**75.3**	**81.9**	**88.86**	**88.92**	**87.63**

TABLE 6 Comparing mPA metrics between different models on the SUIM dataset, where the metrics with the highest rankings are shown in bold.

Method	Backbone	mPA(%)	PA per category(%)					
			BG	RO	FV	HD	RI	WR
PSPNet	MobileNet	84.63	83.21	75.28	84.22	89.91	86.83	86.89
PSPNet	ResNet50	87.27	87.59	77.53	86.58	92.05	89.64	90.55
DeepLabV3	MobileNet	84.9	85.47	76.46	85.71	85.94	89.3	87.08
DeepLabV3	Xception	87.42	86.1	76.47	86.49	90.97	**93.37**	89.78
U-Net	VGG16	84.45	85.23	75.4	84.33	84.23	90.15	88.15
U-Net	ResNet50	86.3	86.66	78.49	87.51	85.66	91.27	88.59
SUIM-Net	VGG	88.25	89.01	75.25	**88.29**	93.52	90.87	89.62
RMP-Net	RepVGG	**92.33**	**93.52**	**80.74**	86.28	**96.31**	92.93	**92.22**

Renderings of the seven images verify the result of concentrated random screening. From the model-predicted results, there can be found two different ambient lights of ocean scene. We put forward the model to forecast the effect of fish, which is still very ideal; at the same time, for different sizes of fish, the model can also be fully used. We need to pick up the goal for fish edge contour feature recovery. From the above results, we can see that subpixel convolution can well-integrate the features of different scales, so that the back-end network can retain and fuse the semantic information of features of different scales extracted from the front-end network, which is very effective for the task of ocean scene segmentation.

3.2.2 Quantitative analysis

We use the mIoU and mPA evaluation indicators to test different models on the DeepFish dataset, and the final test results are shown in Table 7.

It is not difficult to analyze the experimental data: our proposed model is very competitive in both background and fish (foreground) segmentation performance, with an IoU of 99.61% verified on the background and reached 90.90%. The DeepFish dataset contains many fish of different sizes and shapes. Experiments show that our proposed model can achieve very good segmentation results for the above targets, which further confirms the multibranch backbone and subpixel in RMP-Net. The convolution module has a very good effect on feature processing of complex underwater environments and can fully cope with complex ocean scene segmentation tasks.

4 Discussion

4.1 Ablation study

Although the performance of RMP-Net is significantly improved compared with the baseline, it is not clear how each model affects the performance of the whole model. Therefore, ablation experiments are needed to verify the impact of each module on the performance of the network model separately. The ablation results are shown in Table 8.

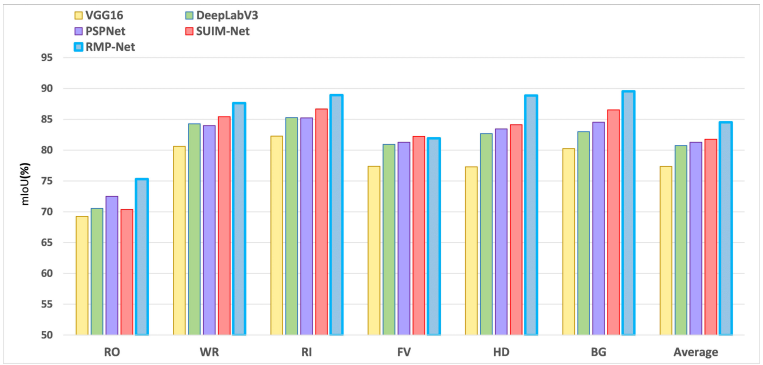


FIGURE 6
Testing different mainstream segmentation networks on the SUIM dataset, RepVGG, and subpixel convolution helps to improve the performance of the segmentation model, illustrating the effectiveness of RMP-Net in the task of ocean scene segmentation.

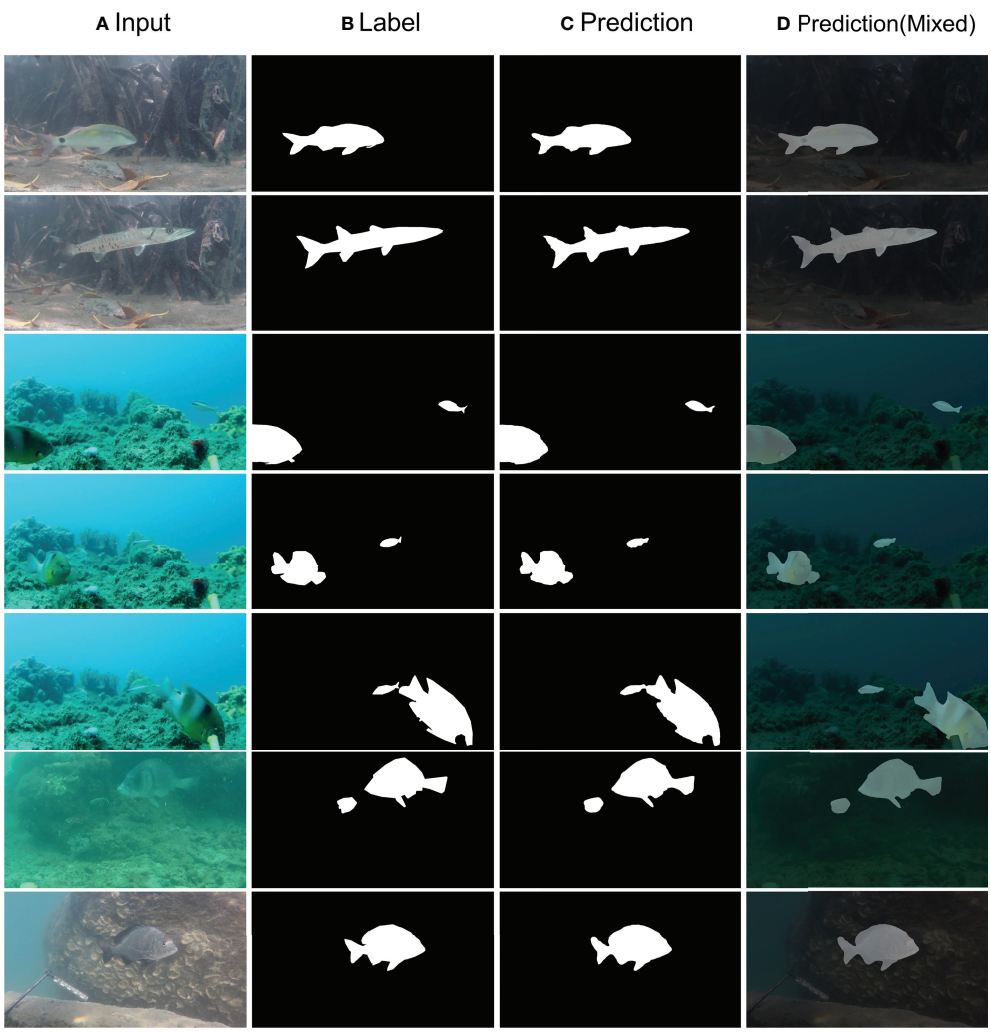


FIGURE 7
Results on DeepFish. **(A)** Input image. **(B)** Label. **(C)** Prediction. **(D)** Mixed image of original image and predicted image.

The data in Table 8 are obtained from the ablation experiment. We can know from the data that RepVGG and subpixel convolution both play a significant role in improving the performance of the network model whether added separately or together. Therefore, in RMP-Net, we apply both structures to the model, which is helpful for the overall performance improvement, and it is not an accidental combination. The above experiments prove the effectiveness of RepVGG and

TABLE 7 Comparing our proposed method with state-of-the-art methods on the DeepFish dataset.

Method	Background IoU(%)	Fish IoU(%)	mIoU(%)
SUIM-Net (Islam et al. (2020))	99.03	78.4	88.71
SegNet (Badrinarayanan et al. (2017))	98.89	68.94	83.91
DeepLabv3 (Chen et al. (2017))	99.11	71.35	85.23
PSPNet (Zhao et al. (2017))	99.15	72.61	85.88
DPANet (Zhang et al. (2021))	99.31	82.86	85.88
MFAS-Net (Haider et al. (2022))	99.15	84.86	92.01
RMP-Net (**Ours**)	**99.61**	**90.9**	**95.26**

The bold type is to emphasize that our algorithm achieved the highest score in the task compared to other state-of-the-art algorithms.

subpixel convolution. This combination may be used in network models of other structures.

4.2 Comparison of model inference speed

We tested and compared different models with a parameter amount, FLOPS, and the inference speed of the model on the test set. It can be seen that compared with other advanced models, RMP-Net has strong backbone feature extraction capabilities and is deployed in inference. When multibranch fusion is performed, the model inference speed can be further accelerated, but the performance in terms of model parameters and inference speed is still not good enough. The test results are detailed in Table 9.

4.3 Future work

With the continuous improvement of computer hardware's computing power and the in-depth exploration and research of machine learning by researchers, deep learning has been widely used in all aspects of human society. Among them, many achievements have been made in object detection, image processing, and natural language processing. These technologies have facilitated people's lives in various aspects. However, there are still very few related applications in ocean observation. In addition to the high cost of equipment required for ocean observation, the main reason is that people do not pay enough attention to the ocean, and there are relatively few research studies related to ocean observation. Through testing, we found that, although the performance of the model is very strong, the operation of the algorithm still requires certain hardware conditions to achieve satisfactory results. Therefore, we aim to compress and optimize the proposed network model based on the work performed in this research, so that it can become more lightweight and can meet the hardware environment with lower requirements.

Realizing that the existing work is far from enough, if there is an opportunity in the future, we will go to the field to conduct ocean observations and will also carry out some data collection, which can also help more researchers to promote ocean observation research and contribute to the development of ocean exploration.

5 Summary

In this paper, a segmentation network RMP-Net based on structural reparameterization and subpixel convolution was proposed for effective and accurate segmentation of seafloor-related targets, which can be used to construct 3D semantic maps to facilitate ocean observation. This study was carried out to enhance the comprehensive performance of the model for segmentation by introducing a structurally reparameterized classification network as the backbone feature extraction network, while redesigning the decoder side of the network to use subpixel convolution for upsampling to recover the information in the deep feature maps, and then using

TABLE 8 Ablation study.

Method	Input size	mIoU(%)	mPA(%)
Baseline	512 x 512	77.36	84.45
+ RV	512 x 512	82.17	89.57
+SP	512 x 512	80.85	88.13
+ RV + SP(**Ours**)	512 x 512	**84.52**	**92.33**

Baseline: U-Net (VGG16). RV, RepVGG; SP, subpixel convolution.
The bold type is to emphasize that our algorithm achieved the highest score in the task compared to other state-of-the-art algorithms.

TABLE 9 Results of comparison of Params, FLOPS, and inference speed of different models (Single Nvidia RTX2080 GPU).

Method	Params(M)	GFLOPS(G)	FPS
PSPNet	46.716	118.44	36.51
DeepLabV3	54.714	167.01	34.10
U-Net(VGG16)	24.892	452.31	23.53
U-Net(ResNet50)	43.934	184.73	44.69
SUIM-Net	12.219	120.58	**48.30**
RMP-Net	78.449	202.46	42.88

Baseline: U-Net (VGG16). RV, RepVGG; SP, subpixel convolution. The bold type is to emphasize that our algorithm achieved the highest score in the task compared to other state-of-the-art algorithms.

multiscale fusion to segment the target features in the images accurately. The backbone network can be compressed by a multibranch fusion algorithm to further improve the model inference and prediction speed without losing performance, making the model simultaneously have high performance and high real-time performance. Several mainstream semantic segmentation networks are tested for evaluation metrics using the SUIM underwater image dataset, and the experimental results show that RMP-Net achieves 84.52% mIoU and 92.33% mPA. Compared with PSPNet, mIoU and mPA are improved by 1.57% and 1.28%, respectively, and the overall performance metrics are the highest compared with other segmentation networks, which also shows that the high performance of RMP-Net is fully capable of meeting the segmentation task requirements in ocean observation scenarios.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: SUIM, https://irvlab.cs.umn.edu/resources/suim-dataset; DeepFish, https://alzayats.github.io/DeepFish/.

Ethics statement

Ethical review and approval were not required for the animal study because, in this research, we used two open databases of DeepFish and SUIM, which can be freely used for academic purposes.

Author contributions

JC completes the main work of this paper. JT, SL and DZ guide JC in the research of this work. WL, JY, BS and YL help complete the part of experiment. All authors are involved in revising the manuscript, proofreading, and approving the submit.

Funding

This work was supported by Special Fund for Scientific and Technological Innovation and Cultivation of College Students in Guangdong Province: pdjh2021 a0625 and the Science and Technology Program of Social Development,Zhuhai,2022 under grant 2220004000195 and the Guangdong Province universities key field special 2022ZDZX4061.

Acknowledgments

Throughout the process of conducting the experiment and writing the paper. I received great support and help, and received many valuable comments. I would like to express my sincere gratitude to all those who have helped me throughout the process. First of all, I would like to thank my supervisor, Tang Jialin, who gave me valuable advice on academic matters. This thesis would not have been possible without his patient guidance, insightful criticism and professional instruction.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (12), 2481–2495 doi: 10.1109/TPAMI.2016.2644615

Bousmalis, K., Trigeorgis, G., Silberman, N., Krishnan, D., and Erhan, D. (2016). Domain separation networks. *Neural Inf. Process. Syst.* 29, 343–351. doi: 10.48550/arXiv.1608.06019

Chen, L.-C., Papandreou, G., Schroff, F., and Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv: Comput. Vision Pattern Recognit.* abs/1706.05587. doi: 10.48550/arXiv.1706.05587

Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V. K., and Le, Q. V. (2018). Autoaugment: Learning augmentation policies from data. *arXiv: Comput. Vision Pattern Recognit.* abs/1706.05587. doi: 10.48550/arXiv.1805.09501

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. *Comput. Vision Pattern Recognit.* 248–255. doi: 10.1109/CVPR.2009.5206848

Ding, X., Zhang, X., Ma, N., Han, J., Ding, G., and Sun, J. (2021). Repvgg: Making vgg-style convnets great again. *Comput. Vision Pattern Recognit.* abs/2101.03697, 13733–13742. doi: 10.1109/CVPR46437.2021.01352

Gonzalez, S., and Miikkulainen, R. (2019). Improved training speed, accuracy, and data utilization through loss function optimization. *congress evol. Comput.* abs/1905.11528, 1–8. doi: 10.48550/arXiv.1905.11528

Haider, A., Arsalan, M., Choi, J., Sultan, H., and Park, K. R. (2022). Robust segmentation of underwater fish based on multi-level feature accumulation. *Front. Mar. Sci.* 9, 1010565. doi: 10.3389/fmars.2022.1010565

He, K., Zhang, X., Ren, S., and Sun, J. (2015a). Deep residual learning for image recognition. *arXiv: Comput. Vision Pattern Recognit.* abs/1512.03385, 770–778. doi: 10.1109/CVPR.2016.90

He, K., Zhang, X., Ren, S., and Sun, J. (2015b). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Int. Conf. Comput. Vision*. abs/1502.01852, 1026–1034. doi: 10.1109/ICCV.2015.123

Islam, J., Edge, C., Xiao, Y., Luo, P., Mehtaz, M., Morse, C., et al. (2020). Semantic segmentation of underwater imagery: Dataset and benchmark. *Intell. Robots Syst*. abs/2004.01241, 1769–1776. doi: 10.1109/IROS45743.2020.9340821

Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv: Learn*. doi: 10.48550/arXiv.1412.6980

Kraft, K., Velhonoja, O., Eerola, T., Suikkanen, S., Tamminen, T., Haraguchi, L., et al. (2022). Towards operational phytoplankton recognition with automated high-throughput imaging, near-real-time data processing, and convolutional neural networks. *Ocean Science Discussions* 2020, 1–20. doi: 10.3389/fmars.2022.867695

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Commun. ACM*. 60, 84–90. doi: 10.1145/3065386

Li, L., Rigall, E., Dong, J., and Chen, G. (2020). Mas3k: An open dataset for marine animal segmentation. *Int. Symp. Benchmarking Meas. Optim.* 12614, 194–212. doi: 10.1007/978-3-030-71058-3_12

Ma, J., Chen, J., Ng, M., Huang, R., Li, Y., Li, C., et al. (2021). Loss odyssey in medical image segmentation. *Med. Image Anal.* 71, 102035. doi: 10.1016/j.media.2021.102035

Pan, S. J., and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. doi: 10.1109/TKDE.2009.191

Rayner, R., Jolly, C., and Gouldman, C. (2019). Ocean observing and the blue economy. *Front. Mar. Sci.* 6, 330. doi: 10.3389/fmars.2019.00330

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *Med. image comput. Comput. assist. intervention*. abs/1505.04597, 234–241. doi: 10.48550/arXiv.1505.04597

Rubinstein, R. Y., and Kroese, D. P. (2004). *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning* Vol. 133 (New York: Springer).

Saleh, A., Laradji, I. H., Konovalov, D. A., Bradley, M., Vazquez, D., and Sheaves, M. (2020). A realistic fish-habitat dataset to evaluate algorithms for underwater visual analysis. *Sci. Rep.* 10, 1–10. doi: 10.1038/s41598-020-71639-x

Shi, W., Caballero, J., Huszar, F., Totz, J., Aitken, A. P., Bishop, R., et al. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. *Comput. Vision Pattern Recognit.* abs/1609.05158, 1874–1883. doi: 10.1109/CVPR.2016.207

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *Comput. Vision Pattern Recognit.*

Sun, Y., Chen, J., Liu, Q., and Liu, G. (2020). Learning image compressed sensing with sub-pixel convolutional generative adversarial network. *Pattern Recognit.* 98, 107051. doi: 10.1016/j.patcog.2019.107051

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2014). Going deeper with convolutions. *arXiv: Comput. Vision Pattern Recognit.* abs/1409.4842, 1–9. doi: 10.48550/arXiv.1409.4842

Xu, G., Xie, W., Dong, C., and Gao, X. (2021). Application of three deep learning schemes into oceanic eddy detection. *Front. Mar. Sci.* 8, 715. doi: 10.3389/fmars.2021.672334

Zhang, W., Wu, C., and Bao, Z. (2021). Dpanet: Dual pooling-aggregated attention network for fish segmentation. *Iet Comput. Vision*. 16, 67–82. doi: 10.1049/cvi2.12065

Zhang, X., Zhou, X., Lin, M., and Sun, J. (2017). Shufflenet: An extremely efficient convolutional neural network for mobile devices. *Comput. Vision Pattern Recognit*. abs/1707.01083, 6848–6856. doi: 10.48550/arXiv.1707.01083

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). Pyramid scene parsing network. *Comput. Vision Pattern Recognit.* abs/1612.01105, 6230–6239. doi: 10.1109/CVPR.2017.660

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 30 November 2022
DOI 10.3389/fmars.2022.1071618

Check for updates

OPEN ACCESS

EDITED BY
Hong Song,
Zhejiang University, China

REVIEWED BY
Zhuang Zhou,
Beijing Institute of Technology, Zhuhai, China
Peng Ren,
China University of Petroleum (East China), China

*CORRESPONDENCE
Chang Liu
liuchang@bistu.edu.cn

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 16 October 2022
ACCEPTED 10 November 2022
PUBLISHED 30 November 2022

CITATION
Hao Z, Qiu J, Zhang H, Ren G and Liu C (2022) UMOTMA: Underwater multiple object tracking with memory aggregation. *Front. Mar. Sci.* 9:1071618. doi: 10.3389/fmars.2022.1071618

COPYRIGHT
© 2022 Hao, Qiu, Zhang, Ren and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

UMOTMA: Underwater multiple object tracking with memory aggregation

Zhicheng Hao[1], Jun Qiu[1], Haimiao Zhang[1], Guangbo Ren[2] and Chang Liu[1]*

[1]Institute of Applied Mathematics, Beijing Information Science and Technology University, Beijing, China, [2]First Institute of Oceanography, Ministry of Natural Resources, Qingdao, Shandong, China

Underwater multi-object tracking (UMOT) is an important technology in marine animal ethology. It is affected by complex factors such as scattering, background interference, and occlusion, which makes it a challenging computer vision task. As a result, the stable continuation of trajectories among different targets has been the key to the tracking performance of UMOT tasks. To solve such challenges, we propose an underwater multi-object tracking algorithm based on memory aggregation (UMOTMA) to effectively associate multiple frames with targets. First, we propose a long short-term memory (LSTM)-based memory aggregation module (LSMAM) to enhance memory utilization between multiple frames. Next, LSMAM embeds LSTM into the transformer structure to save and aggregate features between multiple frames. Then, an underwater image enhancement module M_E is introduced to process the original underwater images, which improves the quality and visibility of the underwater images so that the model can extract better features from the images. Finally, LSMAM and M_E are integrated with a backbone network to implement the entire algorithm framework, which can fully utilize the historical information of the tracked targets. Experiments on the UMOT datasets and the underwater fish school datasets show that UMOTMA generally outperforms existing models and can maintain the stability of the target trajectory while ensuring high-quality detection. The code is available *via* Github.

KEYWORDS

artificial intelligence, underwater multiple object tracking, marine environment, long-short term memory, vision transformer

1 Introduction

Multi-object tracking (MOT) is an important research topic in computer vision and is the basis of many high-level visual semantic understanding tasks. Its primary purpose is to locate multiple targets in image sequences and simultaneously track their trajectories over time. Thus, the same target has unique identification information in the image

sequence. In recent years, MOT has made considerable progress and can be found everywhere in our lives, including autonomous driving (Grigorescu et al., 2020), robot navigation (Luo et al., 2021) and video surveillance (Sreenu and Durai, 2019).

With the powerful discriminative ability of deep neural networks and the huge amount of available training data, the performance of target detection algorithms has been dramatically improved, which makes tracking-by-detection (TBD) a mainstream paradigm in MOT (Wojke et al., 2017). This method decomposes MOT into two subtasks: (1) detection, which uses a detection network to obtain bounding boxes of multiple targets in a single frame, and (2) association, which matches the detected targets with existing trajectories through an association network. However, MOT methods of the TBD paradigm still have some problems. First, the quality of the tracking results depends mainly on the detection results, which weakens the role of the association link. Second, since the whole tracking process is divided into two parts, the tracking speed of the algorithm is generally slow, which makes it challenging to meet the requirements of many application scenarios for real-time performance.

With the maturity of multitask learning methods, the joint detection and tracking (JDT) paradigm has started to attract more attention (Wang et al., 2020; Wu et al., 2021). The JDT paradigm optimizes the two abovementioned subtasks simultaneously over a backbone network. The network can output the results of object detection and the apparent features of the object corresponding to each pixel in the feature map for tracking simultaneously. This method greatly accelerates the speed of MOT, and the frames per second (FPS) can reach the real-time requirement during online tracking. However, compared with the TBD model, the tracking performance of the JDT model is not satisfactory. The feature extractor of JDT is prone to ignore the inherent variability of target localization information and identifying information in sharing embedding learning. As a result, the tracking process of JDT copes with different scale targets and occlusion situations poorly.

Although the above-mentioned MOT methods have been greatly developed, they are all designed based on pedestrian datasets and still face challenges when applied to underwater MOT (UMOT) (Panetta et al., 2021). First, the background is constant for most pedestrian MOT datasets. However, when underwater tracking is performed, the background environment can easily change drastically due to light interference or the movement of water creatures. Second, when pedestrian MOT is performed, the characteristics of most of the tracked targets are more pronounced, and *a priori* knowledge can improve the tracking accuracy. Third, when the tracking scene is switched to underwater, it is difficult for these methods to achieve good results due to the complexity of the motion of underwater creatures and the ambiguity of their characteristics.

A new memory aggregation module is propose to enhance the ability of track algorithms to correlate objects between frames in a complex underwater environment and reduce false tracking and missed tracking due to underwater environment changes. Considering both a convolutional neural network (CNN) (Liu et al., 2022b) and a transformer, improvement of the receptive field is comprehensive with deeper model depth. In contrast, for time series data, the gain from the vast improvement of the receptive field is limited. Meanwhile, more and more experiments (Tolstikhin et al., 2021) have proved that the self-attention layer does not seem to be the reason for the excellent results of the transformer. Therefore, we propose a long short-term memory (LSTM)-based memory aggregation module for historical memory fusion, named LSMAM. The overall structure of the LSMAM module follows the transformer architecture, but an LSTM-based layer replaces the multi-head attention layer. For the characteristics of complex changes in underwater scenes, we use a bi-directional long short-term memory 2D (BiLSTM2D) network Tatsunami and Taki, 2022) as a replacement. This network structure can reduce the length of sequences and produce spatially meaningful sensory fields. In short, this paper proposes an underwater MOT algorithm based on the LSMAM called **UMOTMA**. Our model is an online end-to-end tracking network with a transformer encoding and decoding structure in the main framework. The model integrates the memory module into the tracking process to fully use the location and temporal information contained in the target history information. We introduce this model in detail in Section 3.

The contributions of our proposed UMOTMA can be summarized as follows:

- We propose LSMAM, a BiLSTM2D-based memory aggregation module that is expected to improve the correlation between multiple video frames. The module's architecture follows that of a transformer, replacing the multi-headed attention layer with BiLSTM2D to enhance the modelâ€™s ability to build long temporal sequences.
- To address the problems of blurred underwater scenes and drastic environmental changes, a new end-to-end underwater MOT algorithm called UMOTMA is proposed, which integrates the LSMAM into the tracking process to improve the stability and continuity of target trajectories. The main framework of UMOTMA adopts the transformer encoder-decoder structure. In addition, it integrates an underwater image enhancement module and a memory module into the tracking process to enhance the tracking capability of the model in the complex underwater environment.
- Extensive experiments demonstrate that our method effectively improves the performance of underwater MOT, and ablation experiments show that the memory aggregation module proposed in this paper effectively improves the tracking accuracy of the

algorithm. In addition, comparative experiments are also conducted on the MOT17 datasets to demonstrate the performance and generalization ability of the method in different scenarios.

The remainder of this paper is organized as follows. Section 2 reviews the related work, and Section 3 describes the proposed method. Section 4 presents the experimental results. Section 5 discusses the results, and Section 6 draws the conclusions.

2 Relate work

2.1 Multi-object tracking

Existing MOT work is divided into two main categories: the first is the TBD paradigm, which divides MOT into two separate tasks, i.e., detection and association. The object bounding boxes are first predicted by high-performance detectors in a video frame, and then the appearance and motion features of the target are extracted by a feature extraction module; these features are then used to perform similarity value calculations. In data association, the targets are divided into different groups, and the association problem is solved by a matching algorithm to maintain the maximum global similarity while requiring the targets to achieve the one-to-one association constraint. In 2016, (Bewley et al., 2016) proposed Simple Online and Realtime Tracking (SORT), a simple algorithm framework with a fast operation speed, which has attracted widespread attention since its introduction. However, the algorithm has poor resistance to occlusion and cannot perform longer-term stable tracking. After that, (Wojke et al., 2017) further proposed Deep SORT, which uses a more reliable association metric and association method based on SORT. It can effectively track for a long time and largely reduce identity transformation in the tracking process.

Due to the rapid update of detection algorithms, more and more methods are beginning to utilize powerful detectors to obtain higher tracking performance. The You Only Look Once (YOLO) series of algorithms (Redmon et al., 2016; Redmon and Farhadi, 2018; Wang et al., 2022) has become the most popular detector because of its simplicity, efficiency, and ease of deployment. These detectors have also been adopted in a large number of methods (Chu et al., 2021; Zhang et al., 2021a; Liang et al., 2022). Most of these methods use the detection results from a single image directly for tracking.

The second class, JDT, integrates the detection and tracking modules into a single network for multitask learning to accomplish target detection and tracking simultaneously. (Wang et al., 2020) proposed the Joint Detection and Embedding (JDE) module, which utilizes DarkNet's YOLOv3 framework by adding a re-identification (ReID) branch parallel to the detection branch. The feature vector of the center point of the positive anchor frame is extracted as the apparent feature vector of the target in the feature map output from this branch. (Zhang et al., 2021b) proposed FairMOT, which is based on JDE and which chooses to perform feature extraction at the center of the estimated object. This avoids the problem that the features extracted in a coarse anchor frame may not be aligned with the center of the target, and effectively improves the performance of the tracking algorithm.

2.2 Vision transformer-based MOT

In recent years, vision transformers have been successfully applied to image recognition and video analysis with good results; as such, many works have sought to use apply them to MOT. TrackFormer (Meinhardt et al., 2022) and MOTR (Zeng et al., 2021) input the image into a CNN backbone network first to extract features and then input the extracted features into a transformer encoder. Finally, the output of the encoder and an autoregressive tracking query are used as input to the transformer decoder to perform object detection and association simultaneously. TransCenter (Xu et al., 2021) and Transtrack (Sun et al., 2020) only use transformers as a feature extractor. Their overall structure is based on encoding–decoding to pass the tracking features and learn the aggregated embedding of each object. MeMOT (Cai et al., 2022) was designed as an online tracking algorithm that performs object detection and data association under a common framework. It is capable of linking objects after a long time span, which is realized by storing the identity embeddings of the tracked objects in a large spatiotemporal memory, and by adaptively referencing and aggregating useful information from the memory as needed. Global Tracking Transformers (GTR) (Zhou et al., 2022)global is a global MOT network structure based on transformers, which uses them to encode all target features in the input video sequence and assigns the targets to different trajectories using trajectory queries.

The above works explored the different mechanisms of representing target states as dynamic embeddings. However, compared to CNN models, transformers are not yet sufficiently mature for modeling long-term spatio-temporal observations and adaptive feature aggregation.

2.3 Underwater image enhancement

Underwater image enhancement aims to improve the quality and visibility of underwater images and facilitate the acquisition of more information from the images. Contrast limited adaptive histogram equalization (CLAHE) (Reza et al., 2004) is a traditional and fast method for image enhancement. However, this method may suffer from color distortions. Fusion methods (Ancuti et al., 2012) are another classical approach for image enhancement that consider multiple enhancement techniques to

improve the quality of underwater images. In recent years, deep learning has developed rapidly and many scholars have explored the use of neural network models to improve underwater image enhancement. To avoid the requirement of paired training data, (Zhu et al., 2017) proposed the weakly supervised underwater color correction network UCycleGAN. Based on this, (Fabbri et al., 2018) proposed the underwater generative adversarial network with gradient penalty (UGAN-GP) to deal with the underwater color distortion problem, which uses UCycleGAN to generate a paired data sets for supervised training and combines the Wasserstein GAN-GP loss function to avoid model collapse. (Li et al., 2021)underwater proposed the underwater image enhancement network Ucolor, which uses medium transmission-guided multicolor space embedding. This network enriches the diversity of feature representations by incorporating features from different color spaces into a unified structure. It achieved excellent performance in experiments in various environments.

3 Materials and methods

Given a sequence of video frames $I = \{I^0, I^1, I^2, \cdots, I^T\}$, suppose there are M trajectories in frame $t-1$ and N detection targets in frame t. The goal of MOT is to complete matching between trajectories and detection targets by constructing the associated information between them and finally getting the trajectory of each detection target in the current frame. In this paper, we propose an end-to-end tracking algorithm to learn target detection and association jointly, called UMOTMA. The overall structure of the model is shown in Figure 1, which contains four main parts: (1) an underwater image enhancement module M_E. We use CLAHE, UGAN-GP, and Ucolor to implement the enhancement module, and compare their effects on tracking through experiments (Section 4). (2) A feature extraction module M_F. We use a transformer-based encoder to extract the features of the input frames. (3) A memory aggregation module LSMAM. For the memory stored in the memory buffer, LSMAM will compress it and produce an aggregated representation. (4) A feature association and update module M_A. The aggregated representation output by LSMAM is stitched with the features output by M_F as the candidate embedding for prediction. The candidate embeddings are updated using the transformer decoder, and then the new objects and tracking objects are predicted based on the updated embeddings to get the final trajectory and detection features. Finally, the history information is updated based on the results obtained from the current frame, and tracking is continued in the next frame.

3.1 Underwater image enhancement module M_E

The underwater image enhancement module M_E takes as input the original image captured by the underwater camera. Since underwater scenes are generally turbid, the main purpose of the enhancement module is to reduce color distortion effects and improve visibility. In order to choose the optimal enhancement algorithm for MOT, this paper uses CLAHE, UGAN-GP, and Ucolor as the enhancement module, where CLAHE adopts the default parameters of OpenCV, and UGAN-GP and Ucolor adopt the network model and parameters provided by the original authors. Details of the implementation are described in Section 4.

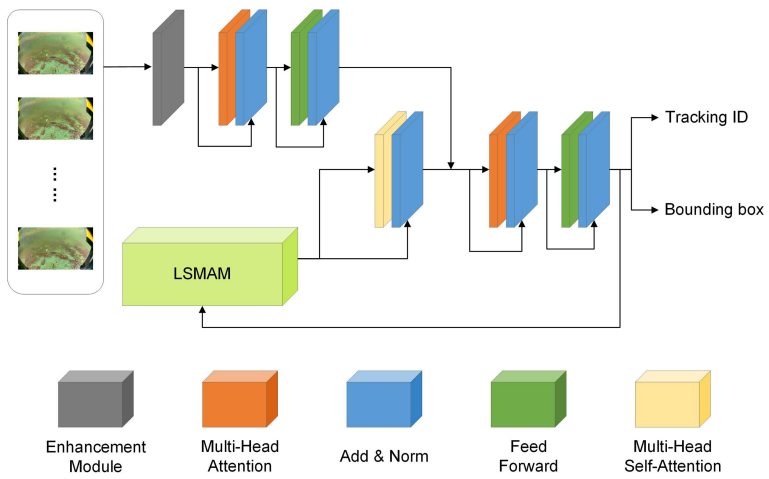


FIGURE 1
The overall structure of the UMOTMA.

3.2 Feature extraction module M_F

The M_F module is built using a transformer-based encoder for the purpose of extracting features from the images. For underwater MOT, the input to M_F is generated by the underwater image enhancement module, M_E, and for the MOT of pedestrians, the original image is fed directly to M_F. The overall framework of M_F is similar to that of MOTR (Zeng et al., 2021), where the feature map is first obtained from the input frame by a CNN, and then the feature map is fed into the transformer-based encoder, which uses the same deformable DETR (Zhu et al., 2020) structure as MOTR, and finally outputs the current frame features.

3.3 LSTM-based memory aggregation module: LSMAM

In order to reduce the length of the module and aggregate as much memory as possible while maintaining efficiency, we propose an LSTM-based memory aggregation module LSMAM, whose structure is shown in Figure 2. The overall structure of LSMAM is based on the transformer structure, in which the self-attentive layer is replaced by an LSTM-based layer called BiLSTM (Graves and Schmidhuber, 2005). In addition, we referred to the literature (Tatsunami and Taki, 2022) to improve the BiLSTM, finally deciding to using a structure similar to the vision permutator (ViP) (Hou et al., 2022), which reduces the length of the sequence, improves the accuracy and efficiency, and produces spatially meaningful sensory fields. BiLSTM consists of two layers that are replaced by combining spatial information with memory-saving memory parameters to reduce the memory cost by mixing the LSTM with memory-saving parameters. The output process is shown as follows

$$\vec{h}_{\text{for}} = \text{LSTM}_{\text{for}}(\vec{x}), \tag{1}$$

$$\overleftarrow{h}_{\text{back}} = \underset{\text{back}}{\text{LSTM}}(\overleftarrow{x}), \tag{2}$$

$$\vec{h}_{\text{back}} = \text{rearrange}\,(\overleftarrow{h}_{\text{back}}), \tag{3}$$

$$h = \text{concat}\,(\vec{h}_{\text{for}}, \vec{h}_{\text{back}})\,. \tag{4}$$

where $\vec{x}$ represents the input sequence, $\overleftarrow{x}$ represents the corresponding reverse sequence, and $\vec{h}_{\text{for}}$ and $\overleftarrow{h}_{\text{back}}$ are the outputs obtained by processing $\vec{x}$ and $\overleftarrow{x}$ the corresponding LSTMs, respectively. Here $\overleftarrow{h}_{\text{back}}$, are the outputs $\vec{h}_{\text{back}}$ rearranged in the original order, so $\vec{h}_{\text{back}}$ and $\vec{h}_{\text{for}}$ are oriented in the same direction, and finally the two are spliced to obtain h.

To parallelize the vertical and horizontal axes, LSMAM introduces two BiLSTMs for parallel processing in the left/right and top/bottom directions, named the horizontal BiLSTM and the vertical BiLSTM, respectively. For input $X \in R^{H\times W\times C}$, H represents the number of sequences in the vertical direction, W represents the number of sequences in the horizontal direction, and C is the channel's dimension. All sequences in the horizontal direction $X_w \in R^{H\times C}, w=0,1,2,\cdots W$ are input into the vertical BiLSTM, sharing the weights and hidden dimension D, and finally the output in the horizontal direction is obtained.

Similarly, all the sequences in the vertical direction $X_h \in R^{W\times C}, h=0,1,2,\cdots H$ are input into the horizontal BiLSTM to obtain the outputs. These processes are formulated as follows

$$\text{Output}_{hor} = \text{BiLSTM}\,(X_w), w = 0, 1, 2, \cdots, W, \tag{5}$$

$$\text{Output}_{ver} = \text{BiLSTM}\,(X_h), h = 0, 1, 2, \cdots, H\,. \tag{6}$$

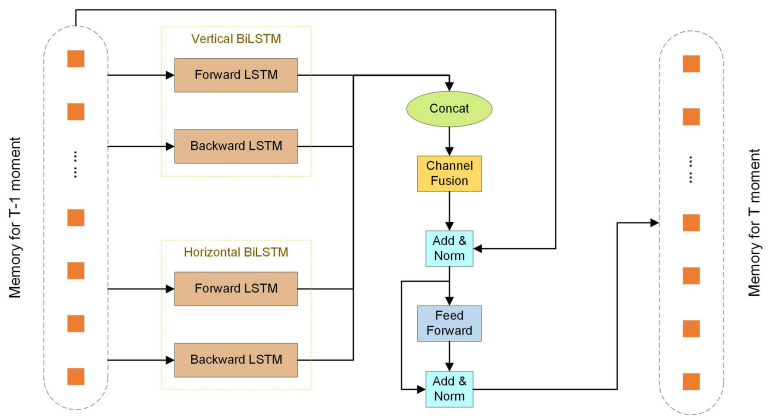


FIGURE 2
The memory aggregation module LSMAM.

We combine the horizontal and vertical results separately to obtain O^{ver} and O^{hor}, and then concatenated O^{ver} and O^{hor} to obtain the final result O

$$O = \text{concat}(O^{ver}, O^{hor}) \quad (7)$$

Note that O^{ver} and O^{hor} have the same hidden dimension $R^{W\times H\times 2C}$, which is determined by the hyperparameter of BiLSTM. Accordingly, vector O has dimensions of $R^{W\times H\times 4C}$.

3.4 Feature association and update module M_A

The overall structure of the M_A module is shown in Figure 3. It consists of multiple stacked transformer decoders, which take the image feature extracted by M_F and the output of LSMAM as the common input, where the output of LSMAM is used as the query of the decoder and the output of M_F is used as the key and value of the decoder. The decoding process produces the tracking result which contains two parts: bounding box prediction and trajectory ID prediction. For the initial frame, we generate a blank embedding to be used as historical information for feature association. In addition, to align the dimensionality of the final output, we pad each output embedding so that they can be fed to the memory aggregation module with the same dimensionality.

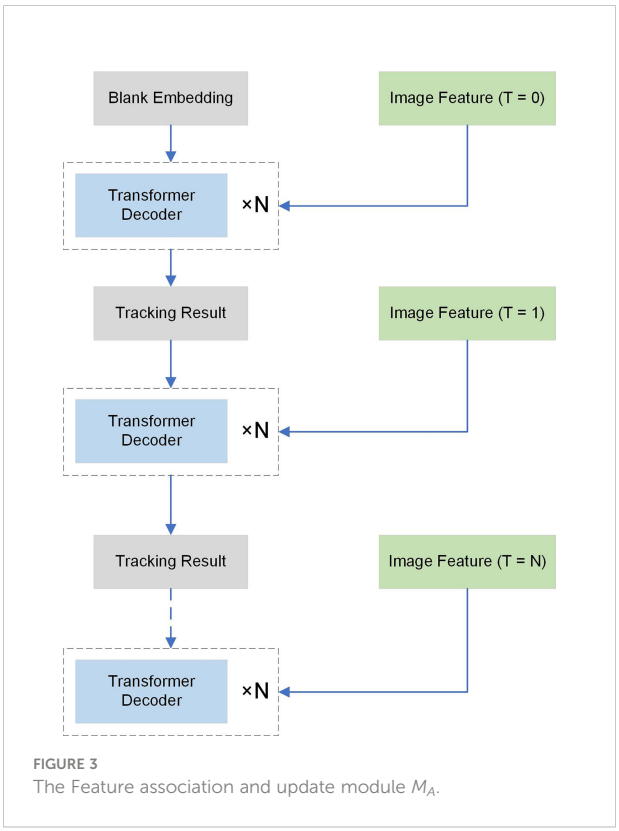


FIGURE 3
The Feature association and update module M_A.

4 Results

4.1 Datasets

To evaluate the MOT performance of UMOTMA fully, we evaluated our model on three benchmark datasets: the UMOT datasets (Zhang et al., 2020), the underwater fish school datasets (Liu et al., 2022a) and MOT17 (Milan et al., 2016). MOT17 is a representative datasets of MOT Challenge, which contains seven training subsets and seven validation subsets. All of the data are collected from the real world and labeled. The the underwater fish school datasets is a recent datasets that the images were all extracted from the observation video of a marine pasture over one year. The UMOT datasets contains four parts, which correspond to the original data set and the data set processed with CLAHE, UGAN-GP, and Fusion. Considering the size of the overall datasets, we chose to remove the Fusion data set and add the Ucolor data set instead.

4.2 Evaluation metrics

We used the same evaluation metrics as MOT Challenge to evaluate our model, where the specific metrics used include the high-order tracking accuracy (HOTA), MOT accuracy (MOTA), identity switching (IDs), recognition score (IDF1), false positives (FP), and false negatives (FN). Among them, MOTA is the most widely used metric and can closely represent human visual assessment; a better MOTA indicates that the proposed method has the ability to balance various factors. HOTA comprehensively evaluates the performance of detection and data association. IDF1 focuses more on association performance, where a higher IDF1 score indicates that the images of an object are mostly mapped to the same identity. FP and FN are defined, respectively, as the number of incorrect targets and the number of missed correct targets. It should be noted that since HOTA is a recently proposed evaluation metric, some authors have not provided this metric during their comparisons.

4.3 Implementation details

We implemented our proposed method in PyTorch 1.11. Our model was trained from scratch with a computer running Ubuntu 20.04 LTS. The entire training process was deployed on two NVIDIA RTX 3090 GPUs with memory of 48 GB. The gradient optimization method was AdamW with batch size of 12. All learning rates were initialized to 2×10^{-4} and decreased to 2×10^{-5} during the training epochs. The GFLOPs of the model is 53.1×10^{6} and the Parameters is 2×10^{11}. The model was initially trained on the UMOT datasets and the underwater fish school datasets with 100 epochs, and then fine-tuned using 60 epochs. It

took 48 hours in total. The initial training on MOT17 consisted of 60 epochs, and fine-tuning used 40 epochs. The total training time was about 36 hours. The fine-tuning started with an initial learning rate, which decreased after 10 epochs. Depending on the nature of the trajectory tracking, the total number of tracks per frame varied. In order to stack the results of multiple frames into a batch, we complemented each frame result with blank trace results to align the lengths of the trace results of all frames.

4.4 Comparison with state-of-the-art methods on the UMOT datasets

Table 1 shows the results of the method proposed in this paper relative to the other tracking methods, including DeepSORT (Wojke et al., 2017), CenterTrack (Zhou et al., 2020), TrackFormer (Meinhardt et al., 2022), GSDT (Wang et al., 2021), and MOTR (Zeng et al., 2021) on the UMOT datasets. Since some of these methods have not been previously applied to the UMOT datasets; so, to be fair, we re-implemented these methods on the experimental equipment and obtained their results for the UMOT datasets. Each metric has an arrow next to it, with "↑" indicating that the higher the metric is, the better, and "↓" indicating that the lower it is, the better.

As can be seen in Table 1, UMOTMA achieves excellent results of 52.3% and 61.1% in terms of the MOTA and IDF1 metrics for the UMOT datasets for underwater scenes, 8.2% and 7.7% better than the second-placed method, respectively. In addition to MOTA and IDF1, other metrics applied to UMOTMA also reflect some improvements relative to the other methods. However, our proposed method does not perform the best in terms of FP and IDs, probably because the extracted features of the false detection targets are very similar to those of the correct targets, which in turn lead to false detections and incorrect associations. The excellent results of MOTA and IDF1 show that our model has good tracking performance and can maintain a stable continuation of the tracking trajectory. This is mainly due to our incorporation of the memory aggregation module, which enables feature extraction of historical information through aggregation of past frame tracking results and improves the accuracy of the associated trajectories with current frame targets. Figure 4 shows some of the results of our underwater tracking.

4.5 Comparison with state-of-the-art methods on the underwater fish school datasets

Recently, the underwater fish school datasets is introduced to provides a better choice to verify the underwater multi-object tracking performance. We further conduct the experiments on the underwater fish school datasets and perform the performance comparison with state-of-the-art methods in Table 2. It shows that UMOTMA achieves much better performance on the underwater fish school datasets. Our method gets a much higher MOTA score, surpassing JDE by 2.3%. For the IDF1 metric, our method also achieves much better performance than JDE (81.1% vs. 72.2%). While for the IDs metric, UMOTMA is inferior to some state-of-the-art methods. It means that UMOTMA performs well on temporal motion learning while the tracking performance is not that stable. The large improvements on MOTA are mainly from the memory aggregation network.

4.6 Ablation studies

4.6.1 UMOTMA components

In this subsection, we research the effectiveness of the different components in UMOTMA, including the underwater enhancement module (M_E), and the memory aggregation module (LSMAM). The experiment results are shown in Table 3. Baseline represents only using the feature extraction module (M_F), and association module (M_A), without using $M_E$and LSMAM, whose results are relatively poor. Baseline + $M_E$means the underwater image enhancement module has been added to the baseline model. The underwater image enhancement method can improve tracking performance by increasing image visibility, so the tracking effect is effectively improved by adding M_E, as

TABLE 1 Comparison of the methods on the UMOT test set.

Methods	IDF1↑	MOTA↑	IDs↓	FP↓	FN↓
DeepSORT	44.7	26.2	18	**312**	2837
GSDT	52.2	39.4	30	430	3045
CenterTrack	50.6	42.4	**15**	914	3101
TrackFormer	53.4	44.1	34	583	3294
MOTR	52.7	42.6	18	483	3516
UMOTMA	**61.1**	**52.3**	28	499	**2809**

In each column, the best result is in bold.

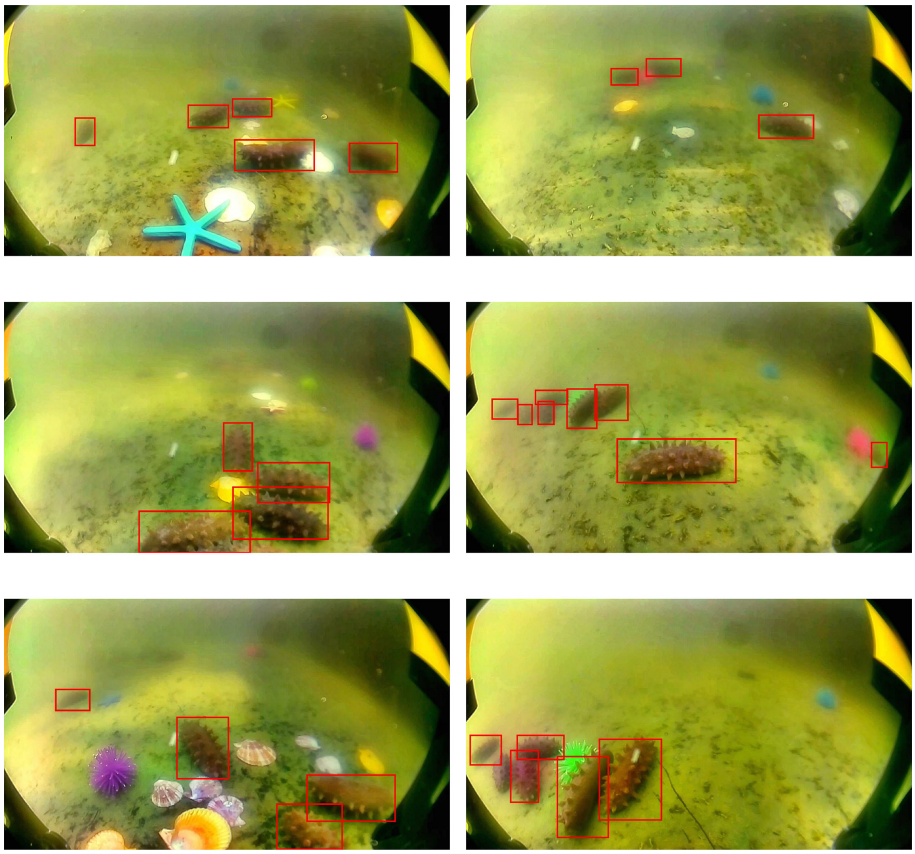


FIGURE 4
Some examples of tracking results produced by our proposed UMOTMA.

indicated by the MOTA index, which has increased by 2.2%. However, due to the increased image visibility, the information captured by the model during feature extraction increases substantially, resulting in a surge of IDs during the association process, which affects the tracking stability of the model. UMOTMA indicates the further addition of the LSMAM to Baseline + M_E. In terms of metrics, the addition of LSMAM increases the MOTA metric by 3.8%, IDF1 by 5.9%, and ID by a factor of 7. The reason it achieves such good results is that LSMAM allows the model to use more historical information to match trajectories with current frame features, effectively reducing the fluctuations of tracking trajectories due to ID, reducing the IDs to a certain extent, and improving the tracking effect.

TABLE 2 Comparison of the methods on the underwater fish school test set.

Methods	IDF1↑	MOTA↑ ‘	IDs↓
SORT	75.4	71.6	342
DeepSORT	77.4	75.4	301
JDE	72.2	76.1	453
Bytetrack	79.3	77.6	**249**
UMOTMA	**81.1**	**78.4**	277

In each column, the best result is in bold.

4.6.2 The influence of different LSTMs on tracking

LSTM is a classical neural network, and there are many variants based on it. To investigate the effect of different LSTMs on the model tracking effect, we used BiLSTM and BiLSTM2D to replace LSTM in the LSMAM module separately. The results of the ablation experiments are shown in Table 4. Compared with the baseline LSTM only, the tracking accuracy of the network with both the BiLSTM and BiLSTM2D structures has significantly improved. The MOTA metric increases by 4.8% and 7.0%, respectively, and the IDF1 metric increases by 3.1% and 8.7%, respectively. The FP and FN metrics also improve to different degrees. These results show that the memory module

TABLE 3 Ablation study on UMOT datasets.

Methods	IDF1↑	MOTA↑	IDs↓	FP↓	FN↓
Baseline	54.7	46.3	**10**	577	3170
Baseline + M_E	55.2	48.5	35	532	3035
UMOTMA	**61.1**	**52.3**	28	**499**	**2809**

In each column, the best result is in bold.

with BiLSTM2D has a significant improvement on the tracking accuracy of the model, and by using a bi-directional 2D structure, the module can perform better aggregation of the information contained in the time series data.

4.6.3 Comparison of underwater image enhancement modules

Underwater image enhancement is a fundamental task in the field of computer vision, and many excellent works have subsequently emerged (Reza et al., 2004; Ancuti et al., 2012; Fabbri et al., 2018; Li et al., 2021). In order to study the enhancement effect of enhancement methods for underwater MOT, we used different algorithms to build enhancement modules and evaluate the amount of enhancement imparted by the different modules *via* experiments. The experimental results are shown in Table 5. From Table 5, we can see that all three image enhancement algorithms have a certain degree of improvement in terms of the MOTA metric, but the IDs metric indicates there are different degrees of degradation, among which UGAN-GP causes the most serious amount of degradation. Figure 5 shows the images produced by the different enhancement modules. Through a comparison of the images, we can find that the image contrast improvement brought by the UGAN-GP algorithm is the highest, which directly changes the color of the image background, thus causing a significant decrease in the ID index. The Ucolor processed image is more consistent with the original image in terms of color, and therefore produced the best IDs metric value.

4.7 The performance on pedestrian datasets

To evaluate the tracking capability of our method UMOTMA in different scenarios, we conducted experiments on the MOT17 datasets and compared the results with those obtained with the other tracking methods. Table 6 lists the metric results of our proposed method UMOTMA against those of the other state-of-the-art MOT methods. It can be seen that UMOTMA obtains competitive tracking accuracy on the MOT17 datasets, achieving the best results for the HOTA, MOTA, IDF1, and FN metrics, and the second-best IDs metric value.

In general, the experimental results for the datasets in the two scenarios show that the method proposed in this paper has obvious advantages in terms of comprehensive performance and tracking accuracy, and performs well in different scenarios. In particular, the results achieved by UMOTMA in the underwater scenario are significantly higher than those of the other existing methods.

5 Discussion

The main purpose of MOT is to assign IDs to detected targets and keep the IDs of the same targets unchanged in the subsequent frames. Most previous works were performed based on pedestrian datasets. However, for underwater scenes, unfavorable conditions such as occlusion, background interference, and motion blurring appear more frequently, so it becomes extremely difficult to keep the tracking stable when performing MOT in underwater environments. In this paper, we propose a new end-to-end MOT algorithm called UMOTMA. The main advantage of UMOTMA over other methods is the introduction of a depth LSTM-based memory aggregation module (LSMAM), which enhances the model in terms of correlation features by fully aggregating the information contained in past frames to maintain stable tracking in complex environments. The experimental results on the UMOT datasets and the underwater fish school datasets showed that our proposed UMOTMA has

TABLE 4 Ablation study about different LSTM.

Methods	IDF1↑	MOTA↑	IDs↓	FP↓	FN↓
LSMAM(LSTM)	52.4	45.3	**19**	813	2999
LSMAM(BiLSTM)	55.1	49.3	26	**485**	3078
LSMAM(BiLSTM2D)	**61.1**	**52.3**	28	499	**2809**

In each column, the best result is in bold.

TABLE 5 Comparison of the different enhancement module.

Enhancement	IDF1↑	MOTA↑	IDs↓	FP↓	FN↓
Origin	51.9	46.7	**12**	484	3011
CLAHE	55.5	49.1	28	482	3060
UGAN-GP	48.3	49.6	34	**312**	2837
Ucolor	**61.1**	**52.3**	28	499	**2809**

In each column, the best result is in bold.

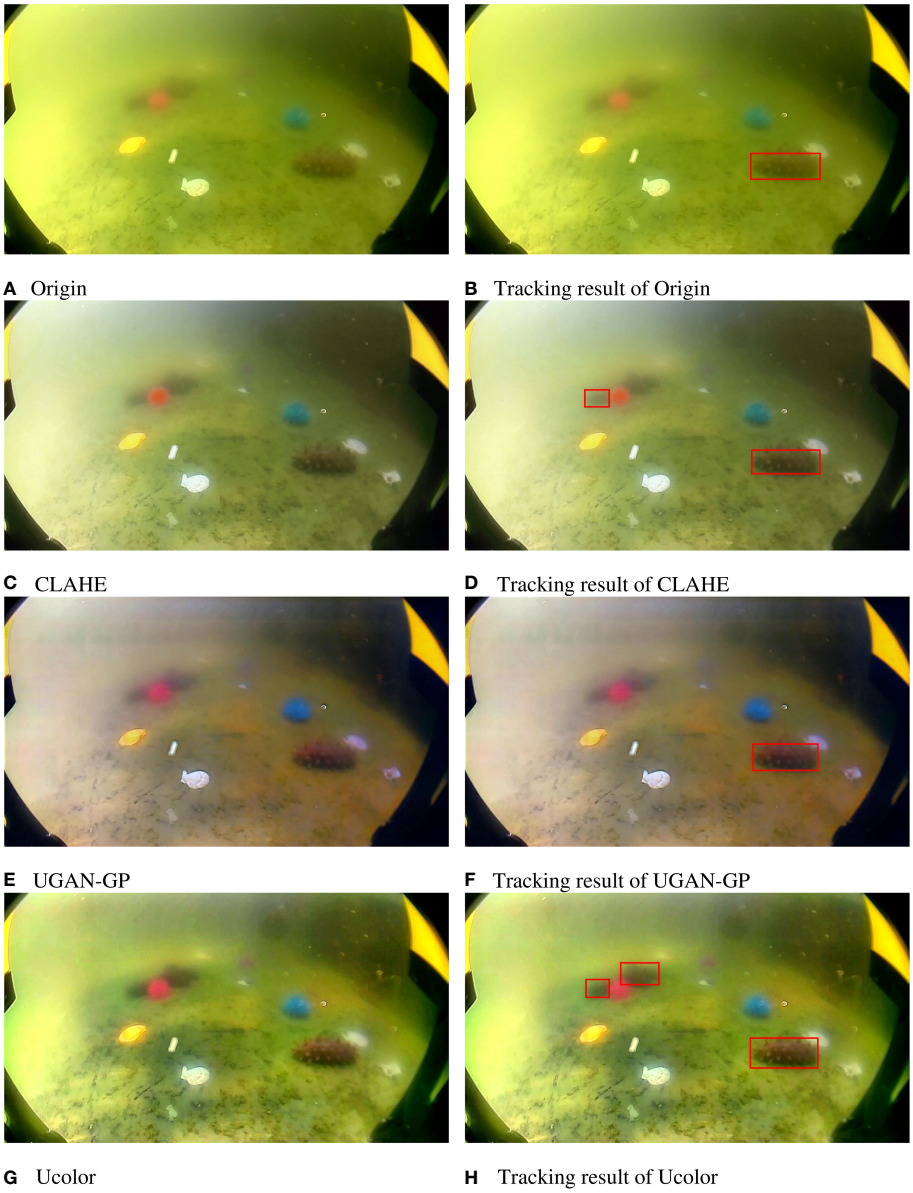


FIGURE 5
Visualization of the results generated by the different enhancement modules.

TABLE 6 Comparison of the state-of-the-art methods on the MOT17 test set.

Methods	HOTA↑	IDF1↑	MOTA↑	IDs↓	FP↓
DeepSORT	–	61.2	60.3	2442	36111
CenterTrack	52.2	64.7	67.8	3039	**18498**
TraDeS	52.7	63.9	69.1	3555	20892
GSDT	55.5	68.7	66.2	3318	26339
TrackFormer	–	63.9	65.0	3528	70443
MOTR	57.2	68.4	71.9	**2115**	32355
UMOTMA	**57.6**	**68.8**	**72.3**	2436	37149

In each column, the best result is in bold.

excellent performance in underwater MOT, and several evaluation metrics reached optimal performance, which proves the effectiveness of the proposed method.

In order to explore the roles that LSMAM performs in the tracking process further, we conducted extensive ablation experiments. The results are shown in Table 3. After adding the LSMAM module, the MOTA and IDF1 metrics of the model increased respectively by 3.8% and 5.9%, and the IDs decreased by a factor of 7, indicating that the memory aggregation module proposed in this paper effectively improves the tracking accuracy of the model. However, We also found that the underwater image enhancement module and the memory aggregation module increase the calculation volume of the model, which leads to slow inference speed of the model and makes it difficult to satisfy some underwater scenarios for real-time MOT. Therefore, it will be an important direction of subsequent work to optimize the model inference speed.

6 Conclusions

This paper proposed a novel deep LSTM-based end-to-end underwater MOT model named UMOTMA. We introduced a memory aggregation module to guide the matching association link between past frame trajectories and current frame features. In the memory aggregation module, we use LSTM for memory aggregation instead of a CNN or transformer, as employed in conventional approaches, which effectively improves the algorithmâ€™s utilization of target information in past frames. Experimental results on the UMOT datasets and the underwater fish school datasets showed that our proposed UMOTMA achieves optimal results in terms of several MOT metrics, and it was significantly better than the second-best method. The experimental results for the MOT17 datasets also showed that our method has a tracking accuracy comparable to other state-of-the-art MOT methods for pedestrian scenes. In addition, we conducted an extensive ablation study to demonstrate the contribution of each component of the proposed MOT framework to the tracking process and briefly discussed the impact of different image enhancement modules on MOT in underwater environments.

In general, our proposed method has good generality and can be adapted to both surface and underwater application scenarios. Especially in the latter, UMOTMA shows exceptionally competitive performance. In the future, we will use our model for MOT of critical underwater scenarios and exploration of marine biological activities.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/Zxl19990529/Underwater-Multiple-Object-Tracking-Dataset.

Author contributions

ZH designed the method with experiments and wrote the first draft of the manuscript supervised by CL and HZ. GR anb JQ critically reviewed the initial manuscript and provided helpful input. All authors contributed to the article and approved the submitted version.

Funding

The authors would like to acknowledge the financial support from the National Natural Science Foundation of China (under Grant Nos. 61871042, 61931003, 62171044, 12101061), the Natural Science Foundation of Beijing (Grants No. 4222004), and QinXin Talents Cultivation Program(Beijing Information Science and Technology University), also from the China High-

Resolution Earth Observation System Program under Grant 41-Y30F07-9001-20/22.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ancuti, C., Ancuti, C. O., Haber, T., and Bekaert, P. (2012). "Enhancing underwater images and videos by fusion," in *Proc. conf. comput. vis. pattern recog* (Providence, RI, USA: IEEE), 81–88.

Bewley, A., Ge, Z., Ott, L., Ramos, F., and Upcroft, B. (2016). "Simple online and realtime tracking," in *Proc. int. conf. image process* (Phoenix, AZ, USA: IEEE), 3464–3468.

Cai, J., Xu, M., Li, W., Xiong, Y., Xia, W., Tu, Z., et al. (2022). "Memot: Multi-object tracking with memory," in *Proc. conf. comput. vis. pattern recog* (New Orleans, LA, USA: IEEE), 8090–8100.

Chu, P., Wang, J., You, Q., Ling, H., and Liu, Z. (2021). Transmot: Spatial-temporal graph transformer for multiple object tracking. *arXiv*. doi: 10.48550/arXiv.2104.00194

Fabbri, C., Islam, M. J., and Sattar, J. (2018). "Enhancing underwater imagery using generative adversarial networks," in *Proc. int. conf. robot. autom* (Brisbane, QLD, Australia: IEEE), 7159–7165.

Graves, A., and Schmidhuber, J. (2005). "Framewise phoneme classification with bidirectional lstm networks," in *Proc. int. joint conf. neural netw*, (Montreal, QC, Canada: IEEE) 4, 2047–2052.

Grigorescu, S., Trasnea, B., Cocias, T., and Macesanu, G. (2020). A survey of deep learning techniques for autonomous driving. *J. Field Robotics* 37, 362–386. doi: 10.1002/rob.21918

Hou, Q., Jiang, Z., Yuan, L., Cheng, M.-M., Yan, S., and Feng, J. (2022). Vision permutator: A permutable mlp-like architecture for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1. doi: 10.1109/TPAMI.2022.3145427

Liang, C., Zhang, Z., Zhou, X., Li, B., Zhu, S., and Hu, W. (2022). Rethinking the competition between detection and reid in multiobject tracking. *IEEE Trans. Image Process.* 31, 3182–3196. doi: 10.1109/TIP.2022.3165376

Li, C., Anwar, S., Hou, J., Cong, R., Guo, C., and Ren, W. (2021). Underwater image enhancement *via* medium transmission-guided multi-color space embedding. *IEEE Trans. Image Process.* 30, 4985–5000. doi: 10.1109/TIP.2021.3076367

Liu, T., He, S., Liu, H., Gu, Y., and Li, P. (2022a). A robust underwater multiclass fish-school tracking algorithm. *Remote Sens.* 14, 2072–4292. doi: 10.3390/rs14164106

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., and Xie, S. (2022b). "A convnet for the 2020s," in *Proc. conf. comput. vis. pattern recog* (New Orleans, LA, USA: IEEE), 11976–11986.

Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., and Kim, T.-K. (2021). Multiple object tracking: A literature review. *Artif. Intell.* 293, 103448. doi: 10.1016/j.artint.2020.103448

Meinhardt, T., Kirillov, A., Leal-Taixe, L., and Feichtenhofer, C. (2022). "Trackformer: Multi-object tracking with transformers," in *Proc. conf. comput. vis. pattern recog* (New Orleans, LA, USA: IEEE), 8844–8854.

Milan, A., Leal-Taixé, L., Reid, I., Roth, S., and Schindler, K. (2016). Mot16: A benchmark for multi-object tracking. *arXiv*. doi: 10.48550/arXiv.1603.00831

Panetta, K., Kezebou, L., Oludare, V., and Agaian, S. (2021). Comprehensive underwater object tracking benchmark dataset and underwater image enhancement with gan. *IEEE J. Oceanic Eng.* 47, 59–75. doi: 10.1109/JOE.2021.3086907

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: Unified, real-time object detection," in *Proc. conf. comput. vis. pattern recog* (Las Vegas, NV, USA: IEEE), 779–788.

Redmon, J., and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv*. doi: 10.48550/arXiv.1804.02767

Reza, A. M. (2004). Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *J. VLSI Signal Process. Syst. signal image video Technol.* 38, 35–44. doi: 10.1023/B:VLSI.0000028532.53893.82

Sreenu, G., and Durai, S. (2019). Intelligent video surveillance: a review through deep learning techniques for crowd analysis. *J. Big Data* 6, 1–27. doi: 10.1186/s40537-019-0212-5

Sun, P., Cao, J., Jiang, Y., Zhang, R., Xie, E., Yuan, Z., et al. (2020). Transtrack: Multiple object tracking with transformer. *arXiv*. doi: 10.48550/arXiv.2012.15460

Tatsunami, Y., and Taki, M. (2022). Sequencer: Deep lstm for image classification. *arXiv*. doi: 10.48550/arXiv.2205.01972

Tolstikhin, I. O., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., et al. (2021). "Mlp-mixer: An all-mlp architecture for vision," in *Advances in neural information processing systems*, (Montreal, QC, Canada: Curran Associates, Inc) 34, 24261–24272.

Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M. (2022). Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv*. doi: 10.48550/arXiv:2207.02696

Wang, Y., Kitani, K., and Weng, X. (2021). "Joint object detection and multi-object tracking with graph neural networks," in *Proc. int. conf. robot. autom* (Xi'an, China: IEEE), 13708–13715.

Wang, Z., Zheng, L., Liu, Y., Li, Y., and Wang, S. (2020). "Towards real-time multi-object tracking," in *Proc. lect. notes comput. sci* (Glasgow, UK: Springer), 107–122.

Wojke, N., Bewley, A., and Paulus, D. (2017). "Simple online and realtime tracking with a deep association metric," in *Proc. int. conf. image process* (Beijing, China: IEEE), 3645–3649.

Wu, J., Cao, J., Song, L., Wang, Y., Yang, M., and Yuan, J. (2021). "Track to detect and segment: An online multi-object tracker," in *Proc. conf. comput. vis. pattern recog* (Nashville, TN, USA: IEEE), 12352–12361.

Xu, Y., Ban, Y., Delorme, G., Gan, C., Rus, D., and Alameda-Pineda, X. (2021). Transcenter: Transformers with dense representations for multiple-object tracking. *arXiv*. doi: 10.48550/arXiv.2103.15145

Zeng, F., Dong, B., Zhang, Y., Wang, T., Zhang, X., and Wei, Y. (2021). Motr: End-to-end multiple-object tracking with transformer. *arXiv*. doi: 10.48550/arXiv.2105.03247

Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., et al. (2021a). Bytetrack: Multi-object tracking by associating every detection box. *arXiv*. doi: 10.48550/arXiv.2110.06864

Zhang, Y., Wang, C., Wang, X., Zeng, W., and Liu, W. (2021b). Fairmot: On the fairness of detection and re-identification in multiple object tracking. *Int. J. Comput. Vision* 129, 3069–3087. doi: 10.1007/s11263-021-01513-4

Zhang, X., Zeng, H., Liu, X., Yu, Z., Zheng, H., and Zheng, B. (2020). *In situ* holothurian noncontact counting system: A general framework for holothurian counting. *IEEE Access* 8, 210041–210053. doi: 10.1109/ACCESS.2020.3038643

Zhou, X., Koltun, V., and Krähenbühl, P. (2020). "Tracking objects as points," in *Proc. lect. notes comput. sci* (Glasgow, UK: Springer), 474–490.

Zhou, X., Yin, T., Koltun, V., and Krähenbühl, P. (2022). "Global tracking transformers," in *Proc. conf. comput. vis. pattern recog* (New Orleans, LA, USA: IEEE), 8771–8780.

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. conf. comput. vis. pattern recog* (Venice, Italy: IEEE), 2223–2232.

Zhu, X., Su, W., Lu, L., Li, B., Wang, X., and Dai, J. (2020). Deformable detr: Deformable transformers for end-to-end object detection. *arXiv*. doi: 10.48550/arXiv.2010.04159

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 01 December 2022
DOI 10.3389/fmars.2022.1039898

Check for updates

OPEN ACCESS

EDITED BY
Hong Song,
Zhejiang University, China

REVIEWED BY
Jamil Hussain,
Kyung Hee University, South Korea
Dildar Hussain,
Korea Institute for Advanced Study, South Korea

*CORRESPONDENCE
Renyou Yang
youngrenyou@sina.com

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 08 September 2022
ACCEPTED 03 November 2022
PUBLISHED 01 December 2022

CITATION
Zheng K, Yang R, Li R, Yang L, Qin H and Li Z (2022) A dual stream hierarchical transformer for starvation grading of golden pomfret in marine aquaculture. *Front. Mar. Sci.* 9:1039898. doi: 10.3389/fmars.2022.1039898

COPYRIGHT
© 2022 Zheng, Yang, Li, Yang, Qin and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A dual stream hierarchical transformer for starvation grading of golden pomfret in marine aquaculture

Kaijian Zheng, Renyou Yang*, Rifu Li, Liang Yang, Hao Qin and Ziyun Li

Southern Marine Science and Engineering Guangdong Laboratory (Zhanjiang), Zhanjiang, China

Introduction: Fish starvation grading can provide feeding information for aquaculture, reducing the cost of lures and helping to promote the unmanned and intelligent process of offshore aquaculture.

Methods: In this study, we used golden pompano as the experimental object to address the fish starvation grading problem in the marine culture vessel environment, and proposed the dual stream hierarchical transformer to provide additional temporal information for the starvation grading task, which improved the grading accuracy. We first built a dual stream dataset with both spatial and temporal channel, and divided the fish school starvation status into five levels (very bloated, a little bloated, modest, a little starving, very starving) according to the feeding time and experience. Based on the marine image characteristics, we proposed a dual stream hierarchical transformer with hierarchical convolutional network, composite fusion convolution and transformer.

Results and discussion: We finally evaluated the efficacy of the model based on qualitative and quantitative analyses, revealing that the proposed dual stream hierarchical transformer achieved the state-of-the-art starvation grading performance with a test accuracy of 98.05%. Our model outperformed other mainstream models, including VGG, ResNet, attentionbased model and other fish status grading related model. Field tests on the vessel further suggested that the model can be applied to the mariculture environment of golden pomfret.

KEYWORDS

neural network, transformer, starvation grading, marine image processing, behavior recognition

Introduction

Feeding is an essential step in the aquaculture procedure. The effectiveness of the feeding plan not only has an impact on fish growth directly, but also has influence on costs and environmental contamination when lures are wasted. The cost of lures makes up about 30% of the entire investment in aquaculture (Hossain et al., 2022), and Lin's environmental study report noted that uneaten fish food would lead to a number of environmental issues, including the production of greenhouse gases (Lin and Lin, 2022). According to experimental data (Jothiswaran et al., 2020), using the AI-controlled machine to feed fish and dynamically modify feeding strategies based on fish status can prevent at least 30% of bait waste. In order to minimize economic losses and pollution, it is crucial to optimize fish condition-based feeding techniques to improve bait waste using AI techniques.

Inspired by deep learning used in image recognition tasks, fish researchers have used these methods to analyze fish behavior. There are individual- and school-based identifying and analyzing approaches. Individual-based methods monitor the individual fish in a clear laboratory environment and use the tracked trajectory for identification, such as swimming speed, direction, acceleration, and body curvature (Jonas et al., 2017). As a result, the validity of the tracked trajectory and the accuracy of fish identification are the two key determinants of study for this type of study.

In real mariculture, problems occur when fish were overlap in the image and when the camera could not capture the whole environment. This made it challenging for the trajectory tracking and fish identification. To avoid the requirement to blindly monitor individual fish, the status or behaviors of the fish school may be deduced from the fish visible in the image using the fish school analysis technique. These techniques are known as school-based techniques (Måløya et al., 2019). The following is an overview of school-based approaches to fish behavior or state recognition: The fish school is first mapped in grayscale using the background and foreground segmentation technique (Ye et al., 2016). The fish's overall optical flow images, such as swimming speed and acceleration, are then obtained by using the grayscale map as a foundation (Zhao et al., 2018). Finally, a prediction network is used for recognition (Zhou et al., 2017). About the fish school status grading, a 5-layers CNN was used to classify the intensity of fish school feeding (Zhou et al., 2019). DSRN (Måløya et al., 2019) used 3D residual convolution and LSTM to classify salmon feeding or non-feeding states. On the other hand, DAN-EfficientNet-B2 (Yang et al., 2021) used an attention-based approach to classify the foraging behavior of Oplegnathus punctatus into five categories. DADSN (Zheng et al., 2022) used the combination of CNN and Transformer for the grading of fish starvation.

Convolutional Neural Network (CNN, Lecun et al., 1998) is proposed and then a lot of CNN-based model sprang up. Krizhevsky et al. proposed a five-layer CNN for image classification and won the ILSVRC competition (Krizhevsky et al., 2012). In the following time, CNN became a hot research topic and developed rapidly. Many classical models were created, such as GoogLeNet (Szegedy et al., 2015) and ResNet (He et al., 2016).

Nowadays, numerous experiments have shown that CNN has the following limitations: First, CNN uses local receptive fields for feature capture and thus cannot directly build the global model; Second, weights learned by CNN are stationary at the time of inference, which makes them unadaptable for various inputs (Raghu et al., 2021). Attention models are a popular area of research in current image recognition studies, with flexible modelling capabilities for Regions of Interest (RoI) that can be used for single image or video recognition. In particular, researchers (Fu et al., 2019) used attention models to study the feeding state of fish in a single image, but unfortunately, this approach was only tested in their laboratory dataset and was not extended to real farming.

In this paper, daily data of fish in commercial aquaculture vessels of golden pomfret was collected. Then a Dual Stream Hierarchical Transformer (DSHT) is proposed. First, a Hierarchical Convolutional Network (HCN) was used as the backbone network to extract mariculture image features. Second, we built spatial and temporal channels, called dual stream structure, to make the input data contain richer information. Among them, the spatial features mainly express the distribution, number, and texture of fish; the temporal features mainly represent the movement changes in time series, such as the swimming speed and acceleration of fish. Finally, the spatial and temporal information is fused and filtered using a Composite Fusion Network (CFN). The method achieved 98.05% accuracy and provided valuable feeding control information for intelligent feeding.

Proposed model

Hierarchical convolutional network

Similar to the process of human vision, CNN learn the ability of feature extraction from surface to deep information. The shallow convolutional layers focus on the detailed information of the object within the image, while the deep convolutional layers focus on the abstract semantic information. The features produced form different depth layers require different visual perception fields. Detailed information such as graphic contours requires a smaller field of view to find these details, so small convolutional kernels are more suitable, while abstract semantic information requires larger convolutional kernels to detect. Therefore, we propose a Hierarchical Convolutional Network (HCN), which divides the features of an image into three categories: shallow, medium, and deep, and uses different convolutional kernels for each layer. As shown in Figure 1D.

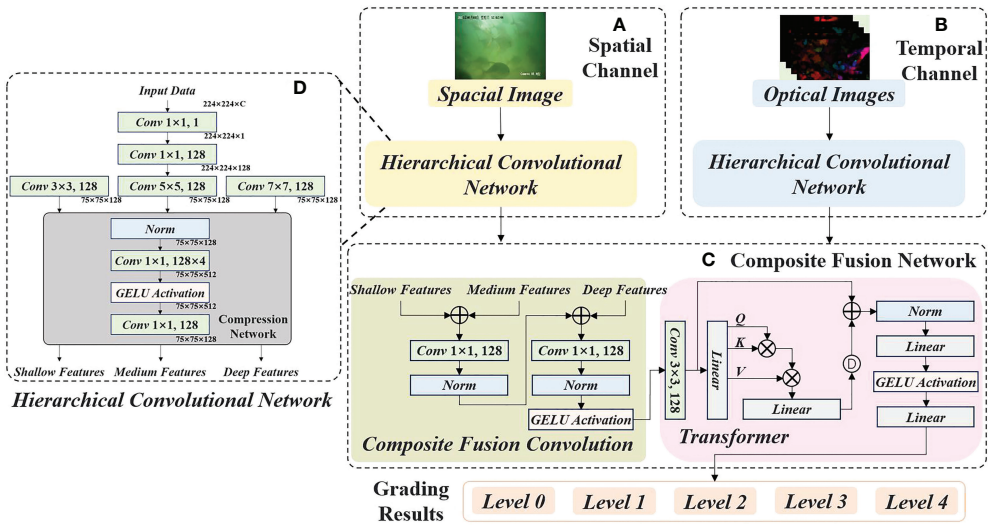


FIGURE 1
The Architecture and Stages of Dual Stream Hierarchical Transformer (DSHT). **(A)** is the spatial channel, and **(D)** is the HCN which is used to process the input and output three hierarchical feature blocks. **(B)** is the temporal channel which uses the similar approach with spatial channel. **(C)** is CFN, which is made up of CFC and Transformer.

The HCN first dimensionally expands the input image. Then three types of CNN with different kernel sizes are used to obtain the hierarchical features. Finally, the obtained features are further refined by a compression network, respectively. In this paper, Gaussian Error Linear Units (GELU, Hendrycks and Gimpel, 2016) is used as the activation function. The formula is expressed as follows:

$$C_i(X) = GELU(W_i * X + b), i = 1,2,3 \quad (1)$$

Where X is the input, C_i is the out, W_i is the weight matrix, b is the bias, the symbol * denotes the convolution operation, and $GELU(x)$ is the activation function:

$$GELU(x) = xP(X \le x) = xN(\mu,\sigma) = 0.5x(1 + \tanh[\sqrt{2/\pi}(x + 0.044715x^3)]) \quad (2)$$

where $N(\mu\sigma)$ is the positive-terrestrial distribution.

Dual stream structure

If only one image is used as the input, the neural network will lack the temporal information, which are pretty importance features for fish school behavior, such as speed and direction. Since the difference between two adjacent images is not obvious, there would be a lot of redundant information if we used consecutive images as input. In this paper, optical flow images were used to represent temporal information. The architecture is referred to as a dual stream structure, where the input of the spatial channel is a spatial image and the inputs of the temporal channel are 20 continuous optical flow images. The optical flow image is an image that records the change of light over time, and it uses the pixels change of neighboring spatial images as the basis to calculate the of motion information (Ranjan and Black, 2017). Assume that the pixel value of (x, y) at time t can be represented by function $I(x, y, t)$, the pixel in the frame $t + 1$ is generated by the pixel $I(x, y, t)$ that has moved a distance (dx, dy) in time dt, so we have:

$$I(x, y, t) = I(x + dx, y + dy, t + dt) \quad (3)$$

If the parameters dx, dy, and dt are small enough, the approximation equation can be obtained by neglecting higher-order infinitesimals:

$$I(x + dx, y + dy, t + dt) = I(x, y, t) + I_x u dt + I_y v dt + I_t dt \quad (4)$$

Where I_x is the bias of the pixel value against x, $I_y$is the bias of the pixel value against y, I_t is the bias of the grey value against t, $u = dx/dt$ and $v = dy/dt$ represents the instantaneous velocity of the pixel in the x direction and $y$direction respectively. Combining the Eq.(3) and Eq.(4), we can obtain the constraint equation:

$$I_x u + I_y v + I_t = 0 \quad (5)$$

By solving the above equations for the variables u and v , we can obtain temporal information about the movement of the pixels within the image. In this paper, we adopt the solution proposed by

Bruce (Lucas, 1985). The visualization of the above calculation results is the optical flow diagram dataset mentioned in this paper.

Feature fusion and transformer

In cognitive science, humans selectively focus on a portion of information due to the bottlenecks in information processing. The transformer is an attention model which imitate the process of human visual observation (Maaz et al., 2022). In this paper, the transformer is used to better perceive the important regions of features between the spatial and temporal channels.

As shown in Figure 1C, the Composite Fusion Network (CFN) consists of two parts: the Composite Fusion Convolution (CFC) and the Transformer. First, the HCN outputs of $\boldsymbol{C}_1(\boldsymbol{X})$ (shallow block) and $\mathbf{C_2}(\boldsymbol{X})$ (medium block) are added, then the features are encoded by a convolutional layer. Then they are added with $\boldsymbol{C}_3(\boldsymbol{X})$ (deep block) and encoded by another convolutional layer.

$$\mathbf{CFC(X) = GRLU(W_2 * (W_1 * (C_1(X) + C_2(X)) + C_3(X)))} \quad (6)$$

$\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are the weight matrix of CFC, $\boldsymbol{C}_1(\boldsymbol{X})$, $\boldsymbol{C}_2(\boldsymbol{X})$, and $\boldsymbol{C}_3(\boldsymbol{X})$ are the output of HCN.

The input of Transformer is the outputs of CFC, which are spread into a two-dimensional matrix. In Transformer, two matrices $\boldsymbol{E}$ and $\boldsymbol{E}_{pos}$ are also created, where $\boldsymbol{E}$ denotes the weight of the transformer values, indicating which features need to be focused on. $\boldsymbol{E}_{pos}$ is the position encoding, and used to represent the position relationship between features. The expression formula is as follows:

$$\mathbf{z_0 = [XE_1; XE_2; XE_3] + E_{pos}} \quad (7)$$

Where $\boldsymbol{z}_0$ denotes the initial result. $\boldsymbol{E}_1$, $\boldsymbol{E}_2$ and $\boldsymbol{E}_3$ are the individual dimension of the weight matrix $\boldsymbol{E}$. $\boldsymbol{XE}_1$, $\boldsymbol{XE}_2$, and $\mathbf{XE_3}$ are the intermediate variables $\boldsymbol{Q}$, $\boldsymbol{K}$ and $\boldsymbol{V}$, respectively.

The $\boldsymbol{Q}$, $\boldsymbol{K}$ and $\boldsymbol{V}$ are transposed and multiplied, and the softmax results are summed with the input to obtain the intermediate results for the current layer.

$$\mathbf{z'_l = softmax\left(\alpha(Q \cdot K^T)\right) \cdot V + z_{l-1}, l = 1, \cdots, L} \quad (8)$$

Where $\boldsymbol{z'}_l$ is the attention results. $\boldsymbol{z}_{l-1}$ is the output of the previous attention. $\boldsymbol{K}^T$ is the transpose of the variable $\boldsymbol{K}$, the symbol is matrix multiplication, $\boldsymbol{\alpha}$ is the normalization factor, and $softmax(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{l} e^{z_j}}$ is the activation function.

The $\boldsymbol{z}_l$ is the final output which obtained by normalizing, bottleneck compressing and add with the attention results $\boldsymbol{z'}_l$.

$$\mathbf{z_l = MLP(LN(normz(z'_l), \gamma dim)) + z'_l, l = 1, \cdots, L} \quad (9)$$

Where $\boldsymbol{norm}(\boldsymbol{z'}_l)$ represents normalize the attention results $\boldsymbol{z'}_l$, $\boldsymbol{LN}(\boldsymbol{x})$ is compress its dimension to the $\boldsymbol{\gamma}$times, which is taken as 0.6 in this paper. $\boldsymbol{MLP}(\boldsymbol{x})$ are dropout, activation, and restore features to the original dimension.

Dual stream hierarchical transformer

The Dual Stream Hierarchical Transformer (DSHT) proposed in this paper is shown in Figure 1. Fist, a spatial image is fed into the spatial channel and extracted by HCN shown in Figure 1D to obtain three hierarchical feature blocks (shallow features, medium features, and deep features). Then, the hierarchical feature blocks are fused by CFC mentioned in Figure 1C to produce the spatial features. For the temporal channel Figure 1B, we take 20 consecutive optical flow images as input and process them with another HCN and CFC like the spatial channel to produce the temporal features. Further, connecting the spatial features and temporal features by the transformer shown in Figure 1C to eliminate useless or duplicate features. Finally, the output of the transformer is converted into a 5-dimensional vector by full concatenation, which represents the grading results of the input samples.

Experimental details

Environment and data collection

We equipped an underwater camera (GW-10, Hikvision) on the aquaculture vessel to record the pomfret video of the starvation status (The yellow box in Figure 2B. As shown in Figure 2A, the vessel has five aquaculture cabins, each of which is a 4×4×3 m^3 rectangle. The farming vessel carries out aquaculture operations in the Beihai, Guangxi Province, about 15 nautical miles from the coastline. The suction pump is used to extract seawater for breeding. Four inlets are configured on each of the four sides of a cabin (The red box in Figure 2B). The bottom of a cabin has a drainage port (The blue box in Figure 2B). It is also equipped with the equipment needed to maintain a suitable environment for aquaculture.

In this paper, fingerlings provided by Guangxi Jingong Marine Technology Co., Ltd., Beihai, China. A breeding cabins has 800 fish, each weighing about 500 g. Feeding was carried out daily in the morning (6:00 - 8:00 am), noon (11:00 am - 13:00 pm), and evening (5:00 - 7:00 pm). Feeding time lasts approximately 10 minutes. the weight of the lures is decided according to the weather and environment. The hunger levels were divided into a total of 5 levels, as shown in Table 1.

Dataset creation

According to the aquaculture data recorded by the breeders, we labelled the fish starvation levels from the videos captured by the underwater camera. A total of 84,000 samples were divided, each containing one spatial image and twenty optical flow images. The training set, validation set, and test set were

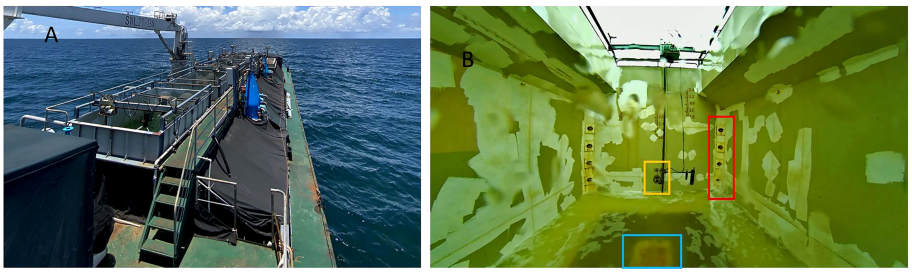


FIGURE 2
(A) is the vessels enviroment. **(B)** is the aquacultural enviroment, and the yellow box, red box and blue box are underwater camera, inlets and outlet, respectively.

divided in the ratio of 6:2:2, and the number of samples in each level was kept consistent. As shown in Table 2.

Metrics

In this paper, three metrics, accuracy, precision, and F1-score, are used to evaluate the model performance. Accuracy indicates whether the model can accurately grade the starvation level of fish. Precision is used to determine how difficult it is to correctly grade samples in that level, and F1-score is used as a comprehensive evaluation index of model performance. It is worth noting that the recall rate is not used in this paper. Since we keep the sample number of each category consistent in order to ensure that training can proceed smoothly. The recall rate and accuracy rate are the same.

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + FN} \times 100\% \quad (10)$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \quad (11)$$

$$\text{F1 score} = \frac{2TP}{2TP + FP + FN} \times 100\% \quad (12)$$

TABLE 1 The Grading Standards.

Hunger Level	Description	Number
Level 0 (very bloated)	Within 1 hour after overfeeding	16,800
Level 1 (a little bloated)	More than 1 hour and less than 2 hours after overfeeding	16,800
Level 2 (modest)	More than 2 hour and less than 3 hours after overfeeding	16,800
Level 3 (a little starving)	More than 3 hour and less than 4 hours after overfeeding	16,800
Level 4 (very starving)	More than 4 hours after overfeeding	16,800

TP is the number of samples labelled with the positive class and correctly classified by the model; TN is the number of samples labelled with the negative class and correctly classified by the model; FP is the number of samples labelled as negative but incorrectly classified by the model; and FN is the number of samples labelled as positive but incorrectly classified by the model.

Experimental setting

The models mentioned in this paper are run on a GPU server in V100. The Python and Pytorch architectures are used. Cross entropy is used as the loss function, the batch size is 32, and 30 rounds are trained. The dynamic learning rate is 0.002 and the dropout rate is 0.01.

Results

Ablation experiments

As shown in Table 3, ablation experiments were designed to verify the feasibility and usefulness of the modules proposed in DSHT. Also, we try to connect these modules in various ways with different input data. There are three type date of the input: only optical flow images (OP), only spatial images (SP) and dual stream images (DS, both of spatial and optical flow images). We

TABLE 2 The distribution of the Dataset.

Subset	Frames	Fraction
Training	48,000	57.14%
Validation	18,000	21.43%
Test	18,000	21.43%
Total	84,000	100%

TABLE 3 The Grading Standards.

Model	Val Acc.	Test Acc.	Pre.	F1-score
HCN (OP)	78.36%	67.97%	68.96%	67.76%
HCN + CFC + global Transformer (OP)	82.91%	80.09%	82.77%	79.91%
HCN (SP)	70.84%	64.18%	66.32%	64.07%
HCN + CFC (SP)	77.96%	71.48%	72.45%	71.21%
HCN + piecemeal Transformer (SP)	85.29%	76.21%	78.41%	75.84%
HCN + piecemeal Transformer + CFC (SP)	86.57%	79.94%	81.73%	79.02%
HCN + global Transformer (SP)	90.01%	89.84%	91.74%	89.56%
HCN + CFC + global Transformer (SP)	95.97%	90.85%	93.49%	90.38%
HCN + CFC (DS)	96.96%	82.46%	83.80%	82.59%
HCN + piecemeal Transformer + CFC (DS)	98.10%	86.57%	87.42%	86.72%
HCN + global Transformer (DS)	99.09%	95.72%	95.86%	95.73%
HCN + CFC + global Transformer (DS) (DSHT)	99.19%	98.05%	98.16%	98.05%

OP, SP and DS are the types of input data, indicating optical flow images only, spatial images only, and dual flow data, respectively. HCN is proposed hierarchical convolutional network. CFC is the composite fusion convolution. The piecemeal Transformer and global Transformer are the connection type of transformer. Val ACC., Test ACC. and Pre. are the validation accuracy, test accuracy, precision, respectively.

test three basic modules in DSHT: Hierarchical Convolutional Network (HCN), Composite Fusion Convolution (CFC) and Transformer, and Transformer have two patterns: piecemeal Transformer and global Transformer. If the feature blocks of HCN are connected before Transformer, the pattern is called global Transformer, and opposite is called piecemeal Transformer. For example, in Table 3, The HCN (OP) model using optical flow images as inputs and only a HCN module to grade the starvation of fish school. On the contrary, the model of HCN + CFC + global Transformer (DS) (DSHT), is the proposed model in this paper.

As shown in Table 3, the HCN (OP) and HCN (SP) are the baseline models, the accuracy of test dataset reaches 67.97% and 64.18%, respectively. By adding the CFC and Transformer to the baseline model, all of the indexes have been improved. The HCN + CFC (SP) model is built by adding the CFC based on HCN, which achieve 71.48%, 72.45% and 71.21% of test accuracy, precision and F1-score. The HCN + piecemeal Transformer (SP) and the HCN + global Transformer (SP) are built by applying the Transformer with different forms, and their test accuracy are 76.21% and 89.84%, respectively. Further, comparing with the single spatial or temporal channel, the performances of dual-stream structure with same methods are more remarkable. The test accuracy of HCN + CFC (DS) model and HCN + global Transformer (DS) model are 82.46% and 95.72%, respectively. If both HCN, CFC and Transformer are used, the model will show even better grading results, e.g. DSHT, which has the best performance with the accuracy, precision and F1-sorce of 98.05%, 98.16% and 98.05%, respectively. Figure 3 shows the accuracy curves of the ablation models in the validation set during the training process. Figure 4 is the visualization of models' output features mentioned in Table 3.

Comparison experiments

For the purpose of validating the reliability of proposed DSHT method, our model is compared with the following baselines in the same condition: VGG 16 (Simonyan and Zisserman, 2015), ResNet 18 (He et al., 2016), Mobile ViT (Mehta and Rastegari, 2022), Triplet Attention (Misra et al., 2021), Residual Attention (Zhu and Wu, 2021), DSRN (Måløya et al., 2019), DAN-EfficientNet-B2 (Yang et al., 2021) and DADSN (Zheng et al., 2022). The results are presented in Table 4. It shows that the DSHT model displays the best performance with the highest grading results in all metrics. The next one is the DADSN model with 83.43%, 84.52%, and 83.21% in the terms of accuracy, precision, and F1-scores, respectively. The worst are the VGG16 and DSRN models, which have only 20% accuracy, 4% precision, and 6.67% F1 score on the test set. The DAN-EfficientNet-B2 model also uses the structure with a CNN-based backbone and attention, but the grading results are not good, with each metrics in the range of 62% to 70%. Figure 5 shows the validation set correctness for each model during training. Figure 6 presents a schematic of the confusion matrix for the better performing models. Figure 7 is the grading results of DSHT model. Figure 8 display the visualization of the output features of each model.

Discussion

In order to provide a more accurate feeding scheme for mariculture, fish starvation grading is one of the prerequisites for AI-feeding. Taking golden pomfret as the research target, this

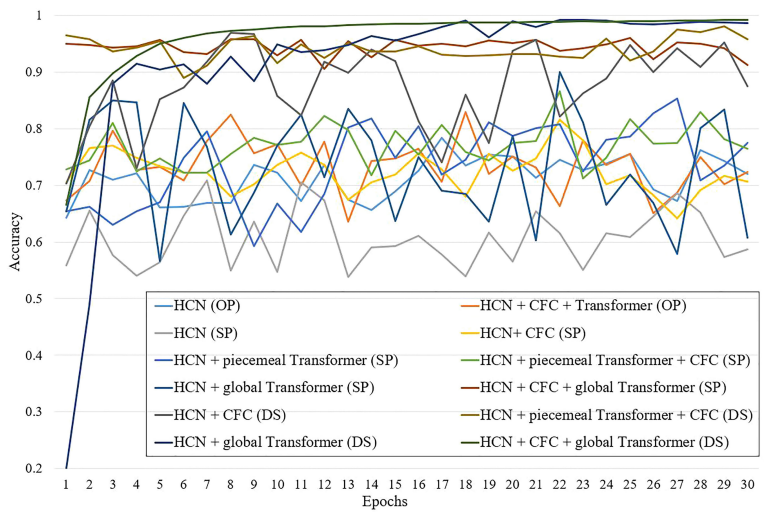


FIGURE 3
Validation Accuracy Curves of Ablation Models.

study attempts to realize automatic identification and monitoring of fish hunger status in the mariculture environment. Our aim is to solve the problem of precise feeding in deep-sea aquaculture and contribute to the development of unmanned and intelligent mariculture equipment. For this purpose, we proposed the Dual Stream Hierarchical Transformer (DSHT).

In the ablation experiment, we tested the modular performance of DSHT, including HCN, CFC, global Transformer and piecemeal Transformer in temporal channel, spatial channel and spatio-temporal channel. It was found that the grading results using a spatio-temporal channel (dual stream, DS) is better than temporal channel only (optical flow images only, OP) or spatial channel only (spatial images only, SP). To be specific, in the structure of HCN + CFC + global Transformer, the test accuracy, precision, and F1-score of DS are 98.05%, 98.16%, and 98.05%, respectively, which is better than the metrics of OP (80.09%, 82.77%, and 79.91%) and SP (90.85%, 93.49%, and 90.38%). The spatial channel provided spatial information, such as the number, distribution, and the mouth movements of fish in the image, while the optical flow images provided temporal information, such as the swimming speed, direction, and acceleration of fish school. As spatial and temporal channel provided unique information for starvation grading, they exhibited complementary effects when combined in the DSHT model. Furthermore, within certain short time interval (e.g., 0.5 seconds), if the object is not moving fast, a spatial image could provide rich detailed data during this time (e.g., whether the fish's abdomen is bulging, whether the fish's excrement appear in the image). In contrast, for consecutive images, because the time interval is extremely short and the images may not change to a great extent, they might provide duplicate information. This made some of them not relevant to the starvation grade (e.g., sunlight in the water), which led to misjudgment. We could not deny that continuous images can provide time-varying information that a single spatial image could not (e.g., swimming speed of the fish school), and this can also have an impact on the starvation grade. Therefore, we used optical flow images to reflect temporal information and experimentally demonstrated that the information provided by optical flow images can be used for starvation grading, with test accuracies of 67.97% and 80.09% for HCN (OP) and HCN + CFC + global Transformer (OP), respectively. In Figure 4 (C1), the temporal information is more concentrated on the tail of the fish, i.e., where the velocity changes faster, while in Figure 4 (C3), the spatial channel is more concentrated on the contour of the fish. Using a dual stream channel is a good way to enrich the input data in mariculture data. We also found that using the global transformer would be more effective than using the piecemeal transformer (the test accuracy of HCN + piecemeal Transformer (SP) is 76.21% and of HCN + global Transformer (SP) is 89.84%). This can be interpreted as the global transformer being calculated after the hierarchical features were integrated for attention, which overall filtering of useless features and marking which features should be attended to. In contrast, the piecemeal transformer is filtered each individual features separately and then integrated them at the end, so it resulted in poor attention to certain features. The local attention shown in Figure 4 (C6) visually demonstrates that the model was

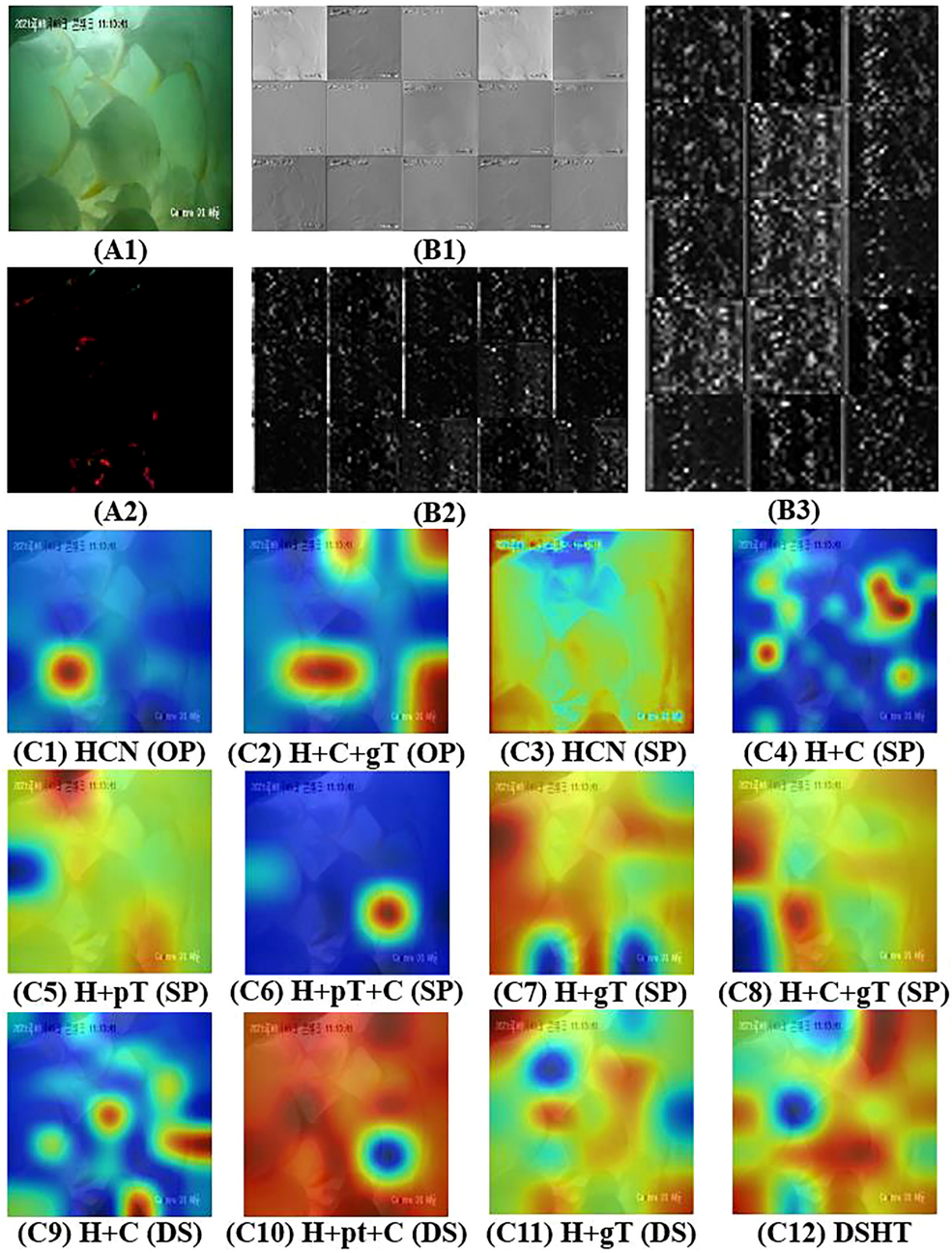


FIGURE 4
Visualization of Output Features. A1 is the spatial image which is fed into the spatial channel. A2 is the optical flow image which is fed into the temporal channle. B1 and B2 are DSHT's hierarchical features in spatial and temporal channels, respectively. B3 is the output features of CFC in DSHT. From C1 to C12, respectively, they represent the Class Activation Map (CAM) of DSHT and other models mentioned in Ablation Experiments. H, C, gT, and pT are abbreviations for HCN, CFC, piecemeal Transformer, and global Transformer, respectively.

limited to a certain point, such as the fish mouth, compared to the global attention shown in Figure 4 (C8), where the model's attention is not limited to a certain place. Compared with the experimental results of HCN + global Transformer (SP, 89.84%) and HCN + CFC + global Transformer (SP, 90.85%), HCN + global Transformer (DS, 95.72%) and HCN + CFC + global Transformer (DS, DSHT, 98.05%), it can be seen that the model using the CFC network outperforms the model without it. The CFC module is a feature fusion module that fuses hierarchical features from the HCN output in three different dimensions, indicating that our proposed DSHT with CFC fusion is more robust. Comparing Figure 4 (C3) and Figure 4 (C4), after adding

TABLE 4 The Grading Standards.

Model	Val Acc. (%)	Test Acc. (%)	Pre. (%)	F1-score
VGG 16	25.00%	20.00%	4.00%	6.67%
ResNet 18	66.90%	59.17%	61.01%	57.80%
Mobile ViT	78.22%	78.44%	86.25%	79.06%
Triplet Attention	70.11%	76.57%	76.24%	75.77%
Residual Attention	64.54%	62.83%	63.32%	61.80%
DSRN	30.51%	20.00%	4.00%	6.67%
DAN-EfficientNet-B2	75.16%	63.84%	69.19%	62.07%
DADSN	86.56%	83.43%	84.52%	83.21%
DSHT (Ours)	99.19%	98.05%	98.16%	98.05%

Val ACC., Test ACC. and Pre. are the validation accuracy, test accuracy, precision, respectively.

CFC, the model starts to convert from the figurative feature of contour to the abstract feature.

Among the CNN-based models, VGG 16, ResNet 18, and HCN, where HCN has the best grading results, this indicated that our proposed model is more suitable for the mariculture environment of pomfret. Because HCN perceives the images with different sizes of kernels and outputs the low-dimensional features (e.g., the boundary of the fish) and high-dimensional features (e.g., the shape of the fish belly) of the images at the same time, HCN is more intuitive and concise than a single convolutional kernel that requires multiple perceptions to transition from low to high dimensions. By comparing the results of the attention-based models, Mobile ViT (78.44% accuracy), Triplet Attention (76.57% accuracy), Residual Attention (62.83% accuracy), and our proposed HCN + Transformer-based models (the accuracy ranged between 76.21% and 98.05%), we found that the performance of the HCN + Transformer performs better because the attention calculation of this model is performed at the feature level, which is the output of the HCN, while other models use attention calculation directly.

The DSRN, DAN-EfficientNet-B2 and DADSN are related researchers' proposed models for fish behavior recognition. DSRN is used to classify the feeding or nonfeeding behavior of salmon in mariculture environments, and it uses 3D residual convolution and LSTM for feature extraction and feature filtering, respectively. The DAN-EfficientNet-B2 model divides the features into position attention and channel attention for recognizing fish feeding behavior. In this paper, we applied the DSRN and DAN-EfficientNet-B2 to the fish starvation grading task and found that their grading accuracy is not satisfactory (20.00% and 63.84% accuracy). We speculate that more fish information is required for the starvation grading task, and these models are difficult to meet the requirement using only 3D

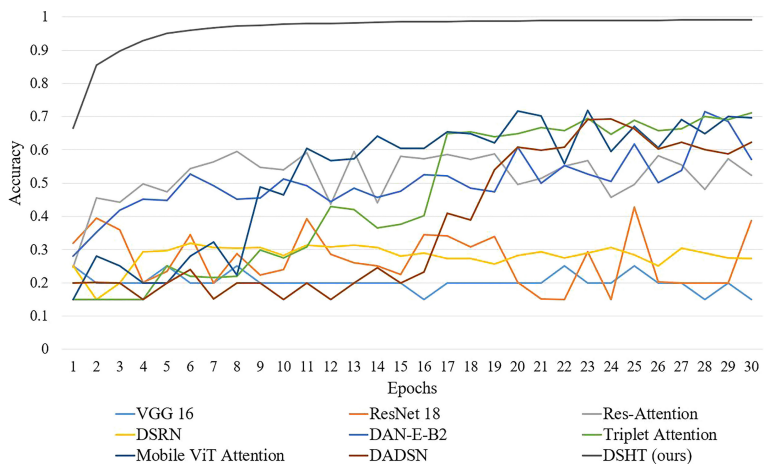


FIGURE 5
Validation Accuracy Curves of Comparison Models.

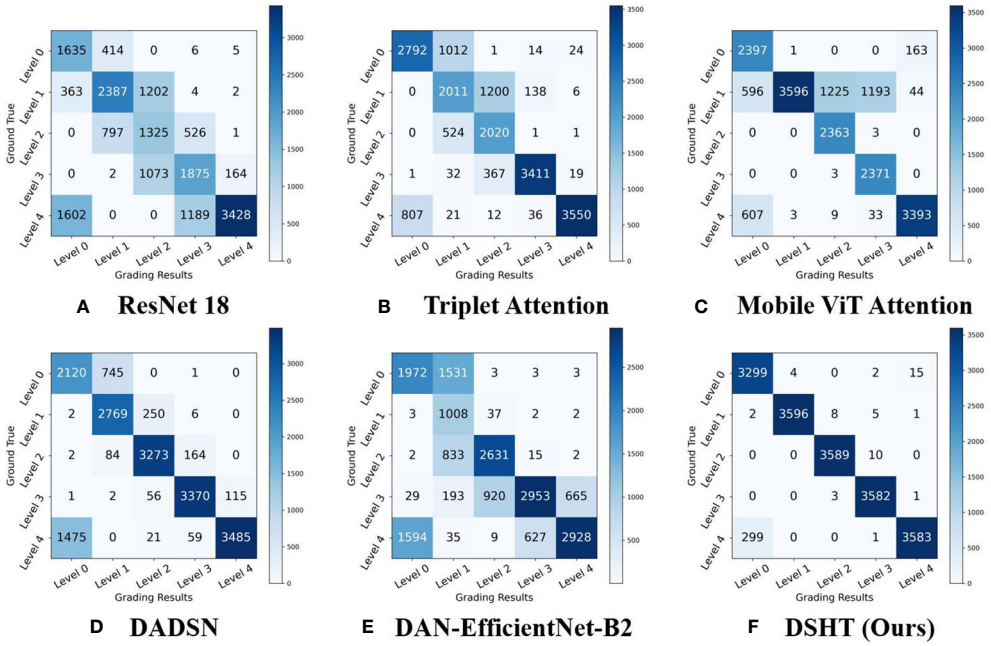


FIGURE 6
The Confusion Matrix. The experimental results of each model mentioned in Table 4 are shown, from **(A–F)**, for ResNet 18, Triplet Attention, Mobile ViT Attention, DADSN, DAN-EfficientNet-B2, and DSHT (Ours). In the figure, each row represents the grading result of the model, and each column represents the true sample category. For example, in A, the first column of the first row indicates that: 1635 samples were grading correctly (true class L0 and graded result L0); the second column of the first row indicates that: 414 samples with true class L1 were graded as L0 by ResNet; the fourth column of the first row indicates that: 6 samples with true class L3 were graded as L0 by ResNet; the first column of the second row indicates that: 363 samples with true class L0 are graded as L1 by ResNet.

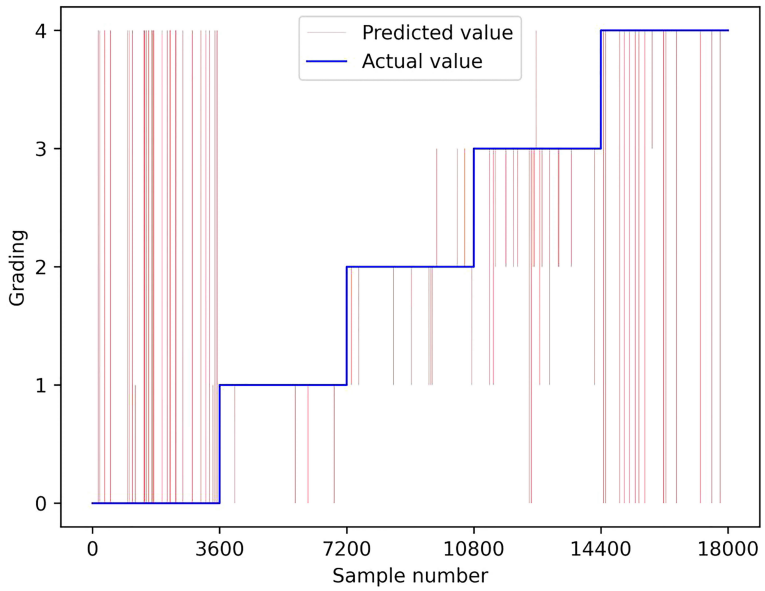


FIGURE 7
Grading Results of DSHT Model. The blue line is the ture value and the red line is the grading results predicted by DSHT.

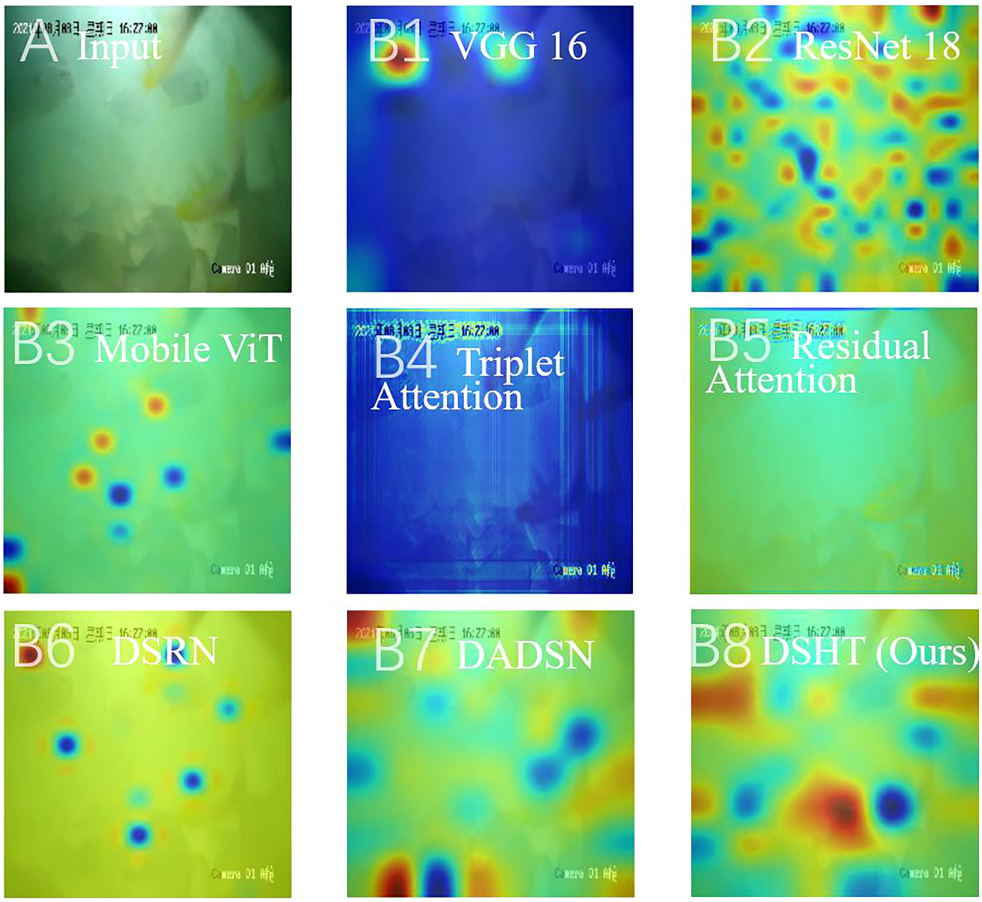


FIGURE 8
Class Activation Map (CAM) of Proposed Method and Other Methods. **(A)** is the input. From **(B1–B8)**, the visualization results of the models in Table 4 are shown separately.

residual convolution or attention. This concept was also verified in the ablation experiment, where the accuracy of the dual stream model HCN + CFC (DS), which only used CNN-based structure, was only 82.46% in accuracy. But the DSHT, which used CNN-based and attention-based structures at the same time, can achieve 98.05%. Notably, the researchers (Zhou et al., 2019) used a 5-layer CNN to grade the intensity level of fish feeding, but when we migrated it to our task for testing, the experimental results were the same as VGG16. The DADSN model (83.43% accuracy) was used for the same starvation grading task as ours for the mariculture of pomfret. It adopted a modified Efficient network and ViT in each channel for feature extraction and attention computation, and uses LSTM to fuse the output features of each channel. Compared with our proposed DSHT model, DSHT does the attention operation in global features after CFC network fusion, while DADSN does it separately in spatial and temporal channels. Moreover, DSHT proposes a more concise multi-scale network HCN, which is another difference with DADSN.

Conclusions

In this paper, we study the starvation grading of golden pomfret school in a marine aquaculture environment. Establishing a dataset and proposing a DSHT model for this task. The DSHT uses a Hierarchical Convolutional Network to extract marine image features, which improves the effectiveness of the neural network in learning the starvation features of golden pomfret school. For the hierarchical features extracted by HCN, CFC is used for fusion, and a transformer is used to increase the weight of some regions that are favorable for starvation classification. To address the problem that fish behavior is temporal in nature, this paper designs a dual-stream structure by introducing optical flow images to increase temporal information. The effectiveness of each module proposed in this paper is demonstrated by ablation experiments. In the results of the comparison tests, DSHT achieved the best performance among all models involved. Experiments on a marine pomfret breeding vessel have shown that DSHT can be effectively applied to pomfret school starvation grading in marine images, with practical

implications. In future research, we will explore how to use AI information for decision-making, control, and management of fish aquaculture.

Data availability statement

The original contributions presented in the study are included in the article/supplementary materials. Further inquiries can be directed to the corresponding author.

Ethics statement

Ethical review and approval were not required for the animal study because the data was collected under normal fish farming conditions.

Author contributions

KZ conceived the study and drafted the manuscript. RY provided critical revision of the article. RL designed the software. LY performed the analysis with the help from HQ. ZY participated in fish farming. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Special Fund for Marine Economic Development (six marine industries) of Guangdong Province, China (GDNRC [2021]42), the Fund of Southern Marine Science and Engineering Guangdong Laboratory (Zhanjiang) (011Z21001), and the Fund of Zhanjiang Municipal Bureau of Science and Technology (2021E05034).

Acknowledgments

We wish to thank the Department of Natural Resources of Guangdong Province, the Zhanjiang Municipal Bureau of Science and Technology and the Southern Marine Science and Engineering Guangdong Laboratory (Zhanjiang) for funding this project. We thank Guangxi Jingong Marine Technology Co., Ltd. For providing the fingerlings and lures.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., et al. (2019). "Dual attention network for scene segmentation," in *The IEEE/CVF International Conference on Computer Vision 2019* (Piscataway, NJ: IEEE). 3141–3149. doi: 10.1109/CVPR.2019.00326

Hendrycks, D., and Gimpel, K. (2016). Gaussian Error linear units. *ArXiv. Prep. ArXiv.* 1606, 8415. doi: 10.48550/arXiv.1606.08415

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *The IEEE/CVF International Conference on Computer Vision 2016*. (Piscataway, NJ: IEEE). 770–778. doi: 10.1109/CVPR.2016.90

Hossain, M. E., Khan, M. A., Saha, S. M., and Dey, M. M. (2022). Economic assessment of freshwater carp polyculture in Bangladesh: Profit sensitivity, economies of scale and liquidity. *Aquaculture* 548 (1), 737552. doi: 10.1016/j.aquaculture.2021.737552

Jonas, J., Wolff, V., Fricke-neuderth, K., and Mothes, O. (2017). "Visual fish tracking: combining a two-stage graph approach with CNN-features," in *Oceans Aberdeen Conference 2017*, (Piscataway, NJ: IEEE). 1–6. doi: 10.1109/OCEANSE.2017.8084691

Jothiswaran, V., Velumani, T., and Jayaraman, R. (2020). Application of artificial intelligence in fisheries and aquaculture. *Biotica. Res. Today* 2 (6), 499–502. Available at: https://www.biospub.com/index.php/biorestoday/article/view/257

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60 (6), 84–90. doi: 10.1145/3065386

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86 (11), 2278–2324. doi: 10.1109/5.726791

Lin, G., and Lin, X. (2022). Bait input altered microbial community structure and increased greenhouse gases production in coastal wetland sediment. *Water Res.* 218, 118520. doi: 10.1016/j.watres.2022.118520

Lucas, B. D. (1985). *Generalized image matching by the method of differences*. PhD. Dissertation, Carnegie Mellon University. doi: 10.5555/912172

Maaz, M., Shaker, A., Cholakkal, H., Khan, S., Zamir, S. W., Anwer, R. M., et al. (2022). EdgeNeXt: efficiently amalgamated CNN-transformer architecture for mobile vision applications. *ArXiv. Prep. ArXiv.* 2206, 10589. doi: 10.48550/arXiv.2206.10589

Måløya, H., Aamodta, A., and Misimi, E. (2019). A spatio-temporal recurrent network for salmon feeding action recognition from underwater videos in aquaculture. *Comput. Electron. Agric.* 167, 1–9. doi: 10.1016/j.compag.2019.105087

Mehta, S., and Rastegari, M. (2022). Separable self-attention for mobile vision transformers. *ArXiv. Prep. ArXiv.* 2206, 2680. doi: 10.48550/arXiv.2206.02680

Misra, D., Nalamada, T., Arasanipalai, A. U., and Hou, Q. (2021). "Rotate to attend: Convolutional triplet attention module," in *The IEEE/CVF Winter Conference on Applications of Computer Vision 2021*. (Piscataway, NJ: IEEE) 3139–3148. doi: 10.1109/WACV48630.2021.00318

Raghu, M., Unterthiner, T., Kornblith, S., Zhang, C., and Dosovitskiy, A. (2021). Do vision transformers see like convolutional neural networks? *ArXiv. Prep. ArXiv.* 2108, 8810. doi: 10.48550/arXiv.2108.08810

Ranjan, A., and Black, M. J. (2017). "Optical flow estimation using a spatial pyramid network. in: Computer vision and pattern recognition," in *The IEEE/CVF International Conference on Computer Vision 2017*. (Piscataway, NJ: IEEE). 4161–4170. doi: 10.1109/CVPR.2017.291

Simonyan, K., and Zisserman, A. (2015). "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations 2015*, (Banff, Canada: OpenReview). 1–14. doi: 10.48550/arXiv.1409.1556

Szegedy, C., Liu, W., Yangqing, J., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *The IEEE/CVF Winter Conference on Applications of Computer Vision 2015*, (Piscataway, NJ: IEEE). 1–9. doi: 10.1109/cvpr.2015.7298594

Yang, L., Yu, H., Cheng, Y., Mei, S., Duan, Y., Li, D., et al. (2021). A dual attention network based on efficientNet-B2 for short-term fish school feeding behavior analysis in aquaculture. *Comput. Electron. Agric.* 187, 106316. doi: 10.1016/j.compag.2021.106316

Ye, Z., Zhao, J., Han, Z., Zhu, S., Li, J., Lu, H., et al. (2016). Behavioral characteristics and statistics-based imaging techniques in the assessment and optimization of tilapia feeding in a recirculating aquaculture system. *Trans. ASABE.* 59, 345–355. doi: 10.13031/trans.59.11406

Zhao, J., Bao, W., Zhang, F., Zhu, S., Liu, Y., Lu, H. D., et al. (2018). Modified motion influence map and recurrent neural network-based monitoring of the local unusual behaviors for fish school in intensive aquaculture. *Aquaculture* 493, 165–175. doi: 10.201016/j.aquaculture.2018.04.064

Zheng, K., Yang, R., Li, R., Yang, L., Qin, H., and Sun, M. (2022). "A deep transformer model-based analysis of fish school starvation degree in marine farming vessels," in *2022 4th International Conference on Control and Robotics (ICCR)*, Guangzhou, China. (Piscataway, NJ: IEEE).

Zhou, C., Xu, D., Chen, L., Zhang, S., Sun, C., Yang, X., et al. (2019). Evaluation of fish feeding intensity in aquaculture using a convolutional neural network and machine vision. *Aquaculture* 507, 457–465. doi: 10.1016/j.aquaculture.2019.04.056

Zhou, C., Zhang, B., Lin, K., Xu, D., Chen, C., Yang, X., et al. (2017). Near-infrared imaging to quantify the feeding behavior of fish in aquaculture. *Comput. Electron. Agric.* 135, 233–246. doi: 10.1016/j.compag.2017.02.013

Zhu, K., and Wu, J. (2021). "Residual attention: A simple but effective method for multi-label recognition," in *The IEEE/CVF International Conference on Computer Vision 2021*, (Piscataway, NJ: IEEE). 184–193. doi: 10.1109/ICCV48922.2021.00025

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 08 December 2022
DOI 10.3389/fmars.2022.1073615

Check for updates

OPEN ACCESS

EDITED BY
Hong Song,
Zhejiang University, China

REVIEWED BY
Masoumeh Zareapoor,
Shanghai Jiao Tong University, China
Irfan Ahmad,
Khurasan University, Afghanistan
Zhuang Zhou,
Beijing Institute of Technology, Zhuhai, China

*CORRESPONDENCE
Jun Qiu
qiujun@bistu.edu.cn

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 18 October 2022
ACCEPTED 17 November 2022
PUBLISHED 08 December 2022

CITATION
Gao S, Liu C, Zhang H, Zhou Z and Qiu J (2022) Multiscale attention-based detection of tiny targets in aerial beach images. *Front. Mar. Sci.* 9:1073615. doi: 10.3389/fmars.2022.1073615

COPYRIGHT
© 2022 Gao, Liu, Zhang, Zhou and Qiu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Multiscale attention-based detection of tiny targets in aerial beach images

Shurun Gao[1], Chang Liu[1], Haimiao Zhang[1], Zhehai Zhou[2] and Jun Qiu[1]*

[1]Institute of Applied Mathematics, Beijing Information Science and Technology University, Beijing, China, [2]Key Laboratory of the Ministry of Education for Optoelectronic Measurement Technology and Instruments, Beijing Information Science and Technology University, Beijing, China

Tiny target detection in marine scenes is of practical importance in marine vision applications such as personnel search and rescue, navigation safety, and marine management. In the past few years, methods based on deep convolutional neural networks (CNN) have performed well for targets of common sizes. However, the accurate detection of tiny targets in marine scene images is affected by three difficulties: perspective multiscale, tiny target pixel ratios, and complex backgrounds. We proposed the feature pyramid network model based on multiscale attention to address the problem of tiny target detection in aerial beach images with large field-of-view, which forms the basis for the tiny target recognition and counting. To improve the ability of the tiny targets' feature extraction, the proposed model focuses on different scales of the images to the target regions based on the multiscale attention enhancement module. To improve the effectiveness of tiny targets' feature fusion, the pyramid structure is guided by the feature fusion module in order to give further semantic information to the low-level feature maps and prevent the tiny targets from being overwhelmed by the information at the high-level. Experimental results show that the proposed model generally outperforms existing models, improves accuracy by 8.56 percent compared to the baseline model, and achieves significant performance gains on the TinyPerson dataset. The code is publicly available *via* Github.

KEYWORDS

tiny object detection, multiscale attention, feature pyramid network, attention mechanism, unmanned aerial vehicle

1 Introduction

Target detection is the key to many computer vision applications, and its importance has gradually increased in the last decade for marine vision tasks, such as ship detection (Chen et al., 2021a; Tian et al., 2021), and maritime rescue (Varga et al., 2022), environmental monitoring (Ribeiro et al., 2019; Lieshout et al., 2020; Cheng et al., 2021). In recent years,

target detection has improved tremendously as deep neural networks are trained faster and more efficiently. The combination of deep learning-based target detection technology and Unmanned Aerial Vehicle (UAV), as an image acquisition system with a large field-of-view and high efficiency, has been extensively employed for marine detection tasks with a large range and ultra-long distance. Compared with natural scene images, most of the targets in aerial marine scenes have tiny scales, and signal-to-noise ratios, and are easily swamped by background noise. Therefore, it is challenging to design a multi-scale tiny target detection model applicable to marine scenes.

The convolutional neural network-based detection model (Yu et al., 2020a; Yu et al., 2020b; Shen et al., 2019) has significantly improved the target detection task. A salient feature of the deep learning models, regarding the ability to generalize, is the quality and quantity of the dataset, and the abundant high-quality data can enhance the robustness and generalization of the model. Kisantal et al. (Kisantal et al., 2019) proposed a copy-and-paste enhancement to increase the number of samples and diversity of the tiny targets by copying and pasting images containing the tiny targets multiple times to ensure that they appear in the correct context. Chen et al. (Chen et al., 2019) proposed an adaptive resampling enhancement strategy to copy-paste the targets considering the contextual information on top of Kisantal's work, to solve the problem of context and scale mismatch in the appearance of targets, thus achieving data enhancement. This method of increasing the number of tiny targets can, to a certain extent, increase the positive samples and better optimize the model for tiny target detection. However, the gains based on the data processing have instead been constrained by the dataset.

Recently, the super-resolution reconstruction of tiny targets based on generative adversarial networks (Li et al., 2017; Na and Fox, 2018; Mehralian and Karasfi, 2018; Deng et al., 2022) has been developed. Bai et al. (Bai et al., 2018b) have proposed a multi-task generative adversarial network that feeds the super-resolution images, generated by up-sampling tiny targets, into a multi-task discriminator network that distinguishes the super-resolution images from the real images and outputs the predicted classes and bounding boxes. Noh et al. (Noh et al., 2019) have proposed a new super-resolution method at the feature level by matching the generated high-resolution features with the perceptual fields of the low-resolution features by utilizing the dilated convolution. This will help in avoiding generation of the incorrect super-resolution features owing to the perceptual field mismatch. However, the generators in the generative adversarial networks generate limited sample diversity, and hence, it is difficult to establish a balance between the generators and discriminations.

In practical applications, data enhancement methods for tiny target features may introduce new noise, which may impair the performance of the model in extracting features. Further, the super-resolution structure may complicate the end-to-end model training. To solve these problems, we designed a multiscale attention-based feature pyramid model of tiny object detection in aerial beach images. First, we addressed the problem of target information loss owing to the down-sampling in convolutional networks. The multiscale attention enhancement module (MAEM) was designed by employing self-attention to obtain the weight of the target location and retain the detailed information and the contextual information. Thus, the proposed model can improve the feature extraction of the tiny targets in the aerial large field-of-view and reduce the interference caused by the complex background. We designed a novel multiscale feature fusion module (MFFM) for the problem of inconsistent gradient computation in the Feature Pyramid Network (FPN) (Lin et al., 2017), which changes the original linear fusion, and employs the attention-guided maps to obtain the weights of the feature maps of different scales. This prevents the target from being overwhelmed by the high-level feature information while giving more semantic information to the low-level feature maps. Further, the proposed model pays more attention to the features of the tiny targets in the fusion process and improves the efficiency of the tiny target feature fusion. The proposed model has been validated on the TinyPerson dataset, and the experimental results show that the accuracy of the model, designed in this work, reaches 59.82%, which can be utilized in personnel search and rescue for seaside security. In summary, the contributions in this work are mainly in the following folds.

- We propose a novel network model for tiny target detection by introducing multiscale attention and feature pyramid networks. The detection performance of tiny targets is improved by enhancing the ability of tiny target feature extraction and fusion.
- We add an attention loss for the convolutional neural network to learn discriminative features to prevent tiny targets from being overwhelmed by complex backgrounds.
- We conducted comprehensive ablation experiments to demonstrate the impact of each of the proposed modules on detection results, and experimentally tested on the TinyPerson dataset with a significant improvement in tiny target detection accuracy over baseline.

The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 describes the proposed model framework, including MAEM, MFFM, and the loss function. The experimental results on TinyPerson datasets are reported in Section4 to validate the performance of the proposed model. Section 5 discusses the results, and Section 6 concludes this paper.

2 Related work

2.1 Small object detection

The advancement of deep learning technology has been improving the accuracy of object detection greatly, researchers

search frameworks for small object detection specifically. The FPN proposed by Lin et al. (Lin et al., 2017) introduces a bottom-up, top-down network structure that achieves feature enhancement by fusing features from adjacent layers. Based on the FPN, Liang et al. (Liang et al., 2018) proposed a deep feature pyramid network using a feature pyramid structure with lateral connections to enhance the semantic features of small targets, and specialized anchors to detect small targets in high-resolution images. Nayan et al. (Nayan et al., 2020) proposed a new real-time detection algorithm for the problem that small targets tend to lose feature information after multi-layer networks. The algorithm uses upsampling and jumps connections to extract multi-scale features of different network depths during the training process, which improves the detection accuracy and speed of small target detection. Rotation equivariant feature image pyramid network (REFIPN) (Shamsolmoali et al., 2022b) improves the ability to focus on small targets in remote sensing images through scale adaptation. REFIPN uses a single detector in parallel with a lightweight image pyramid to extract features at a wide range of scales and orientations and generate regions of interest to improve the performance of small-scale object detection performance. Shamsolmoali et al. (Shamsolmoali et al., 2022a) proposed a weakly supervised approach for object detection in remote sensing images and designed a contextual fine-grained model with significant attention to different objects and target parts. Liu et al. (Liu et al., 2021a) proposed a high-resolution detection network for small targets, which improves the detection performance of small targets with reduced computational cost by using a shallow network for high-resolution images and a deep network for low-resolution images. These methods mentioned above improve the performance of small target detection to some extent.

2.2 Object detection in maritime

In comparison with target detection of natural scenes, aerial images have a wider detection range, so the obtained image field of view is often large. Lee et al. (Lee et al., 2018) modifies the 10 different ship categories in the You Only Look Once (YOLO) algorithm target classification and applies them to maritime video surveillance tasks, thus enabling real-time maritime detection. Ghahremani et al. (Ghahremani et al., 2018) proposed a CNN-based cascaded method for detecting maritime vessels, which takes the candidate target regions in the original image and performs additional processing to improve the accuracy of small target detection. Due to the high complexity of the cascaded method, it is not suitable for real-time monitoring applications. Moon et al. (Moon et al., 2020) proposed a new cascade Region-based CNN (RCNN) method to detect small targets in marine scenes. The improvement of small target detection accuracy is due to retaining its information in all layers.

Soloviev et al. (Soloviev et al., 2020) proposed two datasets for marine ship detection and evaluates the effectiveness of three target detection models, FasterRCNN (Ren et al., 2017), Region-based Fully Convolutional Network (R-FCN) (Dai et al., 2016), and Single Shot multibox Detector (SSD) (Liu et al., 2016), on this dataset. The FasterRCNN with ResNet101 as the backbone has the highest detection accuracy for large targets, but the detection accuracy for small targets is lower.

2.3 Attention-based maritime small object detection

Attention mechanism is widely used for target detection in marine scenarios due to its excellent performance. Woo et al. (Woo et al., 2018) proposed a mixed channel and spatial attention mechanism, which enhances the utilization of spatial and channel information for input features by obtaining attention weights after the spatial attention module and channel attention module are connected in series. For small target detection in the marine environment, Chen (Cheng et al., 2021) proposes a global attention module for sea-level small trash detection by adaptively fusing deep multiscale image and radar data features. Chen (Chen et al., 2021b) proposes an improved ImYolov3 based on an attention mechanism, which integrates spatial and channel attention modules into the network architecture of Yolov3, and improves the representational capability of the network by adjusting the perceptual fields in each branch network. This enables better differentiation between ships and backgrounds. Therefore, how to develop an attention mechanism in aerial image tiny target detection is a very interesting problem.

3 Materials and methods

In the UAV aerial beach images for target detection, the complex background tends to drown tiny targets, which is not conducive to the extraction of tiny target features. Further, the extraction of the contextual information in the images helps the model to differentiate between the target and the background (Zhu et al., 2021). Therefore, we designed the multiscale attention enhancement and fusion network with Swin-T (Liu et al., 2021b) as the backbone, and the network structure is shown in Figure 1. The main contributions are as follows. A key feature attention mechanism has been designed that contains the MAEM and attention loss, since the UAV aerial images contain variable and complex scenes, thus causing a large amount of redundant information in the context information extracted from the backbone. MAEM can guide the model to focus on tiny targets, thus obtaining an enhanced feature map A. Furthermore, we designed a novel feature fusion module MFFM. The height of the UAV aerial photography process varies, and the images of different scale targets have been

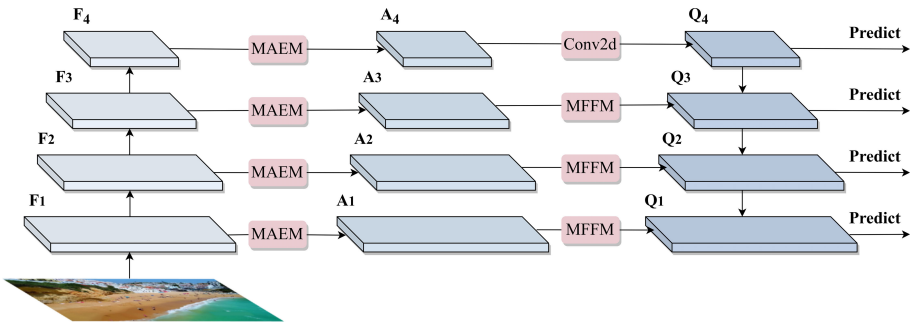


FIGURE 1
Overview of proposed AEFNet for the tiny target detection, it contains a swin-T style architecture of the feature extractor, MAEM for obtaining tiny target feature weights using a self-attention, MFFM for deep and shallow feature map fusion using attention-guided maps.

obtained. The proposed module MFFM introduces an attention mechanism in the feature fusion process to assign attention weights to the feature maps of different scales and put more attention on the tiny targets. Finally, the fused feature map Q has been passed through the Regions with CNN Features (RCNN) (Girshick et al., 2014) prediction network to achieve the final target location and category probability output.

3.1 Multiscale attention enhancement module

According to previous works (Yu and Koltun, 2015; Bai et al., 2018a; Zhang et al., 2021), the appropriate modeling by using contextual information has been beneficial for improving the performance of target detection. MAEM is a basic module in the network as shown in Figure 2, which preserves the detailed information of the target while obtaining contextual information. The module contains two branches, one for computing the global semantic information and the other for computing the local semantic information, which are finally computed to obtain an attention-guided map.

The local semantic information has been divided into $s \times s$ blocks of size $w \times h$ from the input original feature map F . The dependencies among the pixels in the local range have been calculated by the operation of nonlocal (Wang et al., 2018), where all blocks share weights. Thus, all the output feature maps have been gathered together to form a new local association

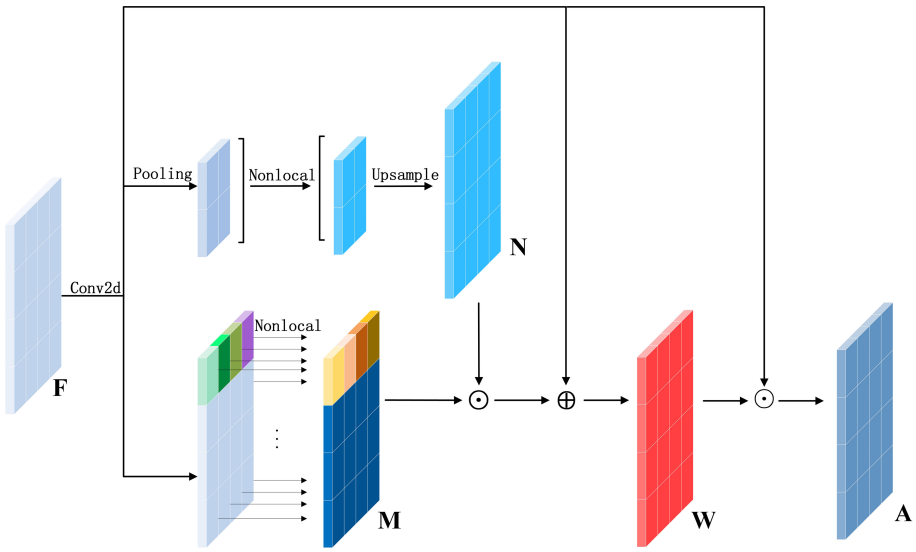


FIGURE 2
Multiscale attention enhancement module. Its upper branch and lower branch represent the global and local semantic information respectively.

feature map M . The purpose is to restrict the perceptual domain of the network to a local range, and then to use the relationships among pixels in the local range to aggregate the pixels belonging to the same class. Concomitantly, this method excludes the influence of structural noise within each patch on the target and computes the probability of the target's appearance. The design of local associations can save computational resources and speed up network training and inference.

The global semantic information has been obtained by first extracting the features of each patch from the input feature map F by adaptive pooling to obtain the pooled features with pixels $s \times s$, where each pixel represents a feature of each patch. Then, the contextual information among each patch has been computed by the non-local operation to travel a new guided graph N . At the global level, noise in the background and targets may have similar associations with respect to local associations, hence we use global semantic information to assist in discerning the location of targets, excluding the interference of similar targets or noise. Further, we calculate the attention-guided graph W of targets by aggregating features between the individual blocks, as shown in the following equation,

$$W_k = \alpha\sigma(N \odot M) + F_k, \tag{1}$$

where $\odot$ denotes element-wise multiplication, σ the Sigmoid activation function, and F_k the feature map at the k^{th} stage. Considering that the attention-guided map M has been employed to guide the enhanced local association features N , in this paper, the elements in M have been directly multiplied with each patch of N . Furthermore, to obtain a more effective representation, setting the learning parameter α will join the feature map F to select more effective semantic features using the adaptive nature of the network.

The attention-guided maps, generated at different stages, have different scale properties. The residual connection has been used to generate the attention-guided map into an enhanced feature map A , as shown in the following equation,

$$A_k = (1 + W_k) \odot F_k . \tag{2}$$

3.2 Multiscale feature fusion module

The high-level semantic and the low-level semantic have been focused on the difference in the target regions, and MFFM guides the higher layer to the shallow layer to select the appropriate features. Further, the appropriate features will be optimized to the same category, which plays the role of the feature selection. Thus, the appropriate target features in the scale range of the current layer will flow into the next layer of computation, whereas other features will be weakened and suppressed, thus enhancing the efficiency of the tiny target feature fusion. Furthermore, if the target has been detected in both the neighboring layers, the higher layer will optimize to the next layer while providing more semantic information, as follows,

$$Q_{k-1} = (I_u(Q_k) + A_{k-1}) \odot (W_{k-1} \odot I_u(W_k)), \tag{3}$$

where W_k denotes the attention-guided map of the k^{th} layer, I_u the upsampling operation, to make the adjacent layer feature maps of the same size, and Q_k the feature map after the k^{th} layer fusion.

3.3 Loss function

The method given in this paper belongs to a two-stage (Ren et al., 2017) detection model, where the first stage generates the proposal frames through a Region Proposal Network (RPN), and the second stage identifies the location of the target class through Regions of Interest (RoI). During the training process, the specific formulas of the respective loss functions of the model are shown below

$$L_{RPN} = \frac{1}{N_{cls}} \sum_{i=1} L_{cls}\left(p_{ri}, p_{ri}^{\star}\right) + \lambda_1 \frac{1}{N_{reg}} \sum_{i=1} P_{ri}^{\star} L_{reg}\left(t_{ri}, t_{ri}^{\star}\right), \tag{4}$$

$$L_{HEAD} = \frac{1}{N_{cls}} \sum_{i=1} L_{cls}\left(p_{hi}, p_{hi}^{\star}\right) + \lambda_2 \frac{1}{N_{reg}} \sum_{i=1} P_{hi}^{\star} L_{reg}\left(t_{hi}, t_{hi}^{\star}\right), \tag{5}$$

$$L_A = \frac{1}{N_{cls}} \sum_{i=1} L_{cls}\left(p_{ai}, p_{ai}^{\star}\right), \tag{6}$$

where L_{RPN} denotes the RPN loss function, L_{HEAD} the RoI Head loss function, and L_A the attention loss function. L_{cls} is the classification loss function, and the binary cross-entropy has been used to compute the classification loss, as shown in Equation (7). L_{reg} is the regression loss function, and smooth L1 has been used to compute the regression loss, as shown in Equation (8). p_{ri} denotes the prediction probability of each bounding box category, and $p_{ri}^{\star}$ denotes the truth value of each ground-truth box category. Since the role of RPN is to select the proposed box and only the foreground needs to be judged, the cross-entropy loss has been employed. Here, t_{ri} the coordinates of the bounding box, $t_{ri}^{\star}$ the coordinates of the ground-truth box, and $P_{ri}^{\star}$ determines the positive example box in the generated detection box to compute the loss. N_{cls} denotes the number of images in each small batch, and N_{reg} denotes the number of anchor box. Classification and regression losses are each normalized by N_{cls} and N_{reg} , and the parameters λ_1 and λ_2 have been used to adjust the balance of the two parts of the loss. The alternative parameter settings in L_{HEAD} and L_A are similar to those of L_{RPN} .

$$L_{cls}(p, p^{\star}) = -\log\ [pp^{\star} + (1-p)(1-p^{\star})], \tag{7}$$

$$L_{reg}(x) = \begin{cases} 0.5x^2, & |x| < 1, \\ |x| - 0.5, & |x| \geq 1. \end{cases} \quad (8)$$

Finally, the three components of the loss have been optimized by a joint loss function as

$$L = L_{RPN} + L_{HEAD} + L_A \,. \quad (9)$$

4 Experiments and results

4.1 Dataset

The numerical experiments in this paper utilize a publicly available dataset for the seaside person target detection, viz., TinyPerson (Yu et al., 2020a). As shown in Table 1.The annotation information of each sample includes the category label, bounding box, and pixel size. The size range is divided into 5 intervals: Tiny1[2, 8], Tiny2[8, 12], Tiny3[12, 20], Tiny[2, 20], and small[20, 32]. We have cropped each image with an overlap of 40 pixels to a resolution of 640 × 512 as the input to the model.

4.2 Experimental setting

We have used Swin-T as the model backbone, and the pre-training parameters loaded during the model training are those obtained by training Swin-T on the ImageNet-1K (Russakovsky et al., 2015) dataset. Experiments have been conducted based on the training and testing sample division of the dataset.

The experimental code in this paper has been implemented based on Pytorch 1.7.1, and the entire training process was deployed on AMD A40 GPU with 48 GB of memory. The code in this paper is based on the design under the MMdetection toolbox. The model is initially trained on the TinyPerson datasets with 12 epochs, and it took 48 hours in total. The gradient optimization method is AdamW (Loshchilov and Hutter, 2017), whose parameters are set as follows; i.e., weight decay at 5e-4, training batch size 4, and the learning rate initialized to 1e-4. The learning rate update is STEP, whose parameters are set as warmup iterations 1000 and warmup ratio 1e-3. The hyperparameters λ_1 and λ_2 are set as 0.6 and 1, respectively. The number of RPN proposal boxes is set as 2000 and 1000 in the training and testing phases, respectively.

To maintain consistency with the TinyPerson benchmark, the evaluation metric in this paper employs Average Precision (AP), Floating point operations (FLOPs), and Parameters (Params). Among them, AP is the evaluation index of the mainstream target detection model, the higher the value the better the model performance; FLOPs is used to measure the complexity of the model; Params is used to evaluate the number of parameters of the model.

4.3 Visualization analysis

This paper visualizes the attention heatmaps by a qualitative method to demonstrate the effect of attention loss on the model performance, which can visually show the part of the region that the model affects. According to Figure 3, the first image is the original image, the second image is the heatmap without the addition of attention loss, and the third image is the heatmap after the addition of attention loss. By comparison, the boundary around the target without adding the attention loss is blurred, and the boundary around the target after adding attention loss is clearer. This makes the model focus more on the tiny target area and avoid the interference of the environment to a certain extent, enhancing the accuracy of target detection, and thus verifying the effectiveness of the attention loss.

To further test the effectiveness of the model design, we present certain detection results in the Tiny Person dataset, including the people scenarios at sea level and on land, as shown in Figure 4.

According to the scene at sea level, the pose of people with their bodies fully exposed on surfboards varies greatly. With respect to the people swimming in the sea with only a part of their bodies exposed, the method in this paper can detect and identify the target and locate it. Furthermore, the scene of people on the land contains dense crowds, cluttered backgrounds, and different-scale crowds. Our method can still detect and identify most of the target people, thus verifying the effectiveness of our method.

4.4 Ablation study

Several sets of experiments have been set up in this paper to demonstrate the effect of patch size on the model in the MAEM. Choosing different patch sizes for different scales has different effects, as shown in Table 2.

We intend to cover both the target area and a certain background area for the selection of patch size so that the

TABLE 1 Details of TinyPerson.

Tiny Person	Training	Testing	Total
Label image	794	816	1,610
Annotations	42,197	30,454	72,651

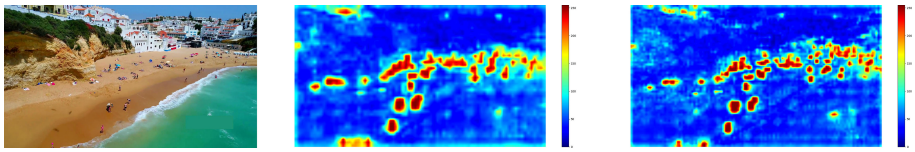


FIGURE 3
Result of attention heatmap. From left to right, the original mage, the heatmap, and the heatmap after adding attention loss.

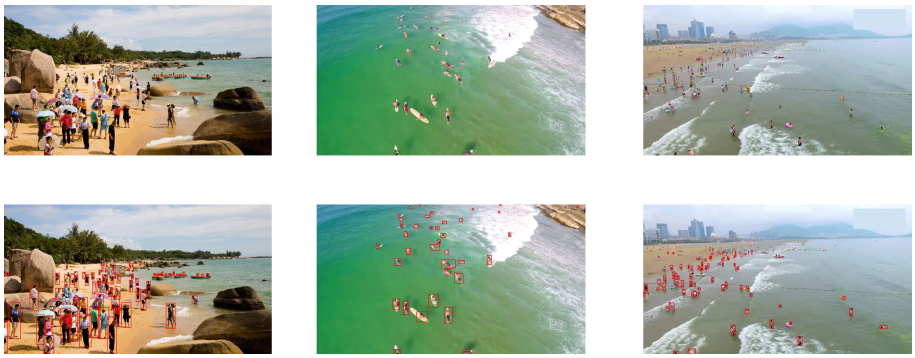


FIGURE 4
Detection results on the TinyPerson dataset. The first row is the original image of different scenes, and the second row is the detection results of the corresponding scenes.

experiments have been conducted for both the single-scale and multiscale combinations. According to Table 2, the experiments with a single size incorporate relatively little information, thus resulting in poorer results. In contrast, better results tend to have more dimensions combined in MAEM. The best results tested in the Tiny Person dataset have been for the combinations of 20,16,15, and 10.

This ablation experiment has been conducted on the Tiny Person dataset to demonstrate the effects of the MAEM, MFFM, and attention loss on the model performance. The results are listed in Table 3.

MAEM has been employed to improve the feature extraction of the tiny targets in the large field-of-view and reduce the interference generated by the complex environments, thus improving detection directly with the base feature map by 0.78% . MFFM optimizes the delivery of suitable features from the deep to shallow levels and improves the detection directly with the base feature map by 1.29% . Attention loss improves the performance by compensating for the shortcomings of the model's focus on the pixel-level classification errors, with the addition of attention loss that improves the model by 1.14% .

TABLE 2 Ablation study on patch size.

Patch Size	AP_{50}^{tiny}
5,5,5,5	57.51
10,6,5,3	57.96
15,10,8,5	59.13
20,16,15,10	59.82

4.5 Comparison to state-of-the-art methods

Table 4 shows the results of the method proposed in this paper relative to the other detection methods, including FasterRCNN

TABLE 3 Ablation study on whole network.

MAEM	MFFM	Attention loss	AP_{50}^{tiny}
×	×	×	55.54
✓	×	✓	56.32
×	✓	✓	56.83
✓	✓	×	58.68
✓	✓	✓	59.82

TABLE 4 Comparison of different methods on TinyPerson.

Method	AP_{50}^{tiny}	AP_{50}^{tiny1}	AP_{50}^{tiny2}	AP_{50}^{tiny3}	AP_{50}^{small}	AP_{25}^{tiny}	AP_{75}^{tiny}	Flops	Params
FasterRCNN	51.26	35.73	56.08	61.25	66.96	71.13	6.46	$2*10^{11}$	$51.1*10^{6}$
RetinaNet	52.43	41.38	57.68	60.54	66.85	76.32	6.73	$5.5*10^{11}$	$31.4*10^{6}$
FoveaNet	54.18	37.97	57.27	64.84	71.63	74.81	7.12	$5.4*10^{11}$	$36.0*10^{6}$
Swin-T	55.54	42.41	58.32	64.56	68.53	74.00	7.03	$2.1*10^{11}$	$44.7*10^{6}$
Yolox	57.12	44.37	59.51	66.03	70.11	76.28	8.13	$1.9*10^{11}$	$54.2*10^{6}$
PVTv2	56.38	43.15	58.73	65.39	69.82	75.12	7.82	$2.1*10^{11}$	$43.2*10^{6}$
SSPNet	58.35	46.37	60.91	66.92	71.90	77.76	8.48	$4.6*10^{11}$	$46.8*10^{6}$
OUR	**59.82**	**47.59**	**61.74**	**68.43**	**72.23**	**78.09**	**9.21**	$2.1*10^{11}$	$51.6*10^{6}$

Bold values indicates the best detection result in each column.

TABLE 5 Comparison of different methods on WSODD.

Method	mAP_{50}^{small}	mAP_{50}^{medium}	mAP_{50}^{large}	mAP_{50}^{max}
FasterRCNN	12.6	17.1	31.6	51.2
SSD	15.3	18.4	28.6	52.8
Yolov3	23.1	26.6	**41.5**	**55.9**
RetinaNet	18.3	20.1	33.8	52.8
CenterNet	10.4	24.1	29.8	43.5
Swin-T	25.8	26.4	32.7	53.6
OUR	**27.1**	**28.2**	40.3	53.7

Bold values indicates the best detection result in each column.

(Ren et al., 2017), RetinaNet (Lin et al., 2020), FoveaNet (Kong et al., 2020), Swin-T (Liu et al., 2021b), Yolox (Ge et al., 2021), PVTv2 (Wang et al., 2022) and SSPNet (Hong et al., 2022) on the TinyPerson dataset. Since some of these methods have not been previously applied to the TinyPerson datasets, all the compared methods have been tested while keeping the same configuration. The left subscript of the evaluation metric AP represents the value of the IoU in the target detection, and the right superscript represents the size of the detection target.

According to Table 4, on the TinyPerson dataset, the accuracy of our proposed method on the evaluation index AP_{50}^{tiny} is 8.56% and 7.39% higher than the anchor-based Faster-RCNN and RetinaNet methods, respectively. Further, it is 5.64% higher than the anchor-free FoveaNet method, 3.28% higher than the Transformer-based Swin-T, and 1.53% higher than the latest method SSPNet. The method in this paper shows a significant improvement compared to the baseline model with comparable Flops and Params. The proposed method in this paper achieves the best results in several other evaluation metrics. The comparison of results on the TinyPerson dataset fully illustrates the effectiveness and superiority of our method for the tiny target detection task.

4.6 The performance on water surface object detection dataset

To evaluate the detection performance of our method in marine scenarios, we conducted experiments under the Water Surface Object Detection Dataset (WSODD) (Zhou et al., 2021), as shown in Table 5. Unlike the TinyPerson dataset, the WSODD dataset contains 14 different kinds of water targets, and we selected seven algorithms (Duan et al., 2019) for testing and used mean Average Precision (mAP) as the average evaluation metric to measure the detection accuracy of all categories. In addition, the proportion of images occupied by each class of targets was divided into four parts: small (≤ 10 %), medium (10-20 %), large (20-30 %), and max (≥ 30 %). The test results show that the method proposed in this paper can achieve higher accuracy on small and medium size targets compared with other methods. The visual inspection comparison is shown in Figure 5. It shows the detection results of the baseline model and the proposed method in this paper under the same training strategy with green and red bounding boxes, respectively. According to Figure 5, it can be obtained that compared with the baseline model, our method detects significantly more ships, especially for small targets, and the recognition rate of obscured targets is significantly improved. In general, the results with different datasets show that the method proposed in this paper has good performance for small target detection in marine scenes.

5 Discussion

The main task of the tiny target detection for aerial beach images is the accurate detection and identification of targets with very few visual features of the image. However, equivalent to common scale targets, tiny targets in aerial beach images usually



FIGURE 5
The detection result on the WSODD dataset. The red and green bounding boxes represent the detection results of the baseline and our model, respectively.

lack sufficient appearance information and are difficult to be extracted from the background as the complex scene changes (Cheng et al., 2022). Compared with the other target detection methods, the primary advantage of the proposed model in this paper is the introduction of a MAEM based on the self-attention mechanism. MAEM enables the model to effectively improve the tiny target feature extraction by learning the relationship between the target and the background. Furthermore, based on MFFM, the effectiveness of tiny target feature fusion is improved by dynamically assigning weights and using self-attention to prevent tiny targets from being overwhelmed by high-level semantic information.

Experiment results reveal that certain tiny targets cannot be accurately detected and recognized, when the scale of the target crowd in the image appears extremely high, owing to the serious tiny target occlusion situation. Therefore, exploring the optimization of the model to cope with the tiny target detection in the crowded situation will be an important direction for the subsequent work. In future work, we will also explore the theory of the attention mechanism for tiny target detection.

6 Conclusion

We proposed a multiscale attention-based feature pyramid network model, which is used for tiny target detection of aerial beach images with large field-of-view. The multiscale attention enhancement module (MAEM) in the model generates multiscale attention-guided maps to obtain contextual information while preserving the detailed information of the target. As a result, MAEM improves the ability of tiny target feature extraction in a large field-of-view and guides the model to focus more on tiny targets. The multiscale feature fusion module (MFFM) employs the attention-guided map to obtain the weights of feature maps at different scales, thus giving more semantic information to the lower-level feature maps, and effectively preventing the target from being overwhelmed by the high-level feature information. Therefore, MFFM improves the efficiency of the tiny target feature fusion. The experimental results show that the accuracy tested on the publicly available dataset Tiny Person reached 59.8% , and the ablation experiments also prove the effectiveness of each module. In future work, we will investigate the application of our proposed model to the vision processing tasks such as target traffic counting and multi-target tracking.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/ucas-vg/PointTinyBenchmark/tree/master/dataset.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements. Written informed consent was not obtained from the individual(s), nor the minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

Author contributions

SG designed the method with experiments and wrote the original draft of the manuscript supervised by JQ. HZ has

overseen and led the planning and execution of research experiments. CL has been responsible for the management and coordination of the planning and execution of research activities. ZZ critically reviewed the initial manuscript and provided helpful input. All authors contributed to the article and approved the submitted version.

Funding

This work is partially supported by the National Natural Science Foundation of China (61931003, 61871042, 62171044, 12101061, 61875237), the Natural Science Foundation of Beijing (4222004), QinXin Talents Cultivation Program(Beijing Information Science and Technology University), and Beijing Great Wall Scholars Program (No.CIT&TCD20190323).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Bai, Y., Zhang, Y., Ding, M., and Ghanem, B. (2018a). "Finding tiny faces in the wild with generative adversarial network," in *2018 IEEE/CVF conference on computer vision and pattern recognition* (Salt Lake City, UT, USA: IEEE), 21–30. doi: 10.1109/CVPR.2018.00010

Bai, Y., Zhang, Y., Ding, M., and Ghanem, B. (2018b). "SOD-MTGAN: Small object detection via multi-task generative adversarial network," in *Computer vision – ECCV 2018*, vol. 11217 . Eds. V. Ferrari, M. Hebert, C. Sminchisescu and Y. Weiss (Cham: Springer International Publishing), 210–226. doi: 10.1007/978-3-030-01261-813

Cheng, Y., Xu, H., and Liu, Y. (2021). "Robust small object detection on the water surface through fusion of camera and millimeter wave radar," in *2021 IEEE/CVF international conference on computer vision (ICCV)* (Montreal, QC, Canada: IEEE). doi: 10.1109/ICCV48922.2021.01498

Cheng, G., Yuan, X., Yao, X., Yan, K., Zeng, Q., and Han, J. (2022). Towards large-scale small object detection: Survey and benchmarks. *arXiv*. doi: 10.48550/arXiv.2207.14096

Chen, L., Shi, W., and Deng, D. (2021b). Improved YOLOv3 based on attention mechanism for fast and accurate ship detection in optical remote sensing images. *Remote Sens.* 13, 660. doi: 10.3390/rs13040660

Chen, D., Sun, S., Lei, Z., Shao, H., and Wang, Y. (2021a). Ship target detection algorithm based on improved YOLOv3 for maritime image. *J. Advanced Transportation* 2021, 1–11. doi: 10.1155/2021/9440212

Chen, C., Zhang, Y., Lv, Q., Wei, S., Wang, X., Sun, X., et al. (2019). "RRNet: A hybrid detector for object detection in drone-captured images," in *2019 IEEE/CVF international conference on computer vision workshop (ICCVW)* (Seoul, Korea (South: IEEE), 100–108. doi: 10.1109/ICCVW.2019.00018

Dai, J., Li, Y., He, K., and Sun, J. (2016). R-fcn: Object detection *via* region-based fully convolutional networks. *Adv. Neural Inf. Process. Syst.* 29. doi: 10.48550/arXiv.1605.06409

Deng, C., Wang, M., Liu, L., Liu, Y., and Jiang, Y. (2022). Extended feature pyramid network for small object detection. *IEEE Trans. Multimedia* 24, 1968–1979. doi: 10.1109/TMM.2021.3074273

Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., and Tian, Q. (2019). "CenterNet: Keypoint triplets for object detection," in *2019 IEEE/CVF international conference on computer vision (ICCV)* (Seoul, Korea (South: IEEE), 6568–6577. doi: 10.1109/ICCV.2019.00667

Ge, Z., Liu, S., Wang, F., Li, Z., and Sun, J. (2021). YOLOX: Exceeding YOLO series in 2021. *ArXiv*. doi: 10.48550/arXiv.2107.08430

Ghahremani, A., Bondarev, E., and De With, P. H. (2018). "Cascaded CNN method for far object detection in outdoor surveillance," in *2018 14th international conference on signal-image technology & Internet-based systems (SITIS)* (Las Palmas de Gran Canaria, Spain: IEEE), 40–47. doi: 10.1109/SITIS.2018.00017

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE conference on computer vision and pattern recognition* (Columbus, OH, USA: IEEE), 580–587. doi: 10.1109/CVPR.2014.81

Hong, M., Li, S., Yang, Y., Zhu, F., Zhao, Q., and Lu, L. (2022). SSPNet: Scale selection pyramid network for tiny person detection from UAV images. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5. doi: 10.1109/LGRS.2021.3103069

Kisantal, M., Wojna, Z., Murawski, J., Naruniec, J., and Cho, K. (2019). "Augmentation for small object detection," in *9th international conference on advances in computing and information technology (ACITY 2019)* (Sydney, Australia: Aircc Publishing Corporation), 119–133. doi: 10.5121/csit.2019.91713

Kong, T., Sun, F., Liu, H., Jiang, Y., Li, L., and Shi, J. (2020). FoveaBox: Beyound anchor-based object detection. *IEEE Trans. Image Process.* 29, 7389–7398. doi: 10.1109/TIP.2020.3002345

Lee, S.-J., Roh, M.-I., Lee, H.-W., Ha, J.-S., and Woo, I.-G. (2018). "Image-based ship detection and classification for unmanned surface vehicle using real-time object detection neural networks," in *The 28th international ocean and polar engineering conference* (Sapporo, Japan: OnePetro), 726–730.

Liang, Z., Shao, J., Zhang, D., and Gao, L. (2018). "Small object detection using deep feature pyramid networks," in *Advances in multimedia information processing – PCM 2018*, vol. 11166 . Eds. R. Hong, W.-H. Cheng, T. Yamasaki, M. Wang and C.-W. Ngo (Cham: Springer International Publishing), 554–564. doi: 10.1007/978-3-030-00764-551

Lieshout, C., Oeveren, K., Emmerik, T., and Postma, E. (2020). Automated river plastic monitoring using deep learning and cameras. *Earth Space Sci.* 7, e2019EA000960. doi: 10.1029/2019EA000960

Li, J., Liang, X., Wei, Y., Xu, T., Feng, J., and Yan, S. (2017). "Perceptual generative adversarial networks for small object detection," in *2017 IEEE conference on computer vision and pattern recognition (CVPR)* (Honolulu, HI: IEEE), 1951–1959. doi: 10.1109/CVPR.2017.211

Lin, T.-Y., Dollar, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). "Feature pyramid networks for object detection," in *2017 IEEE conference on computer vision and pattern recognition (CVPR)* (Honolulu, HI: IEEE), 936–944. doi: 10.1109/CVPR.2017.106

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollar, P. (2020). Focal loss for dense object detection. *IEEE Trans. ON Pattern Anal. AND Mach. Intell.* 42, 10. doi: 10.1109/TPAMI.2018.2858826

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). "SSD: Single shot MultiBox detector," in *Computer vision – ECCV 2016*, vol. 9905 . Eds. B. Leibe, J. Matas, N. Sebe and M. Welling (Cham: Springer International Publishing), 21–37. doi: 10.1007/978-3-319-46448-02

Liu, Z., Gao, G., Sun, L., and Fang, Z. (2021a). "HRDNet: High-resolution detection network for small objects," in *2021 IEEE international conference on multimedia and expo (ICME)* (Shenzhen, China: IEEE), 1–6. doi: 10.1109/ICME51207.2021.9428241

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021b). "Swin transformer: Hierarchical vision transformer using shifted windows," in *2021 IEEE/CVF international conference on computer vision (ICCV)* (Montreal, QC, Canada: IEEE), 9992–10002. doi: 10.1109/ICCV48922.2021.00986

Loshchilov, I., and Hutter, F. (2017). Decoupled weight decay regularization. *arXiv*. doi: 10.48550/arXiv.1711.05101

Mehralian, M., and Karasfi, B. (2018). "RDCGAN: Unsupervised representation learning with regularized deep convolutional generative adversarial networks," in *2018 9th conference on artificial intelligence and robotics and 2nd Asia-pacific international symposium* (Kish Island, Iran: IEEE), 31–38. doi: 10.1109/AIAR.2018.8769811

Moon, S. W., Lee, J., Lee, J., Nam, D., and Yoo, W. (2020). "A comparative study on the maritime object detection performance of deep learning models," in *2020 international conference on information and communication technology convergence (ICTC)* (Jeju, Korea (South: IEEE), 1155–1157. doi: 10.1109/ICTC49870.2020.9289620

Na, B., and Fox, G. C. (2018). "Object detection by a super-resolution method and a convolutional neural networks," in *2018 IEEE international conference on big data (Big data)* (Seattle, WA, USA: IEEE), 2263–2269. doi: 10.1109/BigData.2018.8622135

Nayan, A.-A., Saha, J., Mozumder, A. N., Mahmud, K. R., and al Azad, A. K. (2020). Real time detection of small objects. *ArXiv*. doi: 10.35940/ijitee.E2624.039520

Noh, J., Bae, W., Lee, W., Seo, J., and Kim, G. (2019). "Better to follow, follow to be better: Towards precise supervision of feature super-resolution for small object detection," in *2019 IEEE/CVF international conference on computer vision (ICCV)* (Seoul, Korea (South: IEEE), 9724–9733. doi: 10.1109/ICCV.2019.00982

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster r-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Ribeiro, R., Cruz, G., Matos, J., and Bernardino, A. (2019). A data set for airborne maritime surveillance environments. *IEEE Trans. Circuits Syst. Video Technol.* 29, 2720–2732. doi: 10.1109/TCSVT.2017.2775524

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). Imagenet large scale visual recognition challenge. *Int. J. Comput. Vision* 115, 211–252. doi: 10.1007/s11263-015-0816-y

Shamsolmoali, P., Chanussot, J., Zareapoor, M., Zhou, H., and Yang, J. (2022a). Multipatch feature pyramid network for weakly supervised object detection in optical remote sensing images. *IEEE Trans. Geosci. Remote Sens.* 60, 1–13. doi: 10.1109/TGRS.2021.3106442

Shamsolmoali, P., Zareapoor, M., Chanussot, J., Zhou, H., and Yang, J. (2022b). Rotation equivariant feature image pyramid network for object detection in optical remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* 60, 1–14. doi: 10.1109/TGRS.2021.3112481

Shen, W., Qin, P., and Zeng, J. (2019). "An indoor crowd detection network framework based on feature aggregation module and hybrid attention selection module," in *2019 IEEE/CVF international conference on computer vision workshop (ICCVW)* (Seoul, Korea (South: IEEE), 82–90. doi: 10.1109/ICCVW.2019.00016

Soloviev, V., Farahnakian, F., Zelioli, L., Iancu, B., Lilius, J., and Heikkonen, J. (2020). "Comparing CNN-based object detectors on two novel maritime datasets," in *2020 IEEE international conference on multimedia & expo workshops (ICMEW)* (London, UK: IEEE), 1–6. doi: 10.1109/ICMEW46912.2020.9106019

Tian, L., Cao, Y., He, B., Zhang, Y., He, C., and Li, D. (2021). Image enhancement driven by object characteristics and dense feature reuse network for ship target detection in remote sensing imagery. *Remote Sens.* 13, 1327. doi: 10.3390/rs13071327

Varga, L. A., Kiefer, B., Messmer, M., and Zell, A. (2022). "SeaDronesSee: A maritime benchmark for detecting humans in open water," in *2022 IEEE/CVF winter conference on applications of computer vision (WACV)* (Waikoloa, HI, USA: IEEE), 3686–3696. doi: 10.1109/WACV51458.2022.00374

Wang, X., Girshick, R., Gupta, A., and He, K. (2018). "Non-local neural networks," in *2018 IEEE/CVF conference on computer vision and pattern recognition* (Salt Lake City, UT, USA: IEEE), 7794–7803. doi: 10.1109/CVPR.2018.00813

Wang, W., Xie, E., Li, X., Fan, D.-P., Song, K., Liang, D., et al. (2022). PVT v2: Improved baselines with pyramid vision transformer. *Comput. Visual Media* 8, 415–424. doi: 10.1007/s41095-022-0274-8

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). "CBAM: Convolutional block attention module," in *Computer vision – ECCV 2018*, vol. 11211 . Eds. V. Ferrari, M. Hebert, C. Sminchisescu and Y. Weiss (Cham: Springer International Publishing), 3–19. doi: 10.1007/978-3-030-01234-21

Yu, X., Gong, Y., Jiang, N., Ye, Q., and Han, Z. (2020a). "Scale match for tiny person detection," in *2020 IEEE winter conference on applications of computer vision (WACV)* (Snowmass Village, CO, USA: IEEE), 1246–1254. doi: 10.1109/WACV45572.2020.9093394

Yu, X., Han, Z., Gong, Y., Jan, N., Zhao, J., Ye, Q., et al. (2020b). "The 1st tiny object detection challenge: Methods and results," in *Computer vision – ECCV 2020 workshops*, vol. 12539 . Eds. A. Bartoli and A. Fusiello (Cham: Springer International Publishing), 315–323. doi: 10.1007/978-3-030-68238-5_23

Yu, F., and Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv*. doi: 10.48550/arXiv.1511.07122

Zhang, T., Cao, S., Pu, T., and Peng, Z. (2021). AGPCNet: Attention-guided pyramid context networks for infrared small target detection. *arXiv*. doi: 10.48550/arXiv.2111.03580

Zhou, Z., Sun, J., Yu, J., Liu, K., Duan, J., Chen, L., et al. (2021). An image-based benchmark dataset and a novel object detector for water surface object detection. *Front. Neurorobotics* 15. doi: 10.3389/fnbot.2021.723336

Zhu, X., Lyu, S., Wang, X., and Zhao, Q. (2021). "TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios," in *2021 IEEE/CVF international conference on computer vision workshops (ICCVW)* (Montreal, BC, Canada: IEEE), 2778–2788. doi: 10.1109/ICCVW54120.2021.00312

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 14 December 2022
DOI 10.3389/fmars.2022.1058201

Check for updates

OPEN ACCESS

EDITED BY
Ran Liao,
Tsinghua University, China

REVIEWED BY
Dingtian Yang,
South China Sea Institute of Oceanology, Chinese Academy of Sciences (CAS), China
Qingsheng Xue,
Ocean University of China, China

*CORRESPONDENCE
Xiangqian Quan
quanxq@idsse.ac.cn

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 30 September 2022
ACCEPTED 28 November 2022
PUBLISHED 14 December 2022

CITATION
Quan X, Wei Y, Liu K and Li B (2022) The analysis and design of deep-sea lighting field based on spectral transfer function.
Front. Mar. Sci. 9:1058201.
doi: 10.3389/fmars.2022.1058201

COPYRIGHT
© 2022 Quan, Wei, Liu and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The analysis and design of deep-sea lighting field based on spectral transfer function

Xiangqian Quan[1]*, Yucong Wei[1,2], Kaibin Liu[1] and Bo Li[1]

[1]Institute of Deep-sea Science and Engineering, Chinese Academy of Science, Sanya, China, [2]University of Chinese Academy of Sciences, Beijing, China

Due to the attenuation of light in water, the deep-sea optical imaging system needs an active lighting system to provide the light source. However, because of the nonlinearity of light attenuation in spatial dimension and spectral dimension, the deep-sea lighting differs from terrestrial lighting. In order to quantitatively analyze and design deep-sea lighting system, we proposed a precise deep-sea lighting field simulation model and design method based on spectral transfer function. Firstly, with the analysis of deep-sea lighting-imaging process, the spectral transfer function in lighting field was analyzed and the deep-sea lighting model was built. Then, the platform used to study light attenuation was set up and the attenuation characteristics of light in water were derived. Moreover, the deep-sea lighting field simulation model was built with the computer program. Finally, the experiment platform for testing the underwater lighting field was set up in test pool. The experimental results show that the deep-sea lighting field computational model is accurate. In addition, the optimal deep-sea lighting system design was proposed. This study provides the theoretical basis and experimental data for the design of a deep-sea lighting system which has far-reaching significance for improving the efficiency of deep-sea scientific research.

KEYWORDS

underwater optics, underwater imaging, underwater lighting, marine observation, object recognition

1 Introduction

As the "eye" of deep-sea working platforms such as human-occupied vehicles, remote-operated vehicles and deep-sea landers, etc., deep-sea optical imaging systems can obtain a large amount of information in deep-sea exploration. There is no doubt that deep-sea optical imaging is an important technology for deep-sea resource exploration, deep-sea mineral development, and deep-sea *in situ* biological and chemical activities' observation (Garcia et al., 2013; Jaffe, 2016; Mullen et al., 2016; Giddens et al., 2021; Feng et al., 2022; Zhou et al., 2022). However, because the propagation of light in water is

attenuated with an "exponential" form, the illumination of visible light at a 200m depth is usually less 0.01% than that at the sea surface and the deep ocean environment below 1000m depth is dark. Therefore, deep-sea lighting is a prerequisite for deep-sea optical imaging. However, the nonlinearity of light attenuation in the spatial dimension and spectral dimension makes designing a deep-sea lighting system more difficult (Kocak et al., 2008; Garcia et al., 2013; Jaffe, 2015; Mao et al., 2017; Du et al., 2020).

Many experts and scholars have carried out a series of relevant research on a deep-sea lighting system. To acquire clear deep-sea images, Tommy D. Dickey et al. developed a multi-angle scattering and absorption detector to analyze the attenuation characteristics of light in different sea areas with different water qualities. Then, a theoretical model for analyzing and designing deep-sea lighting system was proposed (Dickey et al., 2011). Ron Gray et al. analyzed the advantages and disadvantages of parabolic, ellipsoidal and spherical as light distribution elements in a deep-sea lighting system and the impacts of a smooth surface, ground-glass surface and diffuse reflective surface on illumination uniformity in the deep-sea lighting system. Then, a series of design schemes for light distribution in a deep-sea lighting system were proposed (Gray, 2004). Considering the high color rendering of white light and high transmittance of blue light in water, Mark S. Olsson et al. integrated a white-light light-emitting diode (LED) and blue light LED in one deep-sea lighting lamp, where two types of light distribution structures were equipped accordingly. The blue light LED was suitable for long-distance and small-field-of-view imaging. The white light LED was suitable for short-range and large-field-of-view imaging (Olsson, 2012). Jules S. Jaffe et al. utilized the structured illumination for deep-sea imaging. The contrast and detection range were enhanced (Jaffe, 2010). Peter C. Y. Chang et al. improved the visibility depth of underwater imaging by use of polarization; the results showed an improvement in contrast and visibility depth (Chang et al., 2003). To detect in situ underwater microplastics, Huang H et al. analyzed the light absorption and scattering in the lighting field for signal correction and calibration. This study is essential for hyperspectral images in a water environment (Huang et al., 2021). To acquire high-resolution seabed images at various depths for scientific research, Jevgenij Guls et al. developed a suitable lighting module for still images and video (Guls et al., 2016).

The above research provides a theoretical basis for deep-sea lighting. However, the attenuation of light is not only related to the light path but also related the wavelength of light. The spectral transfer function in deep-sea lighting process should be considered. Therefore, we proposed a precise analysis model and design method for the deep-sea lighting field based on the spectral transfer function. Firstly, by analyzing the spectral transform function in the deep-sea lighting-imaging process, several key factors affecting deep-sea lighting were considered and a universal deep-sea lighting field computational model based on spectral transfer function was built. Then, the research platform used to study light attenuation was built and the attenuation of light characteristics in water was derived. Moreover, we built the deep-sea lighting field analysis model with a computer program. Finally, an experiment platform for testing the underwater lighting field was set up in the test pool. The experiment results confirmed the accuracy of the deep-sea lighting field computational model and put forward an optimal design for a deep-sea lighting system. This study is significant in fields such as deep-sea artificial intelligence recognition, deep-sea real-time image analysis, deep-sea video-image color correction, deep-sea spectral imaging, deep-sea laser imaging, deep-sea optical communication, deep-sea *in situ* optical detection, etc. It also has far-reaching significance for improving the efficiency of deep-sea scientific research and reducing the cost of deep-sea exploration (Yang and Yang, 2009; O'Byrne et al., 2018; Marques et al., 2019; Han et al., 2019; Tong et al., 2021; Xue et al., 2021).

2 Principle and theory

2.1 The analysis of spectral transfer function in deep-sea lighting-imaging process

The essence of optical imaging is photoelectric conversion. The response function of imaging system $EV(\lambda)$ is shown in Equation 1:

$$EV(\lambda) = t_{integration} \times \frac{I(\lambda) \times \pi\left(\frac{f}{F}\right)^2}{C_{pixel}} \times QE(\lambda) \tag{1}$$

where λ is the wavelength, $t_{integration}$ is the integration time. $I(\lambda)$ is the illumination function of lighting system. f/F is the entrance pupil aperture. F is the F number, f is the focal length, C_{pixel} is the size of the camera pixel and $QE(\lambda)$ is the quantum efficiency. The difference between deep-sea lighting and terrestrial lighting is the light energy transmission process. To quantitatively analyze the illumination distribution of the deep-sea lighting system, we expand the illumination function from $I(\lambda)$ to $I(\lambda)(x,y)$ on the target surface:

$$\begin{cases} I(\lambda)(x,y) = \Phi(\lambda)_0 \cos(\alpha) \cdot e^{-\mu(\lambda)\left(\sqrt{(x-X-L\tan\theta)^2+(y-Y-L\tan\varphi)^2+L^2}+\sqrt{x^2+y^2+L^2}\right)}/S \\ \alpha = arctan\left(\frac{\sqrt{(x-X-L\tan\theta)^2+(y-Y-L\tan\varphi)^2}}{L}\right) \end{cases} \tag{2}$$

where (x,y) is the coordinate on the target surface, $\Phi(\lambda)_0$ is the luminous function of lamp (the luminous function conforms to Lambert's principle), α is the emission angle of light from lamp, L is the distance from the lamp surface to target surface, θ is the angle of light in vertical direction and φ is the angle of light

in horizontal direction, $\mu(\lambda)$ is the attenuation coefficient of different wavelengths, and S is the illumination range. To analyze and design a deep-sea lighting system, the attenuation characteristics of light in water should be studied first.

2.2 The study of light attenuation in water

To analyze and design the deep-sea lighting system, the light attenuation was studied. The experimental setup for analyzing light attenuation coefficient was built as in Figure 1. The light penetrated through the chopper (which, combined with the lock-in amplifier, aims to improve the measurement accuracy) to the monochromator (which aims to decompose the light into monochromatic light). Then, the light passed through the monochromator to the fiber. The light from fiber was collimated by the collimator lens. By penetrating the sample pool, the light struck to the detector of the photo-multiplier tube (PMT).

Analyzing the experimental data in situations of a vacant tank, tap water pool and seawater pool showed that the relationship between the transmittance and wavelength was as shown in Figure 2A. The relationship between the attenuation coefficient and wavelength is shown in Figure 2B. The experimental results indicate that the attenuation of light changes with wavelength in the form of nonlinearity, and the attenuation of light in the blue and green bands is smaller than that in the violet and red bands. In addition, the attenuation coefficient of light in seawater is higher than that in tap water. The design of deep-sea lighting is related to the attenuation characteristics of light in water. Due to the feasibility of the

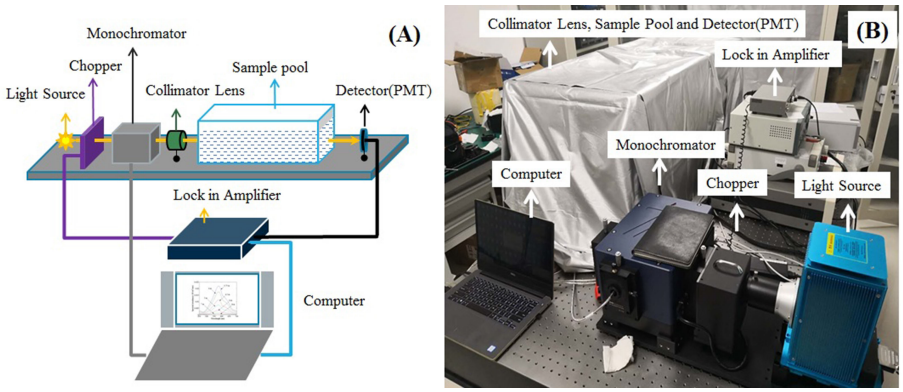


FIGURE 1
The experimental setup of light attenuation analysis: **(A)** the schematic diagram of experimental setup; **(B)** the photo of experimental setup.

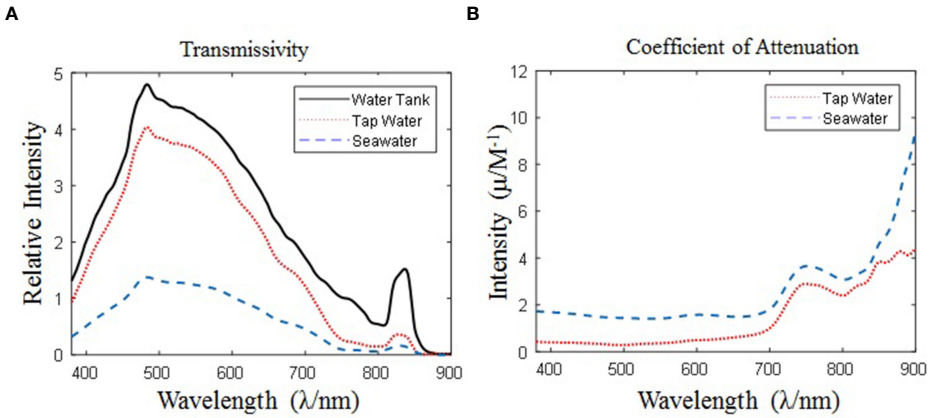


FIGURE 2
The analysis of light attenuation: **(A)** the transmittance of light in the vacant tank, tap water, and seawater; **(B)** the attenuation coefficient of light in tap water and seawater.

experimental verification, we carried out the simulation and experiment with the attenuation characteristics of tap water to verify the reliability of the deep-sea lighting field computational model and used the model to predict the results of the deep-sea lighting design with different water constituents.

3 Simulation

Based on the analysis of light attenuation and deep-sea lighting field computational model, we carried out the simulation design of deep-sea lighting with the computer program. The schematic diagram of deep-sea lighting system design is shown as Figure 3. Three types of lamp, such as warm-light LED, cold-light LED and blue-light LED, were employed in our simulation and experiment. We added the warm-light LED to counteract the heavy attenuation of the red wavelength band. We utilized the cold-light lamp to improve the color rending of deep-sea images combine with the warm-light LED under different distances, and we utilized the blue-light LED to extend the detection range and decrease the energy consumption because of the high transmittance of blue light. The coordinate system was established with the camera as the origin. We installed the lamps separately from the camera to decrease the back-scattering light. If the lamps were installed close to the camera, the imaging quality would deteriorate and affect the results of the experiment. The position coordinates of two warm LED lamps were (20cm, 0) and (-20cm, 0), respectively. The position coordinates of two cold LED lamps were (61cm, 61cm) and (-61cm, 61cm), respectively. The position coordinates of two blue LED lamps were (122cm, 122cm) and (-122cm, 122cm), respectively. The angle of blue light LED lamps was 30° inward in the horizontal direction and vertical direction. The angle of cold light LED lamps was 15° inward in the horizontal direction and vertical direction. Each lamp has a luminous flux of 10000 lm, a beam angle of 75° and a power of 110 W.

According to the deep-sea application scenarios, the lighting distances of the deep-sea lighting system were set to 3m, 6m and 10m. With the lighting distance of 3m, the illumination simulation results of each LED lamps are shown in Figure 4. The illumination mean value of each LED ranged from 43.5 LUX to 124.9 LUX. The illumination uniformities of warm-light LED lamps are better than that of cold-light and blue-light LED lamps.

With a lighting distance of 3m, the illumination distribution of the whole lighting system is shown in Figure 5.

The illumination data of the whole lighting system with a distance of 3m are shown in Table 1. The maximum illumination value is 624.7 LUX, the minimum illumination value is 368.4 LUX, and the mean illumination value is 533.7 LUX. The illumination uniformity is 0.69. According to the real scene lighting situation, illumination may cause overexposure.

The illumination simulation results of each LED lamp with a lighting distance of 6m are shown in Figure 6. The illumination mean value of each LED lamp ranges from 9.7 LUX to 44.4 LUX. The illumination uniformities of warm-light LED lamps are better than those of cold-light and blue-light LED lamps.

With a lighting distance of 6m, the illumination distribution of the whole lighting system is shown in Figure 7.

The illumination data of the whole lighting system with a distance of 6m are shown in Table 2. The maximum illumination value is 233.6 LUX, the minimum illumination value is 177.2 LUX, and the mean illumination value is 214.4 LUX. The illumination uniformity is 0.83. According to the real situation of scene lighting, the illumination meets the requirements.

The illumination simulation results of each LED lamp with a lighting distance of 10m are shown in Figure 8. The illumination mean value of each LED lamp ranges from 3.73

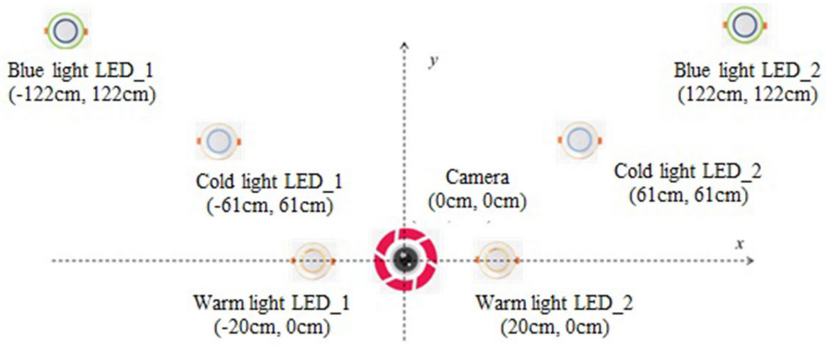


FIGURE 3
The schematic diagram of deep-sea lighting system design.

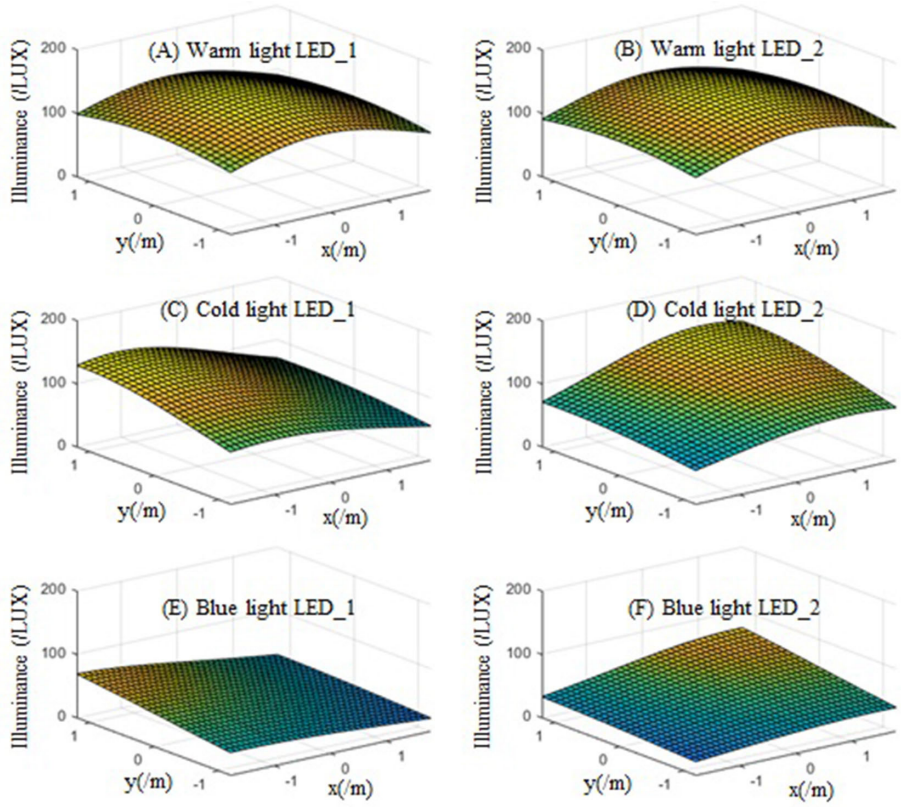


FIGURE 4
The illumination distributions of each single LED lamp with a distance of 3m: **(A)** the illumination distribution of warm light LED_1 on target surface; **(B)** the illumination distribution of warm light LED_2 on target surface; **(C)** the illumination distribution of cold light LED_1 on target surface; **(D)** the illumination distribution of cold light LED_2 on target surface; **(E)** the illumination distribution of blue light LED_1 on target surface; **(F)** the illumination distribution of blue light LED_2 on target surface.

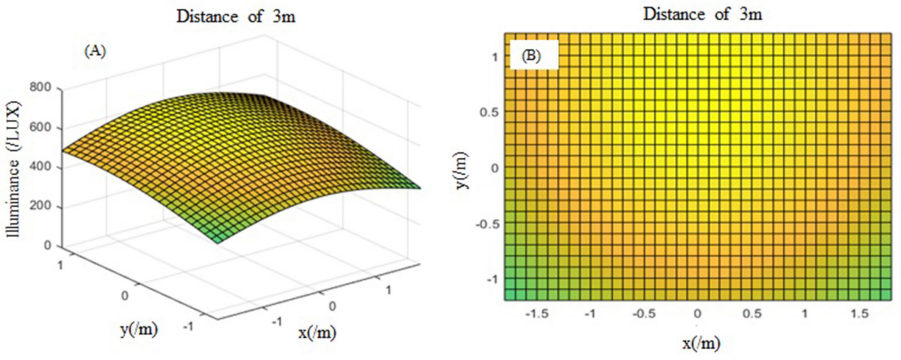


FIGURE 5
The illumination distribution of whole lighting system with a lighting distance of 3m. **(A)** the 3D view of illumination distribution; **(B)** the vertical view of illumination distribution.

10.3389/fmars.2022.1058201

TABLE 1 The illumination distribution analysis with a lighting distance of 3m.

Maximum value (LUX)	Minimum value (LUX)	Mean value (LUX)	Uniformity
624.7	368.4	533.7	0.69

LUX to 13.8 LUX. The illumination uniformities of warm-light LED lamps are better than those of cold-light and blue-light LED lamps.

With a lighting distance of 10m, the illumination distribution of the whole lighting system is shown in Figure 9.

The illumination data of the whole lighting system at a distance of 10m are shown in Table 3. The maximum illumination value is 60.2 LUX, the minimum illumination value is 50.4 LUX, and the mean illumination value is 56.7 LUX. The illumination uniformity is 0.89. According to the real situation of scene lighting, the illumination cannot meet the requirements.

4 Experiment and result

4.1 The confirmatory experiment

In order to verify the simulation model and design the deep-sea lighting system, an experiment platform was set up in the test pool, as shown in Figure 10. Figure 10A shows the experimental process. Figure 10B shows the diffuse reflective plate with the size of 3.6m ×1.8m, which is located in the experimental pool. Figure 10C shows the deep-sea lighting system platform, which consists of lamps, a power supply, control panel, camera, monitor and storage, etc. Figure 10D shows the captured

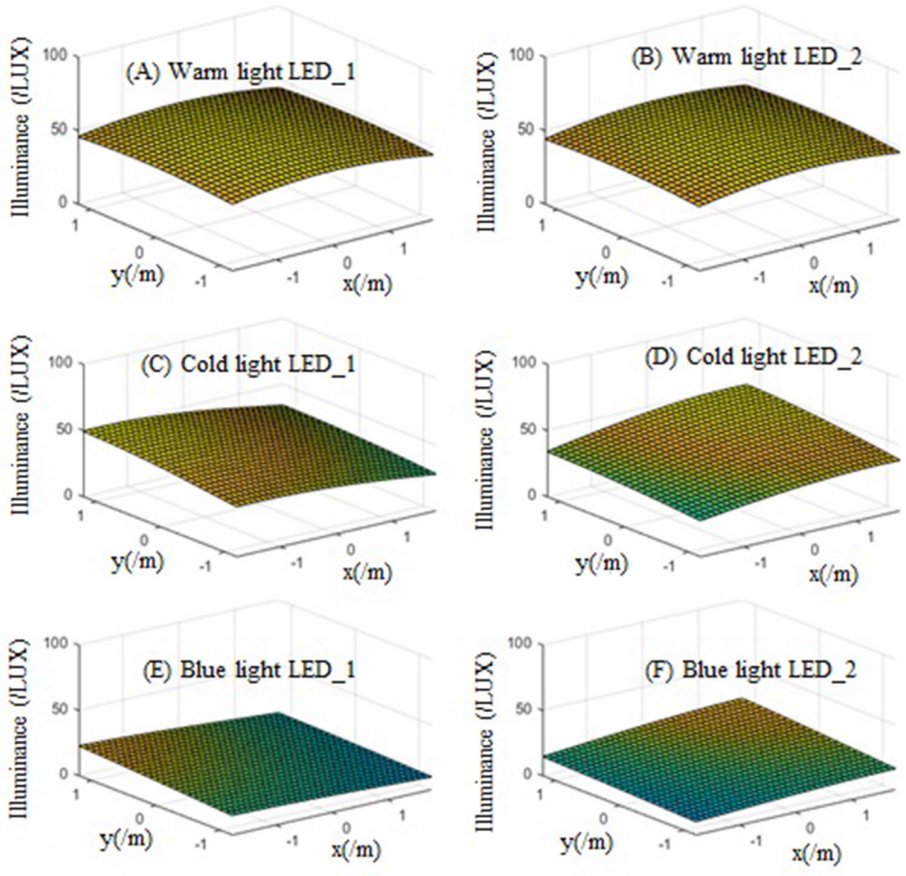


FIGURE 6
The illumination distribution of each single LED lamp with a distance of 6m: **(A)** the illumination distribution of warm-light LED_1 on target surface; **(B)** the illumination distribution of warm-light LED_2 on target surface; **(C)** the illumination distribution of cold-light LED_1 on target surface; **(D)** the illumination distribution of cold-light LED_2 on target surface; **(E)** the illumination distribution of blue-light LED_1 on target surface; **(F)** the illumination distribution of blue-light LED_2 on target surface.

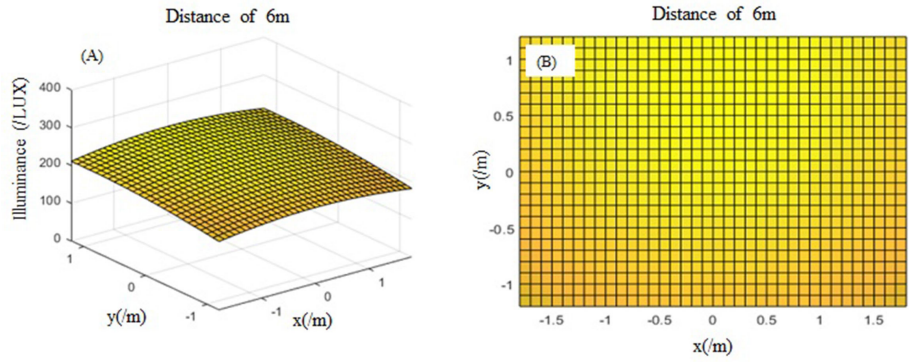


FIGURE 7
The illumination distribution of whole lighting system with a lighting distance of 6m. **(A)** the 3D view of illumination distribution; **(B)** the vertical view of illumination distribution.

experiment pictures. Based on the picture data, the deep-sea lighting system can be quantitatively analyzed.

The experiment was carried out in a dark environment. The diffuse reflective plate was placed on the bottom of the testing pool; the deep-sea lighting system platform is hang up with the crane. By controlling the crane, the distance between the diffuse reflective plate and lighting system can be adjusted. The deep-sea lighting control system consisted of the power system, monitor, and control panel. With the electric cable and optical cable, the lighting system platform can be controlled. There were three types of lamp, including two warm-light lamps, two cold-light lamps and two warm-light lamps. Each lamp had a luminous flux of 10000 lm, a beam angle of 75° and a power of 110 W. By adjusting the lighting lamp power and the lighting distance, the illumination data on the target surface could be derived. The experimental results are shown in Figure 11. The lighting distance was set to 3m, 6m and 10m, respectively. The power of the lighting system was set to 20%, 40%, 60%, 80% and 100%, respectively. Figure 11A shows the raw experiment data. Figure 11B shows the illumination distribution in the red wavelength band. Figure 11D shows the illumination distribution in green wavelength band. Figure 11D shows the illumination distribution in blue wavelength band. The experiment shows that 1) the illumination on target surface decreases with an increase in lighting distance; 2) the illumination on the target surface increases with an increase in lighting LED power; 3) the illumination distribution in the middle of target surface is brighter than that at the edge of the target surface; 4) the uniformity of illumination increases with an increase in lighting distance; 5) the bright spots in the picture are caused by the back-scattered light in the water; and 6) the illumination on the target surface with a lighting distance of 3m is oversized and the pictures are overexposed. This overexposure exists in red, blue and green wavebands. The optimal lighting power for red light is 80%, for green light this is 40%, and for blue light it is 60%, respectively. Overall, the optimal power setting is 40%. 7) the illumination on the target surface with a lighting distance of 6m is suitable and the picture is acceptable. Overexposure exists in the green waveband. The optimal lighting power for red light is 100%, for green light is 80%, and for blue light is 100%, respectively. Overall, the optimal power setting is 80%. 8) the illumination on the target surface with a lighting distance of 10m is low and the picture is underexposed. The optimal lighting power for red light, green light and blue light is 100%. However, this setting, with the power setting of 100%, cannot meet the lighting acquirements. Based on the analysis, the experimental results are consistent with the simulation results.

4.2 The deep-sea lighting field design

After verifying the reliability of the deep-sea lighting field calculation model, we designed the deep-sea lighting based on the attenuation characteristics of light in deep-sea, as in Figure 2. We adopted the same lamp configuration as above. The simulation results are shown in Table 4. This shows that the optimal imaging distance is 3 M. With changes in the imaging distance, the optical power needs to be adjusted accordingly. The

TABLE 2 The illumination distribution analysis with a lighting distance of 6m.

Maximum value (LUX)	Minimum value (LUX)	Mean value (LUX)	Uniformity
233.6	177.2	214.4	0.83

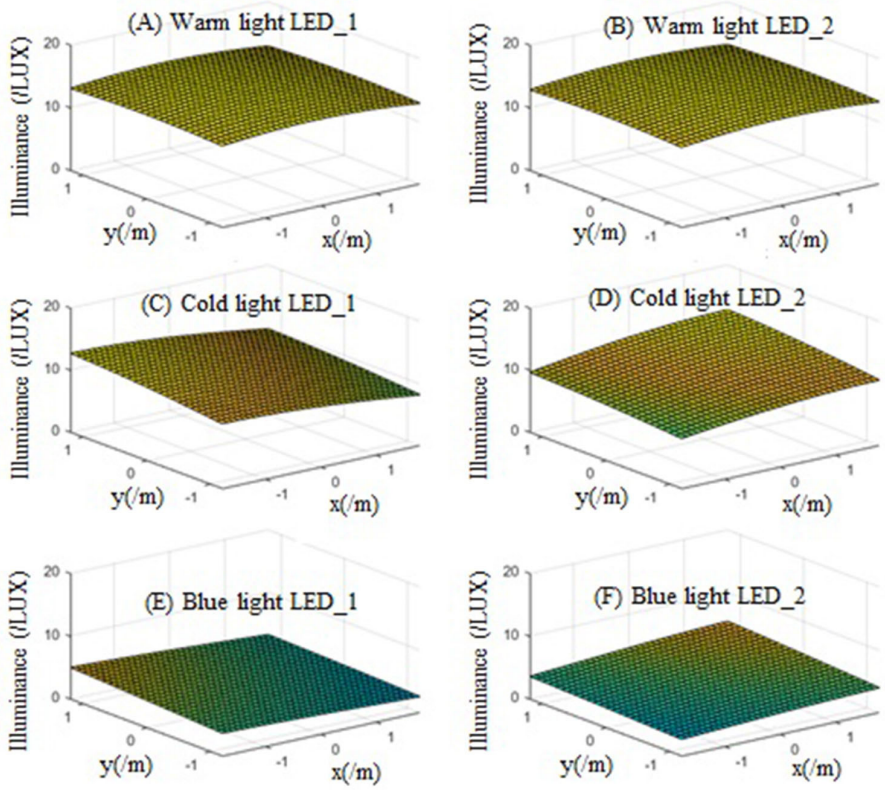


FIGURE 8
The illumination distribution of each single LED lamp with a distance of 10m: **(A)** the illumination distribution of warm-light LED_1 on target surface; **(B)** the illumination distribution of warm-light LED_2 on target surface; **(C)** the illumination distribution of cold-light LED_1 on target surface; **(D)** the illumination distribution of cold-light LED_2 on target surface; **(E)** the illumination distribution of blue-light LED_1 on target surface; **(F)** the illumination distribution of blue-light LED_2 on target surface.

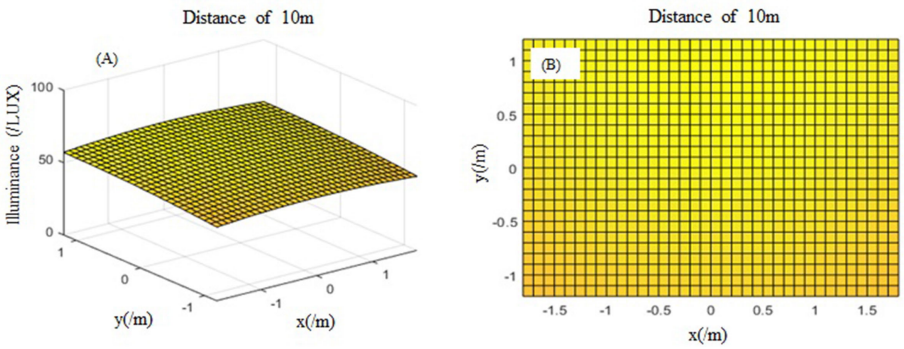


FIGURE 9
The illumination distribution analysis of whole-lighting system with a lighting distance of 10m. **(A)** the 3D view of illumination distribution; **(B)** the vertical view of illumination distribution.

TABLE 3 The illumination distribution analysis with a lighting distance of 10m.

Maximum value (LUX)	Minimum value (LUX)	Mean value (LUX)	Uniformity
60.2	50.4	56.7	0.89

optimal luminous flux under different distances is proposed in the last column of Table 4. With the increase in distance, the illuminance value on the target plane decreases and the uniformity of illuminance gradually increases.

5 Conclusion

Based on the experiment and simulation results, the following conclusions can be drawn: 1) When designing deep-sea lighting system, the attenuation characteristics of different wavelengths, as well as the spectral characteristics of different light sources, should be considered. 2) The simulation model has higher reliability for deep-sea lighting field simulation models, as that based on the spectral response function can overcome the influence of different wavelength attenuations under different water quality environments. 3) Generally, the red wavelength light decays more than blue light. If high-color-rendering pictures are required, the components of warm light should be increased. If an increased detection range and energy conservation are required, the components of blue light should be increased. 4) Due to the nonlinearity of light attenuation in water, the deep-sea lighting field design is also nonlinear. The deep-sea lighting field design should be carried out according to specific environments.

6 Discussion

Deep-sea lighting is a prerequisite for deep-sea optical imaging. It is also the technical basis for a series of deep-sea optical detection scenes. However, there is a huge difference between deep-sea lighting and terrestrial lighting. This paper proposed a deep-sea lighting field simulation model based on spectral response function and an optimal deep-sea lighting design. Firstly, by analyzing the spectral transfer function in the deep-sea lighting-imaging process, several key factors affecting deep-sea lighting were considered and a universal deep-sea lighting field computational model based on spectral transfer function was built. Then, the platform for studying the light attenuation was set up and the light attenuation

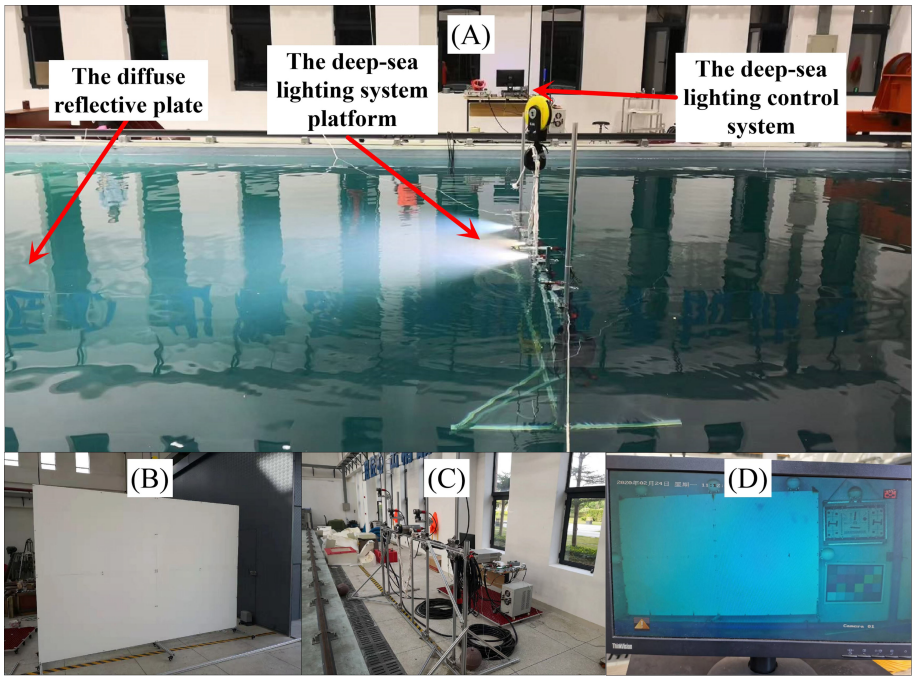


FIGURE 10
The experiment platform for deep-sea lighting system testing. **(A)** the experimental process; **(B)** the diffuse reflective plate with the size of 3.6m ×1.8m; **(C)** the deep-sea lighting system platform; **(D)** the captured experiment pictures.

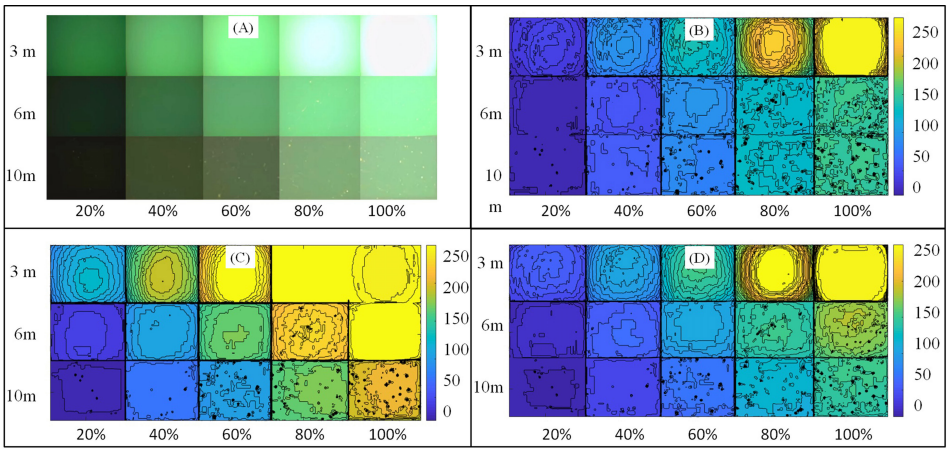


FIGURE 11
The experimental results of deep-sea lighting design. **(A)** the raw experiment picture data; **(B)** the illumination distribution in red wavelength band; **(C)** the illumination distribution in green wavelength band; **(D)** the illumination distribution in blue wavelength band.

TABLE 4 The illumination distribution analysis in deep-sea environment.

Distance	Maximum value (LUX)	Minimum value (LUX)	Mean value (LUX)	Uniformity	Optimum light flux ($\times 10^5$ lm)
1M	748.3	244.0	443.2	0.33	2.7
2M	466.0	165.1	343.9	0.48	3.5
3M	265.1	126.4	214.2	0.59	5.6
4M	147.0	82.9	124.6	0.67	9.6
5M	80.8	50.5	70.5	0.72	17.0
6M	44.2	29.7	39.3	0.75	30.5
7M	24.1	17.0	21.7	0.78	55.3
8M	13.2	9.6	12.0	0.81	100.0
9M	7.2	5.4	6.6	0.82	181.8
10M	3.9	3.0	3.6	0.84	333.3

characteristics were derived. Moreover, the deep-sea lighting field was quantitatively analyzed with a computer program. Finally, an experimental platform was set up to test the underwater lighting system. The result shows that the deep-sea lighting field computational model is accurate. The model can simulate different situations with different lamps, different angles, different lamp array positions, different water qualities and differen\t imaging distances. This study is significant in fields such as deep-sea artificial intelligence recognition, deep-sea real-time image analysis, deep-sea video image color correction, deep-sea spectral imaging, etc. It also has far-reaching significance for improving the efficiency of deep-sea scientific research and reducing the cost of deep-sea exploration.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

Conceptualization, XQ; methodology, XQ and YW; analysis, XQ, KL, BL and YW; manuscript writing, XQ, BL, KL and YW. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the Hainan Provincial Natural Science Foundation of China (No. 2019RC258), the national key research and development program of China (No. 2018YFC0307903), the Youth Innovation Promotion Association CAS (No. 2020361), and the Scientific Instrument Developing Project of the Chinese Academy of Sciences (No. YJKYYQ20180067).

Acknowledgments

The authors thank Kelin Sun, Jingchuan Yang, Bing Zhang and Chen Li (Institute of Deep-sea Science and Engineering, Chinese Academy of Science) for their assistance during the experiment and test.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Chang, P. C. Y., Flitton, J. C., Hopcraft, K. I., Jakeman, E., Jordan, D. L., and Walker, J. G. (2003). Improving visibility depth in passive underwater imaging by use of polarization. *Appl. Optics* 42 (15), 2794. doi: 10.1364/AO.42.002794

Dickey, T. D., Kattawar, G. W., and Voss, K. J. (2011). Shedding new light on light in the ocean. *Phys. Today* 64 (4), 44–49. doi: 10.1063/1.3580492

Du, R., Yang, D., Jiang, G., Song, Y., and Yin, X. (2020). An approach for *In situ* rapid detection of deep-Sea aromatic amino acids using laser-induced fluorescence. *Sensors* 20 (1330), 1–13. doi: 10.3390/s20051330

Feng, J.-C., Yan, J., Wang, Y., Yang, Z., Zhang, S., Liang, S., et al. (2022). Methane mitigation: Learning from the natural marine environment. *Innovation.* 3 (5), 100297. doi: 10.1016/j.xinn.2022.100297

Garcia, R., Nicosevici, T., and Cufi, X. (2002). [*IEEE Oceans 2002 Conference and Exhibition. - Biloxi, MI, USA (29-31 Oct. 2002)] Oceans \"02 MTS/IEEE - On the way to solve lighting problems in underwater imaging[J]* 2, 1018–1024.

Giddens, J., Turchik, A., Goodell, W., Rodriguez, M., and Delaney, D. (2021). The national geographic society deep-Sea camera system: A low-cost remote video survey instrument to advance biodiversity observation in the deep ocean. *Front. Mar. Sci.* 7. doi: 10.3389/fmars.2020.601411

Gray, R. (2004). Light sources, lamps and luminaires-Ronan Gray (Deepsea power & light) provides a brief introduction to a few important basics of underwater illumination. *Sea Technol.* 45 (12), 39–44. doi: 10.1016/j.oceaneng.2004.07.003

Guls, J., Bjarnason, ÓI, Pétursson, Ó, Einarsson, S. Ö., and Foley, J. T. (2016). Application of axiomatic design in designing autonomous underwater photography lighting. *Proc. CIRP* 53, 278–283. doi: 10.1016/j.procir.2016.07.009

Han, B., Zhao, W., Zheng, Y., Meng, J., Wang, T., Han, Y., et al. (2019). Experimental demonstration of quasi-omni-directional transmitter for underwater wireless optical communication based on blue LED array and freeform lens. *Optics Commun.* 434 (2019), 184–190. doi: 10.1016/j.optcom.2018.10.037

Huang, H., Sun, Z., Liu, S., Di, Y., and Wu, J. (2021). Underwater hyperspectral imaging for in situ underwater microplastic detection. *Sci. Total Environ.* 776 (3), 145960. doi: 10.1016/j.scitotenv.2021.145960

Jaffe, J. S. (2010). Enhanced extended range underwater imaging *via* structured illumination. *Optics Express* 18 (12), 12328–12340. doi: 10.1364/OE.18.012328

Jaffe, J. S. (2015). Underwater optical imaging: the past, the present, and the prospects. *IEEE J. Oceanic Eng.* 40 (3), 683–700. doi: 10.1109/JOE.2014.2350751

Jaffe, J. (2016). To sea and to see: That is the answer. *Methods Oceanography* 15-16, 3–20. doi: 10.1016/j.mio.2016.05.003

Kocak, D. M., Dalgleish, F. R., Caimi, F. M., and Schechner, Y. Y. (2008). A focus on recent developments and trends in underwater imaging. *Mar. Technol. Soc. J.* 42 (1), 52–67. doi: 10.4031/002533208786861209

Mao, X., Xu, S., Hu, X., and Xie, Y. (2017). Design of a smooth freeform illumination system for a point light source based on polar-type optimal transport mapping. *Appl. Optics* 56 (22), 6324–6331. doi: 10.1364/AO.56.006324

Marques, T. P., Albu, A. B., and Hoeberechts, M. (2019). A contrast-guided approach for the enhancement of low-lighting underwater images. *J. Imaging.* 5 (10), 79. doi: 10.3390/jimaging5100079

Mullen, A. D., Treibitz, T., Roberts, P. L. D., Kelly, E. L. A., Horwitz, R., et al. (2016). Underwater microscopy for *in situ* studies of benthic ecosystems. *Nat. Commun.* 7 (1), 1–9. doi: 10.1038/ncomms12093

O'Byrne, M., Schoefs, F, Pakrashi, V, and Ghosh, B. I. (2018). An underwater lighting and turbidity image repository for analysing the performance of image-based non-destructive techniques. *Structure Infrastructure Eng.* 14 (1), 1–20. doi: 10.1080/15732479.2017.1330890

Olsson, M. S. (2012). United States Patent Application Publication.

Tong, Z., Yang, X., Chen, X., Zhang, H., Zhang, Y., Zou, H., et al. (2021). Quasi-omnidirectional transmitter for underwater wireless optical communication systems using a prismatic array of three high-power blue LED modules. *Optics Express* 29 (13), 20262–20274. doi: 10.1364/OE.429974

Xue, Q., Tian, Z., Bai, Y., Liang, J., Li, C., Wang, F., et al. (2021). Underwater hyperspectral imaging system using a Prism-Grating-Prism structure. *Appl. Optics* 60 (4), 894–900. doi: 10.1364/AO.415351

Yang, D., and Yang, C. (2009). Detection of seagrass distribution changes from 1991 to 2006 in xincun bay, hainan, with satellite remote sensing. *Sensors* 9 (2), 830–844. doi: 10.3390/s90200830

Zhou, J., Huang, H., Huang, S. H., Si, Y., Shi, K., Quan, X., et al. (2022). AUH, a new technology for ocean exploration. *Engineering.* doi: 10.1016/j.eng.2022.09.007

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 11 January 2023
DOI 10.3389/fmars.2022.1094915

Check for updates

OPEN ACCESS

EDITED BY
Hong Song,
Zhejiang University, China

REVIEWED BY
Muhammad Abdul Rehman Khan,
Air University, Pakistan
Muhammad Basim,
Sungkyunkwan University, South Korea
Danish Khan,
Zhejiang University, China
Muwei Jian,
Shandong University of Finance and Economics, China

*CORRESPONDENCE
Taejoon Kim
ktjcc@chungbuk.ac.kr
Inam Ullah
inam.fragrance@gmail.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 10 November 2022
ACCEPTED 19 December 2022
PUBLISHED 11 January 2023

CITATION
Khan S, Ullah I, Ali F, Shafiq M, Ghadi YY and Kim T (2023) Deep learning-based marine big data fusion for ocean environment monitoring: Towards shape optimization and salient objects detection. *Front. Mar. Sci.* 9:1094915. doi: 10.3389/fmars.2022.1094915

COPYRIGHT
© 2023 Khan, Ullah, Ali, Shafiq, Ghadi and Kim. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Deep learning-based marine big data fusion for ocean environment monitoring: Towards shape optimization and salient objects detection

Sulaiman Khan[1†], Inam Ullah[2*†], Farhad Ali[3], Muhammad Shafiq[4,5], Yazeed Yasin Ghadi[6] and Taejoon Kim[7*]

[1]College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar, [2]BK21 Chungbuk Information Technology Education and Research Center, Chungbuk National University, Cheongju, Republic of Korea, [3]Department of Accounting and Information Systems, College of Business and Economics, Qatar University, Doha, Qatar, [4]Cyberspace Institute of Advanced Technology, Guangzhou University, Guangzhou China, [5]Department of Computer Science, Adjunct Professor at Shenyang Normal University, Shenyang, China, [6]Department of Computer Science, Al Ain University, Abu Dhabi, United Arab Emirates, [7]School of Information and Communication Engineering, Chungbuk National University, Cheongju, Republic of Korea

Objective: During the last few years, underwater object detection and marine resource utilization have gained significant attention from researchers and become active research hotspots in underwater image processing and analysis domains. This research study presents a data fusion-based method for underwater salient object detection and ocean environment monitoring by utilizing a deep model.

Methodology: A hybrid model consists of an upgraded AlexNet with Inception v-4 for salient object detection and ocean environment monitoring. For the categorization of spatial data, AlexNet is utilized, whereas Inception V-4 is employed for temporal data (environment monitoring). Moreover, we used preprocessing techniques before the classification task for underwater image enhancement, segmentation, noise and fog removal, restoration, and color constancy.

Conclusion: The Real-Time Underwater Image Enhancement (RUIE) dataset and the Marine Underwater Environment Database (MUED) dataset are used in this research project's data fusion and experimental activities, respectively. Root mean square error (RMSE), computing usage, and accuracy are used to

construct the model's simulation results. The suggested model's relevance form optimization and conspicuous item prediction issues in the seas is illustrated by the greatest accuracy of 95.7% and low RMSE value of 49 when compared to other baseline models.

KEYWORDS

data fusion, marine big data, ocean environment, underwater saliency detection, underwater image processing

1 Introduction

Oceans cover about 70 percent of the earth's surface and hold enriched natural resources that endeavor in long-term human development. With the emergence of advanced information and communication systems, the researchers paid significant attention to the exploration of mysterious areas of the ocean, marine exploitation, and observation. However, facing relatively hashed and unconstrained marine scenes, there are legion austere inimical factors such as water with high turbidity, dreary color, uneven illustration, and a perilous underwater atmosphere that gravely compromises the accuracy and accessibility of underwater photos in real-world contexts (Jian et al., 2021). To tackle these issues, numerous research reports have been published on how to effectively perform underwater image and visionary processing tasks. Qiao et al. (2022), proposed a novel generative adversarial network (GAN) architecture to address color distortion or under/overexposure problems in underwater images. Their model generates high-quality enhanced images to address the non-uniform illumination problems in images. Moghimi and Mohanna systematically analyzed the extant data and identified different real-time underwater image enhancement models proposed in the literature (Moghimi and Mohanna, 2021). Halimi et al. present an effective technique for boosting underwater image resolution by concurrently reconstructing the reflectivity and complexity of the image using the greatest marginal probability estimate (Halimi et al., 2017).

The aim of salient object recognition, which makes use of image/video segmentation, is to pinpoint the most eye-catching and aesthetically attractive objects or regions in a picture (Wang et al., 2018), image foreground annotation (Cao et al., 2016), image quality assessment (Gu et al., 2016), and video summarization (Cong et al., 2019). In-depth research has been carried out utilizing very superior artificial intelligence (AI) algorithms to identify saliency in photographs of natural settings during the past few years. It is imperative to note that salient object identification differs greatly from studies on anomaly detection and conventional object detection. First of all, unlike anomaly detection and salient object recognition, which only commit to discovering locally notable objects, object detection is a comprehensive job that focuses on detecting everything. Salient object detection produces a saliency probability map at the pixel level as opposed to object detection and anomaly detection, which always identify items with bounding boxes. The researchers used a range of algorithms to discover significant details in images and videos, such as Lee et al. (2018), proposal of a manipulated deep learning model for saliency detection using a convolutional neural network (CNN) and GoogLeNet. This model uses both low-level and high-level features. Google is used to gather high-level data, while CNN architecture is used to extract low-level attributes. By comparing a local region's differences from other areas in an image, these attributes were utilized to assess a region's relative importance. Zhang et al. (2019), proposed a pipelined model for detecting salient objects in underwater photos, using deformable convolutional networks. Using the CNN approach, scientists first reduced noise in underwater photographs to improve contrast. Increase the feature extraction capabilities by implementing a deformable position-sensitive ROI pooling method with RPN and rebuilding the ResNet-101 feature extraction sub-network by using a deformed convolution model.

After studying the literature, it was concluded that most of the models either had a low identification rate, especially in high turbidities, or were highly time-consuming and required more data for the testing and validation process. Our research work addresses these problems with the following key objectives:

- To develop a hybrid deep learning model consists of upgraded AlexNet with Inception v-4 for salient object detection and ocean environment monitoring. AlexNet is considered for the spatial data classification while Inception v-4 is exploited for temporal data (environment monitoring).
- To perform semantic-based data fusion of two publicly available datasets (MUED and RUIE) and perform training and validation tasks of the hybrid model.
- To accurately identify saliencies and perform shape optimization using this hybrid model and fused database. An overall identification rate of 95.7% is

achieved by this model. After comparing its performance with benchmark techniques, our model outperformed even on smaller dataset.

The remainder of this research article is organized by introducing the representative models on different underwater image processing and data fusion-based models in Section 2. Section 3 introduces the experimental setup exploited to accomplish this research work. Results and discussion are outlined in Section 4 of the paper followed by the conclusion in Section 5.

2 Literature review

This portion of the study summarizes recent research reports in the subject area under consideration and several data fusion-based models that have been proposed by academics for deployment in various deep learning application domains.

2.1 Underwater salient objects detection

The congenital capabilities of humans to percept, efficiently perceive, and distinguish salient objects in videos or images keep them ahead of machines (Li et al., 2019). Saliency detection is used in the machine vision and underwater image processing sectors to provide computers the capacity of people to analyze underwater images, which also plays a key role in discovering marine resources. Consequently, many underwater saliency detection models have been developed freshly.

The problem of identifying the prominent components from underwater photos using conventional saliency detection techniques is made more difficult by the uncertainty of underwater settings. The intricacy and diversity of undersea ecosystems have recently sparked a lot of study attention. Xu et al. (2019), proposed Generalized robust principal component analysis (GRPCA) as a unique method for underwater target detection. This model functions by extracting visual characteristics from underwater pictures that have specific goal in recognition and images representation. –, developed an effective underwater target layered background framework based on the visual information perception and processing of a frog's eye, which can distinguish prominent objects from the background of a picture with object contour. Chen et al. (2017), suggested employing a monocular vision sensor to recognize underwater objects using a prominent object identification approach. This technique reduced background noise in order to increase underwater detection accuracy (Ullah et al., 2019; Khan et al., 2019; Su et al., 2020). Jian et al. (Jian et al., 2021; Jian et al., 2021), suggested forward and backward cues for saliency objects detection using visual spatial temporal features. For object localization and tracking they defined a weighted centroid calculation algorithm for center prior generation and tracked it in scene images. A real-time neural network architecture is used for saliencies detection in frames.

2.2 Data fusion background

Data fusion has drawn a lot of interest in data mining and is being used in a variety of research fields, including both civilian and military applications including monitoring animal habitats, risks identification, surveillance, and espionage operations. Data fusion techniques were initially employed to combine data into a single, featured dataset that corresponded to the converted matrix (Maragos et al., 2008). The fusion process is now carried out through data fusion techniques, which collect data characteristics from several sources. For instance, the most basic type of data fusion is merging two one-dimensional datasets. Additionally, it may be done by combining the semantics of the data. In accordance with the characteristics of the data and the machine learning approach employed for the fusion process, many data fusion algorithms produce different optimization solutions. Zheng et al. (Zheng, 2015), categorized data fusion methods into three different classes (feature-based fusion, semantic-based fusion, and stage-based fusion), as shown in Figure 1.

When adopting feature-based data fusion, the same dimensional data characteristics are directly combined after being collected from various sources, and then they are evaluated using machine learning models. Before the direct catenation of data features, careful consideration is given to the elimination of duplicate records before the merging process; second, keep the same dimensions for all the records in the fused database because model performance can be adversely affected during the merging process if you do not; and third, over-fitting is another significant challenge to face during the training process. In the stage-based data fusion models, the characteristics of the data are classified into several classes, and the data from each class is subsequently examined and combined appropriately (Zhu et al., 2018). The foundation of the semantic meaning-based data fusion method is the semantics of the data. The technique of semantic-based data fusion divides data semantics into four categories: transfer learning data fusion, multi-view data fusion, similarity-based data fusion, and probabilistic dependency data fusion. Using data derived from several sources, data values are obtained and assessed in a multi-view-based data fusion system. Co-training, multi-kernel learning, and subspace learning procedures are subcategories of this fusion process. Khan et al. (2021), performed semantic-based data fusion for traffic monitoring and flow predictions in smart cities.

3 Experimental setup

As shown in Figure 2, this study used a hybrid deep learning model made up of Inception v-4 and AlexNet. The AlexNet and

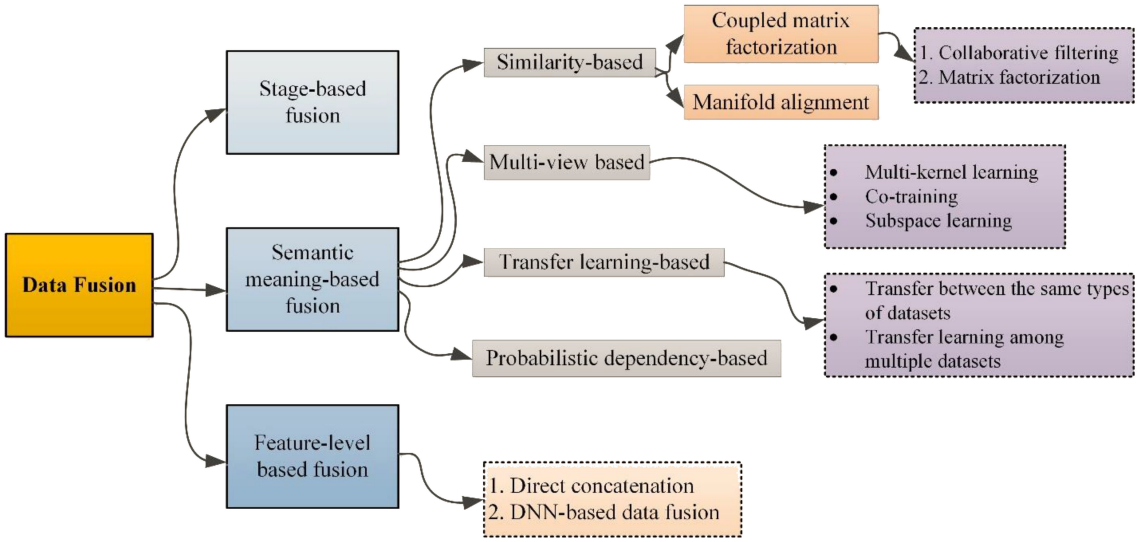


FIGURE 1
Different data fusion techniques.

inception-v4 architecture are merged in this research project. The revised AlexNet structure and the addition of the Inception-v4 module increase the network's reprocessing capacity. The batch normalization layer (BN) is also used to increase generalization abilities, hasten convergence, and stop the gradient from fading. During the training process, mean μ and variance σ is calculated using (Eq.1) and (Eq.2) to normalize each sample in the batch. The normalized bath values are calculated using (Eq.3).

$$\mu = \frac{1}{N}\sum_{i=1}^{k} x_i \tag{1}$$

$$\mu = \frac{1}{N}\sum_{i=1}^{k} (x_i - \mu)^2 \tag{2}$$

$$\widehat{x_i} = \frac{(x_i - \mu)}{\sqrt{(\sigma^2 - \epsilon)}} \tag{3}$$

where N represents total number of samples in a batch and xi represents the input samples. The BN layer's convolution output for a two-dimensional image input is (N, C, W, H), where W and H stand for the dimension of the feature map and C stands for the number of output channels. Afterward, each batch sample may be expressed separately as $x_{c,\ w,\ h}$. Each sample is individually normalized by the BN, thus the resultant number of μ is also $C \times W \times H$.

The BN layer has the ability to manage gradient explosion, restrict the gradient from vanishing, inhibit overfitting, and accelerate network training and convergence. Figure 2 shows how the inception-v4 model's structure was employed in the

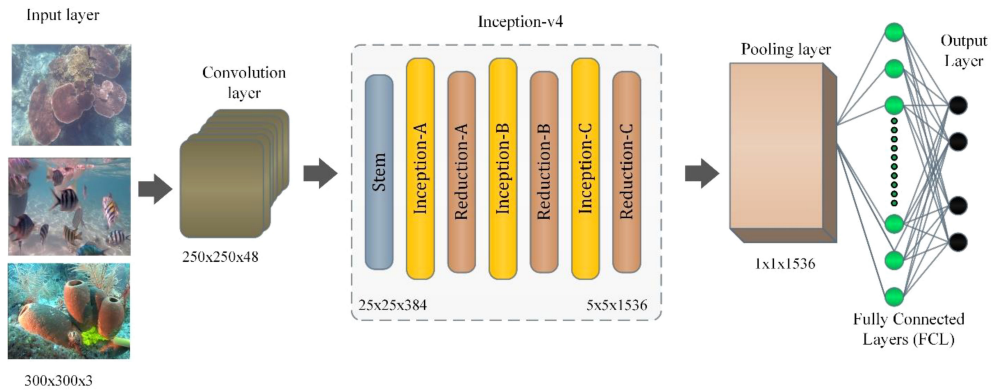


FIGURE 2
Experimental setup.

proposed research project. Input data (images) are provided at the input layer. The number of hidden layers can be defined based on the nature of the research problem and hardware specifications. A huge amount of data and many hidden layers require more simulation time for a simple hardware combination (normal computers with no GPU installed). The output layers, or fully connected layers (FCL), generate the output based on the information generated from hidden layers. It can also be termed the "SoftMax layer." Different activation functions like rectified linear unit (ReLu) and tanh are used for validation purposes. This research work uses transfer learning with five hidden layers. The convolution layer calculates texture information from the input image. Following then, the Inception-V4 continues to feature extraction as a communications system, which is made up of several convolutional and pooling procedures.

The Inception-X module, where X stands for A, B, and C, improves feature utilization by teaching image features through multiple concurrent feature transferring mechanisms. Reduction-X shrinks large feature maps into small feature maps while increasing the number of channels, with X denoting the A and B modules. This research work proposes monocular vision sensor to recognize underwater objects using a prominent object identification approach. This technique reduced background noise in order to increase underwater detection accuracy. This technique avoids high computational complexity without experiencing a significant information loss (Tian et al., 2021). The average pooling layer therefore decreases the number of parameters while increasing the model's robustness while also lowering the deviation from the computed mean. Additionally, the dropout layer is used in the two completely connected layers to prevent the overfitting issues, in which a certain number of neurons are briefly removed from the network during training. The results of the SoftMax regression are finally transferred to the (0, 1) probability interval *via* the output layer (Hang et al., 2019). Jian et al. (2021), reviewed multiple feature extraction and analysis techniques for underwater image processing. They also presented a benchmark underwater image database for identifying the strengths and weaknesses of the existing algorithms for underwater saliencies and images. This database offers offer unparalleled opportunities to researchers in the underwater vision and beyond (Jian et al., 2019). Moreover, the authors in (Ullah et al., 2017; Ullah et al., 2019; Ahmad et al., 2021; Yasir et al., 2022) suggested manipulated feature extraction techniques for object detection underwater, on sea surfaces, etc. When numerous parallel convolution paths are employed, the number of social parameters is reduced. The connectivity with deepening layers can accomplish a similar (or superior) performance with fewer parameters than the connection without deepening layers. The first layer just has to focus on learning the most recent information and can learn successfully with less training data. Feature information may be separated into tiers by growing the network, improving learning efficiency.

3.1 Data acquisition

In this research, we use two different databases that are publicly available for simulations and experimental work, as depicted in Table 1. These two databases are fused together to evaluate and train the model from two different perspectives simultaneously. A few images of the underwater objects are depicted in Figure 3.

3.2 Data fusion

After selecting underwater image databases, the same dimensional data is fused using (Eq. 4). All these varying input values must be multiplied with different weights accordingly. The resultant input map can be obtained by fusing the data as depicted in (Eq. 4).

$$X_{output} = (W_{turb} \times X_{turb}) + \left(W_{pose} \times X_{pose}\right) + \left(W_{variety} \times X_{variety}\right) \quad (4)$$

Where "×" is applied for element wise multiplication and $\boldsymbol{X}_{turb}$, $\boldsymbol{X}_{pose}$, $\boldsymbol{X}_{variety}$ represent the input turbidity of water, diversity in pose, and diversity in variety, while $\boldsymbol{W}_{turb}$, $\boldsymbol{W}_{pose}$, $\boldsymbol{W}_{variety}$ depict the learning parameters classifying numerous impact degree of these factors. The finalized value at t^{th} time

TABLE 1 Databases for underwater image processing.

Database	Type of data	Providers
MUED (Jian et al., 2019)	This collection includes 8600 underwater pictures of 430 distinctive prominent items against noisy backgrounds. It is also quite different in terms of position, variety, stance, water turbidity, etc.	1. Ocean University of China, Qingdao 2. Shandong University of Finance and Economics
RUIE (Liu et al., 2020)	More than 4000 underwater genuine photographs are available in this database, which is divided into three groups based on their classification: underwater image quality sub-aggregate, underwater color cast sub-aggregate, and underwater higher-level task-driven sub-aggregate.	Dalian University of Technology



FIGURE 3
Some sample images of MUED database with a couple of subclasses.

interval is depicted by $\hat{X}_t$. For normalizing the output values a specific range (1, –1) is defined using the hyperbolic function depicted in (Eq. 5).

$$\hat{X}_t = tanh(X_{output}) \quad (5)$$

This fused dataset is used for training and testing purposes. The highest accuracy rate of 95.7% is calculated for salient objects detection in underwater images.

3.3 Training and validation of the model

Different input and performance parameters used for the validation of the proposed hybrid model are depicted in Table 2.

The simulation results are calculated in Python using TensorFlow and Keras libraries for the recognition and classification of reviews. After validating using these different performance metrics an overall accuracy rate of 95.7% is achieved, which reflects the applicability of this hybrid data fusion-based model for the identification of saliencies.

4 Results and discussions

The experimental results and underlined discussion are briefly outlined in this section of the paper. To test the applicability of the proposed algorithm different performance metrics are used that are given below.

4.1 Varying training and test sets

The proposed hybrid deep learning model's recognition characteristics are evaluated using a variety of training and test sets. Figure 4 illustrates that the accuracy results grow together with the training set as can be seen. The suggested model has the greatest accuracy rate, which is 95.7%. This model's ability to recognize important things in underwater photos and video and carry out shape improvement is demonstrated by the high accuracy rate.

TABLE 2 Different parameters used for the validation and experimental work in the proposed research work.

S. No	Parameters	Values
1.	Number of inputs	11800
2.	Activation function	ReLu and tanh
3.	Epoch size	25
4.	Filter	5 × 5
5.	Number of hidden layers	5
6.	Dropout	0.2
7.	Batch size	32 × 32
8.	Optimizer	Adam
9.	Performance metrics	ROC curve, AUC values, precision, accuracy, true-false values, varying training and test sets, miss-classification rate, simulation time.

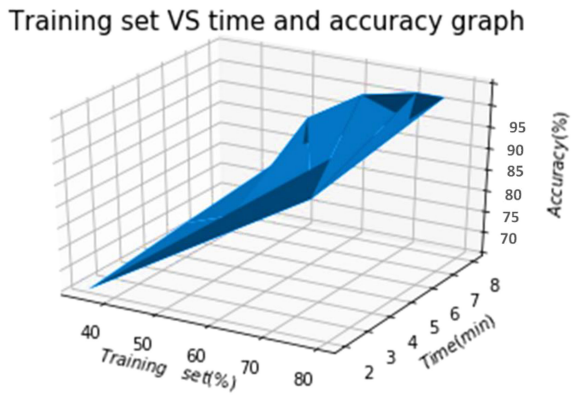


FIGURE 4
Recognition abilities based on varying training and test sets.

4.2 Varying epoch size

To evaluate the performance of the proposed hybrid deep learning system based on different epoch size. In this case an epoch size of 25 is selected for the evaluation purposes. The corresponding results of the model for the sentiment analysis are depicted in Figure 5.

From Figure 5, it is depicted that the graph contains a comparatively lower over-fitting values, which reflects the applicability of the model for the selected sentiment analysis problem.

4.3 ROC curve

A classification model's performance is graphically depicted as a receiver operating characteristic curve (ROC curve), which is valid for all classification thresholds. This graph shows the two parameters:

- True Positive Rate
- False Positive Rate

True Positive Rate (TPR) is an alternate term (synonym) for recall and can be mathematically represented in (Eq. 6) as follows:

$$TPR = \frac{TP}{TP + FN} \tag{6}$$

False Positive Rate (FPR) can be mathematically represented in equation (7) as follows:

$$FPR = \frac{FP}{FP + TN} \tag{7}$$

At different categorization criteria, TPR vs. FPR are shown on a ROC curve. As more items are classified as positive when the classification threshold is lowered, both False Positives and True Positives rise. The hybrid model's proposed ROC curve is shown in Figure 6.

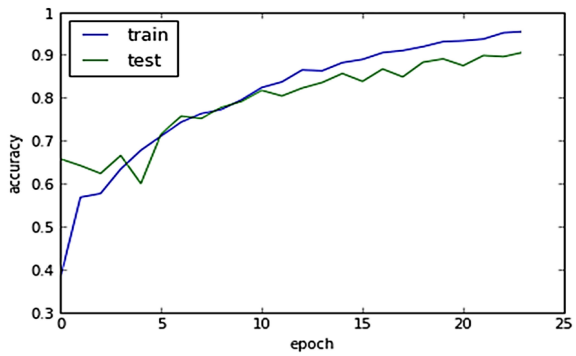


FIGURE 5
Performance evaluation of the proposed model using varying epoch size.

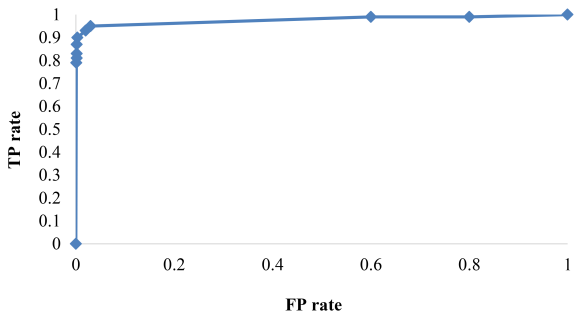


FIGURE 6
Performance evaluation of the proposed hybrid model using ROC curve.

For validating the proposed research problem, the hybrid model is validated using different well-known baseline models reported in the proposed field. These baseline models include:

- **ARIMA** – Auto-Regressive Integrated Moving Average (ARIMA) is extensively used for time series identification problems.
- **Conv1D Net** – Significantly suggested in the extant for temporal dependent problems.
- **LSTM** (Khan and Nazir, 2022) – Long Short-Term Memory (LSTM) models are typically used for sequential data analysis and time prediction problems.
- **St-Res Net** – A sophisticated machine - learning model was developed by Zhang et al. (2017),, When compared to other algorithms, this one delivers correct outcomes for crowd and time forecasts.

The root mean square error rate (RMSE) value is selected as a validating point for comparing these benchmark algorithms with the proposed research work. The generic followed for the RMSE value calculation is depicted in Eq. (8) as follows:

$$RMSE = \sqrt{\frac{1}{X*T}\sum_i (x_i - \bar{x}_i)^2} \quad (8)$$

where T depicts the time interval while X represents the number of inputs observed during the corresponding iteration. x_i represents the predicted value (positive, negative, or neutral response of the user), while $\bar{x}_i$ is the original. The RMSE-based performance results are depicted in Figure 7 below.

The proposed model's lowest RMSE value when compared to the chosen benchmark methods demonstrates the usefulness of the presented study. The usefulness of the suggested approach for detecting prominent objects in background or underwater photos may also be deduced from this RMSE value.

For comparing the performance of the proposed hybrid model with the selected benchmark models. Accuracy is used as the performance metric in this validation process, which uses

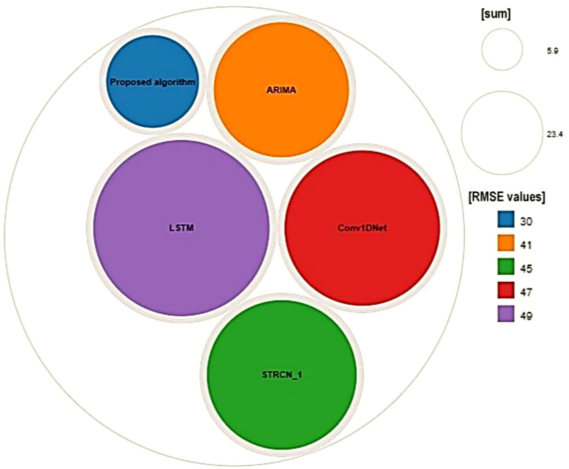


FIGURE 7
RMSE based results of the proposed model with the selected benchmark algorithms.

varying training and test sets. The identification capabilities of different users' emotions are depicted in Figure 8. From Figure 8 it is observed that our model outperformed by producing high identification rates based on varying training and test sets.

4.5 AUC values

The performance of the proposed model is also validated by using the area under the curve (AUC) values based on a fluctuating number of hidden layers. The underlined results are depicted in Figure 9 below.

From Figure 8, one can easily observe that the performance of the proposed model decreases after hidden layer 4 (the maximum AUC value of 95.7 is recorded at h4). This is because of the complexity of the model and the entire high simulation time. So, in our case, a deep learning structure with four hidden layers is an optimum design in our case. Figure 10 shows how much time the suggested hybrid model requires ultimately based on the various hidden layer counts. Figure 10 shows that as the number of hidden layers rises, so does the simulation duration. The circuit complexity and high simulation cost are reflected in the sudden rise in simulation time beyond hidden layer 4.

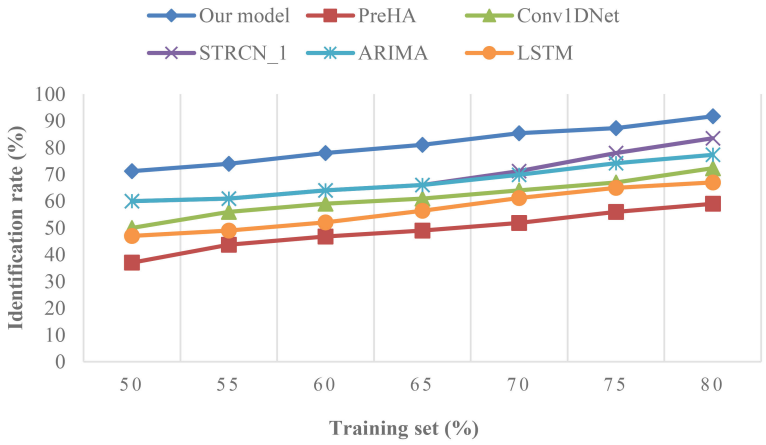


FIGURE 8
Performance analysis of the proposed hybrid model with the selected baseline models.

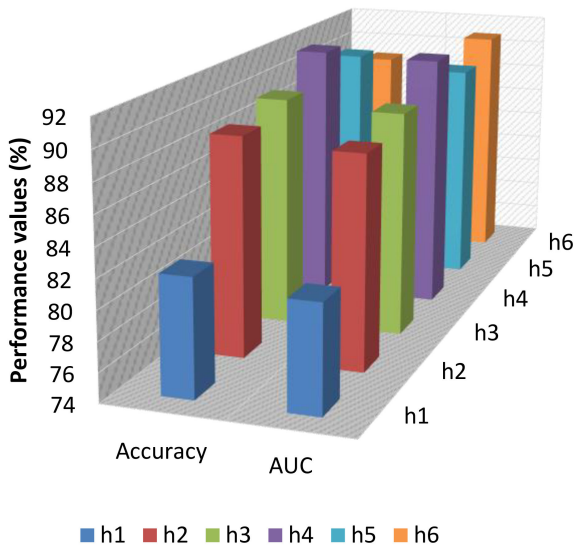


FIGURE 9
AUC values based on the number of hidden layers.

4.6 F-score, precision, and error-rate

The pertinency of the proposed model is also validated using the f-score, precision, and error rate generated. The underlined results are depicted in Figure 11 below.

The small miss-classification rate and high accuracy, precision, recall, and F-score values represented in Figure 11 reflects the pertinency of the proposed hybrid model for the optimum identification of saliencies in underwater images.

5 Conclusion

In this research, we present a hybrid deep learning model for salient object recognition in underwater photos that combines improved AlexNet and Inception v-4. The suggested model can precisely find saliencies with various sizes, water turbidities, and ocean habitats and successfully reduce the crowded backdrops thanks to the data fusion and hybrid architectural design. In order to combine data and conduct experiments, this study combines the MUED and RUIE databases. Model simulation

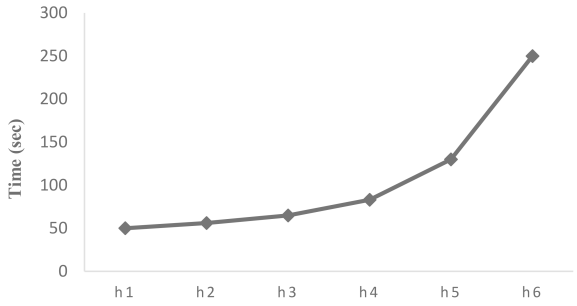


FIGURE 10
Time consumption based on varying number of hidden layers.

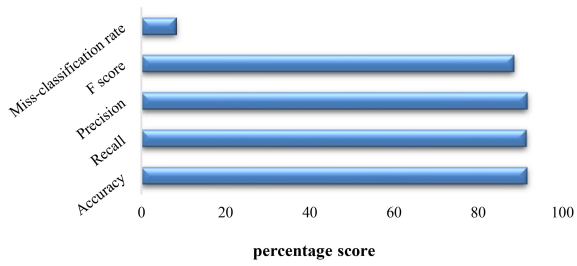


FIGURE 11
Performance evaluation.

results are produced utilizing accuracy, computing usage, and root mean square error (RMSE). The suggested model's suitability for shape optimization and salient item recognition in underwater photos is demonstrated by its high accuracy of 95.7% and low root-mean-square errors (RMSE) values when compared to existing baseline models. The suggested hybrid model's usefulness in the intended study topic is shown by experimental findings. Our model considerably outperformed other benchmark models by producing effective outcomes with minimal training data.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

Conceptualization, SK; methodology, SK, and IU; software, SK, and FA; validation, SK, IU, and FA; formal analysis, IU, and MS; investigation, SK; resources, IU, and FA; data curation, MS, YG, and TK; writing—original draft preparation, SK; and IU; writing—review and editing, IU, and FA; visualization, IU, YG, and TK; supervision, IU and TK; project administration, IU, and TK; funding acquisition, TK. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1I1A3068305).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ahmad, I., Rahman, T., Zeb, A., Khan, I., Ullah, I., Hamam, H., et al. (2021). Analysis of security attacks and taxonomy in underwater wireless sensor networks. *Wireless Commun. Mobile Computing.* doi: 10.1155/2021/1444024

Cao, X., Zhang, C., Fu, H., Guo, X., and Tian, Q. (2016). Saliency-aware nonparametric foreground annotation based on weakly labeled data. *IEEE Trans. Neural Networks Learn. Syst.* 27, 1253–1265. doi: 10.1109/TNNLS.2015.2488637

Chen, Z., Zhang, Z., Dai, F., Bu, Y., and Wang, H. (2017). Monocular vision-based underwater object detection. *Sensors* 17, 1784. doi: 10.3390/s17081784

Cong, R., Lei, J., Fu, H., Cheng, M. M., Lin, W., and Huang, Q. (2019). Review of visual saliency detection with comprehensive information. *IEEE Trans. Circuits Syst. Video Technol.* 29, 2941–2959. doi: 10.1109/TCSVT.2018.2870832

Gu, K., Wang, S., Yang, H., Lin, W., Zhai, G., Yang, X., et al. (2016). Saliency-guided quality assessment of screen content images. *IEEE Trans. Multimedia* 18, 1098–1110. doi: 10.1109/TMM.2016.2547343

Halimi, A., Maccarone, A., McCarthy, A., McLaughlin, S., and Buller, G. S. (2017). Object depth profile and reflectivity restoration from sparse single-photon data acquired in underwater environments. *IEEE Trans. Comput. Imaging* 3, 472–484. doi: 10.1109/TCI.2017.2669867

Hang, J., Zhang, D., Chen, P., Zhang, J., and Wang, B. (2019). Classification of plant leaf diseases based on improved convolutional neural network. *Sensors* 19, 4161. doi: 10.3390/s19194161

Jian, M., Liu, X., Luo, H., Lu, X., Yu, H., and Dong, J. (2021). Underwater image processing and analysis: A review. *Signal Processing: Image Communication* 91, 116088. doi: 10.1016/j.image.2020.116088

Jian, M., Qi, Q., Yu, H., Dong, J., Cui, C., Nie, X., et al. (2019). The extended marine underwater environment database and baseline evaluations. *Appl. Soft Computing* 80, 425–437. doi: 10.1016/j.asoc.2019.04.025

Jian, M., Wang, J., Yu, H., and Wang, G. G. (2021). Integrating object proposal with attention networks for video saliency detection. *Inf. Sci.* 576, 819–830. doi: 10.1016/j.ins.2021.08.069

Jian, M., Wang, J., Yu, H., Wang, G., Meng, X., Yang, L., et al. (2021). Visual saliency detection by integrating spatial position prior of object with background cues. *Expert Syst. Appl.* 168, 114219. doi: 10.1016/j.eswa.2020.114219

Khan, S., and Nazir, S. (2022). Deep learning based pashto characters recognition: LSTM-based handwritten pashto characters recognition system. *Proc. Pakistan Acad. Sciences: A. Phys. Comput. Sci.* 58, 49–58. doi: 10.53560/PPASA(58-3)743

Khan, S., Nazir, S., García-Magariño, I., and Hussain, A. (2021). Deep learning-based urban big data fusion in smart cities: Towards traffic monitoring and flow-preserving fusion. *Comput. Electrical Eng.* 89, 106906. doi: 10.1016/j.compeleceng.2020.106906

Khan, W., Wang, H., Anwar, M. S., Ayaz, M., Ahmad, S., and Ullah, I. (2019). A multi-layer cluster based energy efficient routing scheme for UWSNs. *IEEE Access.* 7, 77398–77410. doi: 10.1109/ACCESS.2019.2922060

Lee, G., Tai, Y. W., and Kim, J. (2018). ELD-net: An efficient deep learning architecture for accurate saliency detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 40, 1599–1610. doi: 10.1109/TPAMI.2017.2737631

Li, C., Cong, R., Hou, J., Zhang, S., Qian, Y., and Kwong, S. (2019). Nested network with two-stream pyramid for salient object detection in optical remote sensing images. *IEEE Trans. Geosci. Remote Sens.* 57, 9156–9166. doi: 10.1109/TGRS.2019.2925070

Liu, R., Fan, X., Zhu, M., Hou, M., and Luo, Z. (2020). Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. *IEEE Trans. Circuits Syst. Video Technol.* 30, 4861–4875. doi: 10.1109/TCSVT.2019.2963772

Maragos, P., Gros, P., Katsamanis, A., and Papandreou, G. (2008). "Cross-modal integration for performance improving in multimedia: A review," in *Multimodal processing and interaction* (Springer), 1–46.

Moghimi, M. K., and Mohanna, F. (2021). Real-time underwater image enhancement: a systematic review. *J. Real-Time Image Process.* 18, 1509–1525. doi: 10.1007/s11554-020-01052-0

Qiao, J., Wang, X., Chen, J., and Jian, M. (2022). Low-light image enhancement with an anti-attention block-based generative adversarial network. *Electronics* 11, 1627. doi: 10.3390/electronics11101627

Shen, J., Fan, T., Tang, M., Zhang, Q., Sun, Z., and Huang, F. (2014). A biological hierarchical model based underwater moving object detection. *Comput. Math. Methods Med.* 2014, 609801. doi: 10.1155/2014/609801

Su, X., Ullah, I., Liu, X., and Choi, D. (2020). A review of underwater localization techniques, algorithms, and challenges. *J. Sensors.* doi: 10.1155/2020/6403161

Tian, Y., Li, E., Liang, Z., Tan, M., and He, X. (2021). Diagnosis of typical apple diseases: A deep learning method based on multi-scale dense classification network. *Front. Plant Sci.*, 2122. doi: 10.3389/fpls.2021.698474

Ullah, I., Chen, J., Su, X., Esposito, C., and Choi, C. (2019). Localization and detection of targets in underwater wireless sensor using distance and angle based algorithms. *IEEE Access.* 7, 45693–45704. doi: 10.1109/ACCESS.2019.2909133

Ullah, I., Liu, Y., Su, X., and Kim, P. (2019). Efficient and accurate target localization in underwater environment. *IEEE Access.* 7, 101415–101426. doi: 10.1109/ACCESS.2019.2930735

Ullah, I., Ming-Sheng, G. A., Kamal, M. M., and Khan, Z. (2017). "A survey on underwater localization, localization techniques and its algorithms," in *3rd Annual International Conference on Electronics, Electrical Engineering and Information Science (EEEIS 2017)*. 252–259 (Guangzhou, Guangdong, China: Atlantis Press).

Wang, W., Shen, J., Yang, R., and Porikli, F. (2018). Saliency-aware video object segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 40, 20–33. doi: 10.1109/TPAMI.2017.2662005

Xu, J., Bi, P., Du, X., Li, J., and Chen, D. (2019). Generalized robust PCA: A new distance metric method for underwater target recognition. *IEEE Access* 7, 51952–51964. doi: 10.1109/ACCESS.2019.2911132

Yasir, M., Hossain, M. S., Nazir, S., Khan, S., and Thapa, R. (2022). Object identification using manipulated edge detection techniques. *Science* 3 (1), 1–6. doi: 10.11648/j.scidev.20220301.11

Zhang, D., Li, L., Zhu, Z., Jin, S., Gao, W., and Li, C. (2019). Object detection algorithm based on deformable convolutional networks for underwater images. *2019 2nd China Symposium Cogn. Computing Hybrid Intell. (CCHI)*, 274–279. doi: 10.1109/CCHI.2019.8901912

Zhang, J., Zheng, Y., and Qi, D. (2017). Deep spatio-temporal residual networks for citywide crowd flows prediction. *Thirty-first AAAI Conf. Artif. Intell* 31 (1), 1655–1661. doi: 10.1609/aaai.v31i1.10735

Zheng, Y. (2015). Methodologies for cross-domain data fusion: An overview. *IEEE Trans. big Data* 1, 16–34. doi: 10.1109/TBDATA.2015.2465959

Zhu, L., Guo, F., Polak, J. W., and Krishnan, R. (2018). Urban link travel time estimation using traffic states-based data fusion. *IET Intelligent Transport Syst.* 12, 651–663. doi: 10.1049/iet-its.2017.0116

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 13 January 2023
DOI 10.3389/fmars.2022.1086140

Check for updates

OPEN ACCESS

EDITED BY
Hong Song,
Zhejiang University, China

REVIEWED BY
Islam Zada,
International Islamic University, Pakistan
Mohammad Faisal,
University of Malakand, Pakistan
Yi Ma,
Ministry of Natural Resources, China

*CORRESPONDENCE
Wan Jianhua
19850014@upc.edu.cn

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 01 November 2022
ACCEPTED 28 November 2022
PUBLISHED 13 January 2023

CITATION
Yasir M, Shanwei L, Mingming X, Hui S, Hossain MS, Colak ATI, Wang D, Jianhua W and Dang KB (2023) Multi-scale ship target detection using SAR images based on improved Yolov5. *Front. Mar. Sci.* 9:1086140. doi: 10.3389/fmars.2022.1086140

COPYRIGHT
© 2023 Yasir, Shanwei, Mingming, Hui, Hossain, Colak, Wang, Jianhua and Dang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Multi-scale ship target detection using SAR images based on improved Yolov5

Muhammad Yasir[1], Liu Shanwei[1], Xu Mingming[1], Sheng Hui[1], Md Sakaouth Hossain[2], Arife Tugsan Isiacik Colak[3], Dawei Wang[1], Wan Jianhua[1]* and Kinh Bac Dang[4]

[1]College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao, China, [2]Department of Geological Sciences, Jahangirnagar University, Dhaka, Bangladesh, [3]National University International Maritime College Oman, Sahar, Oman, [4]Faculty of Geography, VNU University of Science, Vietnam National University, Hanoi, Vietnam

Synthetic aperture radar (SAR) imaging is used to identify ships, which is a vital task in the maritime industry for managing maritime fisheries, marine transit, and rescue operations. However, some problems, like complex background interferences, various size ship feature variations, and indistinct tiny ship characteristics, continue to be challenges that tend to defy accuracy improvements in SAR ship detection. This research study for multiscale SAR ships detection has developed an upgraded YOLOv5s technique to address these issues. Using the C3 and FPN + PAN structures and attention mechanism, the generic YOLOv5 model has been enhanced in the backbone and neck section to achieve high identification rates. The SAR ship detection datasets and AirSARship datasets, along with two SAR large scene images acquired from the Chinese GF-3 satellite, are utilized to determine the experimental results. This model's applicability is assessed using a variety of validation metrics, including accuracy, different training and test sets, and TF values, as well as comparisons with other cutting-edge classification models (ARPN, DAPN, Quad-FPN, HR-SDNet, Grid R-CNN, Cascade R-CNN, Multi-Stage YOLOv4-LITE, EfficientDet, Free-Anchor, Lite-Yolov5). The performance values demonstrate that the suggested model performed superior to the benchmark model used in this study, with higher identification rates. Additionally, these excellent identification rates demonstrate the recommended model's applicability for maritime surveillance.

KEYWORDS

synthetic aperture radar (SAR), ship identification, artificial intelligence, deep learning (DL), YOLOv5S, SAR ship detection dataset (SSDD), AirSARship

1 Introduction

Detection of ships is a crucial task in the maritime industry for controlling maritime fisheries, marine transit, and rescue operations. However, some issues, such as intricate backdrop interferences, numerous ship size fluctuations, and imprecise little ship features, still pose difficulties and frequently thwart advancements in SAR ship recognition accuracy. Accurate position and trajectory determination of the target ship is essential for managing maritime traffic, recovering from maritime accidents, and the economy (Xiao et al., 2020).

According to the kinds of remote sensing technologies used, the two main categories of ship detection study are, respectively, the SAR image-based and optical satellite image-based methodologies. One of the major challenges of ship identification in optical remote sensing images is finding suitable areas from complex backgrounds fast and correctly (Wang et al., 2021). High-resolution capabilities, independence from the weather, and flight altitude independence are all attributes of SAR images. SAR's self-illumination capability ensures that they always produce high-quality images under any circumstance (Chang et al., 2019). SAR has been extensively employed in ship identification (Ma et al., 2018; Xu et al., 2021; Li et al., 2022; Yasir et al., 2022; Xiong et al. 2022), oil spill identification (Yekeen et al., 2020; Wang et al., 2022), change detection (Gao et al., 2019; Chen and Shi, 2020; Zhang et al., 2020b; Wang et al., 2022), and other fields (Niedermeier et al., 2000; Baselice and Ferraioli, 2013). Because of its broad observation range, brief observation duration, great data timeliness, and high spatial resolution (Ouchi, 2013), SAR performing a significant role in ship identification. The amount and quality of SAR data have been steadily improving recently due to the quick development of space-borne SAR-imaging technologies. As a result, many researchers are studying how to identify ships in HR SAR images (Wang et al., 2016; Li et al., 2017b; Salembier et al., 2018; Du et al., 2019; Lin et al., 2019; Wang et al., 2019; Wang et al., 2020b; Wang et al., 2020c; Wang et al., 2020d; Zhang et al., 2020c; Yasir et al., 2022). However, due to the complicated environment and other difficult issues, such as sidelobes and target defocusing (Chen et al., 2019; Han et al., 2019; Xiong et al., 2019; Yuan et al., 2020), identifying ship targets in HR SAR images is still challenging.

Deep learning (DL) technologies has enhanced quickly in recent years, enabling natural image identification. Convolutional neural networks (CNNs) were introduced into the target identification area by R-CNN (Girshick et al., 2014), and as a result, target identification has received new scientific research thoughts, and its use in SAR images has a wide range of potential applications. Currently, two stage identification approaches addressed by R-CNN, Fast R-CNN, and Faster R-CNN (Girshick, 2015; Ren et al., 2015) are the main convolutional neural network-based algorithms employed in ship identification in SAR images. The complexity of their network topologies, the sheer number of parameters, and the slow recognition speed, however, prevent them from being able to complete ship detection tasks in the required amount of time. The target identification problem is also seen as a regression analysis task involving target location and category information by the single stage algorithms from the SSD (Liu et al., 2016) and YOLO (Redmon et al., 2016; Redmon and Farhadi, 2017; Redmon and Farhadi, 2018; Patel et al., 2022) series. They are more suited to ship identification applications that need virtually real-time identification since they output the identification results directly through a neural network model with high accuracy and speed (Willburger et al., 2020).

Although the aforementioned methods have strong detection performance, it is challenging to directly apply them to SAR ship identification. There are still a number of problem with the DL-based ship identification approaches in SAR images (Li et al., 2020; Zhang et al., 2021a). (i) Due to the unique imaging technique utilized by SAR, there is less contrast between the ocean and ship in the SAR images since there are more scattering noise and sea debris and less side flap. (ii) Different ships have various sizes and shapes which are reflected in SAR images as varying numbers of pixels, especially for tiny-scale ships. Smaller ships have less information about their whereabouts than large ships, and since they have fewer pixels, they are more susceptible to being deceived by the speckle noise in SAR images. While this is going on, the detection process becomes more complex, which lowers the accuracy of identification and recognition. (iii) SAR images cannot be directly supplied to the network for identification if the scene is large. It is anticipated that the network has now received the SAR image of the expansive landscape. The ship target will be resampled in this situation to a few or possibly only one pixel, which will significantly reduce the identification accuracy. The main goals of the current study are as follows:

- To identify optimum multi-size ship target in SAR images by modified YOLOv5 model.
- To offer the backbone extraction network a well-designed structure, a set of CSP framework and attention mechanisms have been upgraded, and the output layer has been expanded to four feature layers.
- To improve the overall performance throughout the recognition process, this improved version of the YOLOv5 model also produces effective results in a condensed amount of time with a relatively smaller database.
- To use the SSDD and AirSAR ship Datasets, two distinct and well-known datasets, in these simulations. The SSDD collection contains 1160 SAR images in total, collected by RadarSat-2, TerraSAR-X, and Sentinel-1, with resolutions ranging from 1m to 10m and

polarizations in HH, HV, VV, and VH. Gaofen-3 has collected 31 single-polarized SAR images, which are included in the AirSARship.

- To assess the suggested model's applicability utilizing cutting-edge benchmark convolutional neural network-based techniques.
- To employ several performance indicators for application evaluation reasons, including precision, accuracy, time consumption, and different training and test sets.
- To demonstrate the model's superiority the performance results would be demonstrated to the desired benchmark models (CNN-based SAR ship identification techniques).

The paper is organized as follows; Section 2 shows the study serves as an organizational framework for the remainder of the research, explaining the proposed methodology. The findings and analysis of the suggested research project are described in Section 3. Additionally, by contrasting it with other cutting-edge produced models, it has demonstrated the model's usefulness. Section 4 describes the ablation study and the paper is concluded in Section 5.

2 Proposed methodology

The target of this current study is to develop a ship detection model that could potentially function when there are inadequate hardware resources. Because of its reputation for speed and accuracy, the lightweight version of YOLO has received attention. Open source model YOLO was first presented by Joseph Redmon in 2016 (Redmon et al., 2016). It is suitable as a real-time system since it can identify things at extremely quick speeds. In this research work, the upgraded model lightweight version of YOLOv5 is used. This upgraded model resulted with higher accuracy and efficient identification capabilities (Caputo et al., 2022; Nepal and Eslamiat, 2022). Two datasets that are available in the literature, the AirSARship (Xian et al., 2019) dataset and the SSDD dataset (Li et al., 2017a), have both been considered.

2.1 Data augmentation

In order to train the model for deep learning, a lot of data is typically required. However, in practice, certain data sets are challenging to collect, leading to a small quantity of data in this category that falls short of the required data volume for deep learning. Experts have thus suggested data augmentation approaches to successfully address this issue (Najafabadi et al., 2015). The data augmentation techniques such as random rotation and mosaic was used. Given training data, mosaic randomly crops four images and stitch them together to create one. It has the advantage of enriching the background of the image and enhancing the batch size discretely so that it can help to minimize the model dependence on a large batch size when training (Figure 1).

2.2 Data annotation

The images were annotated using the Labellmg software, which generates a json annotation file and transforms it into a txt file. The type and number of the labeling target, the labeling image's standardized width and height, and the center point's coordinates are all information that can be found in the txt file. Figure 2 displays the labelling outcomes.

2.3 YOLOv5 network

YOLO is a regression-based technique and, despite being less accurate, is actually faster than region proposal-based methods like R-CNN (Girshick et al., 2014). The goal of YOLO is to achieve object identification by approaching it as a regression and classification issue. Identifying the bounding box coordinates for the objects in the images is the first step, and second step is to classifying the objects that are identified in a class. This is accomplished in a single step by first splitting the input images into a grid of cells, then determining the bounding box and relative confidence score for each cell's containing object.

The YOLOv5 network is one of the recent research advancements in the YOLO series of algorithms. Despite sharing a network structure with the YOLOv4 network, it is smaller, has a faster running speed and convergence speed, and uses a lightweight algorithm. Additionally, it improves precision at the same time. As a result, the YOLOv5s algorithm is used in the current study work to detect ships in SAR images. Four components make up the YOLOv5 network structure: input, backbone, neck, and prediction. The Yolov5 framework architecture is displayed in Figure 3A. Networks can be categorized as YOLOv5l, YOLOv5m, YOLOv5x, and YOLOv5s. Their widths and depths may differ significantly, but their network structures are comparable. The network structure of YOLOv5s is the shortest, shallowest, runs the fastest, and has the least accuracy. As a result, the accuracy continues to rise, the speed of operation declines, and the other three network structures increasingly deepen and widen. Adaptive anchor box operation, mosaic data augmentation, image scaling, and CSP structure were used to process the input dataset, while focus and CSP framework were used to build the backbone. Focus increased network speed and cut down on floating-point operations (FLOPs) by clipping the input image. Figure 3A presented the focal structure. The two CSP (Wang et al., 2020a) framework that were used by YOLOv5 were CSP1_X and CSP2_X; CSP1_X was utilized for down

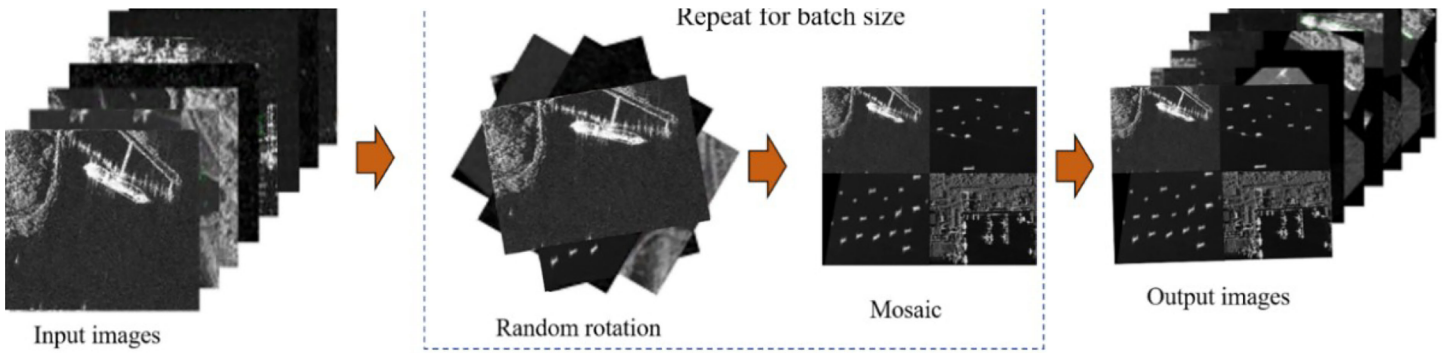


FIGURE 1
Illustrations the random rotation mosaic data augmentation technique.

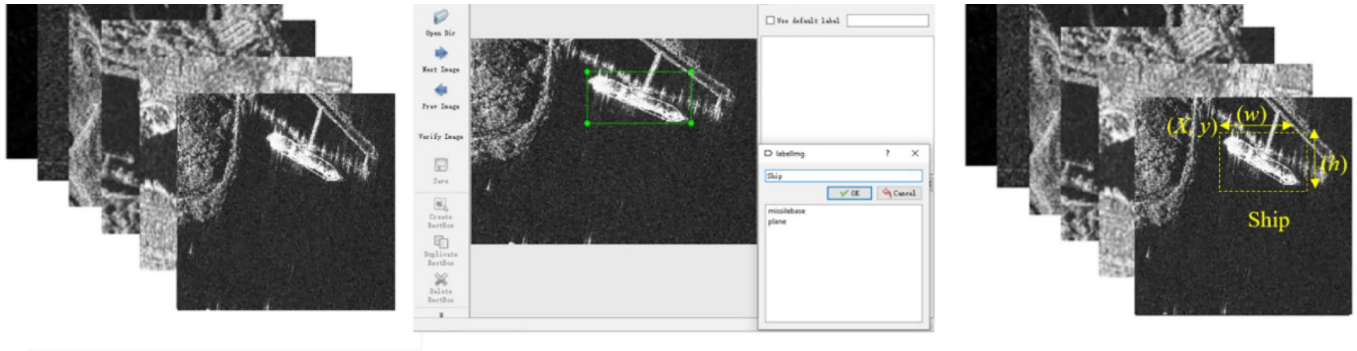


FIGURE 2
SAR images annotation procedures. The parameters of actual ships were obtained using DL algorithms for image annotation, where (x, y) represented for the coordinates of the top left corner of the rectangular box, w for width, and h for height.

10.3389/fmars.2022.1086140

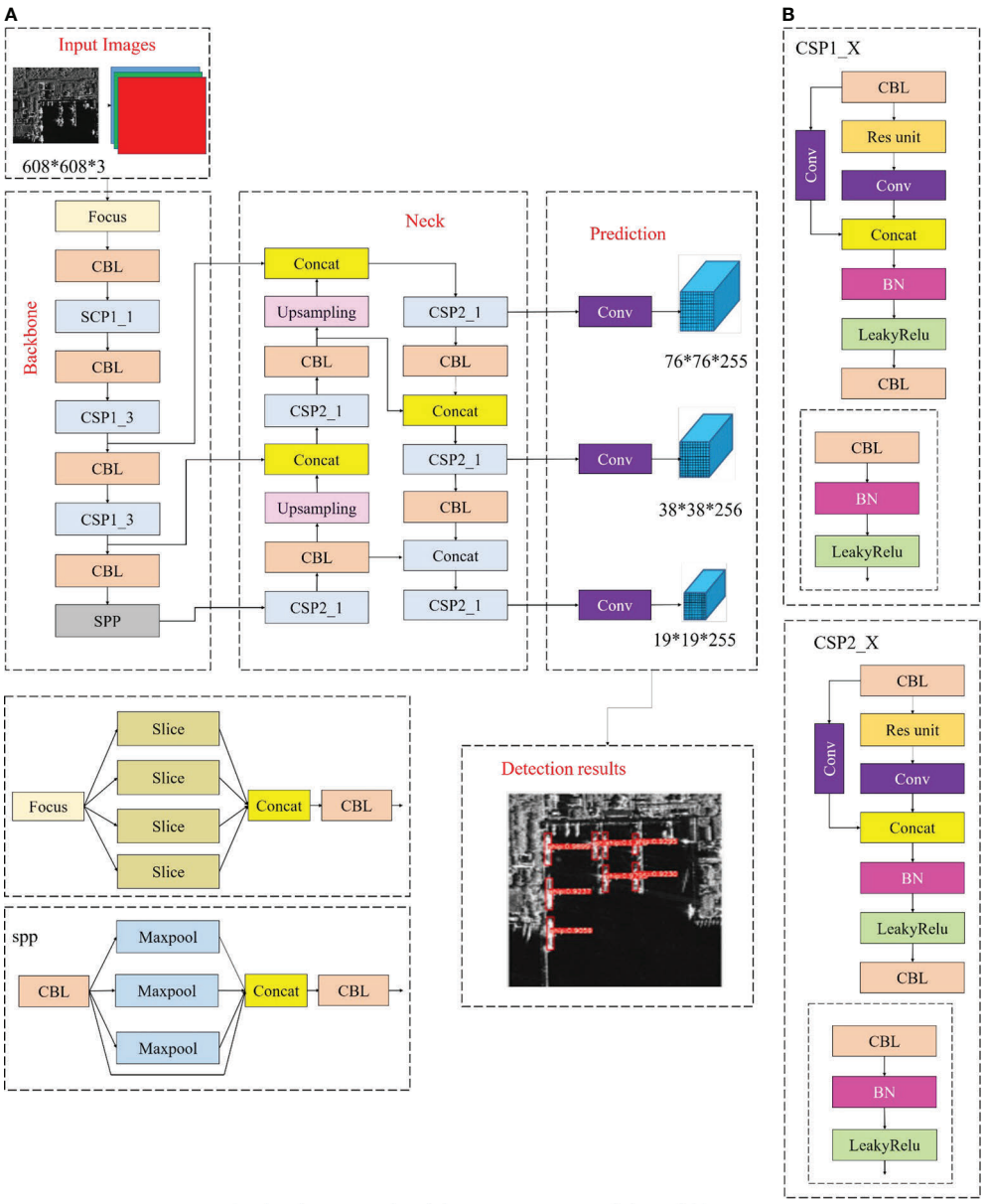


FIGURE 3
(A) The framework of the YOLOv5 Model, and **(B)** CSP structure.

sampling in the backbone while CSP2_ X was utilized in the neck. CSP can reduce operations while increasing the network's capacity for learning and guaranteeing accuracy. Figure 3B depicted the two CSPs' structures; the neck used the SPP-net and FPN + PAN framework to improve the network's feature fusion effect, while the prediction employed the GIOU_ Loss (Rezatofighi et al., 2019), which did not only focus on the overlap between the prediction box and the ground truth but also on the non-overlapping areas. (Yu et al., 2016) found that GIOU maintains the benefits of IOU while solving its issues. The computation for Equation (1-2) is as follows:

$$IOU = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

$$GLOU = IOU - \left| \frac{\frac{C}{(A \cup B)}}{|C|} \right| \tag{2}$$

Four separate networks were used, and YOLOv5 was continuously upgraded as well. Version 5.0 of YOLOv5s was used in this study project; in comparison to version 4.0, this

version modified all functional activation in the framework to SiLU (Elfwing et al., 2018), eliminated the conv in the CSP, and designated it C3 as presented in Figure 4. Additionally, v5.0 has a smaller and faster network structure than v4.0.

2.4 YOLOv5 network improvement

In this section of the study, the improvements made to the YOLOv5 classifier have been described in accordance with the guidelines of the proposed research challenge. The neck and backbone parts are enhanced to produce the greatest identification outcomes.

2.4.1 Backbone improvement

It is frequently possible to combine features from various scales to obtain more meaningful object information. The high-level feature has lesser resolution and poor perception of object information, but the receptive field is bigger, which is suited for identifying huge objects. The low-level feature has higher resolution, a smaller receptive field, more texture information, and more noise. The complex background environment in the SSDD and AirSARship dataset results in some large ground objects having an inadequate detection effect. In the current research work, a set of C3 framework was used to construct the YOLOv5s backbone network. The original three sets of C3 were converted into four sets of C3 to further the network framework as a whole (Figure 4). In turn, the model's detection accuracy may have improved as a result of the network's increased ability to communicate and learn about larger ground objects.

2.4.2 Attentional mechanism

The human visual attention process is referred to as the "attention mechanism," which concentrate on local details and blocks out redundant details. To put it another way, the network is able to identify critical information among a plethora of data due to the attention mechanism. The network performance is enhanced in this way by the addition of a small amount of computing. Figure 5 presented the increased backbone structure.

2.4.3 Neck enhancement

The neck was constructed using the FPN (Lin et al., 2017) + PAN (Liu et al., 2018) framework. This framework incorporated a bottom-up feature pyramid network after the FPN, which improved location information and semantic expression on various scales. The C32_X structure was incorporated into the neck of the YOLOv5s to enhanced the feature fusion impact of the network framework. Because of the development of a set of C3 structures in the current research work, an output layer was updated to the network's neck to increase feature extraction. Figure 6 presented the increase structure of FPN + PAN.

2.4.4 Extending receptive field area

Each pixel in the output feature map must respond to an area in the image that is large enough for it to get information about the large object, which makes the size of the receptive field a major issue in many vision applications. Consequently, a maximum pooling layer has been chosen to be added to the space pyramid in order to improve multiple receptive field fusion and increase the accuracy of identification of tiny targets. The updated architecture is shown in Figure 7.

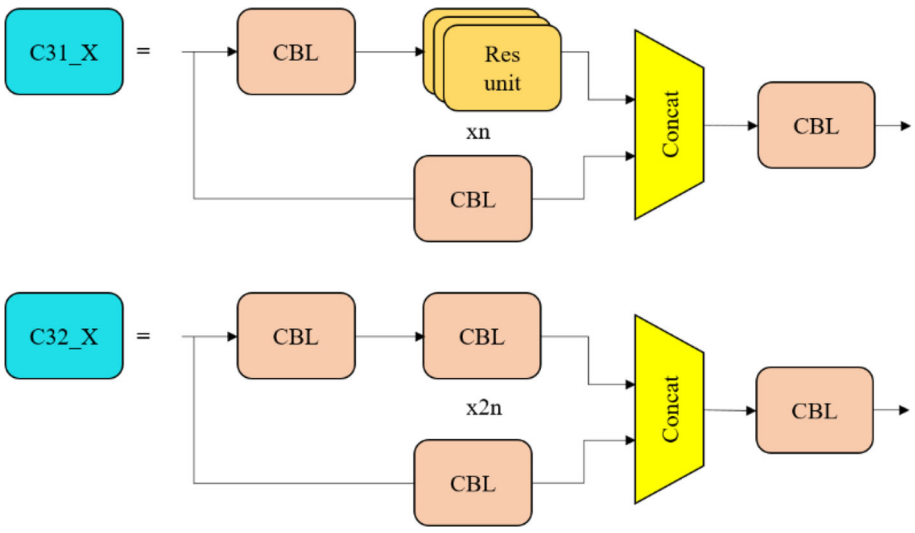


FIGURE 4
The C3 structure.

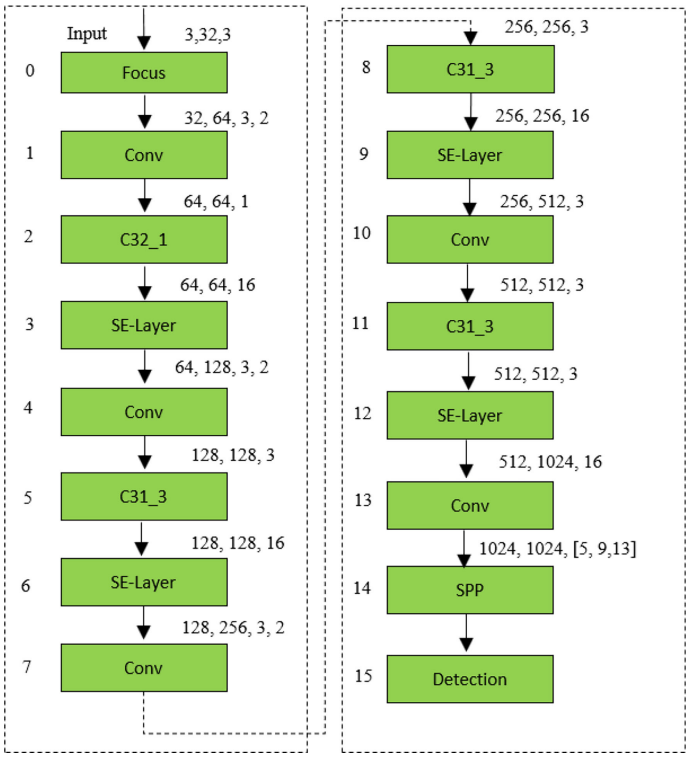


FIGURE 5
The improved backbone Network structure.

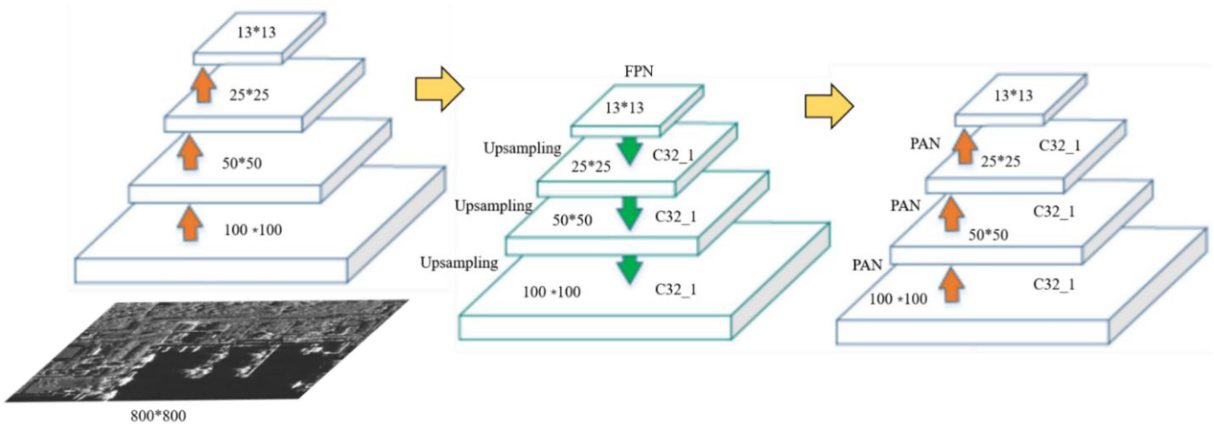


FIGURE 6
Improved FPN + PAN structure.

Figure 7 includes a graphic representation of the contribution of a maximum pooling layer. Figure 7 shows the spatial pyramid pooling module SPP and the combination module CBL, which combines convolutional layers, BN, and activation function layers. The addition of a 3*3 maximum pooling filter has increased the model's receptive field.

3 Results and discussion

This section has a detailed description of the SSDD, AirSARship datasets, experimental settings as well as evaluation metrics and assesses the performance of the technique. The testing set is then separated into two sets, one

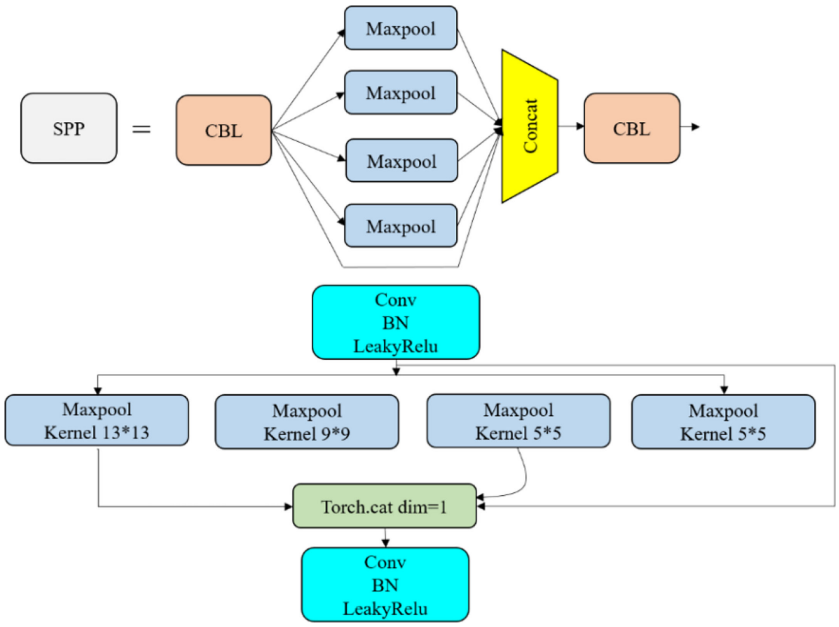


FIGURE 7
Pooling layer improvement structure.

is offshore ships and the other one is inshore ships, and each group has been used to assess the efficacy of the various strategies. The identification outcomes of the current model and a few unique CNN-based models are shown on the two SAR large scene images.

3.1 Dataset introduction

SSDD dataset: The first and most important stage in ship detection using deep learning techniques is the construction of a sizable and representative dataset. Therefore, the experiment utilize the SSDD (Li et al., 2017a) dataset, which have 1160 SAR images with resolutions ranging from 1m to 10m with polarizations in HH, HV, VV, and VH from RadarSat-2, TerraSAR-X, and Sentinel-1. Each sample image has a dimension of roughly 800 x 800, with a ratio of 7:1:2, where the SSDD dataset is divided into three sets for the experiment: a training set, a validation set, and a testing set (Table 1).

AirSARship Dataset: In the present study, experiments also use the AirSARShip-1.0 (Xian et al., 2019) dataset to assess the performance of proposed model utilize high-resolution SAR ship identification dataset. Gaofen-3 provided 31 single-polarized SAR images for AirSARShip-1.0. Most images have a size of 3000 x 3000 pixels, while one has a size of 4140 x 4140 pixels with resolutions ranging from 1 to 3 meters with HH polarization. The large scene image has been split into 1000 x 1000 slices with a ratio of 7:1:2, where the dataset is divided into three sets for the experiment: a training set, a validation set, and a testing set (Table 1).

Two SAR large scene images from the Chinese GF-3 satellite, as shown in Figure 13, further illustrate the efficacy of the suggested strategy for identifying different size ships in SAR large scene images with complicated sceneries. These images contain inshore and offshore scenery as well as ship targets at various scales. In Figure 8, some image slices are presented and offshore and inshore scenes as well as multiscale ship targets are primarily shown in Figure 8A. The dataset clearly shows that

TABLE 1 Reveal the dataset's complete details.

Dataset	Size (pixel)	Image (num)	Mode	Satellite	Resolution(m)
SSDD	390 ×205 500 ×500	1160	SL	RadarSat-2 TerraSAR-X Sentinel-1	1-15
AirSARship	3000 × 3000	31	SL/UFS	GF-3	1/3

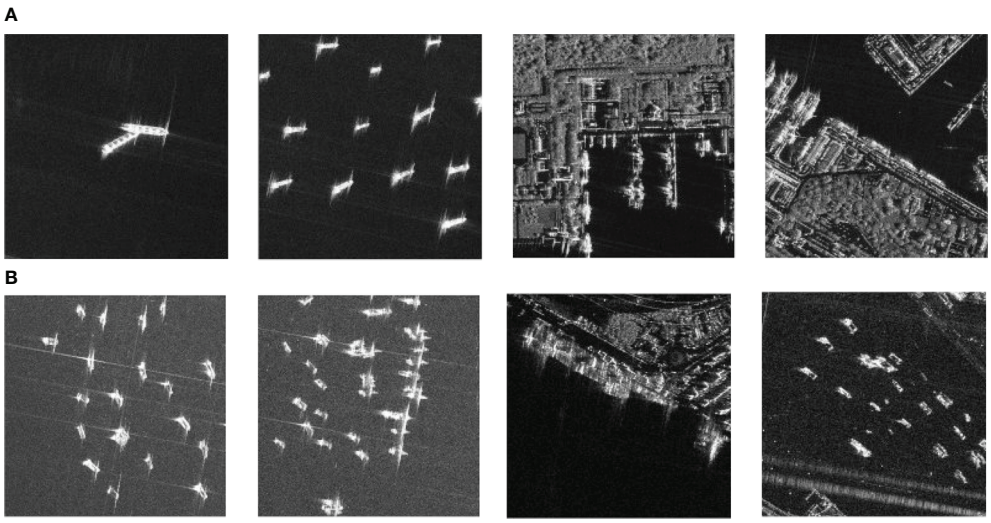


FIGURE 8
Inshore, offshore and different scale ship target on SAR images, **(A)** ships from the first SAR large scene images presented in Figure 13A, and **(B)** ship from the second SAR large images presented in Figure 13B.

both off-shore and inshore scenarios are included, and that the sizes of the ships fluctuate widely.

3.2 Experimental environment

The experiments are all carried out using PyTorch 1.7.0, CUDA 10.1, and CUDNN 7.6.5 on an NVIDIA Geforce GTX 2080Ti GPU and an Intel Core i9-9900KF CPU. The PC, which was equipped with a deep learning environment for our research and was running Windows 10, is depicted in Table 2. Additionally, each model was trained over 100000 iterations utilizing the Stochastic Gradient Descent (SGD) technique on a total of two images per minibatch. The initial learning rate was set at 0.001, while the weight decay was set to 0.00004. In every trial, the detection threshold IOU was set to 0.7. Table 2 displayed the experimental hardware and software configuration. During the experiment, the same platform was used for all comparison techniques.

3.3 Evaluation metrics

Since optical and SAR image object detection tasks are comparable, the effectiveness of various approaches is assessed using a variety of established indicators, such as average precision (AP), recall rate (r), precision rate (p), F score (F1), and these indications are specifically formulated in following equations (3-6):

$$\Pr ecision = \frac{TP}{TP + FP} \quad (3)$$

$$\mathrm{Re}\, call = \frac{TP}{TP + FN} \quad (4)$$

The number of correctly recognized ships, false alarms, and missing ships are denoted by the acronyms FN (false negative), FP (false positives), and TP (true positives). The precision and recall are combined into the F1 score as follows:

$$F1 - score = 2 \times \frac{\Pr ecision \times \mathrm{Re}\, call}{\Pr ecision + \mathrm{Re}\, call} \quad (5)$$

$$AP = \int_0^1 P(R)dR \quad (6)$$

The complete detection effectiveness of the various models is assessed using the AP and F1-score metrics, and a higher number indicates a superior detector performance.

The percentage of ground truth ships that networks correctly predict in all predictions is referred to as the precision rate. The percentage of ground truth ships that networks correctly predicted in

TABLE 2 Environment Configuration.

Project	Model/Parameter
System	windows 10
RAM	32GB
GPU	NVIDIA GTX Geforce 2080 Ti
CPU	Intel i9-9900KF
Framework	CUDA10.1/cudnn7.6.5/torch 1.7.0
Code	python3.7

all ground truth ships is referred to as recall rate. F1 is a comprehensive statistic that combines precision rate and recall rate to assess the effectiveness of various framework. AP outlines the region beneath Precision-Recall (PR) curves and also shows the overall effectiveness of various approaches. Additionally, Frames-Per-Second (FPS), which is derived from Equation (7), is used to assess the detection speed of various approaches. A method achieves a higher speed the higher the FPS.

$$FPS = \frac{1}{T_{per-img}} \quad (7)$$

When processing an image, the inference time (Tper-imgis) is the cost of a method.

3.4 Detection performance of inshore and offshore ships

In this section, the proposed approach and alternative CNN-based approaches, such as Faster-RCNN (Lin et al., 2017), SSD (Liu et al., 2016), R2CNN (Jiang et al., 2017), ARPN (Zhao et al., 2020), DAPN (Cui et al., 2019), Quad-FPN (Zhang et al., 2021b), HR-SDNet (Wei et al., 2020), Grid R-CNN (Lu et al., 2019), Cascade R-CNN (Cai and Vasconcelos, 2018), YOLOv4-LITE (Liu et al., 2022), EfficientDet (Tan et al., 2020), Free-Anchor (Zhang et al., 2019), Lite-Yolov5 (Xu et al., 2022), and yolov5-X, are assessed using offshore and inshore ships of testing sets. Aside from these metrics, F1, AP, and FPS are also employed to investigate the applicability of various methodologies. The suggested model's identification performance against other CNN-based approaches tested on offshore ships and inshore ships based on SSDD dataset and AirSAR ship dataset is presented in Tables 3, 4; Figures 9, 10. The current model provides the best accuracy for offshore SSDD (about 95.36% AP for the offshore scenes). The second-best result is 89.03% from the R2CNN approach, although it still performs better at detecting anomalies than the currently suggested model by 6.33% AP. The studied model also delivers the best accuracy for inshore on SSDD (about 92.27% AP for the inshore scenes). The second-best result is 83.53% from the R2CNN approach, although it still performs better in terms of detection than the currently suggested model by 8.74% AP. The researched model provides the best accuracy for offshore on AirSARship (about 94.57% AP for the offshore scenes). It has the best detection performance, as seen by the second-best result of 88.27% from the Quad-FPN approach, which is still 6.3% AP less than the proposed model. The proposed model also delivers the best accuracy for inshore on AirSARship (about 91.11% AP for the inshore scenes). It also has the best detection performance, with the second-best estimate coming from the Lite-Yolov5 approach at 84.94%, however it is still 6.17% AP lower than the current model.

The suggested model and the other state-of-the-art CNN based approaches include Faster-RCNN (Lin et al., 2017), SSD (Liu et al., 2016), R2CNN (Jiang et al., 2017), ARPN (Zhao et al., 2020), DAPN (Cui et al., 2019), Quad-FPN (Zhang et al., 2021b), HR-SDNet (Wei et al., 2020), Grid R-CNN (Lu et al., 2019), Cascade R-CNN (Cai and Vasconcelos, 2018), YOLOv4-LITE (Liu et al., 2022), EfficientDet (Tan et al., 2020), Free-Anchor (Zhang et al., 2019), Lite-Yolov5 (Xu et al., 2022), and yolov5-X techniques all have detection accuracies that are higher for offshore scenes than for inshore situations. This is reasonable

TABLE 3 The identification outcomes of various state-of-the-art CNN based approaches on offshore and inshore ship scene for SSDD Dataset.

Methods	Off-Shore				In-Shore				FPS
	P (%)	*R (%)*	*AP (%)*	*F1*	*P (%)*	*R (%)*	*AP (%)*	*F1*	
Faster-RCNN (Lin et al., 2017)	88.60	91.81	88.60	0.8770	77.21	79.37	74.20	0.7377	16
SSD (Liu et al., 2016)	85.59	85.65	82.50	0.8170	74.67	71.45	78.25	0.7735	87
R2CNN (Jiang et al., 2017)	91.14	90.07	89.03	0.8853	78.99	70.88	83.53	0.8289	48
ARPN (Zhao et al., 2020)	91.64	90.38	88.10	0.8797	79.63	77.70	77.7	0.7690	21.55
DAPN (Cui et al., 2019)	91.89	89.89	83.37	0.8271	75.76	78.77	70.98	0.6950	20.81
Quad-FPN (Zhang et al., 2021b)	85.97	89.56	80.10	0.7989	52.90	77.78	76.59	0.7551	20.25
HR-SDNet (Wei et al., 2020)	86.47	88.28	84.03	0.8350	72.57	77.97	73.85	0.7225	15.16
Grid R-CNN (Lu et al., 2019)	89.10	91.18	79.73	0.7853	81.08	81.56	74.01	0.7332	7.55
Cascade R-CNN (Cai and Vasconcelos, 2018)	90.10	84.75	87.01	0.8623	69.78	79.03	69.89	0.6814	10.55
YOLOv4-LITE (Liu et al., 2022)	89.59	87.03	79.17	0.7835	54.80	85.70	74.62	0.7323	42.5
EfficientDet (Tan et al., 2020)	92.19	89.15	86.11	0.8517	79.01	77.64	73.63	0.7292	21.44
Free-Anchor (Zhang et al., 2019)	90.31	87.35	87.67	0.8677	75.70	76.07	71.80	0.7055	13.76
Lite-Yolov5 (Xu et al., 2022)	91.32	91.08	80.16	0.7901	75.78	80.62	81.92	0.8045	43.25
Yolov5-X	77.25	82.60	79.56	0.7862	70.89	81.61	76.49	0.7578	140
Our model	94.57	94.06	**95.36**	0.9413	89.44	90.77	**92.27**	0.9140	**157**

TABLE 4 The identification outcomes of various state-of-the-art CNN based approaches on offshore and inshore ship scene for AirSARship Dataset.

Methods	Off-Shore				In-Shore				FPS
	P (%)	*R (%)*	*AP (%)*	*F1*	*P (%)*	*R (%)*	*AP (%)*	*F1*	
Faster-RCNN (Lin et al., 2017)	83.72	89.89	84.12	0. 8370	69.58	86.01	79.21	0.7877	16
SSD (Liu et al., 2016)	92.09	86.66	86.54	0.8550	64.76	88.08	81.41	0.8035	87
R2CNN (Jiang et al., 2017)	84.08	94.07	85.11	0.8653	72.15	90.92	79.22	0.6989	48
ARPN (Zhao et al., 2020)	82.40	80.83	82.22	0.8161	71.74	69.13	68.10	0.6754	21.55
DAPN (Cui et al., 2019)	90.18	87.98	85.83	0.8454	69.50	66.57	72.98	0.7141	20.81
Quad-FPN (Zhang et al., 2021b)	89.70	90.56	88.27	0.8789	62.10	78.15	80.77	0.7945	20.25
HR-SDNet (Wei et al., 2020)	83.40	89.82	79.27	0.7818	70.50	78.71	76.05	0.7506	15.16
Grid R-CNN (Lu et al., 2019)	89.89	89.91	81.03	0.8010	76.81	71.60	69.55	0.6854	7.55
Cascade R-CNN (Cai and Vasconcelos, 2018)	88.75	74.57	79.26	0.7854	70.87	81.03	70.91	0.6998	10.55
YOLOv4-LITE (Liu et al., 2022)	87.01	89.55	79.07	0.7889	68.60	84.09	82. 02	0.8189	42.5
EfficientDet (Tan et al., 2020)	89.10	90.75	81.20	0.8081	87.29	76.91	63.35	0.6260	21.44
Free-Anchor (Zhang et al., 2019)	90.38	77.35	84.60	0.8354	70.65	78.60	74.12	0.7379	13.76
Lite-Yolov5 (Xu et al., 2022)	91.56	90.48	81.17	0.8054	78.90	74.82	84.94	0.8388	66.25
Yolov5-X	79.12	82.70	79.44	0.7822	76.35	79.69	73.46	0.7225	140
Our model	92.83	90.55	**94.57**	0.9313	90.14	89.08	**91.11**	0.9014	**157**

considering that the former has a more complicated background than the latter. Perhaps as a result of their poor small ship identification capabilities, the other alternative approaches have lower precision values than the suggested model. In this current research work, the recall values of the proposed model are occasionally less than those of other offered methods. As a result, a suitable score threshold can be further thought about in the future to balance missed detections and false alarms. Additionally, the current model appears to be faster than other approaches based on the FPS data in Tables 3, 4, and Figures 9, 10, potentially as a result of the separable depth-wise and point-wise convolutions utilized in the backbone network. In conclusion, the offshore scene has greater accuracy, AP, and F1 scores for both datasets than the inshore scenario. This might be due to the inshore scene's densely packed ships and increased backdrop interference from the land. Additionally, it demonstrates that it is more difficult to spot ships in the inshore scene than it is in the offshore environment.

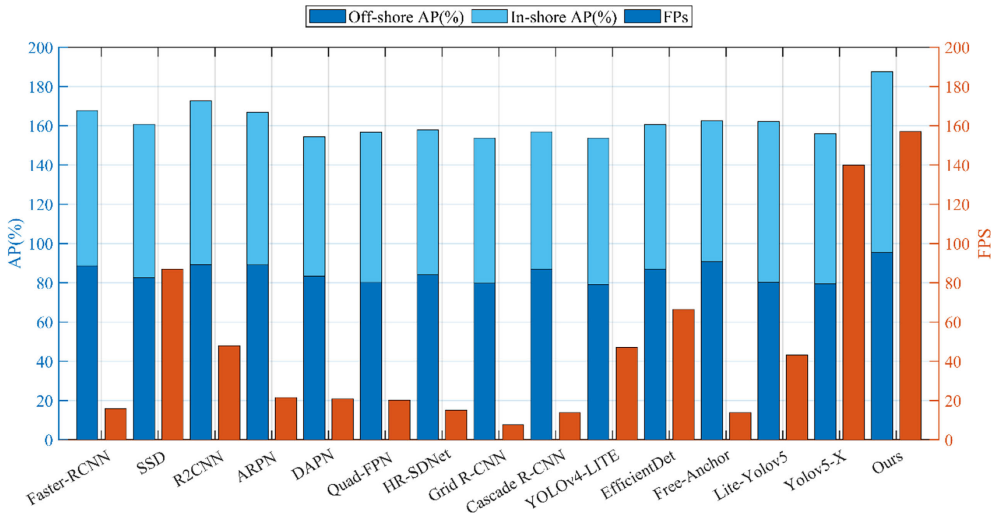


FIGURE 9
Performances of AP and FPS for various CNN-based techniques on offshore and inshore ships for SSDD dataset.

The proposed model's performance in terms of detection results compared to existing convolutional neural network -based approaches tested on offshore ships and inshore ships using SSDD dataset and AirSARship dataset is shown in Figures 11, 12. The suggested model is capable of detecting different SAR ships with multiscale sizes under varied backgrounds. This demonstrates its great scale and scene adaption together with excellent detection performance. The currently proposed model can increase the detection confidence scores when compared to the second-best CNN-based ship detector R2CNN. For instance, the suggested model raises the confidence score in Figure 11 first detection sample from 0.96 to 1.0. This can demonstrate the better trustworthiness of the newly proposed model. It is evident that the inshore scenario contains a sizable number of tightly packed ship targets. The other suggested solutions miss certain closely grouped inshore ships. The proposed model, however, is capable of accurately localizing and detecting these multiscale ships with high probabilities.

The other suggested CNN-based techniques in this research can precisely identify the port's heavily docked ships. However, it can be observed that the suggested method is more accurate and can detect these ships better when comparing the detection outcomes of several proposed CNN-based systems. The detection outcomes of the Faster-RCNN (Lin et al., 2017), SSD (Liu et al., 2016), R2CNN (Jiang et al., 2017), ARPN (Zhao et al., 2020), DAPN (Cui et al., 2019), Quad-FPN (Zhang et al., 2021b), HR-SDNet (Wei et al., 2020), Grid R-CNN (Lu et al., 2019), Cascade R-CNN (Cai and Vasconcelos, 2018), YOLOv4-LITE (Liu et al., 2022), EfficientDet (Tan et al., 2020), Free-Anchor (Zhang et al., 2019), Lite-Yolov5 (Xu et al., 2022), and yolov5-X algorithms contain several false alarms for the offshore and inshore scene (Figures 11, 12). Additionally, there are a few missed ships in the detection findings, which could be a result of how closely docked the ship targets are, making it more challenging for the framework to discriminate between them. Similarly, it is observed by comparing the suggested model's detection results that they are more precise than those produced by existing CNN- based techniques. Figures 11, 12 displays the outcomes of various CNN-based object detection techniques in an offshore scenario created for SAR images. It is evident that the offshore landscape contains a substantial number of dense multi-scale ship targets (the first two column of Figures 11, 12).

A pink color circle denotes false alarms in the identification outcomes of other proposed CNN-based models in this study that are not the current model. This might be due to a small number of false alarms that closely resemble ships, creating it great challenging for the network to successfully recognize. Because there are wakes of ships and surroundings, such ship A in Figure 12, the Faster-RCNN (Lin et al., 2017), SSD (Liu et al., 2016), R2CNN (Jiang et al., 2017), ARPN (Zhao et al., 2020), DAPN (Cui et al., 2019), Quad-FPN (Zhang et al., 2021b), HR-SDNet (Wei et al., 2020), Grid R-CNN (Lu et al., 2019), Cascade R-CNN (Cai and Vasconcelos, 2018), YOLOv4-LITE (Liu et al., 2022), EfficientDet (Tan et al., 2020), Free-Anchor (Zhang et al., 2019), Lite-Yolov5 (Xu et al., 2022), and yolov5-X algorithms can distinguish between wakes of ships and their surroundings. Additionally, several ships are overlooked by the Faster-RCNN (Lin et al., 2017), SSD (Liu et al., 2016), R2CNN (Jiang et al., 2017), ARPN (Zhao et al., 2020), DAPN (Cui et al., 2019), Quad-FPN (Zhang et al., 2021b), HR-SDNet (Wei et al., 2020), Grid R-CNN(Lu et al., 2019), Cascade R-CNN (Cai and Vasconcelos, 2018), YOLOv4-LITE (Liu et al., 2022), EfficientDet (Tan et al., 2020), Free-Anchor (Zhang et al., 2019), Lite-Yolov5 (Xu et al., 2022), and yolov5-X algorithms, as seen by the red circles in Figures 11, 12. The inability to extract distinguishing characteristics of ships and interference may be to blame. However, the suggested

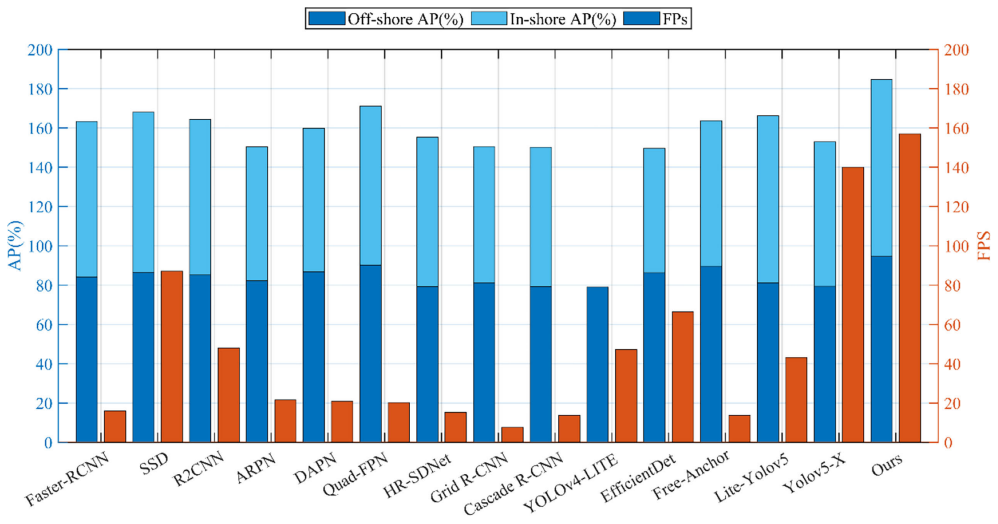


FIGURE 10
Performances of AP and FPS for various CNN-based techniques on offshore and inshore ships for AirSARship dataset.

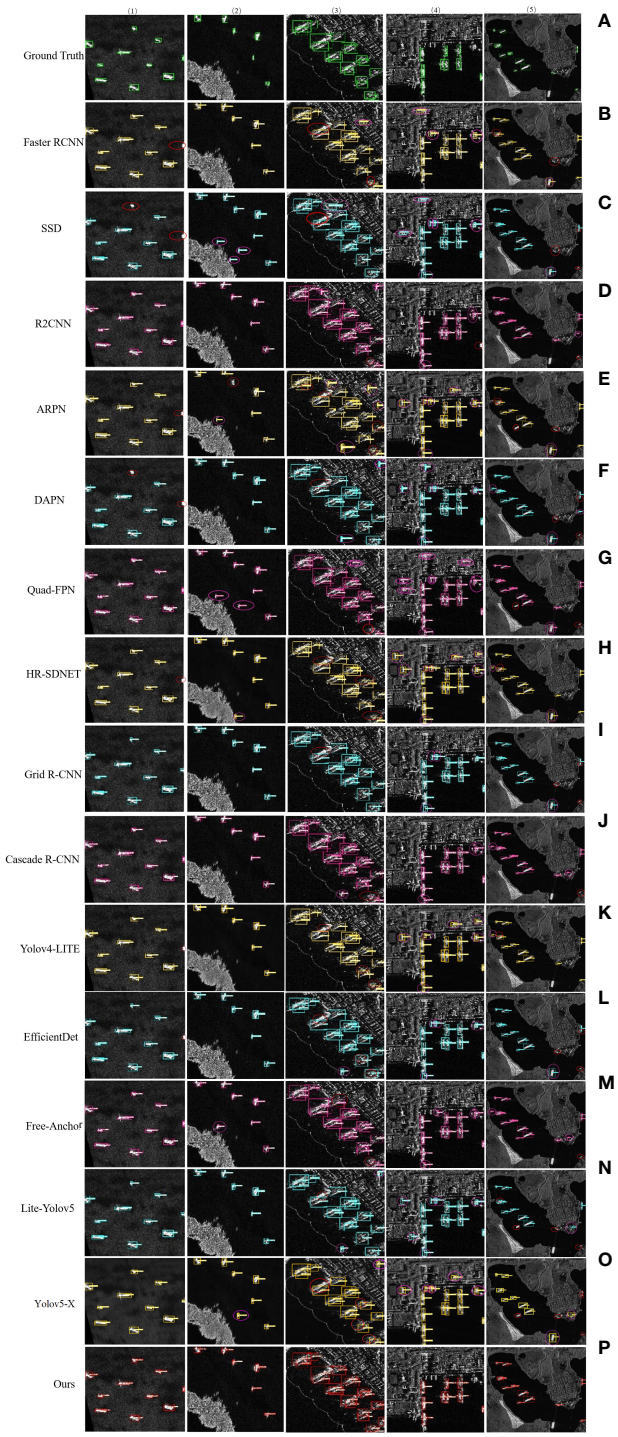


FIGURE 11
The visual detection outcomes of CNN-based approaches for offshore and inshore ships based on SSDD Dataset. Results from **(A)** is ground truth, results from **(B)** is Faster-R-CNN method, results from **(C)** SSD method, results from **(D)** is R2CNN method, results from **(E)** is ARPN, results from **(F)** is DAPN, results from **(G)** is Quad-FPN, results from **(H)** is HR-SDNet, results from **(I)** is Grid R-CNN, results from **(J)** is Cascade R-CNN, results from **(K)** is YOLOv4-LITE, results from **(L)** is EfficientDet, results from **(M)** is Free-Anchor, results from **(N)** is Lite-Yolov5, results from **(O)** is yolov5-X, and results from **(P)** is our proposed method. Note the pink circle show the false detection of ship and red circle is show the missing ship.

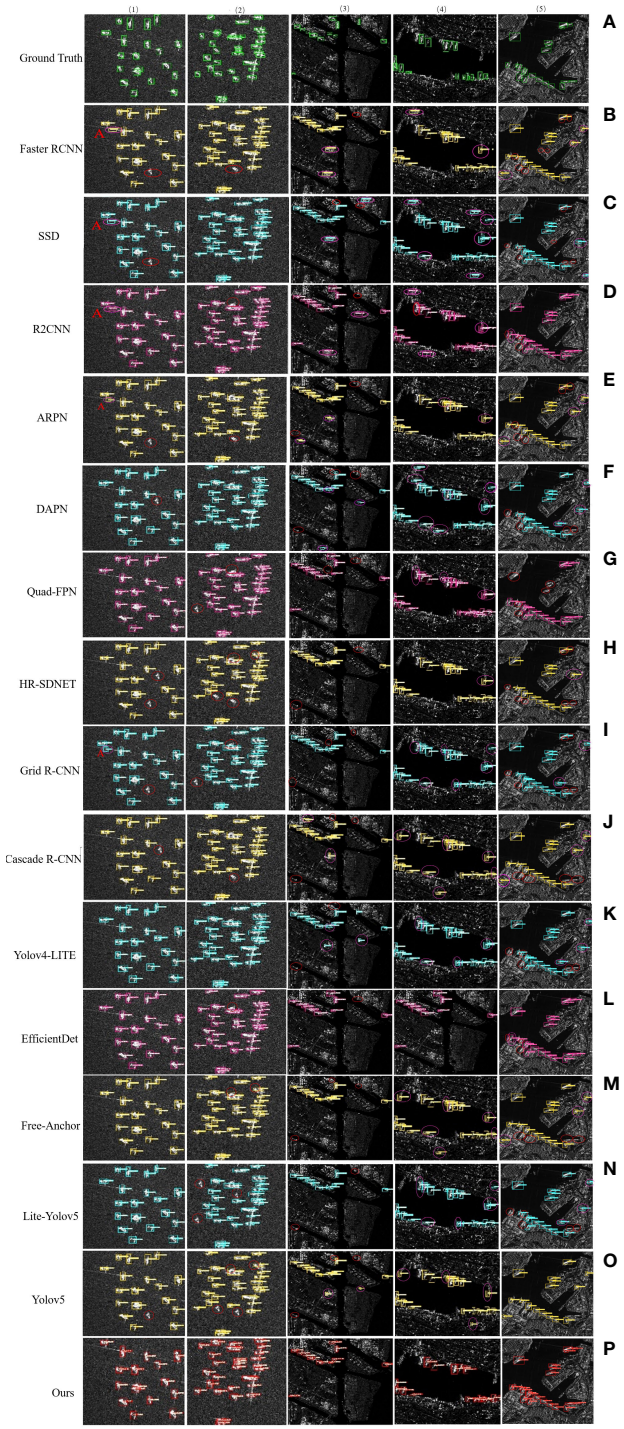


FIGURE 12
The visual detection outcomes of CNN-based approaches for offshore and inshore ships based on AirSARship Dataset. Results from **(A)** is ground truth, results from **(B)** is Faster-R-CNN method, results from **(C)** SSD method, results from **(D)** is R2CNN method, results from **(E)** is ARPN, results from **(F)** is DAPN, results from **(G)** is Quad-FPN, results from **(H)** is HR-SDNet, results from **(I)** is Grid R-CNN, results from **(J)** is Cascade R-CNN, results from **(K)** is YOLOv4-LITE, results from **(L)** is EfficientDet, results from **(M)** is Free-Anchor, results from **(N)** is Lite-Yolov5, results from **(O)** is yolov5-X, and results from **(P)** is our proposed method. Note the pink circle show the false detection of ship and red circle is show the missing ship.

model could identify these ships without any false alarms, which is also one of the driving forces for this paper and highlights the current proposed model's powerful and robust feature representation skills.

The suggested model is compared with other state-of-the-art CNN based methods SAR ship identification techniques. As can be observed, two-stage or multi-stage approaches typically execute detection tasks more effectively than single-stage methods. However, compared to these two stage or multistage detection methods, one stage detection methods clearly have a faster inference efficiency. This might be as a result of the two-stage detection network's sophisticated network architecture and increased computational load. Some one-stage detection techniques that are more effective at detecting ships have recently been proposed, including YOLOv4-LITE (Liu et al., 2022), EfficientDet (Tan et al., 2020), Free Anchor (Zhang et al., 2019), Lite-Yolov5 (Xu et al., 2022), and two-stage detectors like ARPN (Zhao et al., 2020), DAPN (Cui et al., 2019), Quad-FPN (Zhang et al., 2021b), HR-SDNet (Wei et al., 2020), Grid R-CNN(Lu et al., 2019), and Cascade R-CNN (Cai and Vasconcelos, 2018) are the performance comparisons of the current model with other cutting-edge detectors.

The detection accuracies of inshore scenes for the proposed model and the other CNN-based techniques are also all lower than those of offshore scenes. De-formable convolution can lessen the interference of complicated backgrounds, particularly for inshore sceneries, hence the recently presented technique seems to be robust to background interferences. The other state-of-the-art techniques are less precise and have lower recall values than the model now under study because of their poor small ship recognition capabilities. As a result, it will be possible to consider an acceptable score threshold in the future to balance missed detections and false alarms. Additionally, accuracy needs to be further improved, for example, when striking military targets with precision. It might be suggested in the future to choose between speed and accuracy. The suggested model has a higher detection effectiveness. This might be the case because other methods overlook smaller targets since they do not consider the underlying data in the prediction layer. There are some false alarms in the identification outcomes of previous techniques for the complicated inshore scenarios. Particularly, several land features in the inshore scene are wrongly identified as targets by Faster-RCNN, SSD, R2CNN, ARPN, DAPN, Quad-FPN, HR-SDNet, Grid R-CNN, Cascade R-CNN, YOLOv4-LITE, EfficientDet, Free-Anchor, Lite-Yolov5. By doing several experiments using the SSDD and AirSARship datasets, we illustrate the effectiveness of our suggested model. The SSDD dataset ablation studies of FPN +PAN and attention mechanism modules have shown that each of them can enhance ship detection performance, and the combination of both can increase detection outcomes.

3.5 Detection performance on SAR large scene images

This section compares various CNN-based algorithms, such as Yolov5-X, Faster-RCNN (Lin et al., 2017), SSD (Liu et al., 2016), and R2CNN (Jiang et al., 2017), as well as current methods for object recognition in SAR images using large-scale scene SAR images. To validate the good migration ability of the suggested model, the actual ship identification in two more SAR large scene images has been carried out. Figure 13 displays the areas covered by the two SAR large scene images

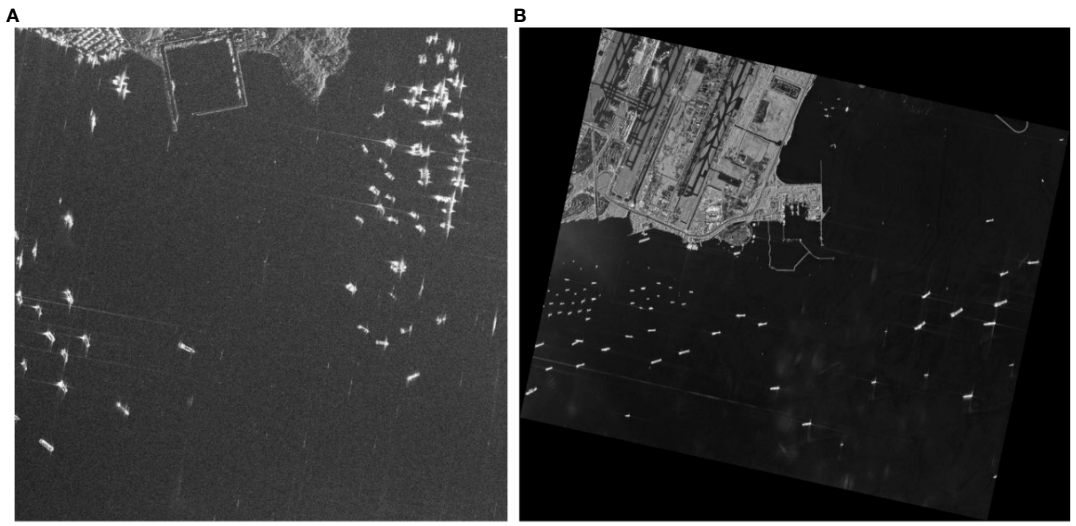


FIGURE 13
In this current research work, the two SAR large scenes images acquired from Chinese GF-3 satellite are utilized for ship detection. **(A)** AirSARship resolution is 1/3m, and **(B)** GF 3 satellite resolution is 1m.

acquired by the Chinese satellite GF3. These two SAR images were selected because they both lie along globally important routes named Malacca strait (Table 5). The VV polarization SAR images from Table 5 are considered because ships often exhibit higher backscattering values in VV polarization (Torres et al., 2012). Due to the restricted GPU memory, they are divided into 800 x 800 (Zhang et al., 2020a) and followed by (Xian et al., 2019) 1000 x 1000 of small sub images before being used for training and testing. SAR ships are ultimately added to the suggested model in order to actually detect them. The outcomes of the sub- image's detection are then added to the original SAR large scene image (Figure 14). The detection accuracy and speed of various approaches are assessed, respectively, using the AP and FPS (Table 6).

The two SAR large scene images are used to depict the results of SAR ship detection using the current model and other CNN-based methods. The current model is able to successfully detect the majority of ships, demonstrating its strong migration application capabilities in ocean surveillance (Figures 15A, B). Features and a clear environment come first. In particular, just a few ships in the big panorama were missed by all of these convolutional neural network -based approaches, which are indicated by pink circles in Figures 15A, B. The identification outcomes on the second SAR large scene image are displayed, and the identification results of various CNN-based approaches on the SAR large scene images are demonstrated on the left side of Figures 15A, B. Figures 15A, B's right side enlarges and displays two particular regions designated with brown and blue rectangles. The comparison of the detection outcomes of various model performances on SAR large scene images is presented in Figures 15A, B. Offshore ships make up the majority of this SAR large-scene image. Inshore scenes have substantial clutter, which could cause false alarms. The results of another CNN-based approach suggested in this study show that there are few false alarms and missed targets in the offshore scenes, and the false alarms are repressed in the inshore scenes as well. However, the proposed approach does not have false alarms or miss target detection, which is one of the motivating factors for this paper. Figures 15A, B illustrates the visualization of the SAR ship identification performance of the suggested model on two SAR large scene images acquired from the Chinese GF3 satellite and employed in the current study. The suggested model can successfully detect the majority of ships, which demonstrates its good migration application capability in ocean surveillance, it can be deduced from Figures 15A, B. In conclusion, the identification outcomes on the two SAR large scene images at

TABLE 5 The details descriptions of two SAR large scene images.

Images	Mode	Satellite	Resolution (m)	Image Size	Format
AirSARship	SL/UFS	GF-3	1/3	3000 × 3000	Tiff
GF3	SL	GF-3	1	15350 × 13592	Tiff

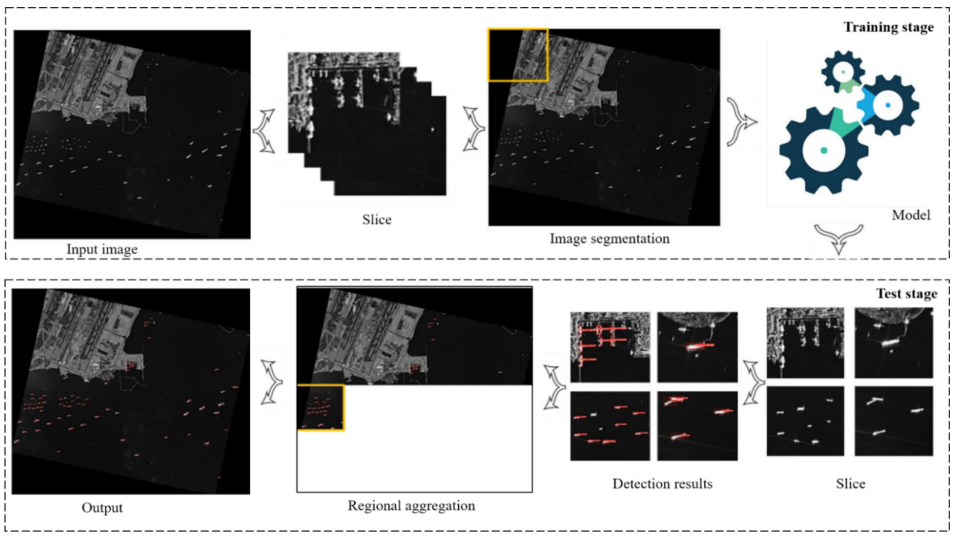


FIGURE 14
Ship target identification Framework in a SAR large scene images.

TABLE 6 Detection outcomes of various CNN-based approaches on two SAR large scene ship images.

Method	Image-1		Image-2	
	AP	Time (s)	AP	Time (s)
Faster-RCNN (Lin et al., 2017)	0.838	37s	0.845	17s
SSD (Liu et al., 2016)	0.790	14s	0.785	25s
R2CNN (Jiang et al., 2017)	0.879	65s	0.830	39s
Yolov5-X	0.765	140s	0.749	140s
Ours	0.956	157s	0.948	156s

various resolutions show that the present system identifies multi size ships with competitive outcomes and has a strong generalization capacity in comparison to the various CNN-based approaches created for identification object in SAR images. It demonstrates that the approach now under study can adapt to SAR images from various sources more effectively.

4 Ablation study

The ablation studies presented in this section, used to demonstrate the suggested FPN+PAN and attention mechanism module's effectiveness through removal and installation to better understand the behavior of the framework.

4.1 Ablation study on the FPN+PAN module

Table 7 illustrates the ablation research of YOLOv5 removal and installation of the FPN+PAN module. Table 7, "✘" denotes YOLOv5 without the FPN+PAN module, while "✓" denotes YOLOv5 with the FPN+PAN module (i.e., our suggested model). Experiments were carried out in offshore and inshore scenes respectively, to evaluate the identification achievement of the suggested approach for offshore and inshore scenes. In Table 7, the identification effectiveness of the model is approximately similar, with little modification in identification indicator, as a result of the relatively simple background of offshore ship and little interference, due to the identification achievement has been outstanding in the simple background, and only limited enhancement can be obtained. In contrast to offshore ships, inshore ships have a more complicated backdrop clutter. Moreover, wharfs as well as other structures on the shore significantly undermine detections, and SAR ship detection effectiveness generally declines. However, the detection performance is significantly enhanced by improving the feature representation capacity and minimizing the aliasing effect of fusion features. In Table 7, by installation of FPN +PAN modules, the result of p, r and AP rate of the model are enhanced by about 2.9%, 4.6% and, 5.5% respectively for offshore ship detection, while for inshore ship detection the p, r and AP rate of the model are enhanced by about 6.08%, 15% and, 8.07% respectively.

4.2 Ablation study on the attention mechanism module

Table 8 presented the ablation research of YOLOv5 removal and installation of the attention mechanism module. Table 8, "✘" denotes YOLOv5 without the attention mechanism module, "✓" denotes YOLOv5 with the attention mechanism module (i.e., our suggested model). The installation of attention mechanism module to our model, the detection performance is significantly enhanced by improving the feature representation capacity and minimizing the aliasing effect of fusion features. In Table 8, through installation of attention mechanism modules, as a result the p, r and AP rate of the model are enhanced by about 7.22%, 1.36% and, 5.31% respectively for offshore ship detection, while for inshore ship detection the p, r and AP rate of the model are enhanced by about 1.03%, 10.97% and, 10.67% respectively.

5 Conclusions and future work

In this modern technological era, the advanced machine learning and artificial intelligence-based models have revolutionized diverse research domains with full spectrum. Due to its automatic feature extraction and strong identification skills, it can be used in a variety of study fields. An improved version of the unique one stage YOLOv5 for SAR ship identification has been proposed in this study work, drawing inspiration from the capabilities of these models in other research domains. The generic YOLOv5 model has been improved to address the major issues with the SAR ship detection process. These issues include complexity (complex background interferences, various size ship feature differences, and indistinct tiny ship characteristics), high-cost effectiveness, poor identification and

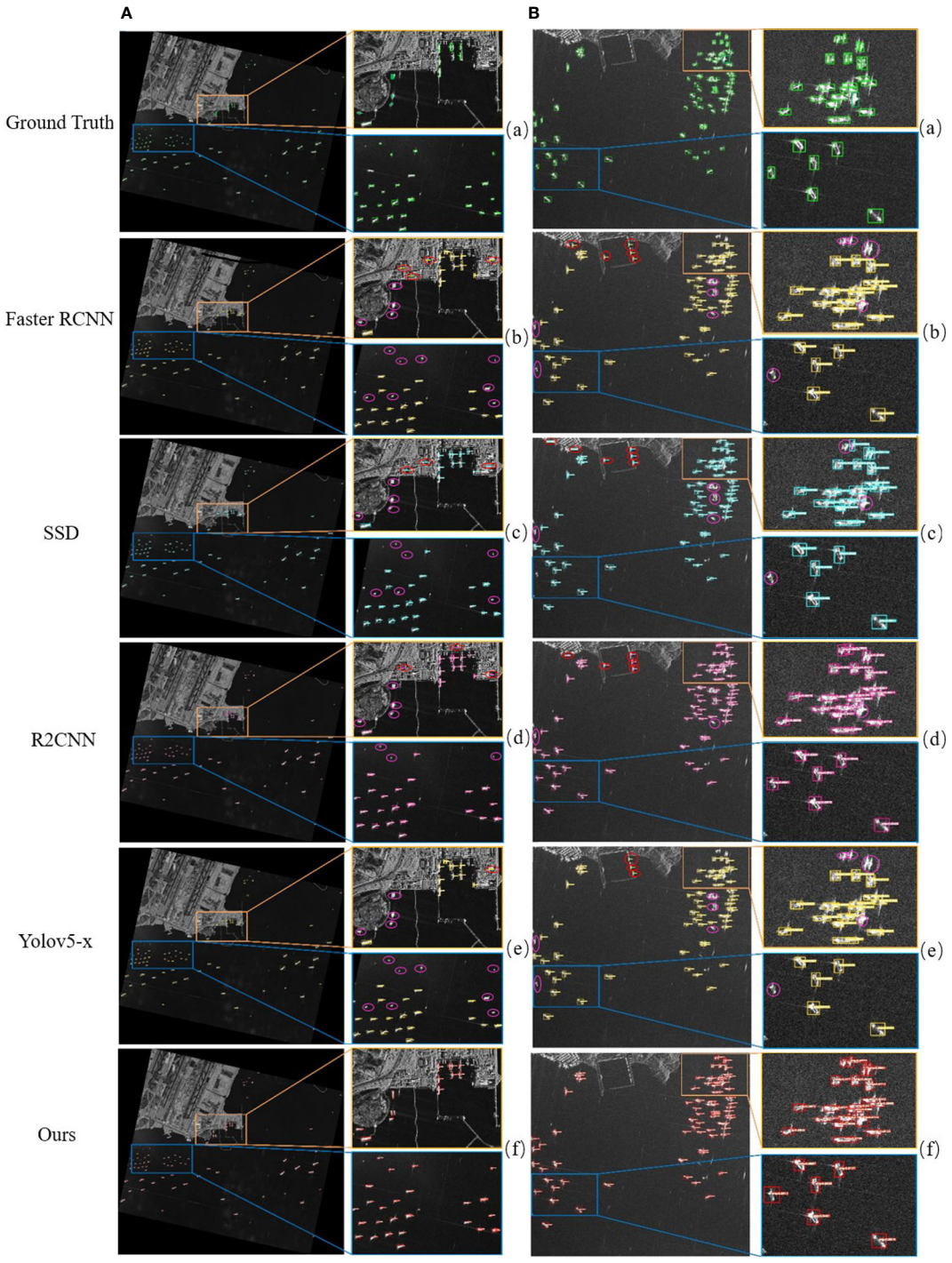


FIGURE 15
The identification outcomes of various CNN based approaches, **(A)** first large-scene SAR image based on SSDD Dataset, **(B)** second SAR large scene image based on AirSARship Dataset. Rectangles with green colors correspond to the ground truth ships while yellow, sky color, pink and red colors are referred to predictions ship, respectively. The red circle shows the false detection of ship and pink circle is showing the missing ship. The right side of an image displays two enlarged special areas that are marked by blue and brown rectangles respectively.

TABLE 7 The ablation study of our proposed model removal and installation the FPN+PAN module on the SSDD dataset.

FPN+PAN	OFF-shore			In-shore		
	p (%)	r (%)	AP(%)	p (%)	r (%)	AP(%)
✗	91.67	89.43	89.78	83.36	75.56	84.20
✓	94.57	94.06	**95.36**	89.44	90.77	**92.27**

TABLE 8 The ablation study of our proposed model removal and installation the attention mechanism module on the SSDD dataset.

Attention Mechanism	OFF-shore			In-shore		
	p (%)	r (%)	AP(%)	p (%)	r (%)	AP(%)
✗	87.35	92.70	90.05	88.41	79.80	81.60
✓	94.57	94.06	**95.36**	89.44	90.77	**92.27**

recognition rates, and implementation complexities. The changes to the generic YOLOv5 model in the neck region and backbone section employing C3 and PAN structure have been designed to address these major issues. The SSDD and AirSARship open SAR ship datasets, as well as two SAR large scene images acquired from the GF-3 Chinese satellite, are utilized to obtain the experimental results. After producing testing findings, it has been determined that the enhancement to the generic YOLOv5 model not only enhanced identification capabilities but also demonstrated that this model is not data-hungry (to provide optimum results even for a small amount of dataset). The applicability of this model is assessed using a variety of validation metrics, including accuracy, different training and test sets, and TF values, as well as comparisons with other cutting-edge classification models (ARPN, DAPN, Quad-FPN, HR-SDNet, Grid R-CNN, Cascade R-CNN, Multi-Stage YOLOv4-LITE, EfficientDet, Free-Anchor, Lite-Yolov5). Based on the performance values, it has been determined that the examined model exceeded the benchmark models targeted in this research work by producing high identification rates. Additionally, these high identification rates show how useful the suggested approach is for maritime surveillance. Recommended and forthcoming future work includes the following:

- To enhance the effectiveness of our model detection in the future, we will consider the challenges in SAR data, such as the azimuth ambiguity, sidelobes, and the sea condition.
- In the future, we will investigate optimizing the detection speed of our model.
- We might suggest merging contemporary deep CNN abstract features with conventional concrete ones to further improve detection accuracy.
- In order to further boost the identification speed and accuracy, we will focus on merging the backscattering characteristics of ships in SAR images with convolutional network architecture and offering a robust constraint, such as a mask.
- Future research on instance segmentation and ship detection will be taken into consideration.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

Conceptualization, MY, LS. Methodology, MY and SH. software, MY, SH and LS. Validation, WJ, LS and XM. Formal analysis, AC and DW. Investigation, XM. Resources, WJ. Data curation, MY, KD and DW. Writing—original draft preparation, MY. Writing—review and editing, MY, XM and LS. Visualization, LS, KD. Supervision, WJ and LS. Project administration, LS. Funding acquisition, WJ and LS. All authors contributed to the article and approved the submitted version.

Funding

Funding National Key Research and Development Program of China (2017YFC1405600).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Baselice, F., and Ferraioli, G. (2013). Unsupervised coastal line extraction from SAR images. *IEEE Geosci. Remote Sens. Lett.* 10, 1350–1354. doi: 10.1109/LGRS.2013.2241013

Cai, Z., and Vasconcelos, N. (2018). "Cascade r-cnn: Delving into high quality object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 6154–6162. doi: 10.48550/arXiv.1712.00726

Caputo, S., Castellano, G., Greco, F., Mencar, C., Petti, N., and Vessio, G. "Human detection in drone images using YOLO for search-and-Rescue operations," in *International conference of the Italian association for artificial intelligence* (Springer), 326–337. doi: 10.1007/978-3-031-08421-8_22

Chang, Y.-L., Anagaw, A., Chang, L., Wang, Y. C., Hsiao, C.-Y., and Lee, W.-H. (2019). Ship detection based on YOLOv2 for SAR imagery. *Remote Sens.* 11, 786. doi: 10.3390/rs11070786

Chen, G., Li, G., Liu, Y., Zhang, X.-P., and Zhang, L. (2019). SAR image despeckling based on combination of fractional-order total variation and nonlocal low rank regularization. *IEEE Trans. Geosci. Remote Sens.* 58, 2056–2070. doi: 10.1109/TGRS.2019.2952662

Chen, H., and Shi, Z. (2020). A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sens.* 12, 1662. doi: 10.3390/rs12101662

Cui, Z., Li, Q., Cao, Z., and Liu, N. (2019). Dense attention pyramid networks for multi-scale ship detection in SAR images. *IEEE Trans. Geosci. Remote Sens.* 57, 8983–8997. doi: 10.1109/TGRS.2019.2923988

Du, L., Dai, H., Wang, Y., Xie, W., and Wang, Z. (2019). Target discrimination based on weakly supervised learning for high-resolution SAR images in complex scenes. *IEEE Trans. Geosci. Remote Sens.* 58, 461–472. doi: 10.1109/TGRS.2019.2937175

Elfwing, S., Uchibe, E., and Doya, K. (2018). Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks* 107, 3–11. doi: 10.48550/arXiv.1702.03118

Gao, Y., Gao, F., Dong, J., and Wang, S. (2019). Change detection from synthetic aperture radar images based on channel weighting-based deep cascade network. *IEEE J. selected topics Appl. Earth observations Remote Sens.* 12, 4517–4529. doi: 10.1109/JSTARS.2019.2953128

Girshick, R. (2015). "Fast r-cnn," in *Proceedings of the IEEE international conference oncomputer vision*. 1440–1448. doi: 10.1109/ICCV.2015.169

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 580–587. doi: 10.48550/arXiv.1311.2524

Han, J., Li, G., and Zhang, X.-P. (2019). Refocusing of moving targets based on low-bit quantized SAR data *via* parametric quantized iterative hard thresholding. *IEEE Trans. Aerospace Electronic Syst.* 56, 2198–2211. doi: 10.1109/TAES.2019.2944707

Jiang, Y., Zhu, X., Wang, X., Yang, S., Li, W., Wang, H, et al. (2017). R2CNN: Rotational region CNN for orientation robust scene text detection. doi: 10.48550/arXiv.1706.09579. arXiv preprint arXiv:1706.09579.

Li, X., Chen, P., and Fan, K. (2020). "Overview of deep convolutional neural network approaches for satellite remote sensing ship monitoring technology," in *IOP conference series: Materials science and engineering* (IOP Publishing), 012071. doi: 10.1088/1757-899X/730/1/012071

Li, T., Liu, Z., Xie, R., and Ran, L (2017b). An improved superpixel-level CFAR detection method for ship targets in high-resolution SAR images. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 11, 184–194. doi: 10.1109/JSTARS.2017.2764506

Lin, H., Chen, H., Jin, K., Zeng, L., and Yang, J. (2019). Ship detection with superpixel-level Fisher vector in high-resolution SAR images. *IEEE Geosci. Remote Sens. Lett.* 17, 247–251. doi: 10.1109/LGRS.2019.2920668

Lin, T.-Y., Girshick, R., He, K., Hariharan, B., and Belongie, ,. S. (2017). "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2117–2125. doi: 10.48550/arXiv.1612.03144

Li, J., Qu, C., and Shao, J. (2017a). "Ship detection in SAR images based on an improved faster r-CNN," in *2017 SAR in big data era: Models, methods and applications (BIGSARDATA)* (IEEE), 1–6. doi: 10.1109/BIGSARDATA.2017.8124934

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al (2016). "Ssd: Single shot multibox detector," in *European Conference on computer vision* (Springer), 21–37. doi: 10.48550/arXiv.1512.02325

Liu, S., Kong, W., Chen, X., Xu, M., Yasir, M., Zhao, L., et al. (2022). Multi-scale ship detection algorithm based on a lightweight neural network for spaceborne SAR images. *Remote Sens.* 14, 1149. doi: 10.3390/rs14051149

Liu, S., Qi, L., Qin, H., Shi, J., and Jia, J. (2018). "Path aggregation network for instance segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 8759–8768. doi: 10.48550/arXiv.1803.01534

Li, J., Xu, C., Su, H., Gao, L., and Wang, T. (2022). Deep learning for SAR ship detection: Past, present and future. *Remote Sens.* 14, 2712. doi: 10.3390/rs14112712

Lu, X., Li, B., Yue, Y., Li, Q., and Yan, J (2019). "Grid r-cnn," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7363–7372. doi: 10.48550/arXiv.1811.12030

Ma, M., Chen, J., Liu, W., and Yang, W. (2018). Ship classification and detection based on CNN using GF-3 SAR images. *Remote Sens.* 10, 2043. doi: 10.3390/rs10122043

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., and Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *J. big Data* 2, 1–21. doi: 10.1186/s40537-014-0007-7

Nepal, U., and Eslamiat, H. (2022). Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVs. *Sensors* 22, 464. doi: 10.3390/s22020464

Niedermeier, A., Romaneessen, E., and Lehner, S. (2000). Detection of coastlines in SAR images using wavelet methods. *IEEE Trans. Geosci. Remote Sens.* 38, 2270–2281. doi: 10.1109/36.868884

Ouchi, K. (2013). Recent trend and advance of synthetic aperture radar with selected topics. *Remote Sens.* 5, 716–807. doi: 10.3390/rs5020716

Patel, K., Bhatt, C., and Mazzeo, P. L. (2022). Deep learning-based automatic detection of ships: an experimental study using satellite images. *J. Imaging* 8, 182. doi: 10.3390/jimaging8070182

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A (2016). "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788. doi: 10.48550/arXiv.1506.02640

Redmon, J., and Farhadi, A (2017). "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7263–7271. doi: 10.48550/arXiv.1612.08242

Redmon, J., and Farhadi, A. (2018). Yolov3: An incremental improvement. doi: 10.48550/arXiv.1804.02767. arXiv preprint arXiv:1804.02767.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* 28. doi: 10.48550/arXiv.1506.01497

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., and Savarese, S. (2019). "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 658–666. doi: 10.48550/arXiv.1902.09630

Salembier, P., Liesegang, S., and López-Martínez, C. (2018). Ship detection in SAR images based on maxtree representation and graph signal processing. *IEEE Trans. Geosci. Remote Sens.* 57, 2709–2724. doi: 10.1109/TGRS.2018.2876603

Tan, M., Pang, R., and Le, Q. V. (2020). "Efficientdet: Scalable and efficient object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10781–10790. doi: 10.48550/arXiv.1911.09070

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., et al. (2012). GMES sentinel-1 mission. *Remote Sens. Environ.* 120, 9–24. doi: 10.1016/j.rse.2011.05.028

Wang, X., Chen, C., Pan, Z., and Pan, Z. (2019). Fast and automatic ship detection for SAR imagery based on multiscale contrast measure. *IEEE Geosci. Remote Sens. Lett.* 16, 1834–1838. doi: 10.1109/LGRS.2019.2913873

Wang, C.-Y., Liao, H.-Y. M., Wu, Y.-H., Chen, P.-Y., Hsieh, J.-W., and Yeh, I.-H. (2020a). "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 390–391. doi: 10.48550/arXiv.1911.11929

Wang, X., Li, G., and Zhang, X.-P. (2020c). "Contrast of contextual Fisher vectors for ship detection in SAR images," in *2020 IEEE international radar conference (RADAR)* (IEEE), 198–202. doi: 10.1109/RADAR42522.2020.9114850

Wang, X., Li, G., Zhang, X.-P., and He, Y. (2020d). Ship detection in SAR images *via* local contrast of Fisher vectors. *IEEE Trans. Geosci. Remote Sens.* 58, 6467–6479. doi: 10.1109/TGRS.2020.2976880

Wang, W., Ren, J., Su, C., and Huang, M. (2021). Ship detection in multispectral remote sensing images *via* saliency analysis. *Appl. Ocean Res.* 106, 102448. doi: 10.1016/j.apor.2020.102448

Wang, S., Wang, M., Yang, S., and Jiao, L. (2016). New hierarchical saliency filtering for fast ship detection in high-resolution SAR images. *IEEE Trans. Geosci. Remote Sens.* 55, 351–362. doi: 10.1109/TGRS.2016.2606481

Wang, D., Wan, J., Liu, S., Chen, Y., Yasir, M., Xu, M., et al. (2022). BO-DRNet: An improved deep learning model for oil spill detection by polarimetric features from SAR images. *Remote Sens.* 14, 264. doi: 10.3390/rs14020264

Wang, R., Xu, F., Pei, J., Zhang, Q., Huang, Y., Zhang, Y., et al (2020b). "Context semantic perception based on superpixel segmentation for inshore ship detection in SAR image," in *2020 IEEE radar conference (RadarConf20)* (IEEE), 1–6. doi: 10.1109/RadarConf2043947.2020.9266627

Wei, S., Su, H., Ming, J., Wang, C., Yan, M., Kumar, D., et al. (2020). Precise and robust ship detection for high-resolution SAR imagery based on HR-SDNet. *Remote Sens.* 12, 167. doi: 10.3390/rs12010167

Willburger, K., Schwenk, K., and Brauchle, J. (2020). Amaro–an on-board ship detection and real-time information system. *Sensors* 20, 1324. doi: 10.3390/s20051324

Xian, S. Z. ,. W., Yuanrui, S., Wenhui, D., Yue, Z., and Kun, F. (2019). Air-sarship–1.0: High resolution sar ship detection dataset. *J. Radars* 8, 852–862. doi: 10.12000/JR19097

Xiao, Q., Cheng, Y., Xiao, M., Zhang, J., Shi, H., Niu, L., et al. (2020). Improved region convolutional neural network for ship detection in multiresolution synthetic aperture radar images. *Concurrency Computation: Pract. Exp.* 32, e5820. doi: 10.1002/cpe.5820

Xiong, G., Wang, F., Zhu, L., Li, J., and Yu, W. (2019). SAR target detection in complex scene based on 2-d singularity power spectrum analysis. *IEEE Trans. Geosci. Remote Sens.* 57, 9993–10003. doi: 10.1109/TGRS.2019.2930797

Xiong, B., Sun, Z., Wang, J., Leng, X., and Ji, K. A. (2022). Lightweight Model for Ship Detection and Recognition in Complex-Scene SAR Images. *Remote Sens* 14, 6053. doi: 10.3390/rs14236053

Xu, P., Li, Q., Zhang, B., Wu, F., Zhao, K., Du, X., et al. (2021). On-board real-time ship detection in HISEA-1 SAR images based on CFAR and lightweight deep learning. *Remote Sens.* 13, 1995. doi: 10.3390/rs13101995

Xu, X., Zhang, X., and Zhang, T. (2022). Lite-yolov5: A lightweight deep learning detector for on-board ship detection in large-scene sentinel-1 sar images. *Remote Sens.* 14, 1018. doi: 10.3390/rs14041018

Yasir, M., Jianhua, W., Mingming, X., Hui, S., Zhe, Z., Shanwei, L., et al. (2022). Ship detection based on deep learning using SAR imagery: a systematic literature review. *Soft Computing*, 1–22. doi: 10.1007/s00500-022-07522-w

Yekeen, S. T., Balogun, A. L., and Yusof, K. B. W. (2020). A novel deep learning instance segmentation model for automated marine oil spill detection. *ISPRS J. Photogrammetry Remote Sens.* 167, 190–200. doi: 10.1016/j.isprsjprs.2020.07.011

Yuan, S., Yu, Z., Li, C., and Wang, S. (2020). A novel SAR sidelobe suppression method based on CNN. *IEEE Geosci. Remote Sens. Lett.* 18, 132–136. doi: 10.1109/LGRS.2020.2968336

Yu, J., Jiang, Y., Wang, Z., Cao, Z., and Huang, T. (2016). "Unitbox: An advanced object detection network," in *Proceedings of the 24th ACM international conference on multimedia*, 516–520. doi: 10.1145/2964284.2967274

Zhang, X., Liu, G., Zhang, C., Atkinson, P. M., Tan, X., Jian, X., et al. (2020b). Two-phase object-based deep learning for multi-temporal SAR image change detection. *Remote Sens.* 12, 548. doi: 10.3390/rs12030548

Zhang, K., Luo, Y., and Liu, Z (2021a). "Overview of research on marine target recognition," in *2nd international conference on computer vision, image, and deep learning* (SPIE), 273–282. doi: 10.1117/12.2604530

Zhang, X., Wan, F., Liu, C., Ji, R., and Ye, Q. (2019). Freeanchor: Learning to match anchors for visual object detection. *Adv. Neural Inf. Process. Syst.* 32. doi: 10.48550/arXiv.1909.02466

Zhang, X., Xie, T., Ren, L., and Yang, L. (2020c). "Ship detection based on superpixel-level hybrid non-local MRF for SAR imagery," in *2020 5th Asia-pacific conference on intelligent robot systems (ACIRS)* (IEEE), 1–6. doi: 10.1109/ACIRS49895.2020.9162609

Zhang, T., Zhang, X., and Ke, X. (2021b). Quad-FPN: A novel quad feature pyramid network for SAR ship detection. *Remote Sens.* 13, 2771. doi: 10.3390/rs13142771

Zhang, T., Zhang, X., Ke, X., Zhan, X., Shi, J., Wei, S., et al. (2020a). LS-SSDD-v1.0: A deep learning dataset dedicated to small ship detection from large-scale sentinel-1 SAR images. *Remote Sens.* 12, 2997. doi: 10.3390/rs12182997

Zhao, Y., Zhao, L., Xiong, B., and Kuang, G. (2020). Attention receptive pyramid network for ship detection in SAR images. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 13, 2738–2756. doi: 10.1109/JSTARS.2020.2997081

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 18 January 2023
DOI 10.3389/fmars.2022.1058019

Check for updates

OPEN ACCESS

EDITED BY
Hong Song,
Zhejiang University, China

REVIEWED BY
Carlos Pérez-Collazo,
University of Vigo, Spain
Jing Lian,
Lanzhou Jiaotong University, China
Viacheslav Voronin,
Moscow State Technological University "Stankin", Russia

*CORRESPONDENCE
Tenghui Wang
vito.wang777@gmail.com

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 30 September 2022
ACCEPTED 01 December 2022
PUBLISHED 18 January 2023

CITATION
Zhao Z, Zhou Z, Lai Y, Wang T, Zou S, Cai H and Xie H (2023) Single underwater image enhancement based on adaptive correction of channel differential and fusion. *Front. Mar. Sci.* 9:1058019. doi: 10.3389/fmars.2022.1058019

COPYRIGHT
© 2023 Zhao, Zhou, Lai, Wang, Zou, Cai and Xie. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Single underwater image enhancement based on adaptive correction of channel differential and fusion

Zefeng Zhao[1], Zhuang Zhou[2], Yunting Lai[2], Tenghui Wang[2]*, Shujie Zou[1], Haohao Cai[1] and Haijun Xie[2]

[1]Faculty of Data Science, City University of Macau, Macau, Macao SAR, China, [2]Key Laboratory of Intelligent Detection in Complex Environment of Aerospace Land and Sea, Beijing Institute of Technology, Zhuhai, China

Clear underwater images are necessary in many underwater applications, while absorption, scattering, and different water conditions will lead to blurring and different color deviations. In order to overcome the limitations of the available color correction and deblurring algorithms, this paper proposed a fusion-based image enhancement method for various water areas. We proposed two novel image processing methods, namely, an adaptive channel deblurring method and a color correction method, by limiting the histogram mapping interval. Subsequently, using these two methods, we took two images from a single underwater image as inputs of the fusion framework. Finally, we obtained a satisfactory underwater image. To validate the effectiveness of the experiment, we tested our method using public datasets. The results showed that the proposed method can adaptively correct color casts and significantly enhance the details and quality of attenuated underwater images.

KEYWORDS

underwater image processing, image enhancement, histogram stretching, color correction, fusion

1 Introduction

Underwater optical imaging is one of the most direct ways to obtain underwater information. Currently, clear underwater images are urgently needed for marine life research, marine environmental protection, and underwater equipment maintenance, among other fields. Unlike that on the ground, however, underwater imaging tends to be less effective due to, for example, color distortion, blurred details, and low contrast and brightness, among others. These issues hinder the use of unprocessed underwater images for research.

Absorption and scattering are the causes of poor underwater imaging. During the propagation of light underwater, the attenuation rates of light at different wavelengths

differ due to water absorption (Adolfson and Berghage, 1974). As the water depth increases, red light is attenuated first, followed by yellow, green and purple, and then blue light. On the other hand, scattering of the underwater medium and particles leads to fogging and the blurring of details (Chao and Wang, 2010). Scattering is classified into forward scattering and backward scattering. Forward scattering is defined as light from the object moving through particles in the same direction before it reaches the camera, while backward scattering is defined as light that is reflected by particles before it reaches the object. Scattering can be considered as chiffon covering the underwater image, reducing its contrast and blurring its details. Therefore, deblurring and color correction are the main challenges to obtaining information from attenuated underwater images.

In this paper, we propose a single underwater image enhancement method based on adaptive correction of channel differential and fusion (ACCDF). In *Section 3*, firstly, we deblur and enhance the underwater image using an underwater physical imaging model and the image pixel-based method separately. Subsequently, we use the pyramid-based fusion method to enhance the underwater image. Finally, we obtain the enhanced underwater image. In *Section 4*, we use the EUVP (Enhancing Underwater Visual Perception) dataset (Islam et al., 2020) and the Raws dataset from the UIEB (Underwater Image Enhancement Benchmark) (Li et al., 2019) for analysis. According to the the results, our method can deblur and adaptively correct color. We also compare our method with other methods using the same dataset, validating the higher performance of our method.

The main contributions of this article are as follows:

1) A method of underwater image blurring was proposed based on the dominant hue of the image combined with dark channel prior to better deal with color degradation in different environments.
2) Based on the contrast limited adaptive histogram equalization (CLAHE) method, the contrast stretching of the dominant hue's channel of the image was restricted better to restore the severely degraded color in the image.
3) It was demonstrated that the fusion of the deblurred and color-corrected images can preserve their respective advantages at the same time.

2 Related work

Underwater image processing includes an underwater physical imaging model and a non-physical model based on image pixel processing. Where the physical method estimates the reasonable degradation model of the image and restores the image quality through the inverse process, the non-physical method improves visual quality by enhancing the values of the parts of interest and weakening the values of the parts of no interest.

Underwater image degradation, also known as fogging, is caused by the scattering of suspended particles and absorption. He et al. (2010) proposed the dark channel prior (DCP) image recovery method to solve image fogging. The authors considered that, in each locality of a picture without fog, there is at least one color channel in each locality with pixel values close to zero, and these darkest points are used to remove fogging evenly. Although DCP performs well in terrestrial images, Chao and Wang (2010) and Chiang and Chen (2011) provided proof that DCP is not effective or even worse when applied directly to underwater scenes due to absorption and the presence of large areas similar to atmospheric light. Considering the attenuation of red light, Galdran et al. (2015) proposed an automatic red channel recovery method to reduce the influence of the red channel and artificial light sources on the estimation of the transmission map. Furthermore, considering that the red channel caused by absorption in the underwater environment is almost dark, Drews et al. (2013) proposed the underwater dark channel prior (UDCP) method. The modified DCP is only applied to the blue and green channels. Peng et al. (2018) proposed the generalization of the dark channel prior (GDCP) method to estimate the transmission map based on the brightness difference between the environment and the object. Similar to UDCP, the authors processed different color channels in the image formation model. Song et al. (2018) estimated the depth of the underwater scene by means of supervised learning. The background light and the transmission map estimated by the depth map have been proven to be able to effectively improve the image restoration effect.

The distribution of the histogram reflects the statistical characteristics of the image gray level. Histogram equalization (HE), as an image enhancement method, can enhance the contrast of the image. The HE method uniformly maps the gray level of the image histogram to the range from 0 to 255. This method can obtain a better contrast enhancement effect when the histogram of the image does not change dramatically. However, it is obvious that roughly stretching the histogram distribution to the range of 0–255 has a limited effect. Based on the HE method, the adaptive histogram equalization (AHE) method was proposed (Ketcham et al., 1974). It uses grids to cut the image into multiple regions, calculates the cumulative distribution function of the region where each pixel is located, and then maps the pixels. However, when the histogram in an area has only one gray level, an obvious "mosaic" phenomenon (Figure 1) will appear in adjacent areas. To avoid tearing caused by the AHE method, Pizer et al. (1987) proposed the contrast limited adaptive histogram equalization (CLAHE) method. This



FIGURE 1
Comparison of the histogram equalization (HE), adaptive histogram equalization (AHE), and contrast limited adaptive histogram equalization (CLAHE) methods.

method includes two innovations. The first is to limit the histogram distribution. When the gray level of an image exceeds the specified threshold, the excess part will be evenly distributed to each gray level. The second is to consider the bilinear interpolation of the image boundary. It finds two adjacent windows for one pixel of the boundary window of the image and four adjacent windows for one pixel of the non-boundary window, calculates the mapping value of the adjacent window histogram cumulative distribution function (CDF) to the pixel point, and finally performs linear and bilinear interpolations. Huang et al. (2018) proposed the relative global histogram stretching (RGHS) method to stretch the histogram of the GB channels according to the histogram distribution and absorption of underwater images.

The above algorithms have their own advantages in terms of color correction or defogging or image detail enhancement. Combining multiple algorithms to process one underwater image may be a good attempt. In order to combine the advantages of various methods, Ancuti et al. (2012) proposed the FUSION framework to solve the problems of a single attenuated underwater image. This framework has two inputs: one is the image processed by white balance and the other the image processed by CLAHE. In the experiment, the authors found that the direct pyramid fusion technology can make the image glow, but the multi-scale filter did not work well. Therefore, the authors used the multi-scale Laplacian pyramid (Burt and Adelson, 1987) to fuse the two processed images. Lu et al. (2015) performed optimization on the basis of FUSION. In order to further strengthen the image details, the authors changed the original FUSION input to the denoised image obtained using the local adaptive filter (Lu et al., 2017) and the denoised and reflected high-resolution image obtained using the subsample super-resolution method (Huang et al., 2015). Bai et al. (2020) used a multi-scale Gaussian filter to denoise the image on the basis of FUSION, conducted global and local histogram processing according to the different color channels of the image, and then fused the two different processing results.

The machine learning method has strong performance and is a powerful technique for underwater image processing. Since machine learning requires ground truth, Perez et al. (2017) believed that the repaired image has details similar to the ground truth, so a new convolutional neural network (CNN) model was proposed, which uses the repaired image as the ground truth to train the model. However, the constructed model cannot adapt to images at different locations. Anwar et al. (2018) considered that the existing machine learning methods would need ground truth for training, which is extremely challenging; therefore, a fully data-driven and end-to-end model was proposed: the underwater image enhancement convolutional neural network (UWCNN). This model trains optimizing the loss of the mean square error (MSE) and structural similarity (SSIM), which was an innovative attempt. Chen et al. (2021) uses the CNN model to estimate the backward scattering and direct transmission. Fabbri et al. (2018) proposed generating enhanced underwater images with generic adversarial networks (GANs). On the other hand, the authors also proposed generating degraded images based on non-degraded images using GANs, which can build more underwater available datasets. Guo et al. (2019) proposed a new GAN model that combines residual learning, dense concatenation, and multi-scaling and uses a variety of loss functions to preserve the textural details.

3 Adaptive correction of channel differential and fusion

In this paper, we propose the enhancement of a single underwater image based on ACCDF. Our main aim was to deblur and correct images based on the degree of color attenuation.

Hitam et al. (2013) stated that red light first degrades in water and is almost absorbed at a depth of 5 m. As the water depth increases, green light is then gradually absorbed, and finally blue light. Due to this phenomenon, it is reasonable to assume that, in most cases, underwater images have the most information on the blue and green channels. Based on a previous work, we have concluded that the main hue of the image correlates with the mean of each channel (Lai et al., 2022) and that, by determining the mean value of the channels, a more appropriate transmission map can be obtained. On the other

hand, the channel with the highest mean value also indicates the least degree of absorption; therefore, its maximum should be limited in color correction and the minimum of the other channels should be increased to better correct the color deviations of the image.

Previous research has proven the good performance of the FUSION method in the field of underwater image enhancement (Ancuti et al., 2012). We therefore assumed that its advantages can be utilized by fusing the deblurring image and the color correction image. The methods we applied in this paper included decomposing the image in the RGB (red, green, blue) color space and correcting the degradation of the various channels of the image due to the differential attenuation of light at different wavelengths in water resulting in varying degrees of degradation of the image channels. Firstly, we compared the mean values of the GB channels in order to determine whether the image is greenish or bluish. For the bluish image, we supposed that only the transmission map obtained from channel B is more suitable for the actual underwater imaging model. On the contrary, for the greenish image, we assumed that only the transmission map obtained from the G channel is more appropriate. We then combined the transmission map and the underwater imaging model to realize image deblurring. Subsequently, we stretched the contrast of the main hue of the deblurred underwater image to the maximum limit, while the contrast of the R channel was also stretched with a minimum increase. Thereafter, we converted the image to the hue saturation value (HSV) color space and stretched the SV channels to improve the contrast and brightness of the image. Finally, deblurring and color correction of the underwater images were combined based on the image pyramid to obtain an enhanced underwater image that can simultaneously achieve deblurring and color correction. The flowchart of the ACCDF is shown in Figure 2.

3.1 Adaptive correction of channel differential for deblurring

The attenuation of light of different wavelengths is different in water, which results in the blue–green hue of most underwater images. After a lot of experiments, we found that the dominant color of the image is directly related to the average size of the RGB channels.

When the image was bluish, we assumed that the information of channel B in the underwater image can better reflect the real information of the target scene.

$$B_{\text{ave}} > G_{\text{ave}} \tag{1}$$

$$J^{\text{ACCDF}}(x,y) = \min_{(x,y)\in\Omega(x,y)}\left(\min_{C=B} J^{C}(x,y)\right) \tag{2}$$

$$\tilde{t}(x,y) = 1 - \min_{(x,y)\in\Omega(x,y)}\left(\min_{C=B}\frac{I^{C}(x,y)}{A^{c}}\right) \tag{3}$$

In the equations, C denotes the color channel, $\Omega(x,y)$ represents a local patch, I^{C} is the input image, J^{C} is the dark channel, A^{c} is the global atmospheric light, and $\tilde{t}$ is the transmission of the patch.

The opposite is true when the image is green.

$$G_{\text{ave}} > B_{\text{ave}} \tag{4}$$

$$J^{\text{ACCDF}}(x,y) = \min_{(x,y)\in\Omega(x,y)}\left(\min_{C=G} J^{C}(x,y)\right) \tag{5}$$

$$\tilde{t}(x,y) = 1 - \min_{(x,y)\in\Omega(x,y)}\left(\min_{C=G}\frac{I^{C}(x,y)}{A^{c}}\right) \tag{6}$$

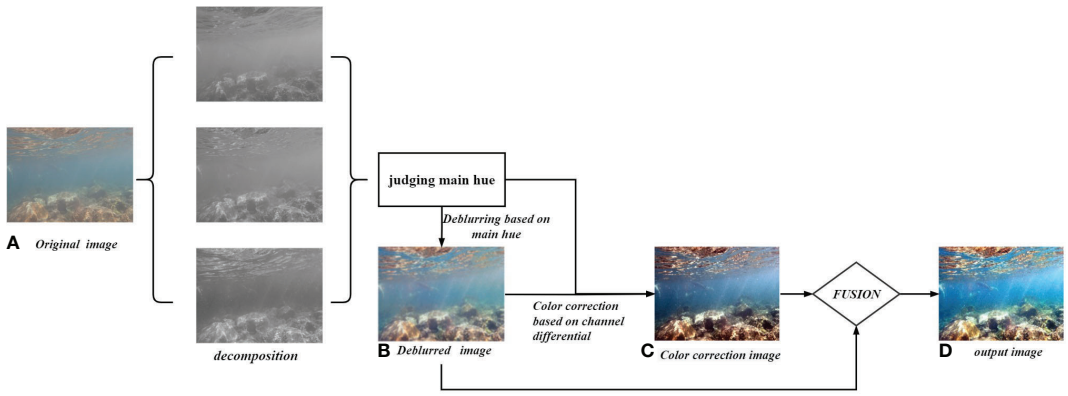


FIGURE 2
Flowchart of adaptive correction of channel differential and fusion (ACCDF). **(A)** Original image. After decomposition, information on the three channels was extracted. **(B, C)** Images obtained by deblurring and color correction based on the mean values of each channel. **(D)** Image obtained from the fusion of **(B, C)** using the FUSION algorithm.

The estimation of channel A also followed the DCP method: the top 10% brightest points were selected from the corresponding dark channel, J^{ACCDF}. Afterward, the pixels at the same position as these points in the original image, $I^C(x,y)$, were determined and the maximum value taken as the estimated value of the background light of channel C. Combining the underwater image and the imaging model, the deblurred underwater image can be obtained.

To avoid over- and underexposure, a threshold for the transmission map was set. According to a large number of engineering experiments, the lower threshold value was set as 0.3 and the upper threshold value as 0.9.

$$\tilde{t}(x,y) = \begin{cases} 0.3, & \text{if} \quad \tilde{t}(x,y) < 0.3 \\ \tilde{t}(x,y), & \textbf{else} \\ 0.9, & \text{if} \quad \tilde{t}(x,y) > 0.9 \end{cases} \tag{7}$$

Based on the formulation below, the deblurred underwater image can be obtained.

$$J^{\mathrm{ACCDFF}}_{\mathrm{deblurred}}(x,y) = \frac{I^C(x,y) - A^C}{\tilde{t}(x,y)} + A^C \tag{8}$$

As can be seen, the first line of Figure 3 shows that both the DCP (Figure 3B) and the ACCDF method (Figure 3C) can produce a clearer image compared to the original image (Figure 3A). However, the image processed using the DCP method had serious blue distortion, which caused the background to appear unclear. On the other hand, the ACCDF method that estimates the transmission map by considering the dominant hue channel significantly and partially eliminated the color distortion of the underwater image. The second line of Figure 3 displays a comparison of the details for selected parts of the first line. In Figure 3D, the fish is not clear because of the foggy blur, and the background and the fish texture are not clear, as if they are hidden. Although the outline of the fish and the background in Figure 3E after DCP processing became clear, the details of the texture on the fish were not obvious because the overall color was blue. Figure 3F shows the elimination of the overall foggy blur of the image after our method processing, and the image does not appear as a blue color cast, showing bright colors and clear details, especially the black stripes on the fish body that were well reproduced.

Figure 4 shows that the estimated transmission map of the ACCDF method was obviously different from that of the DCP method. The transmission map corresponding to the stone part had a clearer texture, reflected in the processing results shown in Figure 4D. The color of the stone part in the image processed using the ACCDF method was clearer and the textural detail richer. In particular, the red boxes were obviously different, which also contributed to the better visual effect of the ACCDF method in the comparison of the processed image background details shown in Figure 3.

3.2 Correcting channel differential for correction of the color cast

In underwater images, the pixel values of the R channel are mostly distributed in a relatively small range. As mentioned in *Section 3.1*, underwater images will show the dominant blue or green color, and the pixel values of the dominant color are usually distributed in the range of large pixel values. For the remaining

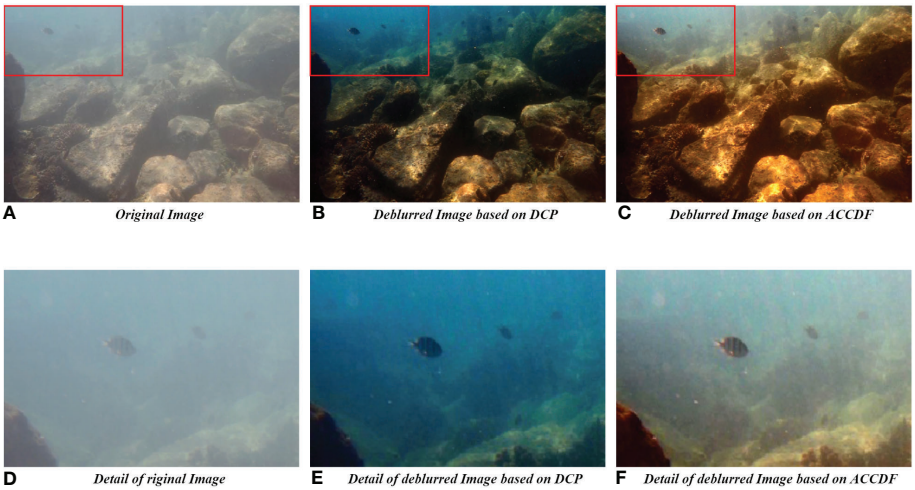


FIGURE 3
(A–C) Images are comparison of the deblurred image based on dark channel prior (DCP) and adaptive correction of channel differential and fusion (ACCDF), respectively. **(D–F)** are the details of **(A–C)**.

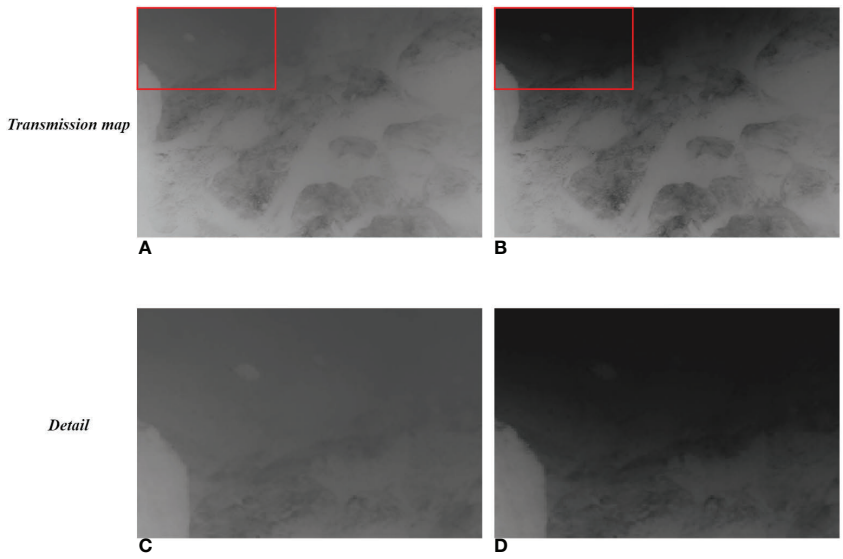


FIGURE 4
Comparison of the transmission maps. **(A, B)** images are the transmission map estimated using the dark channel prior (DCP) and the adaptive correction of channel differential and fusion (ACCDF) method, respectively. **(C, D)** images are the details of **(A, B)**.

channels corresponding to the non-dominant colors, due to the attenuation of the blue and green lights in the underwater environment being smaller than that of red light, the pixel values of these channels corresponding to the non-dominant colors are mostly distributed in the middle (Figure 5A).

We take Figure 5 as an example to show the stretching process of our channels, with the stretching formula as follows:

$$p_{\text{out}} = (p_{\text{in}} - O_{\min})\left(\frac{D_{\max} - D_{\min}}{O_{\max} - O_{\min}}\right) + D_{\min} \quad \text{yellow} \tag{9}$$

where p_{in} and p_{out} are the input and output pixels, respectively; O denotes the intensity of the original image, and D is the intensity limited by our method.

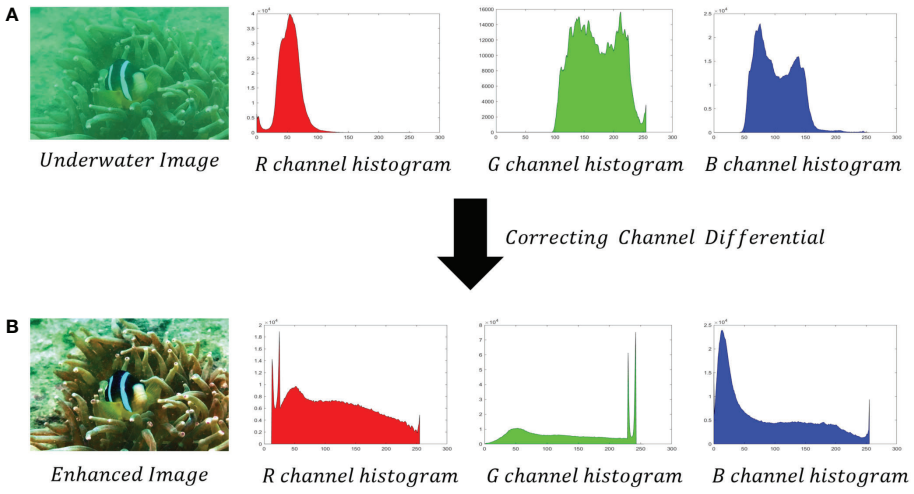


FIGURE 5
(A) image is the original underwater image and it's histogram of the RGB (red, green, blue) channels. **(B)** image is the enhanced image and it's histogram of the RGB (red, green, blue) channels.

The dominant color of the image (Figure 5A) is green, the average value of channel G is G_{mean}=171.19 , the average value of channel B is B_{mean}=103.67 , and G_{mean}>B_{mean} . Therefore, we limited the maximum value of the G channel to 255*0.95=242.5≈242 and used the CLAHE method to stretch the image within [0,242]. The minimum value of the R channel was also limited to 255*0.05=12.75≈13 and CLAHE method used to stretch the image within [13,255]. For channel B, the CLAHE method was used to stretch the image within the range of [0,255]. The stretched image is recorded as $I_S(x,y)$.

$$I_S(x,y) = \text{Stre}\left(J_{deblur}(x,y)\right) \tag{10}$$

where Stre() denotes stretching the limited range of the RGB channels of the image using the CLAHE stretching method. The stretched image is defined as $I(x,y)$.

3.3 Image fusion based on the image pyramid

For underwater images, we have implemented deblurring and color correction based on the imaging model and pixel processing, respectively. Unfortunately, the color-corrected image will again lose a part of the textural detail, even if it has been previously deblurred. Considering that there are a lot of details in the deblurred image, we used the image pyramid fusion method to recover the details from the deblurred image after color correction.

$$\text{OutputImage}(x,y) = \text{Fusion}\left(J^C_{deblur}(x,y), I(x,y)^C_S\right) \tag{11}$$

where Fusion denotes the image pyramid fusion method.

According to the multi-scale Laplacian pyramid decomposition of the input picture, the Laplacian contrast weights, local weight contrast weights, brightness weights, and exposure weights are obtained. These weights are normalized to get $\bar{W}$, and the output image is obtained by fusing the inputs with $\bar{W}$ at every pixel.

$$R(x,y) = \sum_l^L \bar{W}^l(x,y) I^l(x,y)$$

In this paper, although there are a variety of fusion methods, we chose the original fusion method (Ancuti et al., 2012) to verify our adaptive correction method and validate the performance of our method. The flowchart of the fusion framework is shown in Figure 6.

4 Experiment and analysis

In order to ensure effectiveness, we used the Raws dataset from the UIEB (Li et al., 2019) and the EUVP dataset (Islam et al., 2020) for the experiment. The experimental results were compared with the original images, ground truth images, and the images processed by the DCP, CLAHE, RGHS, underwater light attenuation prior (ULAP), FUSION, and deep learning image formation model (DL-IFM) methods (Chen et al., 2021). The comparison comprised two parts: direct visual effect comparison and image quality evaluation index comparison, with the evaluation indices including MSE, SSIM, and underwater color image quality evaluation (UCIQE).

4.1 Datasets and evaluation index

The UIEB dataset included two sub-datasets: the Raws dataset and the Challenge dataset. The Raw dataset included 890 underwater images, each of which has a corresponding high-quality image as a reference, while the Challenge dataset contained 60 challenging underwater images, but without reference standards. Therefore, in this paper, we used the Raws dataset for the experiment. The EUVP dataset has three sub-datasets: Underwater Dark, Underwater ImageNet, and Underwater Scenes. We used the Underwater Dark dataset, which included 5,550 training images and 570 verification images. Each image had a corresponding ground truth as a reference.

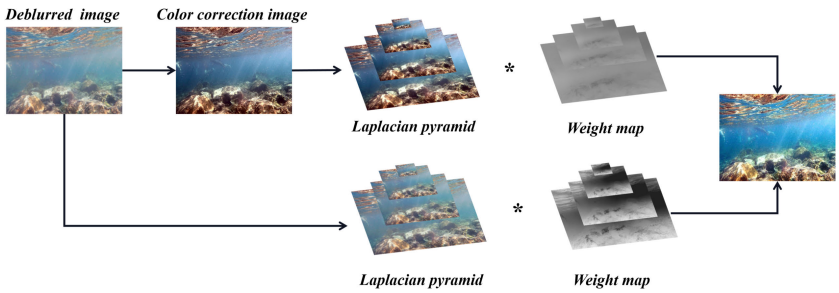


FIGURE 6
Flowchart of fusion processing. The output image is obtained by fusing the inputs with weight map at every pixel, where weight map is obtained by decomposing the input image into multi-scale of the Laplacian pyramid.

In the comparison of the image quality evaluation index, we used MSE, SSIM, and UCIQE. MSE compared the errors between the processed image and the ground truth. The smaller the error, the closer the generated image is to the ground truth. SSIM (Wang et al., 2004) evaluated the similarity of two images by comparing their luminance, contrast, and structure. The SSIM value range was [0,1]. The larger the value, the smaller the image distortion. UCIQE (Yang and Sowmya, 2015) is a combination of the color concentration, saturation, and contrast. The higher the UCIQE value, the better the visual quality of the image.

4.2 Comparison and analysis

In order to observe the differences of the various methods more intuitively, we divided the experimental results into several parts for elaboration. The experimental results of the Raws dataset are shown in Figure 7. It can be clearly seen that the image quality with the use of the ACCDF method was significantly improved compared with that of the original image. The CLAHE method also performed very well in terms of color correction, but lacked deblurring ability. The deblurring ability of the DCP method was good, but its color correction was low. Some images processed using the DCP method appeared red or even black, resulting in the loss of detail of the color distortion parts. Similarly, FUSION performed well, except for the abnormal color in the first image. The DL-IFM method showed mediocre performance as a deep learning method, even weak at deblurring. The overall performance of RGHS was good. Although its color correction ability was stronger than that of CLAHE, its deblurring ability was slightly worse than that of ACCDF. There will be obvious "mosaics" when the original image is fuzzy. Except for its better performance shown in Figures 7A, H, the ULAP method showed abnormal red light in other figures, which may have been caused by incorrect red channel processing.

In Figure 7A, the image quality of the ACCDF, DCP, and RGHS methods had been significantly improved. The images processed using these methods showed an obvious deblurring effect, and the texture of the underwater stones, gravel, and wood had been significantly enhanced. In the upper left corner of the original image, i.e., Figure 8, it can be seen that the stone texture of the DCP image is lost, the images by FUSION, DL-IFM, and ULAP are severely distorted, and the deblurring effect of the RGHS is not obvious. Compared with that of CLAHE, the image processed using ACCDF has a clearer texture.

Figure 7B shows that the deblurring effect of ACCDF was obviously better than that of the other methods. On the premise of a similar deblurring effect, the color correction ability of ACCDF was better than that of CLAHE, RGHS, FUSION and DL-IFM. Although we sacrificed the brightness of the shadow by limiting the stretch range of the red channel, the overall color of the image, especially the color saturation of coral, became better and more visually comfortable.

In Figure 7C, the original image tone is gray and the image has a gray fog blur. However, the ACCDF, CLAHE, RGHS, and FUSION methods achieved good deblurring effects. In the upper



FIGURE 7
Experimental results of the comparison of the Raws dataset. Image **(A–G)** are eight images selected from the Raws dataset. From left to right are the original image, the ground truth image and the image processed by ACCDF, CLAHE, DCP, RGHS, FUSION, DL-IFM and ULAP methods. And the red box indicates the advantages of ACCDF over other methods.

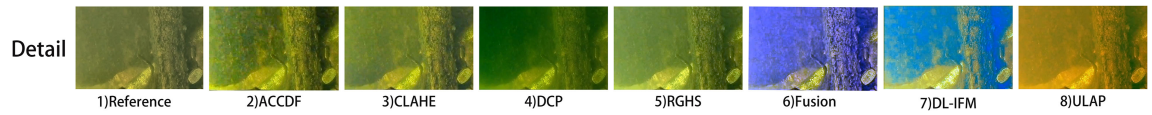


FIGURE 8
Detail comparison of Figure 7A.

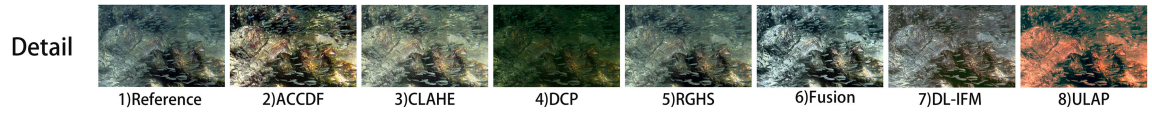


FIGURE 9
Detail comparison of Figure 7C.

right corner of Figure 7C, i.e., Figure 9, it can be seen that the color correction ability of ACCDF was better than that of CLAHE and RGHS, even if the original image is fuzzy. Furthermore, the outline of the small fish in the ACCDF image was clearer than in others.

In Figures 7D–F, it can be seen that the blue background of the three pictures is quite large and that the deblurring effect of the other methods is not very good, especially in the blue background. Due to the adaptive processing of different channels, the ACCDF method was much better at deblurring compared to the other methods, and the textural detail was more obvious. Compared with the reef in the lower part of Figure 7D, the ACCDF method was more effective than the other methods at deblurring, and the features of the texture of the reef were also strengthened. The top left of the picture, i.e., the detailed Figure 10, shows that the deblurring effect of the ACCDF method was better than that of others; furthermore, because the color correction was more reasonable, this made the swimmers in the red box clearer, and it can even be seen that they are wearing red life jackets.

Looking at the lower part of Figure 7E, it can be seen that, due to adaptive color correction, the ACCDF method makes the colors of the coral and the reef closer to reality. On the top right of the picture, i.e., the detailed Figure 11, it can also be seen that ACCDF has a better deblurring effect. In terms of color correction, because we have limited the stretching range of the different channels, the coral processed by the ACCDF method appeared more colorful. The background of ACCDF was brighter than that of DCP and more saturated than that of ALCHE.

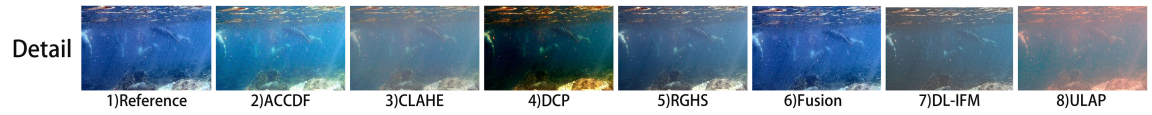


FIGURE 10
Detail comparison of Figure 7D.

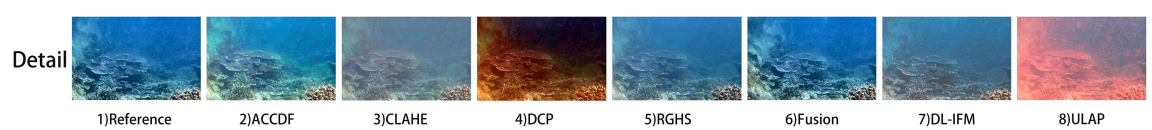


FIGURE 11
Detail comparison of Figure 7E.

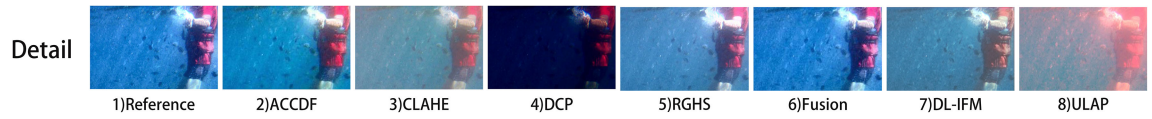


FIGURE 12
Detail comparison of Figure 7F.

Figure 7F displays the excellent deblurring effect of the ACCDF method. The reef at the bottom left of the picture does not appear overexposed, unlike that of RGHS and DL-IFM, and the blue ocean background is clearer than that of the other methods. At the top right of the image, i.e., the detailed Figure 12, due to the stretching range of the red channel being limited, the color of the life jacket in the ACCDF image is perfectly restored and brighter than that of the other methods. It can be said that the image processed using the ACCDF method is almost the same as the ground truth.

In order to further examine the effect of the ACCDF method, we compared the histogram of the original image with that of the processed image (Figure 13). The RGB histogram of the original image was concentrated in the range [100,160], with the blue and green areas overlapping seriously, which made the image gray and the color distinction not obvious, what we call a "fog." After stretching, the "fog" of the image was obviously eliminated, and the blue and green became obvious, with the red also strengthened.

In the original image of Figure 7G, not only was "fog" present but also the scattering of sunlight penetrating the water surface and suspended particles. The ACCDF and DCP methods had obvious deblurring effects. However, the image color of the DCP method was red, and the upper right corner had missing information. The ACCDF, ALCHE, RGHS, and DL-IFM methods showed good color correction effects, but the ALCHE, RGHS, and DL-IFM methods retained a layer of "mist" in the image.

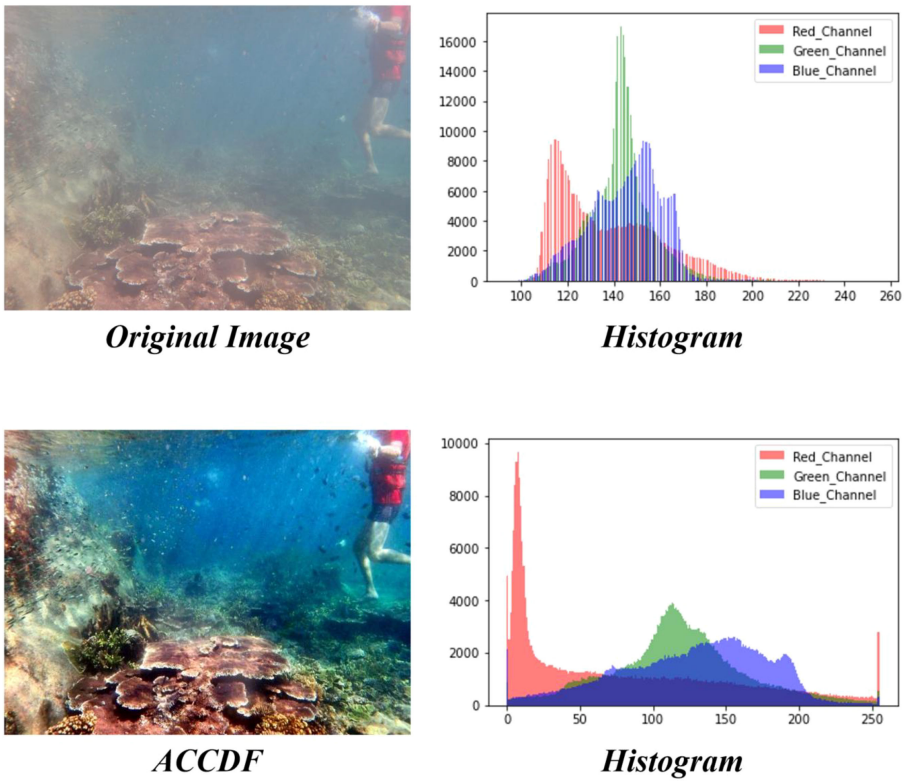


FIGURE 13
Histogram comparison of Figure 7F.

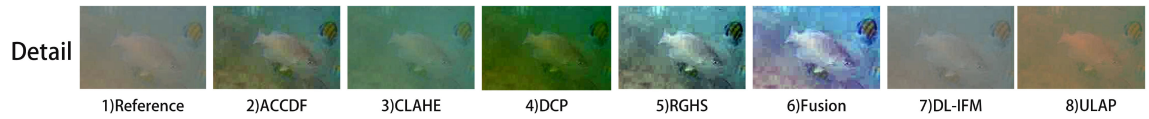


FIGURE 14
Detail comparison of Figure 7H.

The original image in Figure 7H is an extremely challenging one. The image showed extremely heavy "fogging." Except for the fish close to the lens, the overall image hardly displayed other useful information. Surprisingly, all of the methods extracted the ground and fish information in the background from the poor original image. However, the ALCHE method did not remove the blur very well, while the DCP and ULAP methods showed color distortions of green and red. By comparing the details in Figure 14, we found that the RGHS and FUSION methods showed very obvious feature enhancement effects. The feature of the fish and reef in the background were very noticeable, but the image appeared as a mosaic, which means that the deblurring ability was slightly weaker. In addition to removing the fog from the original image, the ACCDF image made the fish and the rocks in the upper right corner of the image appear sharper. Furthermore, adaptive color channel processing of the ACCDF method made the two fish colors closer to the ground truth.

In order to further examine the effect of the ACCDF method, we compared the histogram of the original image with that of the processed image (Figure 15). The histogram of the original image was sparsely distributed, and the red and blue channel pixels were within the range [60,100], making the image dark as a whole. After stretching using the ACCDF method, the pixel distribution of the RGB channels turned uniform, which allowed the information of the image to be reproduced.

The experimental results of the EUVP dataset are shown in Figure 16. Since the images in this dataset have been seriously attenuated, the original color of the image cannot be restored by non-deep learning methods; therefore, in this part, we mainly compared the deblurring ability of each method.

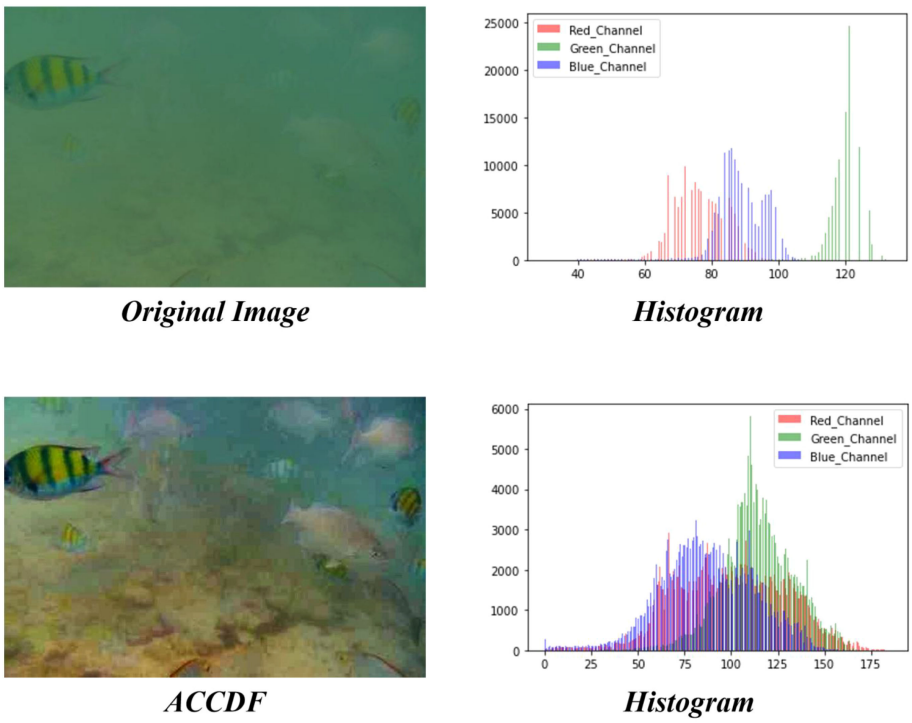


FIGURE 15
Histogram comparison of Figure 7H.

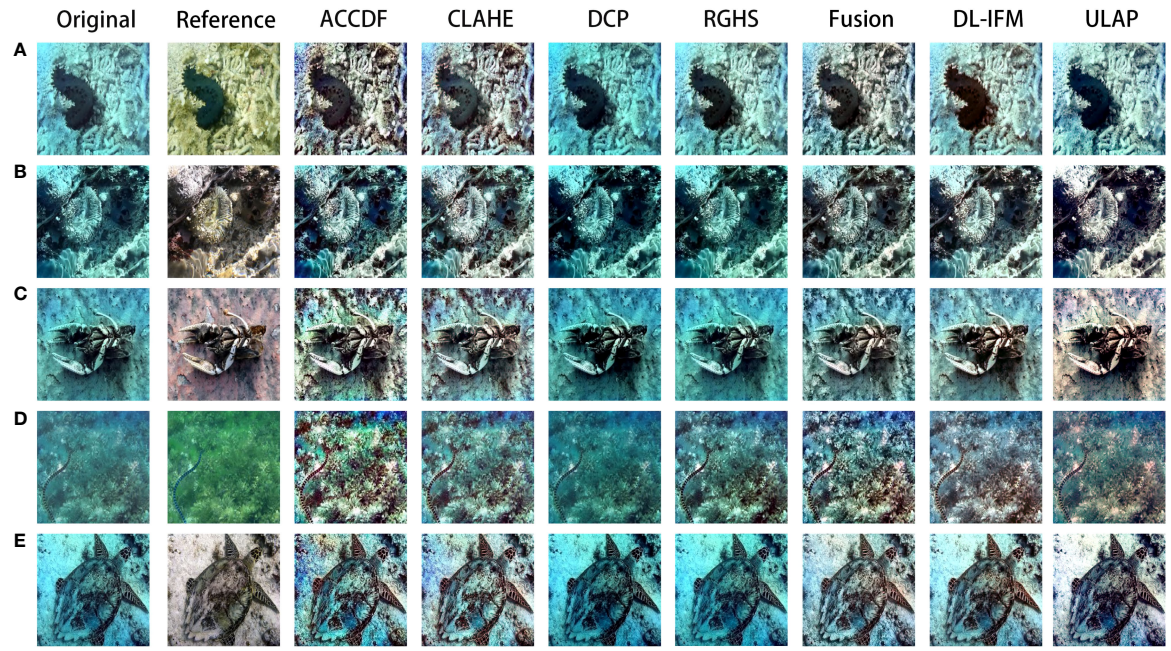


FIGURE 16
Image **(A–E)** are five images selected from the EUVP dataset. From left to right are the original image, the ground truth image and the image processed by ACCDF, CLAHE, DCP, RGHS, FUSION, DL-IFM and ULAP methods.

Figure 16A presents the better deblurring effects of the ACCDF, CLAHE, and FUSION methods compared to the other methods. The contour of the sea cucumber in the middle of the ACCDF image was clearer, while the sea cucumber in the middle of the DL-IFM image turned blurred. Externally, it can be seen that the texture of the gravel in the ACCDF image was more obvious and the outline clearer by comparing the background part of the image. The background texture of the DCP, RGHS, DL-IFM, and ULAP images was blurred, especially in the upper left corner of the image.

In Figure 16B, the images processed using DCP and RGHS are shown to be almost the same as the original pictures. Details were missing in the upper left corner of the DL-IFM and ULAP images. CLAHE had a good deblurring effect, but compared with the ACCDF method, especially in the texture comparison between the sea anemone in the middle of the image and the coral in the upper left corner, the enhancement ability for the textural features was slightly weaker.

In Figure 16C shows that the ACCDF, CLAHE, and FUSION methods removed most of the blue cast in the original image, resulting in obvious image details. The ULAP method removed most of the blue cast and had good visual effect, but lost some textural details on the lobster head in the middle of the image. The ACCDF image showed not only enhanced details of the lobster head but also enhanced ground texture of the image background.

In Figures 16D, E, although the ACCDF and FUSION images had abnormal colors compared with the ground truth images, the details of the image were enhanced. For example, in Figure 16D, the seaweed contour of the ACCDF method was clearer than that of the other methods, while the FUSION image showed a mosaic-like blur despite its high contrast. The pattern of the turtle shell in the ACCDF image in Figure 16E had higher contrast and is clearer than that of the other methods. The sand and stone on the bottom of the DL-IFM and ULAP images were partially exposed and became blurred, while the sand and stone texture in the ACCDF image was clearer.

We combined the image quality evaluation indices for analysis. The quality evaluation of Figure 7D is shown in Table 1. CLAHE and DL-IFM performed slightly poorly in the quality evaluation because of their slightly weaker deblurring ability. DCP and ULAP has unusually large MSE values due to severe color distortion and loss of detail in the processed pictures. The overall performance of ACCDF and FUSION was good. The MSE values of ACCDF were the smallest, while those of SSIM and UCIQE were relatively large. As displayed in Figure 7D, ACCDF and FUSION provided the best deblurred and color-corrected images. However, compared with FUSION, ACCDF had a warmer color orientation; hence, the image details were more easily observed. This has no relevance in the image quality metrics evaluation, but it has resulted in a better visual experience.

The quality evaluation of Figure 7F is shown in Table 2. Based on its evaluation index, the FUSION method performed very well. It can also be seen in the picture that the FUSION method performed well in deblurring and color correction. The CLAHE, RGHS, and DL-IFM methods performed well in the evaluation criteria, but a layer of "mist" can still be seen in the image, which results in slightly poor visual performance. The DCP method performed poorly in the evaluation criteria because the images lost a lot of information. Surprisingly, the ACCDF method had the largest MSE value, which did not correspond to the visual perception of the image, probably because the method removed bubbles from the image during the deblurring process. Comparison of the SSM and UCIQE values indicated that the ACCDF method performed well. It can be seen that the image color saturation of ACCDF in Figures 7F, 12 was higher and the textural detail clearer compared to the other images.

Comparison of the quality evaluation indicators in Figure 16C is shown in Table 3. The MSE value of ULAP was the lowest, and the image processing results also showed that the image color using this method was the closest to the ground truth. The SSIM and the UCIQE value of the RGHS and the DL-IFM method, respectively, was the largest. The evaluation index of the ACCDF method was average due to the original image having been seriously attenuated, and the non-machine learning method cannot accomplish restoration. The ACCDF method focused on the blue dominant tone of the image to stretch the histogram, which resulted in color distortion. However, when the textural features were examined closely, it was found that, in the ACCDF image, the details of the lobster head were obviously enhanced, and the textural contrast of the ground sand and stone at the edge of the image was higher and the outline clearer.

5 Conclusion

In this paper, we proposed an adaptive color correction and deblurring method for underwater images with different color casts. This method adaptively deblurs the dark channel and stretches the histogram by perceiving the main hue of the image. It then fuses the images processed by the two techniques into a better image. The results of the experiment showed that this method can handle underwater images in different green and blue distortions and has better deblurring and color correction capabilities compared to other classical algorithms. We also

TABLE 1 Quality evaluation of Figure 7D.

	ACCDF	CLAHE	DCP	DL-IFM	FUSION	RGHS	ULAP
MSE	**15.1173**	62.8496	185.4159	84.3816	**23.7198**	100.2109	60.3795
SSIM	**0.8664**	0.7766	0.8065	**0.8652**	**0.9366**	**0.8996**	0.6302
UCIQE	**0.6376**	0.5102	**0.7238**	0.5602	**0.6515**	**0.6545**	0.473

MSE, mean square error; SSIM, structural similarity; UCIQE, underwater color image quality evaluation; ACCDF, adaptive correction of channel differential and fusion; CLAHE, contrast limited adaptive histogram equalization; DCP, dark channel prior; DL-IFM, deep learning image formation model; RGHS, relative global histogram stretching; ULAP, underwater light attenuation prior. Bold indicates good performance.

TABLE 2 Quality evaluation of Figure 7F.

	ACCDF	CLAHE	DCP	DL-IFM	FUSION	RGHS	ULAP
MSE	90.2184	46.3438	243.7132	75.5761	**23.4494**	67.1776	46.7848
SSIM	**0.8784**	0.7586	0.2474	**0.8495**	**0.9386**	**0.8888**	0.4997
UCIQE	**0.6771**	0.5047	0.5482	0.5731	**0.6587**	**0.6677**	0.4977

MSE, mean square error; SSIM, structural similarity; UCIQE, underwater color image quality evaluation; ACCDF, adaptive correction of channel differential and fusion; CLAHE, contrast limited adaptive histogram equalization; DCP, dark channel prior; DL-IFM, deep learning image formation model; RGHS, relative global histogram stretching; ULAP, underwater light attenuation prior. Bold indicates good performance.

TABLE 3 Quality evaluation of Figure 14C.

	ACCDF	CLAHE	DCP	DL-IFM	FUSION	RGHS	ULAP
MSE	95.25	82.054	127.7532	76.2379	92.6004	82.8455	**64.9118**
SSIM	0.745	0.8411	0.8891	0.9525	0.8487	**0.9732**	0.8335
UCIQE	**0.5922**	0.5916	0.6047	**0.6115**	0.5998	0.5971	0.6043

MSE, mean square error; SSIM, structural similarity; UCIQE, underwater color image quality evaluation; ACCDF, adaptive correction of channel differential and fusion; CLAHE, contrast limited adaptive histogram equalization; DCP, dark channel prior; DL-IFM, deep learning image formation model; RGHS, relative global histogram stretching; ULAP, underwater light attenuation prior. Bold indicates good performance.

found that the images processed using this method had warmth, which made the visual experience better, especially with a severely attenuated red. In addition, this method enhanced the textural detail extremely well, finding more hidden details and effectively improving the quality of the blurred underwater image.

On the other hand, when the color of the image has been seriously attenuated, the ACCDF method performed poorly in terms of color correction. This scenario is more suitable for color filling using deep learning methods. In the future, we will consider using a deep learning algorithm to supplement and correct the color of images, adding it to the fusion framework to make it stronger.

Data availability statement

Publicly available datasets were analyzed in this study. The data can be found here: EUVP: https://irvlab.cs.umn.edu/resources/euvp-dataset; UIEB: https://li-chongyi.github.io/proj_benchmark.html.

Author contributions

ZeZ completed the main work of this paper. TW provided guidance during the research. ZeZ, ZhZ, YL, TW, SZ, and HC were involved in proofreading and manuscript revision. HX help complete the part of experiment. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by: National Natural Science Foundation of China and Macao Science and Technology Development Joint Fund (0066/2019/AFJ): Research on knowledge-oriented probabilistic graphical model theory based on multi-source data; the MOST-FDCT Projects (0058/2019/AMJ,2019YFE0110300; Research and Application of Cooperative Multi-Agent Platform for Zhuhai-Macao Manufacturing Service); the Key Platform and Scientific Research Project for General Universities, Guangdong, 2022 (grant no. 2022ZDZX4061); Science and Technology Program of Social Development, Zhuhai, 2022 (grant no. 2220004000195); Key Platform and Scientific Research Project for General Universities, Guangdong, 2022, (grant no.2022ZDZX4061; Science and Technology Program of Social Development, Zhuhai, 2022, (grant no.2220004000195); Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, BNU-HKBU United International Collegeproject code 2022B1212010006; Guangdong Higher Education Upgrading Plan (2021-2025) with UIC research grant R0400001-22 and R201902; R&D plan project in key fields of Guangdong Province (No. 2021B0707010001); R&D plan project in key fields of Guangdong Province (No. 2018B010109001).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Adolfson, J., and Berghage, T. (1974). *Perception and performance under water* (John Wiley & Sons).

Ancuti, C., Ancuti, C. O., Haber, T., and Bekaert, P. (2012). "Enhancing underwater images and videos by fusion," in *2012 IEEE conference on computer vision and pattern recognition* (IEEE), 81–88. doi: 0.1109/CVPR.2012.6247661

Anwar, S., Li, C., and Porikli, F. (2018). Deep underwater image enhancement. *arXiv preprint arXiv*, 1807.03528. doi: 10.48550/ARXIV.1807.03528

Bai, L., Zhang, W., Pan, X., and Zhao, C. (2020) *Underwater image enhancement based on global and local equalization of histogram and dual-image multi-scale fusion* (IEEE) (Accessed Access 8). doi: 10.1109/ACCESS.2020.3009161

Burt, P. J., and Adelson, E. H. (1987). "The laplacian pyramid as a compact image code," in *Readings in computer vision* (Elsevier), 671–679. doi: 10.1109/TCOM.1983.1095851

Chao, L., and Wang, M. (2010). "Removal of water scattering," in *2010 2nd international conference on computer engineering and technology*, vol. 2. (IEEE), V2–35. doi: 10.1109/ICCET.2010.5485339

Chen, X., Zhang, P., Quan, L., Yi, C., and Lu, C. (2021). Underwater image enhancement based on deep learning and image formation model. *arXiv preprint arXiv*, 2101.00991. doi: 10.48550/ARXIV.2101.00991

Chiang, J. Y., and Chen, Y.-C. (2011). Underwater image enhancement by wavelength compensation and dehazing. *IEEE Trans. image Process.* 21, 1756–1769. doi: 10.1109/TIP.2011.2179666

Drews, P., Nascimento, E., Moraes, F., Botelho, S., and Campos, M. (2013). "Transmission estimation in underwater single images," in *Proceedings of the IEEE international conference on computer vision workshops*, 825–830. doi: 10.1109/ICCVW.2013.113

Fabbri, C., Islam, M. J., and Sattar, J. (2018). "Enhancing underwater imagery using generative adversarial networks," in *2018 IEEE international conference on robotics and automation (ICRA)* (IEEE), 7159–7165. doi: 10.1109/ICRA.2018.8460552

Galdran, A., Pardo, D., Picón, A., and Alvarez-Gila, A. (2015). Automatic red-channel underwater image restoration. *J. Visual Communication Image Representation* 26, 132–145. doi: 10.1016/j.jvcir.2014.11.006

Guo, Y., Li, H., and Zhuang, P. (2019). Underwater image enhancement using a multiscale dense generative adversarial network. *IEEE J. Oceanic Eng.* 45, 862–870. doi: 10.1109/JOE.2019.2911447

He, K., Sun, J., and Tang, X. (2010). Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 2341–2353. doi: 10.1109/TPAMI.2010.168

Hitam, M. S., Awalludin, E. A., Yussof, W. N. J. H. W., and Bachok, Z. (2013). "Mixture contrast limited adaptive histogram equalization for underwater image enhancement," in *2013 international conference on computer applications technology (ICCAT)* (IEEE), 1–5. doi: 10.1109/ICCAT.2013.652201

Huang, J.-B., Singh, A., and Ahuja, N. (2015). "Single image super-resolution from transformed self-exemplars," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5197–5206. doi: 10.1109/CVPR.2015.7299156

Huang, D., Wang, Y., Song, W., Sequeira, J., and Mavromatis, S. (2018). "Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition," in *International conference on multimedia modeling* (Springer), 453–465. doi: 10.1007/978-3-319-73603-7_37

Islam, M. J., Xia, Y., and Sattar, J. (2020). "Fast underwater image enhancement for improved visual perception," in *IEEE Robotics and automation letters*, vol. 5. , 3227–3234. doi: 10.1109/LRA.2020.2974710

Ketcham, D. J., Lowe, R. W., and Weber, J. W. (1974). "Image enhancement techniques for cockpit displays," in *Tech. rep* (Hughes Aircraft Co Culver City Ca Display Systems Lab).

Lai, Y., Zhou, Z., Su, B., and Xuanyuan, Z. (2022). Single underwater image enhancement based on differential attenuation compensation. *Front. Mar. Sci.*, 2216. doi: 10.3389/fmars.2022.1047053

Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., et al. (2019). An underwater image enhancement benchmark dataset and beyond. *IEEE Trans. Image Process.* 29, 4376–4389. doi: 10.1109/TIP.2019.2955241

Lu, H., Li, Y., Nakashima, S., Kim, H., and Serikawa, S. (2017). Underwater image super-resolution by descattering and fusion. *IEEE Access* 5, 670–679. doi: 10.1109/ACCESS.2017.2648845

Lu, H., Li, Y., Zhang, L., and Serikawa, S. (2015). Contrast enhancement for images in turbid water. *JOSA A* 32, 886–893. doi: 10.1364/JOSAA.32.000886

Peng, Y.-T., Cao, K., and Cosman, P. C. (2018). Generalization of the dark channel prior for single image restoration. *IEEE Trans. Image Process.* 27, 2856–2868. doi: 10.1109/TIP.2018.2813092

Perez, J., Attanasio, A. C., Nechyporenko, N., and Sanz, P. J. (2017). "A deep learning approach for underwater image enhancement," in *International work-conference on the interplay between natural and artificial computation* (Springer), 183–192. doi: 10.1007/978-3-319-59773-7_19

Pizer, S. M., Amburn, E. P., Austin, J. D., Cromartie, R., Geselowitz, A., Greer, T., et al. (1987). Adaptive histogram equalization and its variations. *Comput. vision graphics image Process.* 39, 355–368. doi: 10.1016/S0734-189X(87)80186-X

Song, W., Wang, Y., Huang, D., and Tjondronegoro, D. (2018). "A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration," in *Pacific rim conference on multimedia* (Springer), 678–688. doi: 10.1007/978-3-030-00776-8_62

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. image Process.* 13, 600–612. doi: 10.1109/TIP.2003.819861

Yang, M., and Sowmya, A. (2015). An underwater color image quality evaluation metric. *IEEE Trans. Image Process.* 24, 6062–6071. doi: 10.1109/TIP.2015.2491020

frontiers | Frontiers in Marine Science

TYPE Methods
PUBLISHED 30 January 2023
DOI 10.3389/fmars.2023.1074428

Check for updates

OPEN ACCESS

EDITED BY
Rizwan Ali Naqvi,
Sejong University, South Korea

REVIEWED BY
Jules S. Jaffe,
University of California, San Diego, United States
Peng Ren,
China University of Petroleum (East China), China
Zibang Zhang,
Jinan University, China
Fahim Mannan,
Algolux, Canada

*CORRESPONDENCE
Jianping Li
jp.li@siat.ac.cn

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 19 October 2022
ACCEPTED 06 January 2023
PUBLISHED 30 January 2023

CITATION
Chen T, Li J, Ma W, Guo G, Yang Z, Li Z and Qiao J (2023) Deep focus-extended darkfield imaging for *in situ* observation of marine plankton.
Front. Mar. Sci. 10:1074428.
doi: 10.3389/fmars.2023.1074428

COPYRIGHT
© 2023 Chen, Li, Ma, Guo, Yang, Li and Qiao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Deep focus-extended darkfield imaging for *in situ* observation of marine plankton

Tao Chen[1,2], Jianping Li[1,2*], Wenqi Ma[3], Guannan Guo[1,2], Zhenyu Yang[1,2], Zhenping Li[1,2] and Jianping Qiao[3]

[1]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, [2]University of Chinese Academy of Sciences, Beijing, China, [3]School of Physics and Electronics, Shandong Normal University, Jinan, China

Darkfield imaging can achieve *in situ* observation of marine plankton with unique advantages of high-resolution, high-contrast and colorful imaging for plankton species identification, size measurement and abundance estimation. However, existing underwater darkfield imagers have very shallow depth-of-field, leading to inefficient seawater sampling for plankton observation. We develop a data-driven method that can algorithmically refocus planktonic objects in their defocused darkfield images, equivalently achieving focus-extension for their acquisition imagers. We devise a set of dual-channel imaging apparatus to quickly capture paired images of live plankton with different defocus degrees in seawater samples, simulating the settings as in *in situ* darkfield plankton imaging. Through a series of registration and preprocessing operations on the raw image pairs, a dataset consisting of 55 000 pairs of defocused-focused plankter images have been constructed with an accurate defocus distance label for each defocused image. We use the dataset to train an end-to-end deep convolution neural network named IsPlanktonFE, and testify its focus-extension performance through extensive experiments. The experimental results show that IsPlanktonFE has extended the depth-of-field of a 0.5× darkfield imaging system to ~7 times of its original value. Moreover, the model has exhibited good content and instrument generalizability, and considerable accuracy improvement for a pre-trained ResNet-18 network to classify defocused plankton images. This focus-extension technology is expected to greatly enhance the sampling throughput and efficiency for the future *in situ* marine plankton observation systems, and promote the wide applications of darkfield plankton imaging instruments in marine ecology research and aquatic environment monitoring programs.

KEYWORDS

underwater imaging, deep learning, focus extension, ocean observation, marine plankton

1 Introduction

Marine plankton are abundant underwater drifters widely distributed in the world's oceans (Batten et al., 2019). They are mainly phytoplankton and zooplankton with weak swimming ability. They are also starters of food, material and energy cycles in the ocean, and foundations of the marine ecosystems and food webs (Steinberg and Landry, 2017; Lombard

et al., 2019; Suthers et al., 2019). Their regional proliferation in a short period of time can form blooms, which are often accompanied with negative physical and toxicological effects, threatening the safety of nearby aquaculture (Cowen and Sponaugle, 2009), coastal facilities (Zhanhui et al., 2020), and even human health (Suthers et al., 2019). Therefore, the observation of marine plankton is important for understanding the impact of anthropogenic activities and global change on marine ecosystems, and conversely the responses of marine ecosystems to global change (Alvarez-Fernandez et al., 2018). It is also an indispensable means in operational oceanography applications such as marine environment monitoring, biodiversity investigation, fishery resources assessment, and harmful organism breakout warning (Lombard et al., 2019).

As early as in the era of film photography, people tried to use underwater optical imaging for *in situ* observation of marine plankton (Ortner et al., 1979). With the maturity of solid-state lighting and digital camera technologies since the 1990s, a variety of underwater plankton imagers have been developed (Benfield et al., 1996; Cowen and Guigand, 2008; Schulz et al., 2009; Picheral et al., 2010; Bi et al., 2013; M. Rotermund and Samson, 2015; Gallager, 2019; Orenstein et al., 2020; Li et al., 2022), through which digital images of plankters are captured in natural seawaters. By further analysis of the obtained images using digital processing and machine learning algorithms, people can achieve automatic observation of plankton taxonomy and various functional traits (Orenstein et al., 2022). Compared with traditional methods, *in situ* imaging has the advantages of longer observational time and continuity, and higher spatio-temporal resolution. Its non-contact property also makes it more suitable for observing fragile gelatinous organisms. These advancements have greatly expanded our knowledge on related marine sciences (Gorsky et al., 2000; Hirche et al., 2014; Campbell et al., 2020).

However, *in situ* plankton imaging has always faced the trade-off between imaging quality and sampling efficiency. On the one hand, the complexity of seawater composition (Davies and Nepstad, 2017) and plankton attributes makes the optical properties of imaging medium and targets variable and heterogeneous, which easily leads to the deterioration of imaging quality (Lombard et al., 2019; Cheng et al., 2020). On the other hand, high magnification is necessary for sufficient resolution to identify and measure tiny plankters. This leads to a shallow depth-of-field (DOF) and a small volume of seawater sampled by a single frame, which in turn leads to low throughput and efficiency in seawater sampling (Lombard et al., 2019).

To increase the sampling throughput, existing imagers have adopted different strategies and methods. Imaging flow cytometers actively pump seawater into their instruments and implement microscopy (mainly on microphytoplankton) as the seawater is flowing through an interrogation volume to improve the sampling rate (Olson and Sosik, 2007; Grcs et al., 2018). However, this is no longer a strictly "*in situ*" strategy and is inefficient for sampling larger and scarcer mesoplankton. Many other underwater imagers use more direct strategy of imaging targets in the seawater outside their housings through transparent ports to achieve *in situ* snapshot imaging sampling. For example, silhouette imagers such as ISIIS (Cowen and Guigand, 2008) and ZooVIS (Bi et al., 2013) use parallel light beam illumination and shadowgraphic imaging to enlarge their DOF. Combined with towed deployment, they can improve the sampling throughput to 70L/s (Cowen and Guigand, 2008). However, shadowgraphy seriously compromises the resolution, texture, and color information of the acquired images, and is easy to lose target edge sharpness and signal-to-background ratio (SBR) by underwater light scattering, resulting in great plankton detection difficulty in turbid seawater (Cheng et al., 2020; Panaïotis et al., 2022). Digital holography is another typical *in situ* method that uses coherent illumination to record holograms and refocusing computation to achieve focus extended imaging. Known instruments include LISST-HOLO (Graham and Nimmo Smith, 2010), HOLOCAM (Katz and Sheng, 2010), and 4Deep (Rotermund and Samson, 2015), among others. However, this method has the shortcomings of intensive reconstruction computation, speckle noise, and loss of color etc. More seriously, in the turbid seawater environment, the coherence of illumination light is easily reduced, leading to refocusing and target detection difficult (Nayak et al., 2021).

Darkfield imaging is also a popular method for *in situ* plankton observation. Its images own excellent resolution and rich color information for better representation of the planktonic target, and are more intuitive to human vision. Additionally, it often achieves good image SBR to facilitate object detection in subsequent processing. These features are beneficial for finer plankton taxonomy and quantification, especially favorable for biodiversity study (Lombard et al., 2019). Therefore, darkfield imaging has been adopted by many underwater plankton cameras (Benfield et al., 1996; Gorsky et al., 2000; Schulz et al., 2009; Gallager, 2019; Orenstein et al., 2020; Li et al., 2022). For example, the Imaging Plankton Probe (IPP) developed by (Li et al., 2022) is a darkfield plankton imager that supports long-time near-shore buoy deployment. It features a compressed orthogonal white-light illumination to reduce stray light scattered from outside the imaging DOF, enabling high-contrast true-color *in situ* imaging of plankton and suspended particles in a wide size range of 200μm-40mm and finer plankton identification.

However, darkfield imaging also has very shallow DOF. For example, the DOF of a 0.5× lens used in IPP is only ~3mm, corresponding to a sampling volume of just ~1.5 mL by a single frame capture. Although low-magnification lenses have thicker DOF, using them as replacement of high magnification lens in any imager will sacrifice resolution. To overcome this drawback, (Wang et al., 2020) installed different magnification lenses onto a rotatable nosepiece in their darkfield imager to expand the imaging range. (Merz et al., 2021) simply put a high and a low magnification imaging optics into one darkfield imager housing for simultaneous acquisition of wider size range. But both made the instrument cumbersome, expensive, and unreliable. Exploring other possible strategies by hardware modifications, one can easily think of methods for darkfield imaging focus-extension by using liquid lenses (Cheng et al., 2021), diffractive optical elements (Xu et al., 2019), or light-field cameras (Martínez-Corral and Javidi, 2018). However, these methods are essentially at the cost of sacrificing temporal or spatial freedom, and will increase the complexity and cost of instrumentation, which is not favorable for long-term work in the harsh oceanic underwater environment.

It will become very attractive if the focus of underwater darkfield imager could be effectively extended without hardware modification. This directs us to image restoration algorithm as an alternative for this

goal, which can be roughly classified into two categories: physical modeling-based method (Krishnan et al., 2011; Nishiyama et al., 2011; Karaali and Jung, 2018) and data-driven method (Abuolaim and Brown, 2020; Lee et al., 2021; Luo et al., 2021). In the scenario of underwater imaging, due to the extremely complex properties of seawater and targets, and inconsistent imaging characteristics of different instruments, the accurate prior for physical modeling is difficult to estimate or measure. Therefore, relevant works were mostly carried out in laboratories (Fan et al., 2010; Makarkin and Bratashov, 2021). The data-driven method trains deep convolution neural network (DNN) models through a large-scale dataset, and can achieve end-to-end focus restoration for images with a certain degree of defocus. However, the large-scale high-quality real-world data needed to train such deep-learning models is difficult to obtain. In few successful DNN-training reports, (Luo et al., 2021) and (Rai Dastidar and Ethirajan, 2020) used microscopic image stack datasets collected by expensive precision automatic microscopes and (Abuolaim and Brown, 2020) and (Lee et al., 2021) relied on a special dual-pixel camera chip to collect natural scene image datasets with defocus distance information embedded. Obviously, these datasets construction patterns have high time and money costs and are difficult to duplicate for building up live marine plankton image datasets. Moreover, in order to train DNNs to generalize well in actual ocean observation, the dataset is necessary to have considerable example quantity and diversity. This undoubtedly makes the modeling of learning-based algorithm for underwater plankton imaging focus-extension face great challenge in data availability.

To this end, we used the off-the-shelf optical components to customize a set of dual-channel darkfield imaging apparatus, which efficiently facilitated us to construct a dataset consisting of 55 000 pairs of defocused-focused marine plankter images. The acquisition of this dataset mimicked the settings as in real oceanic *in situ* imaging, and all defocused images are provided with accurate defocus distance labels. Using this dataset, we trained a self-guided focus-extension DNN named IsPlanktonFE, which, to the best of our knowledge, achieved end-to-end defocus restoration of real *in situ* plankton darkfield images for the first time. The idea of using IsPlanktonFE for darkfield plankton imager focus-extension is shown in Figure 1. We used a standard SiO_2 bead and real plankton as targets to test and calibrate its performance. The results show that it can significantly improve the accuracy of plankton images classification and bead size measurement. IsPlanktonFE has extended the DOF of a 0.5× darkfield imaging system to ±10mm range, which is ~7 times of its original value of ~3mm. Its performance was further verified on a lot of *in situ* images collected by different IPPs from the actual sea sites. The results show that the network has good generalizability, and plays a significant role in improving the efficiency of marine plankton observation in practice.

The contributions of this article are emphasized as follows.

1. A data-driven method that can achieve focus-extension for single-image acquired by underwater darkfield imaging systems is proposed, which can greatly improve their efficiency for *in situ* observing marine plankton and suspended particles.
2. A simple dual-channel imaging apparatus is devised and a complete protocol for using the apparatus to conveniently and efficiently building up large-scale defocused-focused image pair datasets of live plankton is introduced. The accurate defocus distance labels contained in the datasets can provide quantitative reference to the training of imaging focus-extension models.
3. A self-guided end-to-end convolution neural network that can effectively extend the DOF of underwater darkfield imaging with proven instrument and content generalizability is designed and trained.

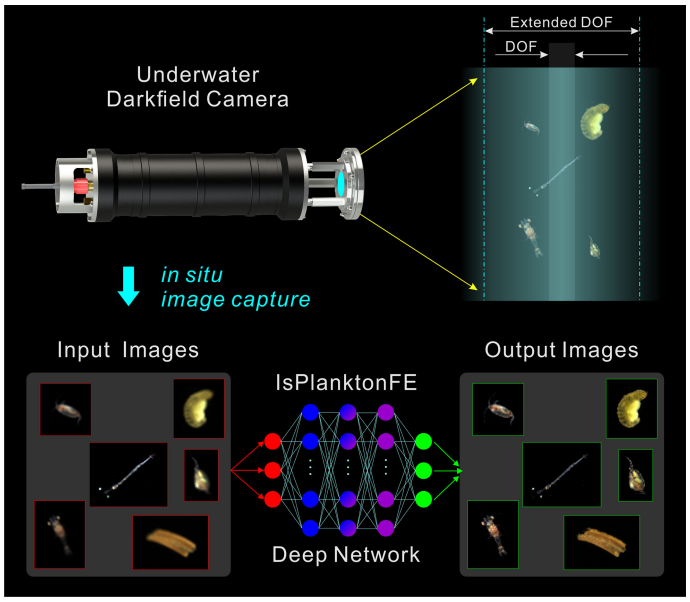


FIGURE 1
Concept of a deep learning-based focus-extension for *in situ* darkfield imaging of marine plankton.

2 Materials and methods

2.1 Sample preparation

10 liters of coastal seawater sample containing live plankton was collected from Dapeng Ao Cove (22°34'4" N, 114°31'53" E), Shenzhen of China by light trapping on December 9th, 2021 and July 28th, 2022, respectively. The seawater samples were kept at ambient temperature in a bucket when they were returned to laboratory in one hour time after collection. Every time we used a pipet to select some active plankters from the samples and added them into the container of the dual-channel darkfield imaging apparatus for dataset construction. According to plankton expert's identification from the recorded images, the taxa in the samples were mainly *Arthropoda* and *Annelida*.

2.2 Dataset construction

In order to acquire defocused-focused image pairs of live plankton in their natural state, we built a set of dual-channel darkfield imaging apparatus in laboratory to simulate the real *in situ* imaging settings in the ocean. Its composition and principle are shown in Figure 2. When the plankters freely swim through the illuminated space in the transparent container, they are illuminated by a white-light ring LED illuminator. Part of the scattered or refracted light by the plankton is split into two pathways by a cube beam-splitter and enters two imaging sub-systems for simultaneous imaging. Each sub-system consists of a telecentric lens (0.5×, DOF~3mm) and a CMOS camera (4096×3000 pixels, FLIR BFS-PGE-122S6C-C) mounted on a high-precision translation stage. During image acquisition, the focal plane of one channel was fixed to a certain position behind the container's window, and the focal plane of the other channel was adjusted to different axial positions with a spacing Δz relative to the previous focal plane. At each Δz, the cameras of the two channels were triggered to synchronously capture paired images with very short exposure time of 400μs to avoid motion blur, and their framerate was both set to 4fps. After enough images were acquired at one Δz position, one of the focal planes was translated stepwise (1mm) to next axial position to continue capturing a new set of image pairs. By repeating such process while Δz was varied in a range of 0mm-10mm, we finally acquired enough raw darkfield image pairs of live plankton with discrete defocus spacing. Note that multiple clear and blurry planktonic targets may coexist in all the raw images at this stage.

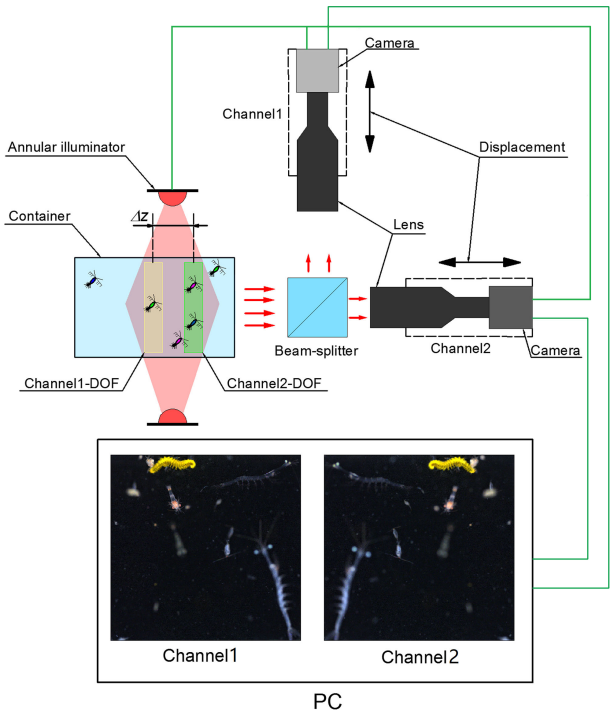


FIGURE 2
Setup of the dual-channel darkfield imaging apparatus for defocused-focused live plankton darkfield image pair acquisition.

Next, we applied a set of image registration and preprocessing operations to process all the raw images. Figure 3 illustrates the steps of this process to generate multiple defocused-focused plankton region-of-interest (ROI) pairs from one pair of raw images *I*-*I'* with focal spacing Δz between them. The steps include performing (1) affine transformation of image *I* relative to the reference image *I'* to obtain a registered image I_R (refer to Supplementary Materials for image registration details), (2) background subtraction and white balancing of the registered image pair I_R-*I'* (Li et al., 2022), (3) ROI extraction by thresholding from the processed image pair I_P-I'_P (Li et al., 2022), (4) focus evaluation of all ROIs extracted from I_P to select coordinates of all the in-focus ROIs (Yang et al., 2021), (5) defocused-focused ROI pairs cropping from the raw image pair I_P-I'_P (note all the cropped ROIs from I_P are in-focus as evaluated by the algorithm, and their corresponding ROIs cropped from I'_P are defocused with a defocus distance Δz), (6) repeating steps (4) and (5) to process ROIs extracted from I'_P in step (3) to obtain defocused-focused ROI pairs (note the cropped ROIs from I'_P are in-focus, and their corresponding ROIs cropped from I_P are defocused with a defocus distance of -Δz). Applying the above processing, we obtained the raw data of defocused-focused ROI pairs with different defocus distances.

Finally, the raw ROI pair data was further cleaned by removing some unwanted ROIs selected based on human visual determination. We firstly selected the ROIs still with visual blur, ROIs containing multiple objects, and ROIs with both defocused and focused parts in one planktonic object from all the ROIs determined by the algorithm as "clear". Then we discarded all these selected ROIs together with their corresponding defocused counterparts. For the remained ROI pairs after cleaning, we denoted *d* as the defocus distance label to the defocused ROI in the pair. When the position of a target is between the focal plane and the lens, *d*=Δz; while when the target's position is outside the focal plane, *d*=-Δz. Thus, the construction of a large-scale defocused-focused marine plankton ROI image pair dataset with defocus distance label *d* was eventually completed.

2.3 IsPlanktonFE modeling

2.3.1 Network structure

Based on the characteristics of the defocused-focused plankton ROI pair dataset, we designed a self-guided DNN network

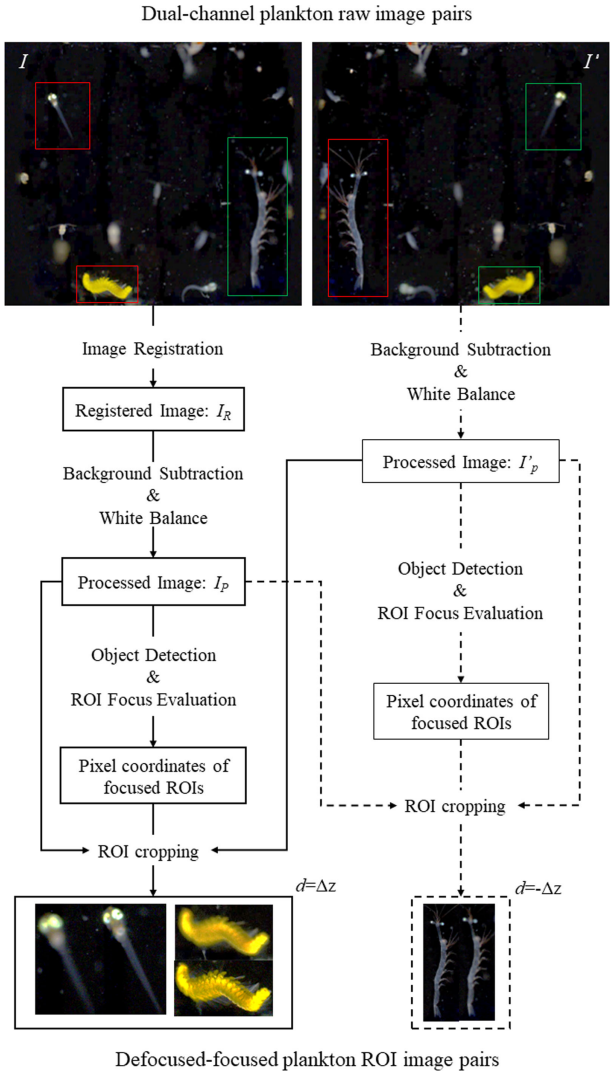


FIGURE 3
Flow chart of generating defocused-focused plankton ROI pairs from processing a pair of dual-channel raw plankton images. Red and green frames represent focused and defocused plankton ROIs, respectively.

IsPlanktonFE to achieve end-to-end focus-extension for underwater darkfield plankton imagers. As Figure 4 shows, the structure of IsPlanktonFE includes a defocus distance estimation sub-network DDE-Net and a focus-extension sub-network FE-Net. The DDE-Net can estimate the defocus distance of an input ROI image and encode the estimated value for the FE-Net to guide its focus-extension. The FE-Net extracts the features of the input ROI, and then encodes and decodes it with the reference of the estimated defocus distance. Finally, it reconstructs and outputs a refocused image.

The DDE-Net is composed of a feature extractor, an aggregator, a regressor and an encoder. A ResNet-34 network is trained as feature extractor to extract features of the input ROI image in patches (refer to Supplementary Materials for details). Then, the mean and standard deviation, quantile, and moment of the extracted features are aggregated into three vectors. Next, partial least squares regressor (PLSR) is used to regress the three vectors, and the mean of the regression value is taken as the estimate of the defocus distance of the input defocused ROI. Finally, the estimate is transformed by the encoder into two one-dimensional vectors α and β to meet the input requirements of the FE-Net combiner. The encoder consists of five convolutional layers with 1×1 cores.

FE-Net is the backbone of IsPlanktonFE, which is composed of an extractor, a combiner, and a decoder. The extractor extracts the feature map fm_{ex} from the input, which consists of four e-blocks and each contains three convolutional layers with batch-normalization and ReLU. The combiner is responsible for adding α and β to fm_{ex}. It consists of two f-blocks and each has a convolution and an FILM layer (Perez et al., 2018). The FILM layer implements the function as expressed in the equation below, where fm_{ref} is a feature map with reference information.

$$fm_{ref} = \alpha fm_{ex} + \beta$$

The decoder is composed of only one convolutional layer, which performs dimension transformation of the combined feature map fm_{ref} to obtain and merge the R, G and B channel values to output the final refocused image.

2.3.2 Loss

IsPlanktonFE adopts a weighted sum of contextual loss and MSE loss to guide its training optimization, as is formulated by the equation:

$$Loss = k_1 * L_{cont} + k_2 * L_{mse}$$

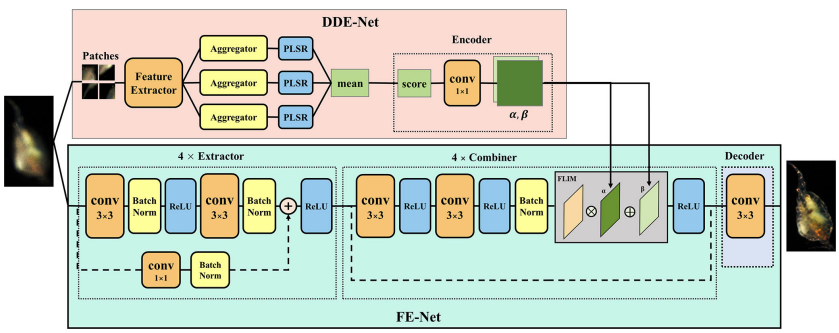


FIGURE 4
Schematic structure of the deep focus-extension network IsPlanktonFE.

The contextual loss (Li et al., 2017) is known to be insensitive to image misalignment, but it does not consider sufficiently on the global distribution of features and often generates mosaic artifacts in the output images (Odena et al., 2016). Conversely, the MSE loss takes the global features into account, but it is very sensitive to image registration errors. Therefore, the combinational loss is expected to complement their strengths and weaknesses to direct the IsPlanktonFE training towards more accurate feature recovery and better visual effect optimization for its ultimate focus-extension performance.

2.3.3 Network training

IsPlanktonFE was implemented using PyTorch on a server with 6 NVIDIA GeForce RTX3090 GPUs. Its training was divided into two phases: (1) training DDE-Net with defocused ROIs and the absolute value of their corresponding defocus distances in the dataset (more details are provided in the Supplementary Materials); (2) training FE-Net by using the trained DDE-NET and the defocused-focused ROI pairs in the dataset. In this phase, the model was trained with Adam optimizer and an initial learning rate of 6*10^ (-3). The initial weights k_1 and k_2 for the contextual loss and MSE loss were empirically set to 0.97 and 0.03. The batch size was set to 48 and patch size was set to 128. To improve training efficiency, only the number of foreground pixels in a patch exceeded 820 (empirical value) would it be used for training. After 200 epochs of training, the learning rate decreased to 0.316 times of its initial, and k_1 and k_2 became 0.95 and 0.05, respectively.

3 Results

3.1 Image dataset

A dual-channel darkfield imaging apparatus (Figure 2) has been home-built to capture raw image pairs of live plankton in seawater samples with focal distance difference between the channels adjusted in a range of 10 mm at a step of 1 mm. Afterwards, the raw image pairs were processed to finally construct a dataset containing 55 000 defocused-focused ROI pairs of marine plankton. We randomly selected 2500 ROI pairs recorded at each defocus distance as training set and reserved extra 250 ROI pairs as test set for testing and validating the IsPlanktonFE and the DeblurGAN-v2 deep networks (Kupyn et al., 2019). The DeblurGAN-v2 is a state-of-the-art deep-learning model for image restoration, which is selected for focus-extension performance comparison with our proposed IsPlanktonFE network. Its modeling details are provided in the Supplementary Materials.

3.2 Visual evaluation

The focus-extension results of the two DNN models are shown in Figure 5. From top to bottom, each column lists an input marine plankton ROI with different defocus distances in the test set, output images from the two DNN models, and the in-focus image of the input target (ground truth, GT), respectively. From visual perception on this comparison, we can see that both DNNs have restored

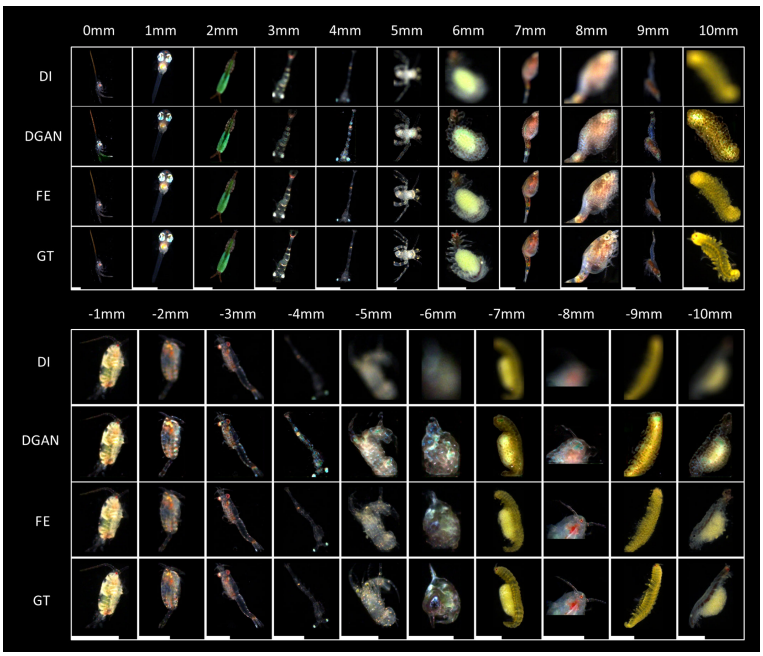


FIGURE 5
Visual comparisons of the focus-extension performances of two DNNs on representative marine plankton images. DI, DGAN, FE and GT represents the defocused images as network input, their corresponding restored images by the DeblurGAN-v2 and IsPlanktonFE models, and the focused ground truth images, respectively. The value above each column indicates the defocus distance label of the DI relative to the GT images. The length of the scale bars at the bottom is 1mm. The abbreviations of DI, FE, GT and DGAN are also applicable to other figures in this work.

considerable details of the targets from their defocused images. As the defocus distance increases, the defocus blur of the input image increases, the recovered details in the two network outputs become less, and the artifacts start to emerge and gradually become serious. But the artifacts in the IsPlanktonFE outputs are much weaker than those from the DeblurGAN-v2 model, and the restoration effect by IsPlanktonFE is visually closer to the corresponding sharp ground truths.

3.3 Quantitative evaluation

To make more quantitative evaluation of the DNNs' focus-extension performance, we used structural similarity (SSIM), root mean square error (RMSE) and a self-defined focus score (Yang et al., 2021) as three metrics, and calculated their values on all the input images in the test set and their outputs from the two DNN models relative to their in-focus GT images. We compared the mean and standard deviation of these scores at each defocus distance as the results displayed in Figure 6.

On all the metrics, the IsPlanktonFE network has achieved obvious improvements compared to its inputs and relatively stable standard deviations, indicating its efficacy in focus-extension. Moreover, IsPlanktonFE has also achieved higher scores than DeblurGAN-v2 on the metrics of SSIM (Figure 6A) and RMSE (Figure 6B). In comparison, DeblurGAN-v2 has not exhibited RMSE improvement (Figure 6B) and even obtained peculiarly worse scores than its inputs on SSIM (Figure 6A), though it marginally outperformed IsPlanktonFE on the focus score (Figure 6C). Upon visual inspection of the DeblurGAN-v2 outputs shown in Figure 5, we found these images contain quite some high-frequency artifacts.

Note in Figures 6A, B that the average SSIM and RMSE values of the gradually defocused input images decrease very slowly with the increase of defocus distance. Differently, as shown in Figure 6C, their focus scores decrease sharply with the increase of defocus distance, indicating that the evaluation of focus score is more consistent with human vision (Yang et al., 2021). This is not surprising, since this metric was designed to incorporate group evaluations from human vision.

3.4 Quantitative calibration

To reduce evaluation variance caused by morphological irregularity of diverse real planktonic organisms when characterizing IsPlanktonFE performance, we used a 2mm SiO_2 bead as target to further calibrate its achievable DOF through comparative image analysis of the bead. As shown in Figure 7, we used the dual-channel imaging apparatus to collect a series of focused and defocused image pairs of the bead in the defocus distance range of -12mm-12mm.

It is well known that defocused contour of a target leads to its inaccurate size measurement. Therefore, we calibrated the focus extended range by IsPlanktonFE based on the measurement accuracy of the bead's diameter from its darkfield images. As the results shown in Figure 8, the diameter measured from the in-focus images of the bead is 2.02mm, which is very close to its nominal value of 2mm. From the defocused images, the measurement error gradually increases with the increase of defocus distance, and reaches 0.25mm at ±10mm (>10%). In comparison, the measurement error from the focus extended images is almost negligible within a defocus range of -5mm-5mm. Only when the defocus distance is outside this range, the error begins to increase

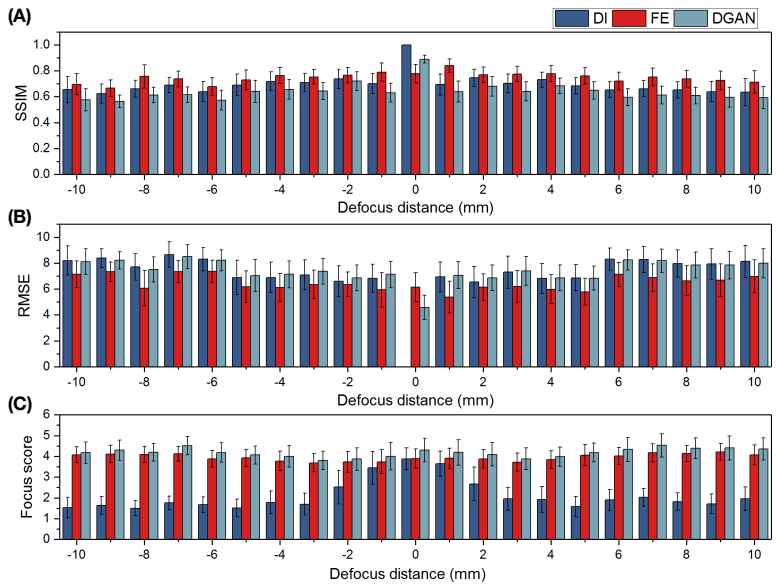


FIGURE 6
Quantitative comparisons of the focus-extension performances of the DeblurGAN-v2 and IsPlanktonFE networks on restoring marine plankton images with metrics of SSIM **(A)**, RMSE **(B)** and focus score **(C)**. Here larger SSIM (∈[0,1]), smaller RMSE, and higher Focus score indicates sharper image focusing state, respectively.

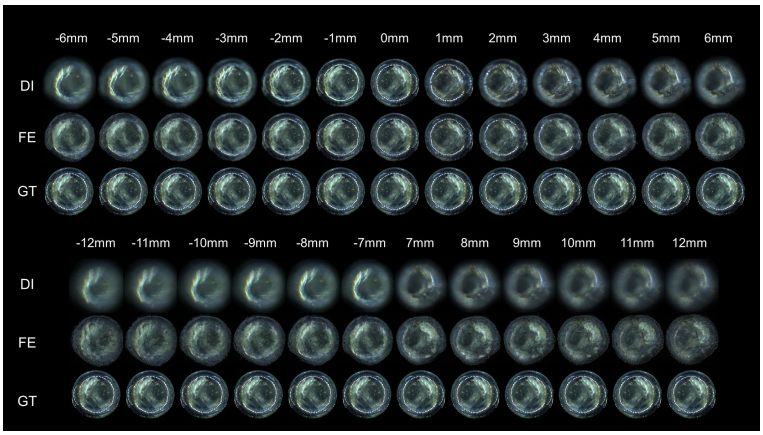


FIGURE 7
Visual comparisons of the focus-extension performance of the IsPlanktonFE network on darkfield images of a SiO_2 bead target. The value above each column indicates the defocus distance between the DI and the GT images. The nominal diameter of the SiO_2 bead is 2mm.

gradually. Even at a defocus distance of ±10mm, the error is only 0.1mm (~5%).

On the other hand, the details inside the contour of a planktonic target, e.g., the high-frequency features of some biological structures, often play a decisive role for its recognition. Defocus obviously blurs the faithful representation of such features, resulting in serious decline in the accuracy of subsequent recognition. Therefore, we further calibrated the focus-extension performance of IsPlanktonFE by evaluating its ability in restoring the internal features of the bead's images. Using the RMSE, SSIM and focus score metrics again, we quantified and compared the features in different defocused and the focus-extended images of the bead as the results shown in Figure 9. We found that the measurement from the focus-extended images were significantly better than from the defocused images in the range of -12mm-12mm. IsPlanktonFE has not only achieved good restoration effect in the trained defocus range of -10mm-10mm, but also considerable improvement outside this range.

To further confirm the focus-extended range by IsPlanktonFE, we designed another experiment to compare the performance of a ResNet-18 network, which had been pre-trained only on 20 classes of focused plankton ROIs, on classifying marine plankton from their defocused and corresponding focus-extended images. For each class, 200 ROIs were subsampled from the dataset constructed in this work. The results in Figure 10A show that within the defocus range of -10mm-10mm, the accuracy of the ResNet-18 network on classifying the blurred images decreases significantly with the increase of defocus distance, which is consistent with common sense. In contrast, the accuracy on classifying the focus-extended images decreases much slower and remains better than 50% even at defocus distances of ±10 mm. Figure 10B further details the changes of the ResNet-18's accuracy on classifying focus-extended images with different plankton sizes. It can be seen that for the images collected by the 0.5× imaging optics, the accuracy on classifying plankton with body length > 2mm decreases very slowly with the increase of defocus distance, reaching more than 70% within the defocus range of -10mm-10mm. However, for smaller plankton with body length< 2mm, the accuracy decreases rapidly. Even though, such accuracy is still higher than that achieved on the defocused images. These results indicate that within the defocus range of -10mm-10mm, the focus-extension by IsPlanktonFE has contributed effective improvement to a machine classifier's performance on recognizing plankton images. And the larger the target, the better such performance.

The above experimental results from the SiO_2 bead size measurement, bead image feature restoration assessment, and machine classifier performance evaluation on real plankton images have proved that our trained IsPlanktonFE deep network at least extended a 0.5× darkfield imaging system's DOF of 3mm to a wider range of -10mm-10mm, which is equivalent to ~7 times focus-extension.

3.5 Model generalizability

We selected a batch of *in situ* acquired ROIs by different IPPs deployed separately from coastal sites near Shenzhen, Pingtan, and Changjiang of China, and Hobart of Australia, and used the metric of

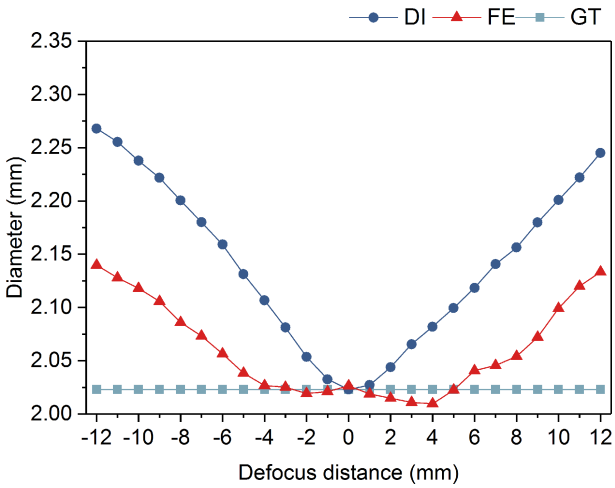


FIGURE 8
Diameter comparison of a SiO_2 bead measured from its defocused images, focus-extended images by IsPlanktonFE, and focused GT images.

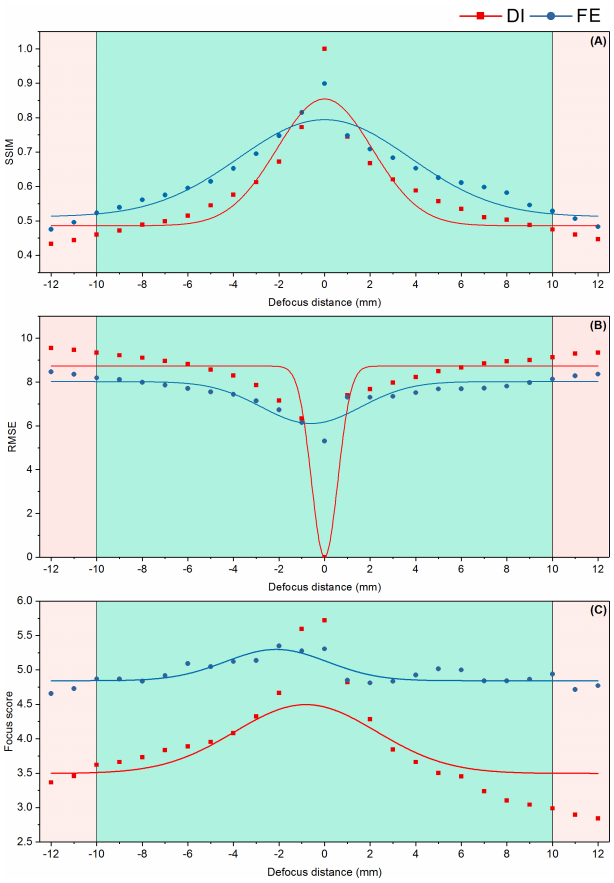


FIGURE 9
Quantitative focus-extension evaluation of the IsPlanktonFE network using a 2mm SiO_2 bead by the metrics of SSIM **(A)**, RMSE **(B)** and focus score **(C)**, respectively. The cyan areas in the panels indicate a defocus distance range from which the recorded SiO_2 bead images are used for training the IsPlanktonFE model. The pink areas indicate defocus distance ranges where the recorded SiO_2 bead images were NOT used to train the IsPlanktonFE model.

focus score to evaluate their focus states before and after the processing by IsPlanktonFE to test its generalization performance on real-world test data.

Figure 11 shows the test results of IsPlanktonFE on some ROIs containing in-distribution (i.e., plankton classes already in the training dataset) plankton of *Copepod*, *Polychaete* and *Mysid*, in which the defocused and refocused ROIs of these representative common plankters and their focus scores are given for comparison. Note that the principles of IPP and the dual-channel imaging apparatus are both darkfield imaging, but their optics has certain differences in lens magnification, pixel size and illumination. These make the selected test images have feature variations from those in the training set, although biologically they belong to the same taxa. Judging from the resultant visual evaluation and quantitative focus score comparisons of the input and output images, we see that IsPlanktonFE has evidently refocused the blurred images.

Then we selected another subset of out-of-distribution ROIs (biological classes and/or defocus states not included in the training set) to further validate the generalizability of IsPlanktonFE on different image contents. The test results are shown in Figure 12, in which the *Lucifer* sp. (Figure 12A) and the *Polychaete* sp. and the *copepod* (Figure 12B) exhibits more complex defocus status although their taxa were already included in the training set. Specifically, different parts of the *Lucifer* sp. defocus differently, and the *Polychate* sp. and the *copepod* has different defocus degree. While in Figures 12C, D, the defocus status of the *Monstroid* with eggs and the *Cresis Acicula* is also very complex, and their taxa are not included in the training set. Observing the refocused ROIs and their focus scores, we can see IsPlanktonFE has achieved excellent refocusing effect on them with never-trained contents. Remarkably, as shown in Figures 12E, F and more in Figure S3 in the Supplementary Materials, IsPlanktonFE also turned out to be very effective in refocusing the suspended particle images, which were never trained and are known to be extremely heterogeneous in morphology.

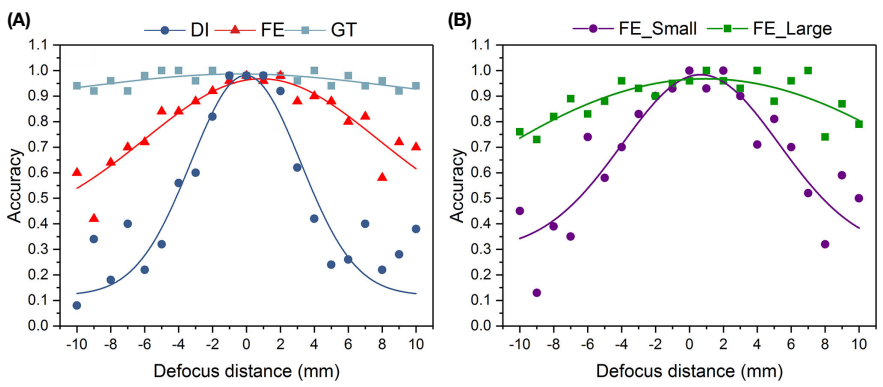


FIGURE 10
Performance comparison of a trained ResNet-18 deep network on classifying defocused plankton images, their focus-extended counterparts by IsPlanktonFE, and corresponding focused GT images. The ResNet-18 classifier has been trained on focused plankton images before this test. The value of each point is the average accuracy of 50 randomly selected testing plankton images. **(A)** plots the classification accuracy of the ResNet-18 model versus defocus distance on DI, FE, and GT testing plankton images. **(B)** further plots the classification accuracy on the FE plankton images grouped in two different sizes. FE_Small refers to plankton length (diagonal of ROI)<2mm. FE_Large refers to plankton length >2mm.

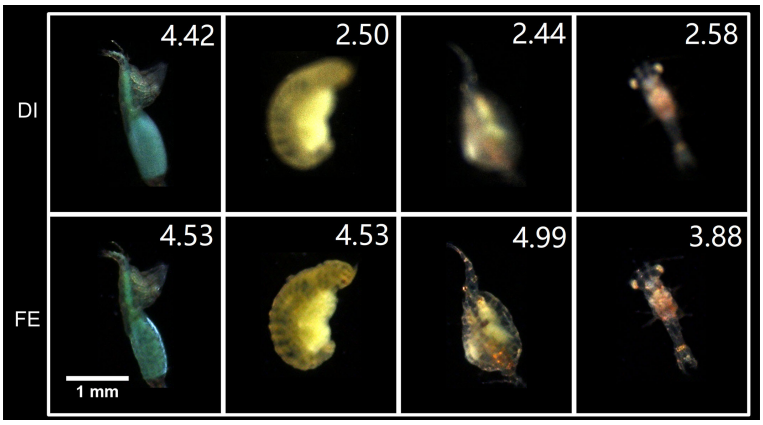


FIGURE 11
Focus-extension comparisons of IsPlanktonFE on some representative *in situ* marine plankton ROIs acquired by IPP. The value at the top-right corner of each ROI indicates its focus score.

3.6 Observation efficiency improvement

We used the raw ROIs acquired from a trial IPP deployment underwater at the wharf of CSRIO Ocean and Atmosphere of Australia during December 25-29, 2019 as testing data to verify the efficiency improvement by applying IsPlanktonFE for *in situ* marine plankton observation. In this data, an average of 26,414 ROIs were taken per day. Because the focus evaluation filtration algorithm (Yang et al., 2021) was not deployed at the time during their collection, the data contains many defocused ROIs.

We first screened the focus scores of all the original ROIs and found that the average daily proportion of the clear images (focus score>4.0) in the raw ROIs was only 2.15%. After applying focus-extension to all the raw ROIs by IsPlanktonFE, such proportion rose to 12.75%. The proportional change of clear ROIs before and after IsPlanktonFE processing and the absolute number of raw ROIs collected daily are shown in Figure 13. The proportional fluctuations of the focused ROIs were mainly caused by the natural change of the plankton and suspended particles abundance and spatial distribution with time. This result further confirmed the application of IsPlanktonFE can greatly increase the fraction of in-focus images acquired in the raw data. The observation efficiency of IPP has been further lifted by ~6 times on average.

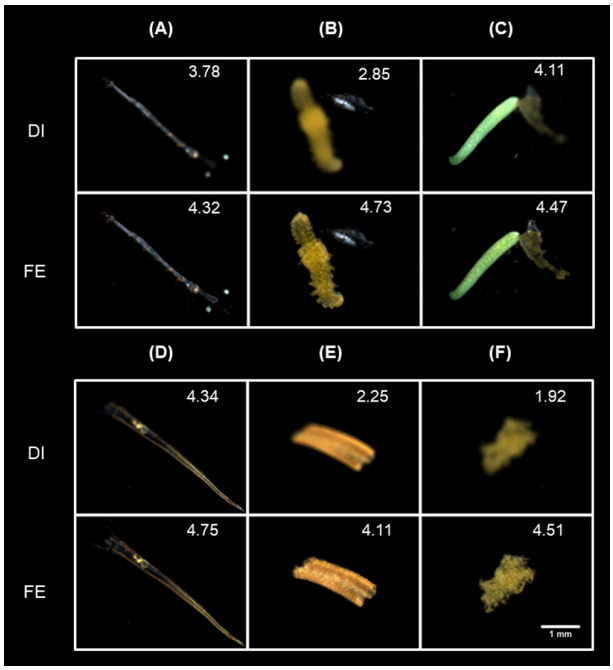


FIGURE 12
Focus-extension comparisons of IsPlanktonFE on some special *in situ* darkfield ROIs acquired by IPP. **(A)** A ROI contains an in-distribution plankter with different degrees of defocus in different body parts. **(B)** A ROI contains two different in-distribution plankters with different degrees of defocus. **(C)** and **(D)** contains an out-of-distribution (OOD) plankter, respectively. **(E)** and **(F)** contains an OOD suspended particle, respectively. The value at the top-right corner of each ROI indicates its focus score.

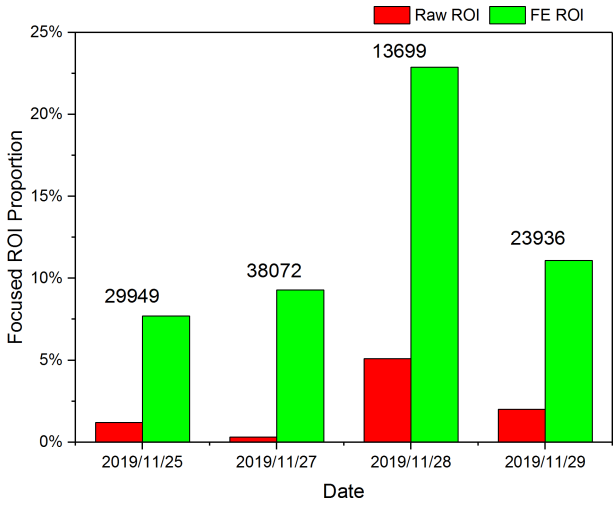


FIGURE 13
Proportion change of the clear ROI images in a 4-day *in situ* image data acquired by an IPP camera deployed underwater before and after all the raw ROIs processed by IsPlanktonFE.

4 Discussion

4.1 Model performance

Single image refocusing is an ill-posed problem. There are many possible solutions to the mapping between a defocused image and its focused GT. The DDE-Net essentially provides a unique descriptor for the defocus degree of the input, which assists IsPlanktonFE to learn more definitely the optimal mapping between blurred and clear images. The practicality of this module has obviously benefited from our dual-channel imaging apparatus, which provided accurate defocus distance label to the training image pairs. In contrast, DeblurGan-v2 cannot utilize defocus distance prior and only relied on image data to learn such complicated mapping. Limited by network layer depth and parameters, DeblurGan-v2 is difficult to handle volatile defocus situations in the input images, and generated serious artifacts for many images. This is why the output of DeblurGAN-v2 even appear lower SSIM and RMSE scores than its inputs. The DNN used in the focus score evaluation (Yang et al., 2021) is only sensitive to the content rather than the distribution of high-frequency features, so DeblurGAN-v2 has achieved higher scores on this metric as it is prone to generate falsely distributed high-frequency artifacts in the output images.

With the increase of defocus distance, each pixel in a defocused image receives contributions from more neighboring object features, which renders it more challenging to remap them correctly to their original pixel locations. In this limit, it gradually becomes difficult for IsPlanktonFE to obtain an optimal solution to the mapping even it is provided with the defocus distance information.

Figure 14 shows the focus-extension effect of IsPlanktonFE on planktons with different sizes at different defocus distances. The comparison between Figures 14A, B shows that at the same defocus distances, the larger plankter containing more features in the defocused image achieved higher SSIM score and better image quality after the restoration. The comparison between Figures 14A, C reflects that at a larger defocus distance, the model is also more likely to perform better for larger plankton. These results indicate that the focus-extension performance of IsPlanktonFE will decline with the decrease of planktonic object size. The resolution of an optical system with fixed magnification will become insufficient when imaging smaller targets. Under such situation, defocus blur will result in further loss of high-frequency features in the low-resolution images of tiny objects, thus disadvantaging the restoration of any focus-extension models without exception of IsPlanktonFE. When the resolution of acquired images is not enough to entertain the computation requirements of IsPlanktonFE, higher magnification of the imaging instruments should be considered.

It is worth discussion that the darkfield ROIs have offered IsPlanktonFE relatively simplified foreground and background contents compared to the more complex natural scene images. On the one hand, it is rare in the ROIs that multiple foreground targets overlap or a single target has inconsistent defocus situations at different focal positions. This means the label *d* can depict the defocus distances of the foreground targets accurately and uniquely. On the other hand, darkfield ROIs have almost zero backgrounds with little interference to the foregrounds. These data characteristics are all favorable to the training of IsPlanktonFE for performance. In addition, the dual-channel imaging apparatus has an inborn advantage of reciprocity in acquiring the image pairs, which means the defocused and focused targets in one channel image are counterparts to those in the other channel image, so it can acquire 2n image pairs by just stepping one focal plane of the two channels for only n positions. Note that this is very useful to speed up the data acquisition to keep the fragile planktonic organisms' bioactivity after being caught out from the marine environment. It certainly also saves time to upscale the live plankton image dataset for IsPlanktonFE training.

4.2 Model generalizability

Currently, the IsPlanktonFE dataset has reached a scale of 55 000 ROI pairs, containing at least 20 classes of plankton from the phyla of *Arthropoda* and *Annelida*. Admittedly, compared with the actual number of plankton in the ocean and the high complexity of the underwater environment, the scale and diversity of this dataset is far from sufficient. Fortunately, IsPlanktonFE has learnt the low-level

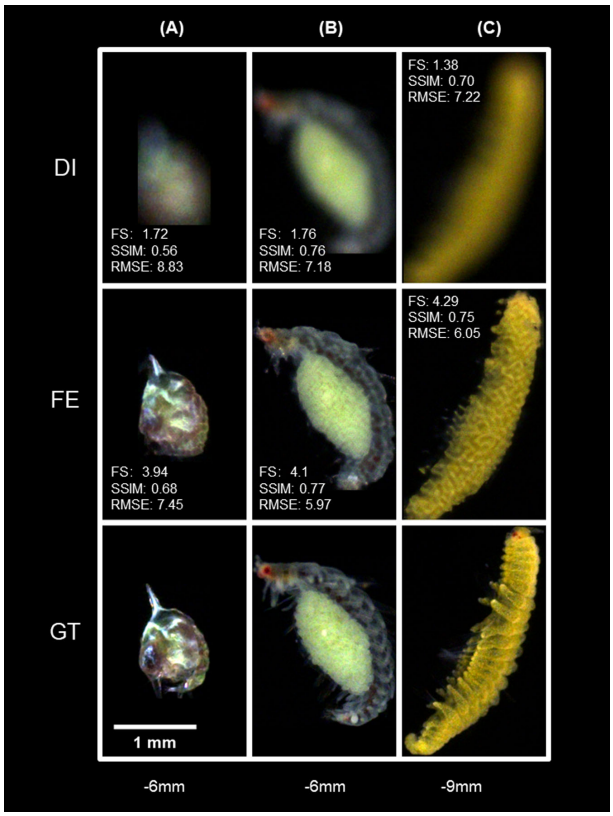


FIGURE 14
Effects of plankton size on the focus-extension performance of IsPlanktonFE. **(A)** a *Megalopa larva* at *d*=-6mm, **(B)** a *Polychate* with eggs at *d*=-6mm, **(C)** a different *Polychate* at d=-9mm.

pixel-by-pixel feature mapping between defocused-focused image pairs, which is essentially different from an image classifier learning mappings between images and their categorical labels, so it is less sensitive to high-level image semantics such as classes, morphologies, etc. When the ROI pairs in the IsPlanktonFE dataset were acquired, the plankton were alive so they kept on swimming. Compared with the limited FOV and DOF of the imaging apparatus and the confined space in the container, their motion is not finite at all. In addition, we only used a slow frame rate during the raw image acquisition, so even for the same plankton target, its defocus position, orientation and morphology can vary significantly in temporally close frames and spatially close positions. This allows our dataset to have a very rich "defocused image-focused image" mapping variety, *i.e.*, defocus diversity, even though its construction is based on a limited number of taxonomic groups. As a result, even trained and tested by a simple random 10:1 split of the datasets, the obtained IsPlanktonFE model has achieved good generalizability on the images of untrained organism classes, images collected by different instruments, or even suspended particles, as demonstrated in the results and the Figure S3 in the Supplementary Materials. In principle, we can reasonably infer that this generalizability has more potentials.

Firstly, the good performance of IsPlanktonFE on *in situ* images acquired by different IPPs suggests its application potential for a variety of other *in situ* darkfield plankton imagers. Although the optics of the dual-channel imaging apparatus has various differences from the other underwater darkfield imagers besides IPP, their principle is all snapshot darkfield macrophotography or microscopy (Benfield et al., 1996; Schulz et al., 2009; Picheral et al., 2010; Gallager, 2019; Orenstein et al., 2020; Li et al., 2022). The representation of plankton and particles by their images exhibits very similar morphological features. Therefore, the modeling process of IsPlanktonFE could be repeated to optimize different hyperparameters for these instruments. In fact, several sets of hyperparameters at different magnifications can also be multiplexed to further improve the universality of the model. This perhaps is the most attractive aspect of IsPlanktonFE, as it requires no hardware changes to existing imagers, but can computationally extend their DOFs and make observation efficiency improvements.

Secondly, although the marine plankton species is extremely diverse, the number of high-level taxonomic groups become much smaller. It is known that organisms belong to the same phylum have many similar morphological features but different between different phyla (Batten et al., 2019), e.g., *Copepoda* and *Eumalacostraca* share quite some alike exoskeleton features as they both belong to *Arthropoda*. Due to this biological property, the training scale of IsPlanktonFE might not necessarily need to cover all plankton species, but only needs to cover enough higher-order phyla, and then it is possible to generalize to the images of untrained species under the same phyla. The examples of *Monstrilloid* and *Creseis acicula* in Figures 12C, D exactly justified this hypothesis. It is known that *copepods* alone account for 90% of all marine zooplankton abundance (Suthers et al., 2019), so it is anticipated that IsPlanktonFE has the potential to generalize to a considerable number of plankton classes. However, the dataset still lacks examples from *Chordata*, *Cnidaria* and etc. The performance of IsPlanktonFE is likely to drop if it were directly applied to *in situ* plankton images of these species. But users can simply apply the apparatus and methods described in this paper, combined with live sampling of these plankton, their images can be conveniently supplemented into the training set. The model can then be retrained to quickly regain the focus-extension ability on these images.

Surprisingly, IsPlanktonFE also achieves excellent performance on various suspended particle images, although none of them are used in the training. This may be because the kaleidoscopic particles in their darkfield images also contain many features similar to plankton morphology that have been learnt by the model. This speculation does not lack of scientific reasoning as many particulates are the debris of plankton or fecal pellets (smaller planktonic debris). On the other hand, marine plankton might evolve their morphology to resemble the suspended particles in seawater, because this is preferable for their hiding and survival. It is well known that the quantity of suspended particles in the ocean is extremely large [>80% *in situ* images are particles (Panaïotis et al., 2022)], and their size and volume are the most concerned measurands in the study of oceanic carbon cycle (Lombard et al., 2019; Giering et al., 2020). IsPlanktonFE has achieved remarkable performance on refocusing the contour of defocused suspended particles, which is obviously beneficial for their more accurate measurement.

4.3 Impact on marine plankton observation

Generally, the density of plankton in ocean is low, and larger organisms are in lower abundance. For more efficient observations, the seawater sampled per unit time by any plankton imaging method is always favored as more as possible to obtain statistically representative plankton information and more undersea space coverage. Extending the focus of an imaging instrument with fixed magnification (i.e., fixed field of view) is obviously a straightforward way for this purpose. Taking the results of this study as an example, IsPlanktonFE can extends the DOF of a 0.5× darkfield imaging system from ~3mm to ~20mm, equivalent to a seawater sampling volume improvement of ~7 times. Based on our experience in IPP nearshore deployment, this will result in a total water volume enlargement from 388.8L to 2592L per day at an imaging frame rate of 3FPS. This is expected to further fill the gap in seawater sampling throughput and plankton quantification between the traditional methods and the *in situ* imagers (Barth and Stone, 2022; Le et al., 2022).

The value of IsPlanktonFE for plankton observation is also reflected in its generalization for future darkfield imaging instrumentation. When an imager works underwater, the space outside its DOF will inevitably be illuminated, and the stray light from these spaces can enter the camera and reach the imaging chip. This is the fundamental cause for defocus blur in darkfield imaging. The existing darkfield imagers generally assess the focus of the target in a ROI by simple edge gradient calculation routine, and directly discard the defocused ROIs to just retain the "clear" ones (Gallager, 2019; Orenstein et al., 2020; Li et al., 2022). This made their high-quality image yield and light energy utilization rate very low. IPP improves such yield by physically compressing the illumination into a layer of thickness to ~6.89mm to reduce illumination outside the DOF. However, it is not trivial to further compress incoherent light beams thinner than such thickness. This will not only increase the

complexity of the optomechanical structure of the illuminator and lose instrumentation flexibility, but also greatly waste illumination energy. If the DOF is extended to ~20mm by IsPlanktonFE, the difficulty of matching the illumination layer with the imaging DOF can be greatly alleviated, in favor of simplifying the instrumentation and can also improve the high-quality image yield of darkfield imaging to its limit. Moreover, the process of IsPlanktonFE establishment has provided users references to solve their respective research problems of interest, in which the users may have to deal with different challenges from complicated observational objects and application environments. All these will certainly help marine scientists to explore more of the unknown ocean at better cost.

One shortcoming of IsPlanktonFE lies in its deep network structure, leading to high computation cost and long training time. It took 120 hours and six RTX3090 GPU cards in training the model to achieve convergence. Fortunately, the computation is greatly reduced in its production phase, during which IsPlanktonFE was verified to achieve an average speed of 1-2 ROI/s using a single RTX3090 GPU. This is much faster than the average speed of ROI generation by an IPP deployed nearshore according to our experience (Li et al., 2022). In the future, the model can be light-weighted by network pruning (Molchanov et al., 2019), quantification (Wang et al., 2022), knowledge distillation (Yim et al., 2017) techniques to further reduce its demand on computation for training and inference. These will enable the deployment of IsPlanktonFE on a cloud computing platform to facilitate next-generation *in situ* real-time marine plankton observations.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

TC and JL conceived the idea, designed the dual-channel darkfield imaging apparatus and JL supervised the project. TC collected the *in situ* marine plankton image raw data. TC, WM, GG, ZY and ZL built the IsPlanktonFE dataset. TC, GG and ZY developed computer programming for the IsPlanktonFE model. WM developed computer programming for the DeBlurGan-v2 model and JQ supervised this work. TC and JL designed the experiments, analysed the findings, and drafted the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by International Partnership Program of Chinese Academy of Sciences No. 172644kysb20210022, Shenzhen Science and Technology Innovation Program No. JCYJ20200109105823170 and Key Project of the Guangdong Basic and Applied Basic Research Foundation No. 2022B1515120030.

Acknowledgments

The authors are grateful to Mr. Mark Luk and Mr. Tom Xu from Shenzhen Xcube Technology Co. Ltd. for their technical support during IPP deployment at Hobart, Australia, and Pingtan, China, and to Dr. Haifeng Gu from Third Institute of Oceanography, Ministry of Natural Resources, China for his logistics support during IPP deployment at Pingtan. T.C. acknowledges Dr. Tim Yue Him Wong from Shenzhen University for his inspiring discussion and language advice.

Conflict of interest

TC and JL have three pending China invention patent applications on the presented dual-channel darkfield imaging apparatus and the IsPlanktonFE algorithm.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1074428/full#supplementary-material

References

Abuolaim, A., and Brown, M. S. (2020). "Defocus deblurring using dual-pixel data," in *European Conference on computer vision*. (Springer, Cham), 111–126. doi: 10.1007/978-3-030-58607-2_7

Alvarez-Fernandez, S., Bach, L. T., Taucher, J., Riebesell, U., Sommer, U., Aberle, N., et al. (2018). Plankton responses to ocean acidification: The role of nutrient limitation. *Prog. Oceanogr.* 165, 11–18. doi: 10.1016/j.pocean.2018.04.006

Barth, A., and Stone, J. (2022). Comparison of an *in situ* imaging device and net-based method to study mesozooplankton communities in an oligotrophic system. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.898057

Batten, S. D., Abu-Alhaija, R., Chiba, S., Edwards, M., Graham, G., Jyothibabu, R., et al. (2019). A global plankton diversity monitoring program. *Front. Mar. Sci.* 6. doi: 10.3389/fmars.2019.00321

Benfield, M. C., Davis, C. S., Wiebe, P. H., Gallager, S. M., Lough, R. G., and Copley, N. J. (1996). Video plankton recorder estimates of copepod, pteropod and larvacean distributions from a stratified region of georges bank with comparative measurements from a MOCNESS sampler. *Deep-sea Res. Part 2. Topical Stud. oceanography.* 43 (43), 1925–1945. doi: 10.1016/S0967-0645(96)00044-6

Bi, H., Cook, S., Yu, H., Benfield, M. C., and Houde, E. D. (2013). Deployment of an imaging system to investigate fine-scale spatial distribution of early life stages of the ctenophore *Mnemiopsis leidyi* in Chesapeake Bay. *J. Plankton Res.* 35 (2), 270–280. doi: 10.1093/plankt/fbs094

Campbell, R. W., Roberts, P. L., and Jaffe, J. (2020). The prince William sound plankton camera: a profiling *in situ* observatory of plankton and particulates. *ICES J. Mar. Sci.* 77 (4), 1440–1455. doi: 10.1093/icesjms/fsaa029

Cheng, Y., Cao, J., Tang, X., and Hao, Q. (2021). Optical zoom imaging systems using adaptive liquid lenses. *Bioinspir. Biomim.* 16 (4), 041002. doi: 10.1088/1748-3190/abfc2b

Cheng, X., Cheng, K., and Bi, H. (2020). Dynamic downscaling segmentation for noisy, low-contrast in situ underwater plankton images. *IEEE Access.* 8, 111012–111026. doi: 10.1109/access.2020.3001613

Cowen, R. K., and Guigand, C. M. (2008). *In situ* ichthyoplankton imaging system (ISIIS): system design and preliminary results. *Limnol. Oceanogr.: Methods.* 6 (2), 126–132. doi: 10.4319/lom.2008.6.126

Cowen, R. K., and Sponaugle, S. (2009). Larval dispersal and marine population connectivity. *Annu. Rev. Mar. Sci.* 1 (1), 443–466. doi: 10.1146/annurev.marine.010908.163757

Davies, E. J., and Nepstad, R. (2017). *In situ* characterisation of complex suspended particulates surrounding an active submarine tailings placement site in a Norwegian fjord. *Regional Stud. Mar. Sci.* 16, 198–207. doi: 10.1016/j.rsma.2017.09.008

Fan, F., Yang, K., Fu, B., Xia, M., and Zhang, W. (2010). "Application of blind deconvolution approach with image quality metric in underwater image restoration," in Proceedings of the International Conference on Image Analysis and Signal Processing. (Zhejiang, China: IEEE), 236–239. doi: 10.1109/IASP.2010.5476122

Gallager, S. M. (2019). Continuous particle imaging and classification system. Washington, DC: U.S. Patent and Trademark Office, U.S. Patent No 10222688.

Giering, S. L. C., Cavan, E. L., Basedow, S. L., Briggs, N., Burd, A. B., Darroch, L. J., et al. (2020). Sinking organic particles in the ocean–flux estimates from *in situ* optical devices. *Front. Mar. Sci.* 6 (834). doi: 10.3389/fmars.2019.00834

Gorsky, G., Picheral, M., and Stemmann, L. (2000). Use of the underwater video profiler for the study of aggregate dynamics in the north Mediterranean. *Estuarine Coast. Shelf Sci.* 50 (1), 121–128. doi: 10.1006/ecss.1999.0539

Graham, G. W., and Nimmo Smith, W. A. M. (2010). The application of holography to the analysis of size and settling velocity of suspended cohesive sediments. *Limnol. Oceanogr.: Methods.* 8 (1), 1–15. doi: 10.4319/lom.2010.8.1

Grcs, Z., Tamamitsu, M., Bianco, V., Wolf, P., Roy, S., Shindo, K., et al. (2018). A deep learning-enabled portable imaging flow cytometer for cost-effective, high-throughput, and label-free analysis of natural water samples. *Light Sci. Appl.* 7, 66. doi: 10.1038/s41377-018-0067-0

Hirche, H. J., Barz, K., Ayon, P., and Schulz, J. (2014). High resolution vertical distribution of the copepod *Calanus chilensis* in relation to the shallow oxygen minimum zone off northern Peru using LOKI, a new plankton imaging system. *Deep Sea Res. Part I: Oceanographic Res. Papers.* 88, 63–73. doi: 10.1016/j.dsr.2014.03.001

Karaali, A., and Jung, C. R. (2018). Edge-based defocus blur estimation with adaptive scale selection. *IEEE Trans. Image Process.* 27 (3), 1126–1137. doi: 10.1109/TIP.2017.2771563

Katz, J., and Sheng, J. (2010). Applications of holography in fluid mechanics and particle dynamics. *Annu. Rev. Fluid Mechanics.* 42 (1), 531–555. doi: 10.1146/annurev-fluid-121108-145508

Krishnan, D., Tay, T., and Fergus, R. (2011). "Blind deconvolution using a normalized sparsity measure," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* (Colorado Springs, CO, USA: IEEE), 233–240. doi: 10.1109/CVPR.2011.5995521

Kupyn, O., Martyniuk, T., Wu, J., and Wang, Z. (2019). "Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better," in *Proceedings of the IEEE/CVF International Conference on Computer Vision.* (Seoul, South Korea: IEEE), 8878–8887. doi: 10.1109/ICCV.2019.00897

Lee, J., Son, H., Rim, J., Cho, S., and Lee, S. (2021). "Iterative filter adaptive network for single image defocus deblurring," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (Nashville, TN, USA: IEEE), 2034–2042. doi: 10.1109/CVPR46437.2021.00207

Le, K. T., Yuan, Z., Syed, A., Ratelle, D., Orenstein, E. C., Carter, M. L., et al. (2022). Benchmarking and automating the image recognition capability of an *In situ* plankton imaging system. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.869088

Li, J., Chen, T., Yang, Z., Chen, L., Liu, P., Zhang, Y., et al. (2022). Development of a buoy-borne underwater imaging system for *In situ* mesoplankton monitoring of coastal waters. *IEEE J. Oceanic Eng.* 47 (1), 88–110. doi: 10.1109/JOE.2021.3106122

Li, D., Jiang, T., and Jiang, M. (2017). "Exploiting high-level semantics for no-reference image quality assessment of realistic blur images," in *Proceedings of the 25th ACM international conference on Multimedia*. (Mountain View, California USA: ACM), 378–386. doi: 10.1145/3123266.3123322

Lombard, F., Boss, E., Waite, A. M., Vogt, M., Uitz, J., Stemmann, L., et al. (2019). Globally consistent quantitative observations of planktonic ecosystems. *Front. Mar. Sci.* 6. doi: 10.3389/fmars.2019.00196

Luo, Y., Huang, L., Rivenson, Y., and Ozcan, A. (2021). Single-shot autofocusing of microscopy images using deep learning. *ACS Photonics.* 8 (2), 625–638. doi: 10.1021/acsphotonics.0c01774

Makarkin, M., and Bratashov, D. (2021). State-of-the-Art approaches for image deconvolution problems, including modern deep learning architectures. *Micromachines.* 12 (12), 1558. doi: 10.3390/mi12121558

Martínez-Corral, M., and Javidi, B. (2018). Fundamentals of 3D imaging and displays: a tutorial on integral imaging, light-field, and plenoptic systems. *Adv. Optics Photonics.* 10 (3), 512–566. doi: 10.1364/aop.10.000512

Merz, E., Kozakiewicz, T., Reyes, M., Ebi, C., Isles, P., Baity-Jesi, M., et al. (2021). Underwater dual-magnification imaging for automated lake plankton monitoring. *Water Res.* 203, 117524. doi: 10.1016/j.watres.2021.117524

Molchanov, P., Mallya, A., Tyree, S., Frosio, I., and Kautz, J. (2019). "Importance estimation for neural network pruning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (Beach, CA, USA: IEEE), 11264–11272. doi: 10.1109/CVPR.2019.01152

Nayak, A. R., Malkiel, E., McFarland, M. N., Twardowski, M. S., and Sullivan, J. M. (2021). A review of holography in the aquatic sciences: *In situ* characterization of particles, plankton, and small scale biophysical interactions. *Front. Mar. Sci.* 7. doi: 10.3389/fmars.2020.572147

Nishiyama, M., Hadid, A., Takeshima, H., Shotton, J., Kozakaya, T., and Yamaguchi, O. (2011). Facial deblur inference using subspace analysis for recognition of blurred faces. *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (4), 838–845. doi: 10.1109/TPAMI.2010.203

Odena, A., Dumoulin, V., and Olah, C. (2016). Deconvolution and checkerboard artifacts. *Distill.* 1 (10), e3. doi: 10.23915/distill.00003

Olson, R. J., and Sosik, H. M. (2007). A submersible imaging-in-flow instrument to analyze nano-and microplankton: Imaging FlowCytobot. *Limnol. Oceanogr. Methods.* 5, 195–203. doi: 10.4319/lom.2007.5.195

Orenstein, E. C., Ayata, S.-D., Maps, F., Becker, É.C., Benedetti, F., Biard, T., et al. (2022). Machine learning techniques to characterize functional traits of plankton from image data. *Limnol. Oceanogr.* 67 (8), 1647–1669. doi: 10.1002/lno.12101

Orenstein, E. C., Ratelle, D., Briseño-Avena, C., Carter, M. L., Franks, P. J. S., Jaffe, J. S., et al. (2020). The Scripps plankton camera system: A framework and platform for *in situ* microscopy. *Limnol. Oceanogr. Methods* 18 (11), 681–695. doi: 10.1002/lom3.10394

Ortner, P. B., Cummings, S. R., Aftring, R. P., and Edgerton, H. E. (1979). Silhouette photography of oceanic zooplankton. *Nature.* 277 (5691), 50–51. doi: 10.1038/277050a0

Panaïotis, T., Caray–Counil, L., Woodward, B., Schmid, M. S., Daprano, D., Tsai, S. T., et al. (2022). Content-aware segmentation of objects spanning a Large size range: Application to plankton images. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.870005

Perez, E., Strub, F., De Vries, H., Dumoulin, V., and Courville, A. (2018). "Film: Visual reasoning with a general conditioning layer," in *Proceedings of the AAAI Conference on Artificial Intelligence*. (Hilton New Orleans Riverside, New Orleans, Louisiana, USA: ACM), 3942–3951. doi: 10.1609/aaai.v32i1.11671

Picheral, M., Guidi, L., Stemmann, L., Karl, D. M., Iddaoud, G., and Gorsky, G. (2010). The underwater vision profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton. *Limnol. Oceanogr. Methods* 8 (9), 462–473. doi: 10.4319/lom.2010.8.462

Rai Dastidar, T., and Ethirajan, R. (2020). Whole slide imaging system using deep learning-based automated focusing. *BioMed. Opt Express.* 11 (1), 480–491. doi: 10.1364/BOE.379780

Rotermund, L. M., and Samson, J. (2015). A submersible holographic microscope for 4-d in-situ studies of micro-organisms in the ocean with intensity and quantitative phase imaging. *J. Mar. Science: Res. Dev.* 06 (01). doi: 10.4172/2155-9910.1000181

Schulz, J., Barz, K., Mengedoht, D., Hanken, T., Lilienthal, H., Rieper, N., et al. (2009). Lightframe on-sight key species investigation (LOKI). (Bremen, Germany: OCEANS 2009-EUROPE, IEEE), 1–5. doi: 10.1109/OCEANSE.2009.5278252

Steinberg, D. K., and Landry, M. R. (2017). Zooplankton and the ocean carbon cycle. *Annu. Rev. Mar. Sci.* 9 (1), 413–444. doi: 10.1146/annurev-marine-010814-015924

Suthers, I., Rissik, D., and Richardson, A. (2019). "Plankton: A guide to their ecology and monitoring for water quality," in *Plankton: A guide to their ecology and monitoring for water quality*. (Boca Raton, FL: CRC Press, CSIRO Publishing). doi: 10.1093/plankt/fbp102

Wang, Z., Liu, X., Huang, L., Chen, Y., Zhang, Y., Lin, Z., et al. (2022). QSFM: Model pruning based on quantified similarity between feature maps for AI on edge. *IEEE Internet Things J.* 9 (23). doi: 10.1109/jiot.2022.3190873

Wang, N., Yu, J., Yang, B., Zheng, H., and Zheng, B. (2020). Vision-based *In situ* monitoring of plankton size spectra *Via* a convolutional neural network. *IEEE J. Oceanic Eng.* 45 (2), 511–520. doi: 10.1109/joe.2018.2881387

Xu, J., Kong, Y., Jiang, Z., Gao, S., Xue, L., Liu, F., et al. (2019). Accelerating wavefront-sensing-based autofocusing using pixel reduction in spatial and frequency domains. *Appl. Opt.* 58 (11), 3003–3012. doi: 10.1364/AO.58.003003

Yang, Z., Chen, T., Li, J., and Sun, J. (2021). "Focusing evaluation for in situ darkfield imaging of marine plankton," in *OCEANS 2021: San Diego – Porto*. (Diego, CA, USA: IEEE), 1–8. doi: 10.23919/OCEANS44145.2021.9706029

Yim, J., Joo, D., Bae, J., and Kim, J. (2017). "A gift from knowledge distillation: Fast optimization, network minimization and transfer learning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Honolulu, HI, USA: IEEE), 4133–4141. doi: 10.1109/CVPR.2017.754

Zhanhui, Q., Rongjun, S., Ming, D., Tingting, H., and Honghui, H. (2020). A review on ecological of *Creseis acicula* and preliminary analysis on its outbreak triggers in daya bay. *J. Trop. Oceanography.* 40 (5), 147–152. doi: 10.11978/2020112

frontiers | Frontiers in Marine Science

TYPE Methods
PUBLISHED 02 February 2023
DOI 10.3389/fmars.2023.1117787

Check for updates

OPEN ACCESS

EDITED BY
Rizwan Ali Naqvi,
Sejong University, Republic of Korea

REVIEWED BY
Abdul Waheed,
Women University Swabi, Pakistan
Irfan Ahmad,
Khurasan University, Afghanistan
Ali Wagdy Mohamed,
Cairo University, Egypt

*CORRESPONDENCE
Salil Bharany
salil.bharany@gmail.com

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 06 December 2022
ACCEPTED 11 January 2023
PUBLISHED 02 February 2023

CITATION
Bharany S, Sharma S, Alsharabi N, Tag Eldin E and Ghamry NA (2023) Energy-efficient clustering protocol for underwater wireless sensor networks using optimized glowworm swarm optimization. *Front. Mar. Sci.* 10:1117787. doi: 10.3389/fmars.2023.1117787

COPYRIGHT
© 2023 Bharany, Sharma, Alsharabi, Tag Eldin and Ghamry. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Energy-efficient clustering protocol for underwater wireless sensor networks using optimized glowworm swarm optimization

Salil Bharany[1*], Sandeep Sharma[1], Naif Alsharabi[2,3], Elsayed Tag Eldin[4] and Nivin A. Ghamry[5]

[1]Department of Computer Engineering & Technology, Guru Nanak Dev University, Amritsar, India, [2]Department of Computer Engineering, College of Computer Science and Engineering, University of Hail, Hail, Saudi Arabia, [3]Computer Science, College of Engineering and Information Technology, Amran University, Amran, Yemen, [4]Faculty of Engineering and Technology, Future University in Egypt, New Cairo, Egypt, [5]Faculty of Computers and Artificial Intelligence, Cairo University, Giza, Egypt

In the past few decades, cutting-edge information and communication technology has been used in several ways to keep an eye on the marine environment. Underwater wireless sensor networks (UWSNs) can measure the amount of water and soil conditions, such as soil salinity, moisture, and movements, to predict landslides. UWSNs are made up of many wireless underwater sensor nodes (WSNs) that are spread out across the thalassic environment. These networks have several uses, including data collection, navigation, resource analysis, surveillance, disaster prediction, etc. Nowadays, energy efficiency becomes a complex issue to handle in the design of the UWSN due to the limited battery capacity and the challenges associated with changing or charging the integrated batteries. According to previous research, clustering and routing have already been effective methods of improving energy efficiency in the UWSN, as unreplaceable batteries and long-distance communication delays are particularly vulnerable. As a result, one of the UWSN's critical issues is determining how to extend the network's lifespan while balancing its energy consumption and shortening transmission distances. In UWSN clustering, the most important considerations are acquiring a suitable count of clusters, constituting the clusters, and picking the most satisfactory cluster head (CH) for each cluster. Based on several factors, such as residuary energy, total energy consumption, and other considerations, our proposed approach picks CHs and arranges them into clusters. Also, the proposed SS-GSO method constructs a fitness function by including various sources of information, like total energy, residual energy, and luciferin value. Several simulation runs were executed to test how much better the SS-GSO approach worked. The comparison results showed that while evaluating clustering time, our proposed SS-GSO technique performs 22.91%, 50.03%, 42.42%, 58.06% better, in case of Total energy consumption 27.02%,14%,33.76%,41.97% more energy efficient, in Cluster lifetime 9.2%,19.88%,35.91%,40.54% less and in Packet delivery rate 8.29%,14.05%,17.67%,23.97% better as compared with other heuristic techniques, such as ACO, GWO, MFO and LEACH.

KEYWORDS

UWSN, energy, GSO, routing, clustering, bio-inspired

1 Introduction

The oceans encompass more than three-quarters of the planet's surface and have focused on human concerns. The oceans are home to many animals and other living things, and humans rely on them for their primary sources of food, transportation, and other resources (Khan et al., 2019). The Internet of Things (IoT) has recently gained widespread acceptance as a viable paradigm that can revolutionize our culture and economy. As a result, it is possible to network together a wide variety of sensors, identifiers, processors, communicators, actuators, and networks. It is a network of dispersed sensors for physical and environmental monitoring that communicate wirelessly with one another.

On the other hand, big data is a new technology that has spawned a flurry of academic interest in data mining, machine learning, databases, and distributed computing. Smart homes, smart buildings, smart transportation, intelligent industrial automation, innovative healthcare, smart grids, and smart cities are examples of how IoT-assisted WSNs have been used in the past several decades. The monitoring and defense of marine ecosystems is another natural domain for IoT-based WSNs. As our civilization and economy have progressed, so has the marine environment become a focal point of study for academics and scientists. It will be more expensive when it comes to traditional monitoring of the maritime environment, including oceanographic and hydrographic research vessels. The data they collect and analyze could be better quality, and the processes involved take too much time. With UWSN, data processing is far more potent than within WSNs, allowing for more intelligent object control. Water temperature, pressure, wind direction, speed, salinity, turbidity, pH, oxygen density, and chlorophyll levels are just a few of the physical and chemical parameters that can be measured by sensors employed in an Internet of Things-based marine environment monitoring system. Depending on the level of sophistication, an IoT-based marine environment monitoring and protection system may be able to manipulate special equipment or devices in the marine ecosystem. The health of the oceans and the marine environment has gained more and more attention as it is a rich source of food and essential transportation.

Since the turn of the twentieth century, WSNs have become more prevalent in this research field (Bharany et al., 2021; Khan et al., 2021). Initially, wireless sensor networks (WSNs) were exclusively employed in terrestrial situations. Nevertheless, as shown in Figure 1, improvements in marine communication technology have made it possible for wireless sensor networks to work underwater. Compared to land-based situations, the implementation and maintenance of underwater wireless sensor networks are more complicated due to the environment (Khan et al., 2019; Bharany et al., 2021). Monitoring and communicating with an underwater environment are critical for various reasons like vast areas, salty water, and high-water pressure. The extent of these applications is broad and includes everything from temperature monitoring, undersea environments, marine ecology, natural disasters, and navigation to the investigation, surveillance, etc. In recent years, WSNs have been seen as an alternative way to study marine environments because they are easy to find, can be tracked quickly, and do not cost too much (Bharany et al., 2021).

1.1 Background

UWSNs were divided into sensor nodes (SNs) and data collection stations (collectively referred to as sinks) (Awan et al., 2019; Bharany

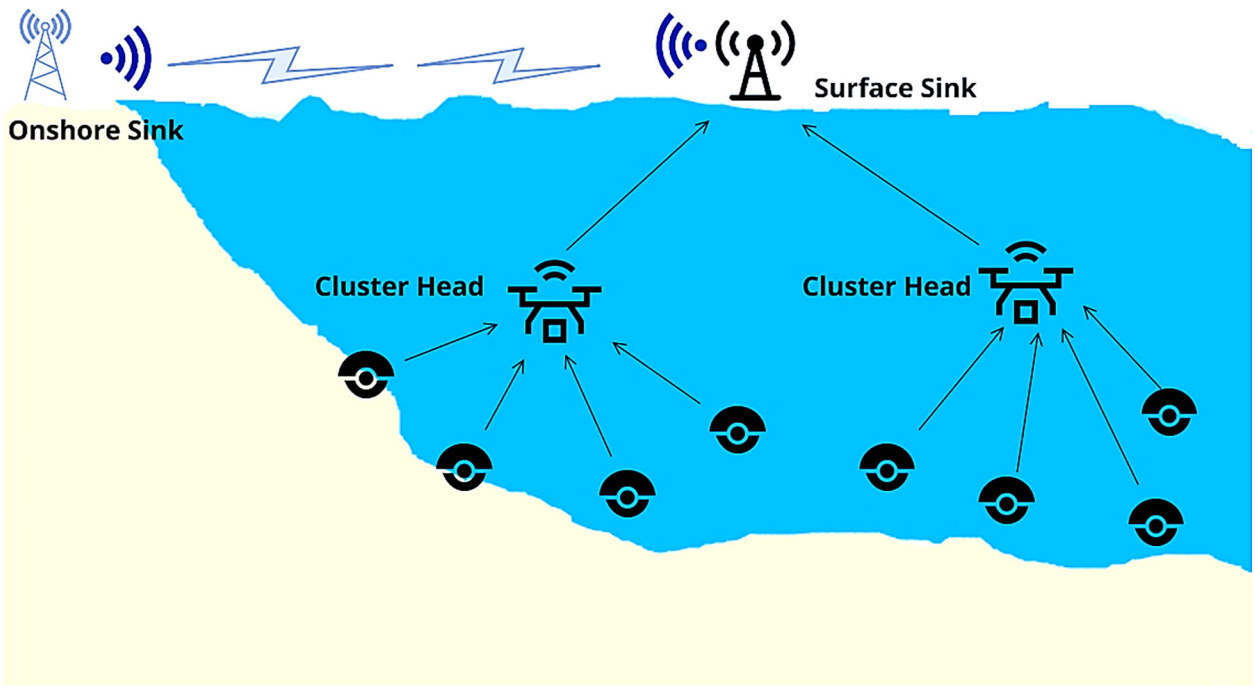


FIGURE 1
UWSN Structure.

et al., 2021; Khan et al., 2021). Sensor nodes are small devices that track and demonstrate physical characteristics and transmit the information to a base station using wireless communication protocols. Radiofrequency waves are not suggested in underwater sensor networks because their signal attenuation is excessive and limited in distance range (Awan et al., 2019; Khan et al., 2021; Nguyen et al., 2021). Due to the limited capacity and frequency of the UW channel, its bandwidth is significantly smaller than that of a conventional radio channel. Due to the high absorption, attenuation, and propagation, UWSNs have a restricted bandwidth, which piqued researchers' curiosity. Another important consideration is the failure of SNs, which might be due to environmental conditions or some other. Energy conservation is another major limitation of UWSNs that significantly impacts network longevity. There is no way to replace or recharge network nodes in the harsh environment of the UWSN, which means the network will run out of nodes (Bharany et al., 2021) with time. Tsunami predictions, offshore exploration, tactical surveillance, oil and gas spill monitoring, aided navigation, pollution monitoring, and several commercial uses are all possible with UWSNs, which are further visualized to enable a wide range of applications (Nguyen et al., 2021). Underwater devices need to communicate with each other optimally for these applications to work efficiently (Bharany et al., 2021). Differentiating from the terrestrial environment, the marine environment is far more complicated and variable, making ocean monitoring more challenging. A UWSN is an *ad-hoc* network designed for underwater communication and data transmission in which several surface stations are required for better communication (Awan et al., 2019; Bharany et al., 2021; Khan et al., 2021; Nguyen et al., 2021) so that Sensor nodes can easily transmit and receive data between a source and a destination.

Sensor nodes are responsible for three primary functions: data monitoring, processing, and reception of data. Each node is designed to sense and monitor a particular physical or chemical characteristic underwater and transmit the information gathered to a central point (Heinzelman et al., 2002). The following are some of the most significant issues related to undersea applications:

- First, there is a substantially longer propagation delay than in the terrestrial environment.
- Because of the multi-path fading issue, the channel has a dynamic character.
- Since channel characteristics change constantly, dynamic channel characteristics allow for high bit error rates and short disconnections.
- Batteries have a certain amount of power, and most are not rechargeable.
- Pollution and rust can cause underwater sensors to malfunction.

Electronic Magnetic waves at high frequencies suffer from considerable attenuation due to the water's saltiness. They cannot be used underwater because of their high frequency (Reddy and Khare, 2017). Several obstacles still make underwater communication more difficult, even after considering the mentioned challenges (Wan et al., 2018). One such issue is a need for more available bandwidth combined with significant energy usage. It has an immediate impact on the transmission rate of a network. UWSNs use energy quickly due to the current in the water and the many sources of noise present. Because of these issues, clustering in UWSNs has become problematic and expensive (Zhang et al., 2017). Developing an effective clustering procedure to handle the issues outlined above is necessary. The unique characteristics of UWSNs have attracted the attention of researchers interested in exploring their designs and networking protocols. Grounded on this interest, we developed an energy-efficient technique for clustering UWSNs that outperformed LEACH and other bio-inspired clustering algorithms (Heinzelman et al., 2002; Singh and Lobiyal, 2012; Rao et al., 2017; Alhazmi et al., 2018; Wan et al., 2018). Among the most well-known and energy-efficient clustering algorithms, those employed in terrestrial wireless sensor networks cannot deploy in the UWSN network. We also try to figure out why the WSN clustering protocols do not work well in the UWSN (Heinzelman et al., 2002).

According to the proposed technique, UWSN nodes will be separated into clusters using a dynamic cluster head selection process, as seen in Figure 2. In sensor network design, clustering organizes sensor nodes into logical groups called a cluster. Several nodes assemble information and transfer it to the central station as the Base station. Only one node in the relay is designated as the cluster head, while the leftover nodes are referred to as member nodes (Nguyen et al., 2021). The cluster head's responsibility is to ensure that all inter- and intra-cluster communication is appropriately coordinated. When choosing the cluster head for each node, the amount of energy left and other parameters of the nodes will be considered. Additionally, the distance between CH and BS may be considered in this decision. The cluster size and the cluster heads' election are already important parameters for decision-making (Kazmi et al., 2019).

Additionally, the correct number of clusters in UWSNs impacts the universal energy performance of the network. It will be necessary to select a new CH when the energy of the cluster head is less than a particular threshold value. Optimizing the glowworm swarm optimization (GSO) algorithm (Javadpour et al., 2018; Reddy and Khare, 2017; Rajakumar et al., 2017; Potthuri et al., 2018) is the basis for developing the proposed protocol. An evolutionary-based optimizer ensures enough clusters to keep the network running for a long time, resulting in fewer hops and lower energy usage.

1.2 Our contribution

The following are the most significant contributions made by this research:

- Improved communication in the underwater environment by developing an energy-efficient clustering technique for sensor nodes.
- UWSN's clustering can be optimized by using an evolutionary algorithm. These algorithms are built upon various insects' and animals' behavioral patterns to find the most efficient solution. These algorithms also give optimum clustering solutions. They have been utilized effectively in MANET, VANET, and FANET but not in the underwater environment. This study applies the optimized glowworm swarm

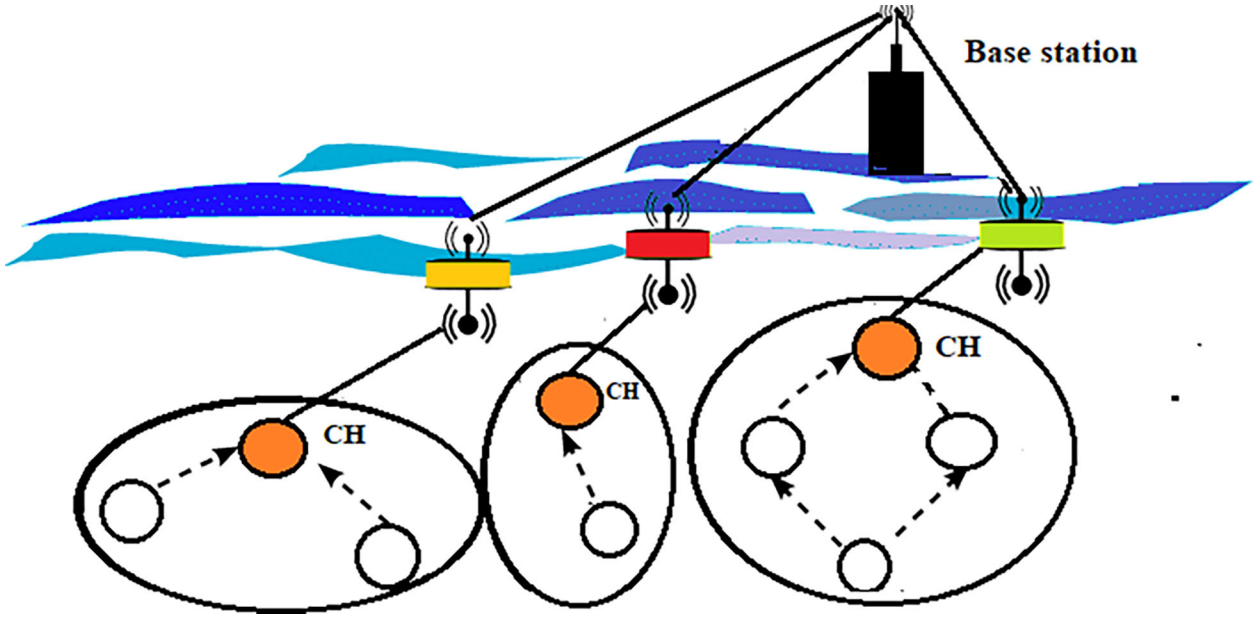


FIGURE 2
UWSN Clustering structure.

optimization approach, which includes various addition to the algorithm to survive in a new environment for clustering optimization to improve UWSN communication efficiency.

- This article compares the results of our proposed algorithm with other popular evolutionary algorithms based on various parameters. which shows that the proposed research is better than the other state-of-art.
- In the proposed method, we use a method to determine the correct number of clusters, choose the best cluster head, and use an aggregation method to get the best information to the base station with least redundancy.
- We build a new fitness function that considers luciferin value while considering residual energy and the entire network's energy usage. Afterward, the fitness function is used to choose cluster Heads.
- Experimental and statistical graphs demonstrate the suggested technique's effectiveness in cluster count consistency, energy consumption, packet transmission, and network lifetime.
- In the proposed algorithm, we utilize the TDMA table for the priority of allotting a slot to transfer data from CH to the base station, which is determined by the UAV's neighbor range, residual energy, and location, that will lead to reducing overhead for load balancing and congestion management purposes.

1.3 Related work

Underwater sensor networks have emerged as a new area of interest for academics looking for new and novel approaches. Several researchers have discovered their distinguishing characteristics for optimal network communication and have handled the setup challenges that arise (Javadpour et al., 2018). Compared to wireless sensor networks, the underwater medium presents an exceptionally high degree of unpredictability and difficulty. The most significant distinctions that make UWSN different from WSN are as follows: (1) amount of energy consumption; (2) propagation delay; (3) a low bandwidth; (4) dynamic topology operation; (5) propagation speed; (6) efficiency; (7) data transmission rate; and (8) noise interference. Several academics have developed a new energy-efficient clustering methodology for UWSNs (Nguyen et al., 2021). Because of the specific constraints of the environment of the underwater scene, the clustering technique developed for terrestrial networks cannot be easily used for UWSNs. Conserving energy is a critical goal for UWSNs. The clustering method is a well-known technology that has been successfully applied to cut down the energy usage of UWSNs (Bharany et al., 2021). A considerable amount of research has been done on clustering and the election of cluster heads to lower the energy consumption of the UWSN. However, there needs to be more research carried out on UWSN. This part addresses some of the research findings in even more detail.

The research in (Heinzelman et al., 2000) proposed a Low-Energy Adaptive Clustering Hierarchy as a low-energy replacement for traditional clustering techniques. Selecting cluster leaders randomly and cycling between them is a common tactic in hierarchical networks to ensure fair energy distribution. LEACH considers the nodes' available data, including their residual energy, communication energy usage, and the number of neighboring nodes, to select the most energy-efficient CH nodes. The LEACH-C algorithm is a significant improvement over its forerunner. Sending information about the sensor nodes' locations and remaining battery life to the base station can increase the system's overall efficiency. Centralized deployment of LEACH-C lowers the network's total energy

consumption and each node's energy consumption. However, LEACH-C also has some problems. If we take the LEACH-C network as an example, every node has an equal opportunity to become a CH. If the network runs out of power, the least powerful nodes will have to act as CHs, reducing the network's overall efficiency and placing an undue burden on the system. To create an energy-conscious CH selection approach, Singh et al. (Singh and Lobiyal, 2012) employed PSO for clustering the nodes. The fitness function is derived by multiplying the fitness metric by the sum of the residual energy, the distance, and the node density. This approach requires consideration of inefficient cluster formation, which drains the network's resources. In order to solve the issue brought up in (Heinzelman et al., 2002; Singh and Lobiyal, 2012), Rao et al. (2017) presented a PSO-based energy-efficient cluster head selection (PSOECHS) that includes journey time to the base station (BS) in its fitness function.

PSOECHS also affects the CH selection mechanism by changing the total number of nodes in each cluster, which extends the network's lifespan. Contrarily, computing and selecting CHs at each node consumes much more power than that. Known as UMOD-LEACH (Alhazmi et al., 2018), it is a highly-improved variant of the Leach protocol. When applied to water, this technique outperforms LEACH by 30%. It uses localization and time-division multiple-access (TDMA). According to (Wan et al., 2018), the Adaptive Clustering Underwater Network (ACUN) for underwater networks was a successor to the aforementioned adaptive clustering routing system. The amount of remaining energy at the CH node and the distance between the CH node and the sending node are the two primary factors in multipath routing.

Zhang et al. (2017) developed a clustering scheme for a medium-sized UWSN using discrete PSO and a genetic algorithm (GA) to increase the network's durability. However, the methodology is not advised because of its low stability, and the clustering model needs to be more precise to use UWSN. Pengwei Li et al. (2017) created an improved particle swarm optimization technique for clustering UWSNs that considers individual nodes' transmission power needs. However, fixing the issues from the previous study and having the network last even longer is challenging because particle coding is much more complex for sensor nodes. The Energy Center Searching model employing PSO (EC-PSO) was proposed in the literature (Wang et al., 2019) to prevent these energy gaps and to search for energy centers for CHs selection. Extra measures were taken to ensure that no two CHs would ever be close to one another. A safeguard with a threshold value was implemented to prevent low-energy nodes from delivering data. However, because the distance between nodes, CHs, and the sink node is not considered while making the fitness function, it affects how much energy the system uses.

The fuzzy enable clustering strategy was used to improve the routing method. In the literature (Alia, 2014), it has been suggested that decentralized fuzzy C-means are used as part of an energy-efficient routing protocol to reduce the total amount of energy a network uses while also making it last longer. CHs are chosen iterative through each cluster one at a time. Bhatti et al. (2016) developed the fuzzy c-means clustering and energy-efficient CH selection mechanism. The FCM method makes clusters, and CHs are chosen based on the SNs' location, signal-to-noise ratio, and residual energy.

On the other hand, the sensor network can only be built in a two-dimensional space in terms of its topology. Researchers who worked with wireless sensor networks came up with a clustered routing method called BECR that used energy in a balanced way (Zhao et al., 2018). The first cluster head chosen was the node closest to the group's center. When the energy level of a cluster head fell below 20%, the FCM was used to choose a new cluster head for that cluster. A fresh round of cluster head elections was performed if the energy level of each cluster head declined by 10%; otherwise, no elections for cluster heads were performed. The randomness feature makes it difficult to avoid falling into the trap of local optimality while employing this technique. So far, a variety of researchers have used the hybrid strategy, which has the advantage of resolving the weaknesses of one methodology while simultaneously incorporating the benefits of another. To generate clusters, the literature (Sharma et al., 2019) utilized FCM and selected the best node to act as a CH. It used a fuzzy inference technique to determine how well each node performed. This technique was used in conjunction with an evolutionary method to choose the CH to make better decisions. The literature (Shokouhifar and Jalali, 2017) reported using an artificial bee colony approach to update the fuzzy rules of LEACH-SF, which was subsequently corrected. It was made to double the network's lifetime by following the rules, and the algorithm's fitness function makes this happen.

We have widely deployed sophisticated algorithms to ensure efficient data routing in WSNs (such as the PSO, GWO, and MFO). However, now we are considering these techniques for UWSNs. In (Javadpour et al., [[NoYear]]), fuzzy clustering was used to connect sensors in a network, and the initial value of the cluster heads was estimated using the PSO technique. It started by clustering using the fuzzy technique and then selected the best cluster head by applying the PSO algorithm to the fuzzy table created by the fuzzy approach. The algorithm had a considerable influence on reducing the amount of energy consumed. Afterward, various clustering methods were proposed by taking this as a base, including differential evolution and simulated annealing (Potthuri et al., 2018), fuzzy c-means, and the Genetic algorithm (Reddy and Khare, 2017). An optimization approach known as the Grey Wolf Optimization Method (GWO) is also used in many UWSNs to optimize clustering as an optimization methodology. GWO-LPWSN (Rajakumar et al., 2017) employs the grey wolf optimization technique to find nodes and fix placement faults. For wireless sensor networks (WSNs), the proposed method (Kazmi et al., 2019) offers a GWO-based transmission rate management mechanism. The grey wolf optimizer is used to fine-tune the support vector machine's (SVM) transmission rate, which controls the rate data is sent between nodes. In particular, a cluster head is chosen by the sink node, and the optimum path for data transmission is determined by the nodes using GWO (Al-Aboody and Al-Raweshidy, 2016).

Similarly, optimal grouping is achieved with the help of grey wolf optimization and evolutionary algorithms (Lipare et al., 2020). The moth flame optimizer is only one of several innovative algorithms that can be used to solve optimization problems. Taking cues from moths in the wild can help find optimal solutions to NP-complete problems. Optimal clustering is the goal of a method proposed in (Mittal, 2018) that uses the MFO algorithm. Moth-flame and genetic algorithms are

used for efficient and effective clustering (Sharma et al., 2020). Due to its high-quality results, it can also be implemented in UWSNs (Bharany et al., 2022). Another routing protocol designed for innovative ocean applications is SOSNET (Durrani et al., 2019). The moth flame optimizer (MFO) is employed to efficiently perform optimal clustering at a low cost (Patel et al., 2020; Bharany et al., 2022). Similarly, the technique proposed in (Kumari et al., 2019) addresses the problem of fault-tolerant routing by employing MFO, which determines the optimal data transmission path (Namasudra and Roy, 2018; Fattah et al., 2020; Bharany et al., 2022). A comparison of algorithms in Table 1 shows how they deal with different problems and have different goals.

2 Methods

While using research into the collective optimization behavior of glowworm swarms, this section suggests a self-organizational approach to clustering. Using this strategy, UWSNs can manage

TABLE 1 Comparison of various algorithms.

Protocol	Category	Objective	Issues
LEACH (Heinzelman et al., 2000)	probabilistic	reducing energy usage	When selecting CHs, it does not take into account the information about the node, resulting in network energy imbalance.
LEACH-C (Heinzelman et al., 2002)	probabilistic	increasing the lifespan of a network	Inefficient in energy consumption and uneven energy load for equal possibility to select the CHs
EACHS-PSO (Singh and Lobiyal, 2012)	Heuristic	Selecting an energy-efficient CH to increase the network's lifespan	No cluster formation process is involved
PSOECHS (Rao et al., 2017)	Heuristic	Energy efficient Cluster formation and elects the CHs for energy efficiency	energy dissipation increases while CH selection process
UMOD-LEACH (Alhazmi et al., 2018).	probabilistic	Less energy usage	Clustering is not so much optimal
ACUN (Wan et al., 2018)	Heuristic	To reduce the energy usage of the network	Complex processor for clustering
RCA-UASN (Zhang et al., 2017)	Heuristic	Process of clustering in energy-efficient way for medium-sized scale UWSNs	Lack of stability of algorithm and not fits into the environment of UWSN
IPSOC-UASN[13	Heuristic	Optimizing the network lifespan using improved version of PSO	Highly complex particle coding for SNs
EC-PSO (Wang et al., 2019)	Heuristic	for selecting and searching energy efficient CHs and avoiding energy holes	fitness function ignores the distance between SNs, CHs and sink which leads to lack of efficiency
DFCMER (Alia, 2014)	Non-deterministic	Increase of network lifetime using FCM protocol	No optimization techniques involved and Ch selection process is executed locally
FCM-EECHS (Bhatti et al., 2016)	Non-deterministic	Energy efficient clustering	Suitable for only 2-D space
BECR (Zhao et al., 2018)	Non-deterministic	Make an even balance between energy consumption and network longevity	Easily slips to local optimality issue
EEFCM-DE (Sharma et al., 2019)	Hybrid	Energy efficient clustering and better throughput	Consumption of energy increase is we apply to all SNs
OFC (Shokouhifar and Jalali, 2017)	Hybrid	Balance clustering technique and increase network lifetime	Only for static environment not suitable for dynamic
CFC-PSO (Javadpour et al., 2018)	Hybrid	Optimal CHs selection in energy efficient way	Initial CH is selected randomly and cluster count is fixed
DESA (Potthuri et al., 2018)	Hybrid	Efficient clustering process	No distance between nodes is consider for fitness function
GWO-LPWSN (Rajakumar et al., 2017)	Heuristic	spot the correct position of unknown nodes	Suitable for static and fixed number of SNs
GWO-SVM (Kazmi et al., 2019)	Hybrid	control the congestion	Complex in execution
MLHP-GWO (Al-Aboody and Al-Raweshidy, 2016)	Heuristic	Improve efficiency, lifetime, and stability period.	High energy consumption
TECP-MFO[(Mittal, 2018)	Heuristic	load balancing	Not efficient for UWSN
SOSNET (Durrani et al., 2019)	Heuristic	To find optimal number of clusters required for routing	In some case results does not perform in an optimal way
FCM-PSO (Krishnaswamy and Manvi, 2019)	Hybrid	To avoid the early dying of CH node	While selecting CH convergence speed is very low

their networks better and have more reliable communications. GSO analyses the area around the glowworm to calculate its luciferin value. The fundamental justification for using GSO for an underwater WSN (Bharany et al., 2022) is its ability to provide an appropriate optimization solution in the face of a glowworm's variable luciferin value. This GSO feature could be helpful for a UWSN built on a cluster architecture. After that, we pick the shortest path between the Sensor nodes and the BS. Each node's residual energy and the average residual energy are determined as the first steps in our suggested algorithm. A node is not considered for CH selection if its residual energy is lower than the mean residual energy. Afterward, the fitness function is determined with the help of methodical equations, which will be detailed in the following section and used in selecting CHs. After then, information travels from CH to BS *via* a TDMA slot. The XOR operation is used during an aggregation process to eliminate duplicate information. Figure 3 depicts the sequence of events before data is transmitted to BS. Therefore, our method is divided into three parts.

2.1 Overview of mathematical representation of GSO

The GSO algorithm assigns a luciferin value to each glowworm and a neighborhood range for making decisions locally. The objective function and its position establish the luciferin value of a glowworm. Compared to other glowworms, a higher luciferin value indicates that the glowworm is located in an exceedingly bright area. You can see the entire SS-GSO process in Figure 3, which is the algorithm.

The following equation is used to keep track of the glowworm's Luciferin value:

$$Luci_i(t + 1) = (1 - \rho)Li(t) + \gamma\ F\ (pi(t)) \quad (1)$$

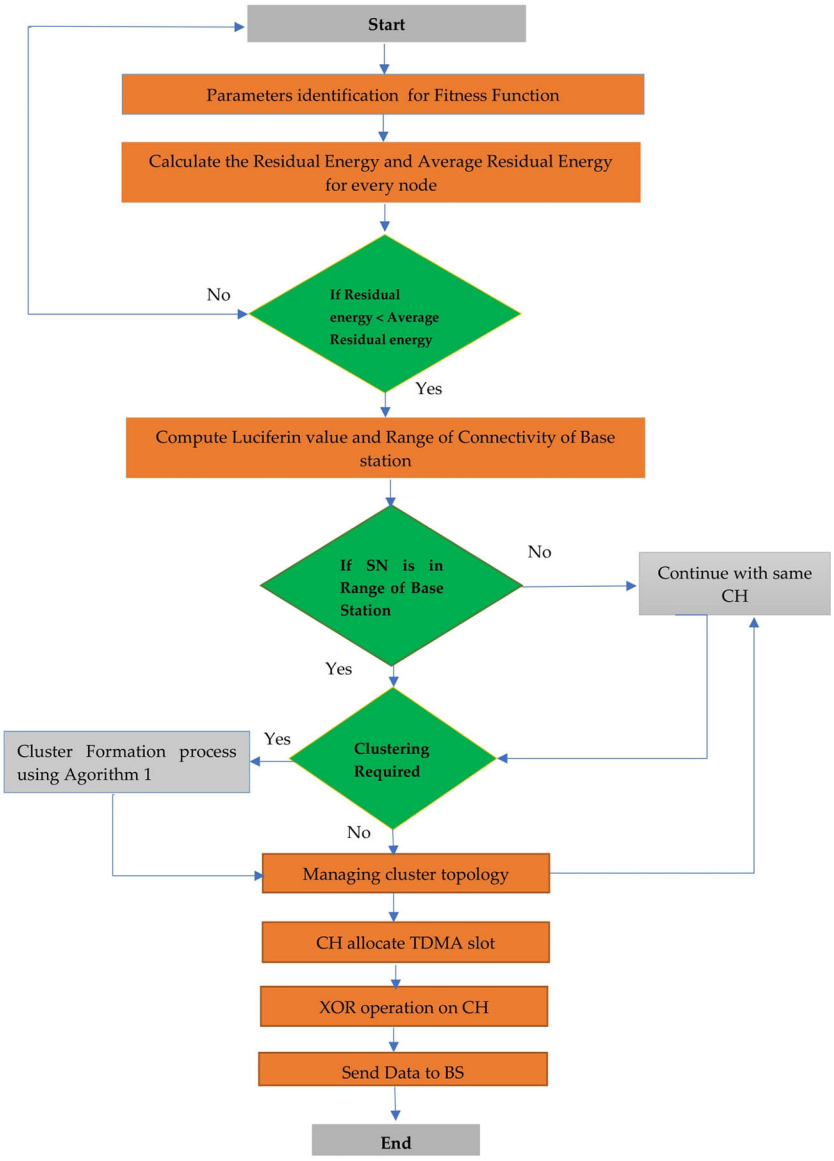


FIGURE 3
Flowchart for proposed algorithm.

With respect to each individual glowworm i, $Luci_i(t)$ symbolizes the luciferin value of that glowworm. In addition, the luciferin decay constant, announced by ρ which ranges from zero to one, the luciferin enhancement fraction, which is acted by γ, and $F\ (p_i(t))$ which represents the objective function, given for current position p_i of glowworm i.

Then, using the following rule, each glowworm i investigates its immediate surroundings in search of the neighbour with the greatest luciferin value:

$$z \in Ni(t)\ \ if\ \ f\ \ Dist_{iz}\ \ < rang_i(t)\ \ and \quad (2)$$

$$Luci_z\ (t) > Luci_i(t)$$

Ni(t) is the group of glowworms surrounded by glowworm i, while z is i's nearest neighbor. Glowworm i's local decision range is denoted by $rangi(t)$, and the luciferin concentrations of glowworm z and i is represented by $Luciz\ (t)$ and $Luci\ i(t)$, respectively. Using the probabilities of each neighboring glowworm, as shown in Equation 3, a best-neighbor glowworm is selected.

$$\mathbf{Prob_{iz}} = \frac{\mathbf{Luci(t) - Luci_i(t)}}{\sum_{\mathbf{x}} \in \mathbf{N_i(t)}\ \ \mathbf{L_x(t) - L_i(t)}} \quad (3)$$

The location of the glowworm is updated in accordance with the best position of an adjacent glowworm that has been identified and computed as follows:

$$\mathrm{p}_i(\mathrm{t}=1) = \mathrm{p_i(t)} + s\frac{p_z\ (t) - p_i(t)}{Dist_{iz}} \quad (4)$$

where s > 0 denotes the distance travelled by a glowworm in order to reach another glowworm. The decision range $randec_i(t)$ is computed by the following equation:

$$randec_i\ (t\ +\ 1)$$

$$=\ min\{rad_s,\ max[0,\ randec_i\ (t)\ +\ \beta(nt\ -\ |Ni(t)|)]\} \quad (5)$$

Where $rads$ is the constant representing the radial sensor range, β represents the model constant, and n_t controls the maximum allowed number of neighbors.

2.2 CH selection and cluster establishment

In our proposed clustering strategy the selection of CH is considered based on connectivity with the base station and fitness function. Fitness function is contingent the luciferin value, total energy consumption, and residual energy value of the UWSN to ensure effective communication and data transfer (Khan et al., 2019).

The fitness functions are represented by the following equation:

$$\text{Fitness function} = \mathrm{w_1} \times \mathrm{f_1} + \mathrm{w_2} \times \mathrm{f_2} + \mathrm{f_3} \times \mathrm{w_3} \quad (6)$$

where w_1, w_2, $w_3 = 0.5$

First of all, we calculate the residual energy of the every UWSN node by:

$$ReSEi\ \ =\ \ (initialE_i\ \ -\ \ CurrentE_i) \quad (7)$$

The residual energy is represented by ReSEi, the starting energy of the ith UWSN is represented by initialEi, and the current energy level of the ith UWSN is represented by $CurrentE_i$. After calculating residual energy of every node, we make an average of all nodes residual energy and then compare the residual energy of a node with average energy of all nodes. If current energy of a node is less than the average energy then this node is being skipped for the current round of CH selection.

Total amount of transmission and reception energy of a node is calculated by:

$$f_1 =\ \ Total_{Econ_i}\ \ =\ \ TranB_i\ \times\ \ Engy_{Tran}\ +\ \ RecB_i\ \times\ \ Engy_{Rec} \quad (8)$$

The number of transmitted bits and received bits in node i are denoted by the $TranB_i$ and $RecB_i$, respectively. $Engy_{Tran}$ and $Engy_{Rec}$ are characterized as transmission energy and reception energy correspondingly are calculated as:

$$\mathrm{Engy_{Tran}} = \text{Energy for radio signal transmitter} + \mathrm{E_A} * \mathrm{Dist^2} \quad (9)$$

ET_radio and ER_radio are the energy needed for radio transmission by the transmitter and the receiver, respectively. EA is the energy needed to boost(amplify) a signal, and Dist [Khan et al., 2019; Bharany et al., 2021] is the distance between two nodes, m and n.

The function for luciferin value is given by:

$$f_2 = Luci_i(t+1)by\mathbf{Eq}.(\mathbf{1}) \quad (10)$$

$$f_3 =\ \ ReSEi \quad (11)$$

In our proposed algorithm, Eq. 6 is used to calculate the fitness of each UWSN in the cluster formulation mechanism, as seen in Figure 4. The weighted sum of the luciferin value and the total energy spent and residual energy from Equations. 8, 10, and 11 is used to compute the fitness. In addition to its fitness, each UWSN sends a "Hello Message" along with its fitness to all the other nodes. When the SNs get a Hello message, the received fitness is compared by the nodes with its own. When a "Hello" message is received, the UWSN creates and updates a neighbour table using the UWSN's entries. The UWSN then sorts the neighbour table by decreasing fitness values. The closest node to the base station is the "Cluster Head". This would be the best place for CH. However, when there is just one node in the range of the base station or more than one, the UWSN with the most significant fitness value is designated as the cluster head and sends a Cluster Formation message to all nearby nodes. By sending a cluster-joining message from all nodes with lower fitness values, they can recognize their CH. The UWSN coalition of *ad-hoc* nodes is created on-demand when a UWSN wants to transfer data but is out of range of the base station. In order to get the information to its final destination, the Nodes use an intermediate UWSN to make many hops (Khan et al., 2019; Bharany et al., 2021; Khan et al., 2021). The remainder of the UWSN nodes will become superficial nodes or cluster members after the election of the CH, which will be purely based on a fitness function. The Luciferin level values of each glowworm in each cluster are updated according to their position and fitness function. Members of the cluster must keep track of the CH's movement and alter their locations accordingly. The CH keeps track of the cluster's topology by receiving location data from all UWSNs and updating the topology

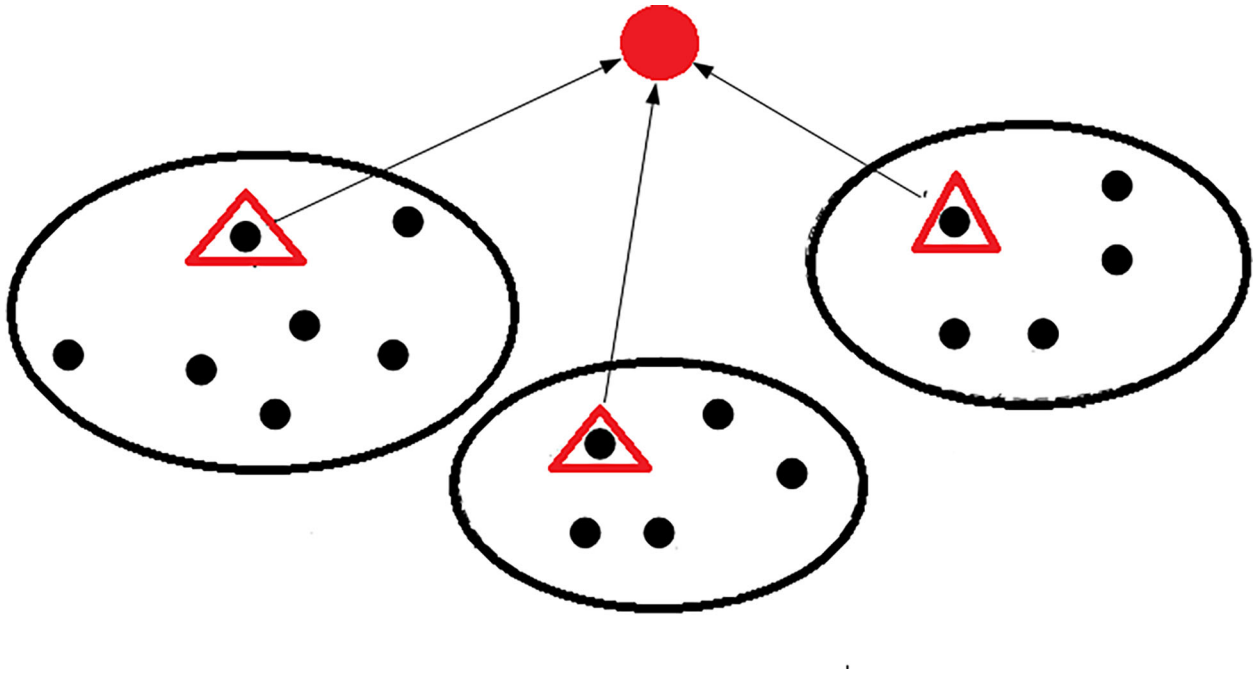


FIGURE 4
Clustering Basic structure.

database accordingly. Suppose a UWSN travels out of the cluster's range based on its updated position (as determined by the luciferin value). In that case, it is no longer considered a part of the cluster. In the next step, the CH transmits the updated topology table to the cluster members and maintains the cluster (Bharany et al., 2021). This method also includes a technique for managing the topology of clusters, which is described in detail below. Using a topology configuration message, each CM communicates its luciferin value to the other CM UWSNs. Upon receipt of the topology configuration message, the cluster head of UWSN modifies the location of the cluster member nodes following the luciferin values stored in the cluster topology table. Once the position information for each CM of UWSN has been updated, the cluster head sends this cluster topology table to all the cluster members. This way, all the CM behave like a swarm and move to keep up with the CH.

2.3 Optimal path selection and UWSN communication

The best path selection for information transmission in an efficient routing mechanism result in lower energy consumption in the network and enhances the cluster's lifetime. Every UWSN node in the proposed system follows the location of CH and adjusts its position in response to that of CH. Once the sensor nodes are in clusters, each group's head node (CH) makes a TDMA schedule and sends it to the other cluster members. In order to save energy, each node sends its data to the cluster head during the time slot defined in the TDMA table and then switches off its radio transmitter while waiting for the next transmission slot as stated in the TDMA table. In the TDMA table, the priority of allotting a slot is determined by the UAV's neighbor range, residual energy, and location, which are used for load balancing and congestion management purposes. When allocating TDMA slots, the UWSN in the neighborhood range of the source UWSN (close to the source UAV) and having the lowest residual energy will be prioritized over the other UWSNs, as this would be necessary to retrieve the data from this node. Therefore, a node with low residual energy will be assigned to the first available TDMA slot before any other node with high residual energy. On the other hand, the cluster heads keep their radios turned on to receive data from cluster members, and they perform an XOR operation on that data to reduce the amount of duplicate data. As soon as a cluster head gets data from its members, it performs data aggregation and sends it to the surface station near the water's edge. As the primary base station on the ground is typically located a long distance away, employing these strategies will prevent a high-energy transmission since sending it straight to the primary base station on the ground would result in high energy consumption and a high risk of data loss. This method sends information through the surface station to the primary base station to save energy and transmission costs.

3 Results

This section investigates the performance of the SS-GSO method, utilizing other modern approaches in a range of scenarios. Table 2 has some simulation parameters considered in our experimentation setup.

3.1 Network lifetime

In Figure 5, we can see the number of alive nodes (NAN) analysis of the SS-GSO method compared to older methods is done. The

TABLE 2 Simulation parameters.

Parameters	SS-GSO	ACO	GWO	MFO	LEACH
Number of UWSN	20–200	20–200	20–200	20–200	20–200
Max Iterations	100	100	100	100	100
Simulation Runs	20	20	20	20	20
Nodes Positions	Fixed	Fixed	Fixed	Fixed	Fixed
Simulation Area	500 × 2000 m^2	500 × 2000 m^2	500 × 2000 m^2	500 × 2000 m^2	500 × 2000 m^2
Mobility	Static	Static	Static	Static	Static
Inertia Weight	0.694	0.694	0.694	0.694	0.694
Transmission Range	Fixed	Fixed	Fixed	Fixed	Fixed
Distance between Nodes(max)	5 m	5 m	5 m	5 m	5 m
Distance between Nodes(min)	2m	2m	2m	2m	2m
W_1, W_2, W_3	0.5	0.5	0.5	0.5	0.5

experiment was carried out using the 2018a version of MATLAB programming, and it was completed on a 7th generation core i5 machine with DDR4 8 GB of RAM. An experimental grid size ranging from 500 meters to 2000 meters was used for this investigation. The number of nodes used ranged from 0 to 300, and each node's transmission range was between 25 m and 200 m. Nodes are made to stay in the same place or move very slowly so they do not get in the water flow (Bharany et al., 2022). It was equated to different state-of-the-art evolutionary clustering algorithms, such as the ant colony optimizer (ACO), the grey wolf optimizer (GWO), the moth flame optimizer (MFO), and LEACH. In Table 2, we can see all the simulation metrics. The experimental results suggest that the low-energy adaptive clustering hierarchy (LEACH) technique delivers poor outputs with the least feasible NAN. The MFO technique obtained a more significant NAN value than the LEACH procedure. Accordingly, the ACO techniques do not produce a favorable outcome either. While the GWO approach sought to create a respectable NAN compared to the other, Figure 5 revealed that the SS-GWO strategy outscored them all in terms of NAN. We can also see in the figure that the LND (the last node died) is best for our proposed algorithms. In Table 3, we can see the results of the network node analysis of all the techniques.

3.2 Clustering time

It is common practice to refer to the computational complexity of a clustering algorithm as the time it takes to complete the clustering process using that approach. Typically, clustering algorithms input a set of nodes and associated fitness ratings and output a subset of CH nodes and other related nodes. The time it takes for an algorithm to go from receiving inputs to generating outputs is known as its "cluster

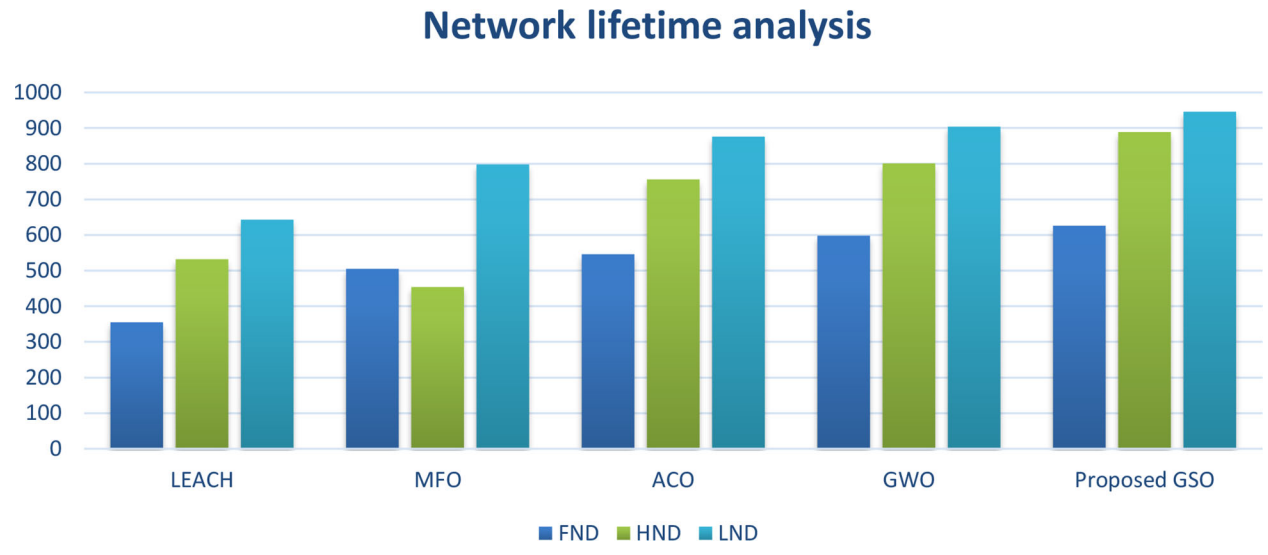


FIGURE 5
Network lifeline analysis.

TABLE 3 Network lifespan analysis of various algorithms.

Methods	FND	HND	LND
LEACH	355	532	643
MFO	515	724	804
ACO	578	784	898
GWO	599	802	904
PROPOSED SS-GSO	704	899	987

building time." Since the UWSN has little memory and processing power, a longer cluster formation time will harm WSN performance. Increasing the time required to create a cluster uses more energy, reducing the time SNs in a network can remain in operation. Figure 6 shows that our proposed SS-GSO performs better than LEACH, ACO, MFO, and GWO because, in contrast to ACO and GWO, which begin with several solutions and converge to the optimal through iteration, our SS-GSO always quickly converges towards the best solution. Using the SS-GSO method, the time it takes to build a cluster and the time it takes to choose a path is kept to a minimum. This makes it possible for SNs to do complex calculations with much less energy.

3.3 Total energy consumption

It is shown in Figure 7 that our suggested method, SS-GSO, compares favorably to other strategies in terms of total energy consumption (TEC). The LEACH procedure could have been more efficient than alternative techniques with a maximum TEC. It was also seen in Figure 7 that the MFO approach had a lower TEC than the LEACH method. Regarding TECs, MFO approaches were slightly better throughout several iterations than the LEACH method. It is worth noting that while the GSO performed better in all other elevation parameters, in TEC, the ACO strategy had the second-lowest TEC and surpassed the GWO strategy. Finally, our projected SS-GSO executes better than all other techniques (He et al., 2016; Bharany et al., 2022), as only ACO competes with SS-GSO.

3.4 Cluster head count

The number of cluster head nodes substantially influences the efficiency of the protocol, as there will be an increase in node energy consumption and overall energy consumption. As a result, if the count of CHs is low, the number of SNs per CH will be high, and the time of SN data transmission will be affected (Okoth et al., 2022; Srivastava et al., 2022). The total energy consumed by each round of networks increases and is directly proportional to the number of CH. Figure 8 shows that our proposed SS-GSO clustering protocols', outperform LEACH, ACO, GWO, and MFO protocols in headcount fluctuation frequency. In this case, we are considering only 1000 rounds for evaluation. It is simple to see that the cluster headcount in the LEACH protocol swings between $4 \leq T \leq 17$ and $5 \leq T \leq 14$ in GWO, $6 \leq T \leq 14$ in ACO, and $3 \leq T \leq 17$ in MFO, and that the cluster headcount in our suggested protocol swings between $3 \leq k \leq 13$ which is the most desirable.

3.5 Number of packets received

This section compares the number of packets received (NOPR) analysis of the SS-GSO technique to other methods in various iterations. The experimental findings indicated that the LEACH strategy produced inferior results with a lower NOPR than the conventional approach. The MFO approach obtained a NOPR marginally more significant than the LEACH procedure. After that, ACO approach rounds resulted in a NOPR that was only marginally closer to the original last NOPR. While the GWO method was trying to get a reasonable NOPR, the SS-GSO

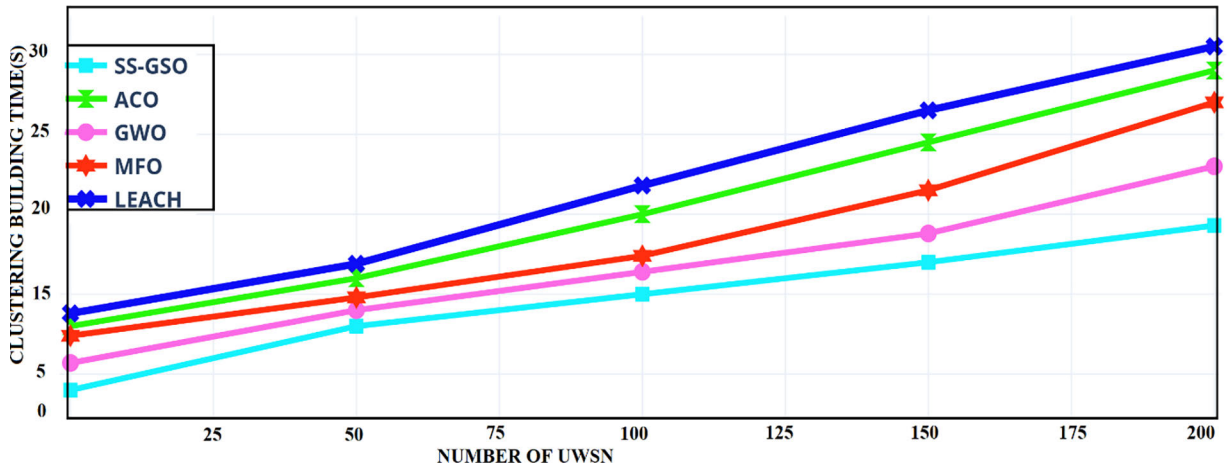


FIGURE 6
Clustering time Vs number of USWN.

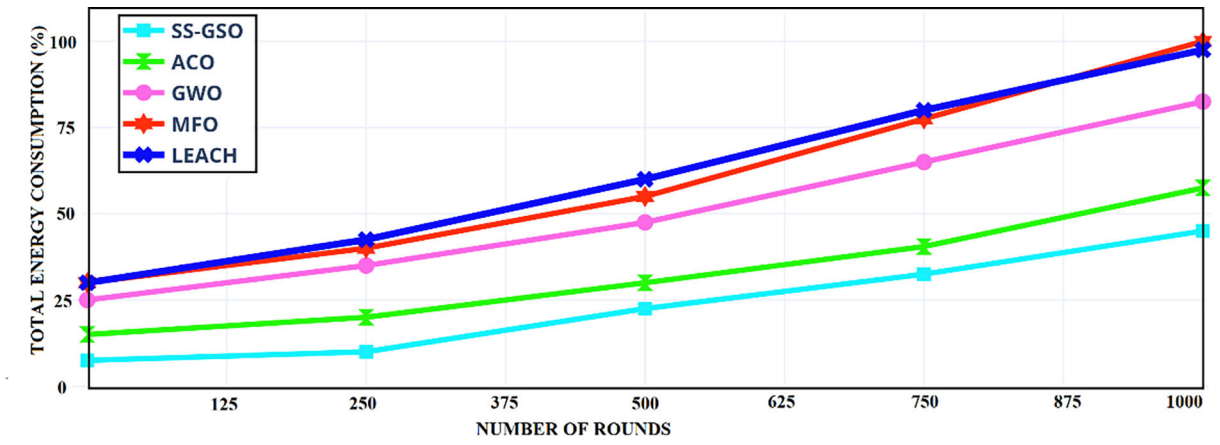


FIGURE 7
Total energy consumption by USWN.

method could get a higher NOPR than other methods (Mohamed et al., 2019; Arshad et al., 2020; Asif et al., 2020; Mehmood et al., 2020; Ahmad et al., 2022; Mazhar et al., 2022; Zeb et al., 2022).

As seen in Figure 9 and Table 4, the SS-GSO approach is a good strategy for maximizing energy efficiency and longevity in the UWSN context. Overall, the results of the experiments showed that the LEACH method performed poorly across all comparison criteria (Mohamed and Mohamed, 2017; Mohamed and Suganthan, 2017; Ahmad et al., 2021; Mehmood et al., 2021; Saleem et al., 2021; Ahmad et al., 2022). Finally, our proposed method has performed better than all other techniques.

4 Discussion

From the information shown in the Figure 10 tables and figures above, it is clear that the SS-GSO approach is very effective at getting the highest possible energy efficiency and durability in the UWSN context. Based on the initial results, the number of packets received (NOPR) analysis of the SS-PSO technique is compared to that of other methods in multiple iterations. The experimental data showed that the LEACH method performed poorly in all rounds, with a lower NOPR, the longest time to build clusters, and the most energy used. After that, the NOPR for the GWO model was higher than that for the LEACH method. In all

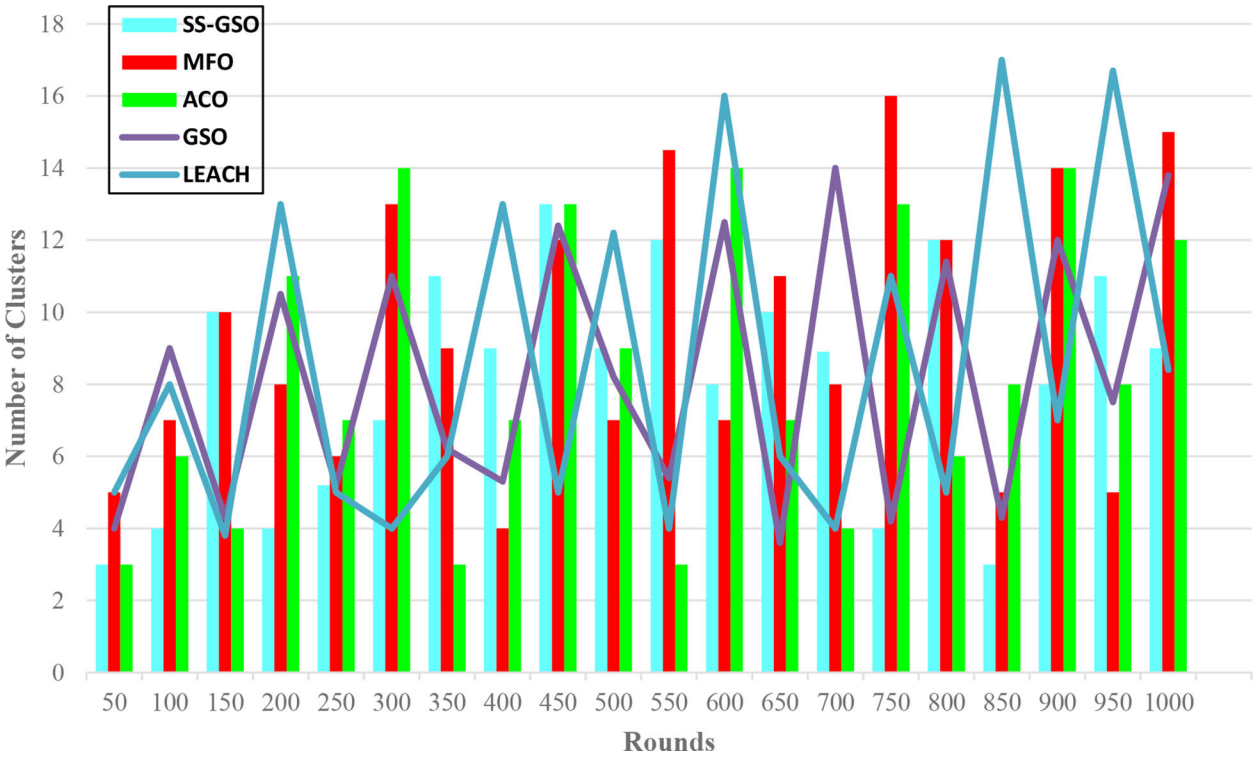


FIGURE 8
consistency of USWN cluster head cou.

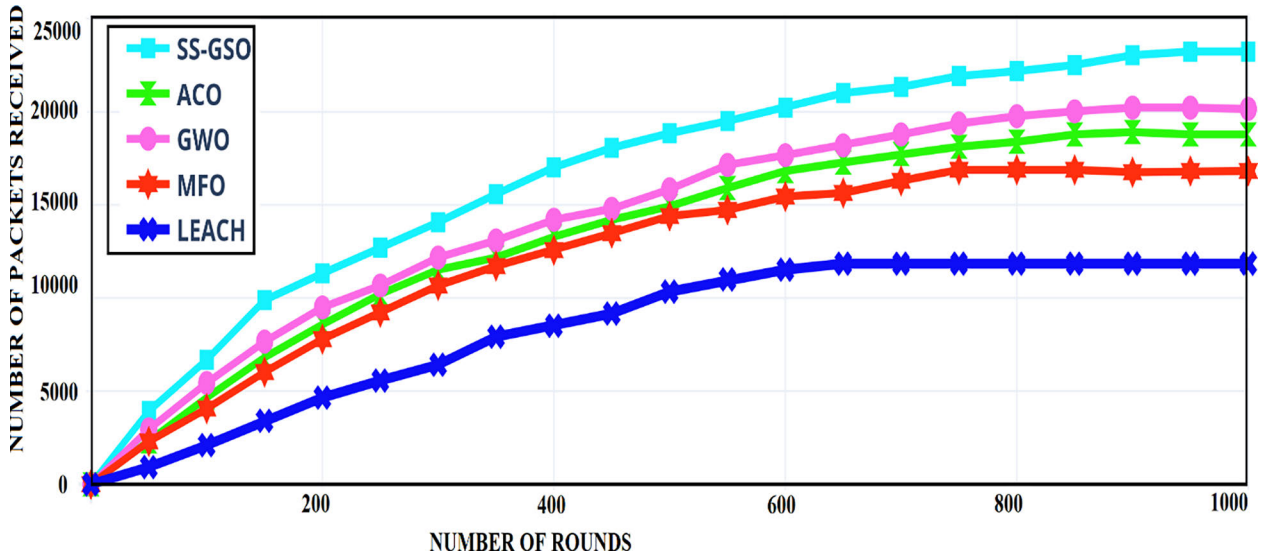


FIGURE 9
Total packet received successful.

TABLE 4 Number of packets received analysis.

No. of Rounds	LEACH	MFO	ACO	GWO	Proposed SS-GSO
0	0	0	0	0	0
50	945	2302	2312	2951	3989
100	2215	4175	4756	5512	6765
150	3548	6245	6901	7785	9912
200	4846	7845	8645	9501	11,402
250	5894	9345	10,456	10,745	12,801
300	6589	10,785	11,715	12,223	14,156
350	8012	11,845	12,256	13,189	15,645
400	8614	12,678	13,313	14,312	17,156
450	9234	13,745	14,345	14,856	18,178
500	10,545	14,512	15,001	15,956	18,934
550	11,012	14,845	15,945	17,238	19,536
600	11,522	15,678	16,856	17,759	20,400
650	11,945	15,785	17,345	18,223	21,102
700	11,995	16,345	17,786	18,978	21,512
750	11,995	16,915	18,245	19,450	21,989
800	11,995	16,989	18,465	19,885	22,307
850	11,995	16,887	18,856	20,156	22,645
900	11,995	16,989	18,923	20,148	23,185
950	11,995	16,989	18,779	20,148	23,348
1000	11,995	16,989	18,779	20,139	23,234

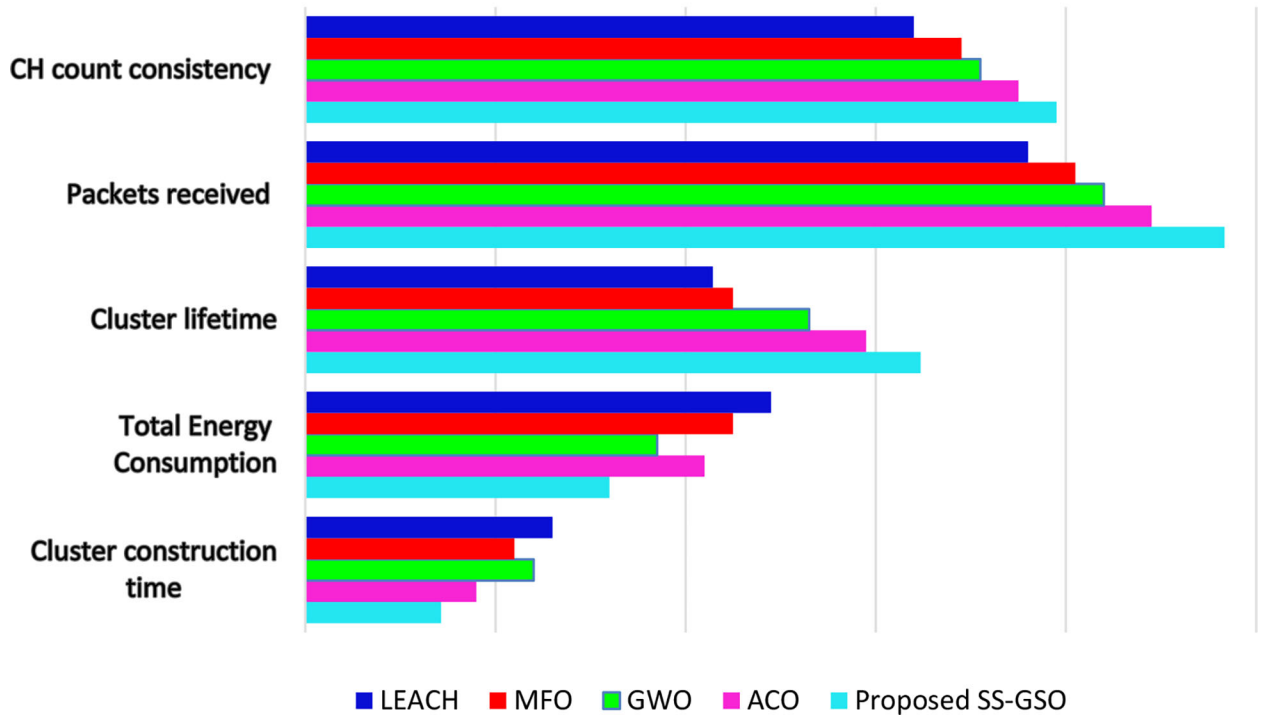


FIGURE 10
Overall performance analysis.

of the evaluated parameters except for the total amount of energy used, GWO is behind our proposed algorithm. The ACO works better than GSO, so the level of acceptance for GWO changes. We cannot put any algorithm at the second position from all the above experiments.

An SS-GSO for UWSN has been proposed in the current study. It has proven to be a very effective and scalable clustering protocol that works in a search space and chooses the optimal number of clusters to be deployed. It makes use of the GSO approach to accomplish this. For UWSN, GSO works iteratively in a defined search space to discover a better potential solution to improve its energy efficiency and longevity. As the number of clusters required decreases, the amount of energy consumed eventually reduces the cost of routing and the conservation of power in the nodes. On average, ten simulation runs were executed to see how well the suggested method worked. The algorithm was tested using a variety of assessed parameters. The findings demonstrate that SS-GSO is an excellent clustering technique for underwater environments. Clustering techniques such as GWO, MFO, and ACO are compared with our suggested algorithm for clustering. The results reveal that the SS-GSO is superior to all these algorithms in terms of performance, energy consumption, and other evaluating parameters. Incorporating data aggregation tactics early on in the design process has been shown to further increase the energy efficiency of SS-GSO approaches. Also, metaheuristic methods could be utilized to maximize resource utilization after being distributed to other nodes.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

Conceptualization SB. Methodology SB. Software NA, EE and NG. Validation SB, SS. Formal analysis NA, EE and NG. Investigation SB. Resources SS, EE. Data Curation NA, EE and NG. Writing - Original Draft SB. Writing - Review & Editing SB, NA, EE and NG. Visualization SB, EE. Supervision SS, EE. Project administration NA, EE and NG. Funding acquisition NA, EE and NG. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer AM declared a shared affiliation with the author NG to the handling editor at the time of review.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ahmad, I., Rahman, T., Khan, I., Jan, S., Musa, S., and Uddin, M. I. (2022). RACE-SM: Reliability and adaptive cooperation for efficient UWSNs using sink mobility. *Front. Mar. Sci.* 9, 1030113. doi: 10.3389/fmars.2022.1030113

Ahmad, I., Rahman, T., Zeb, A., Khan, I., Othman, M. T. B., and Hamam, H. (2022). Cooperative energy-efficient routing protocol for underwater wireless sensor networks. *Sensors* 22, 6945. doi: 10.3390/s22186945

Ahmad, I., Rahman, T., Zeb, A., Khan, I., Ullah, I., Hamam, H., et al. (2021). Analysis of security attacks and taxonomy in underwater wireless sensor networks. *Wirel Commun Mob Comput.*, 1–15. doi: 10.1155/2021/1444024

Al-Aboody, N. A., and Al-Raweshidy, H. S. (2016). "Grey wolf optimization-based energy-efficient routing protocol for heterogeneous wireless sensor networks," in *2016 4th international symposium on computational and business intelligence (ISCBI)* (IEEE). doi: 10.1109/iscbi.2016.7743266

Alhazmi, A. S., Moustafa, A. I., and AlDosari, F. M. (2018). "Energy aware approach for underwater wireless sensor networks scheduling: UMOD_LEACH," in *2018 21st Saudi Computer Society National Computer Conference (NCC). 2018 21st Saudi Computer Society National Computer Conference (NCC)* (IEEE). doi: 10.1109/ncg.2018.8593112

Alia, O. M. (2014). A decentralized fuzzy c-Means-Based energy-efficient routing protocol for wireless sensor networks. *Sci. World J.* 2014, 1–9. doi: 10.1155/2014/647281

Arshad, J., Rehman, A., Rehman, A. U., Ullah, R., and Hwang, S. O. (2020). Spectral efficiency augmentation in uplink massive MIMO systems by increasing transmit power and uniform linear array gain. *Sensors* 20, 4982. doi: 10.3390/s20174982

Asif, R. M., Ur Rehman, A., Ur Rehman, S., Arshad, J., Hamid, J., Tariq Sadiq, M., et al. (2020). Design and analysis of robust fuzzy logic maximum power point tracking based isolated photovoltaic energy system. *Eng. rep.* 2, 12234. doi: 10.1002/eng2.12234

Awan, K. M., Shah, P. A., Iqbal, K., Gillani, S., Ahmad, W., and Nam, Y. (2019). Underwater wireless sensor networks: A review of recent issues and challenges. *Wirel Commun Mob Comput.* , 1–20. doi: 10.1155/2019/6470359

Bharany, S., Badotra, S., Sharma, S., Rani, S., Alazab, M., Jhaveri, R. H., et al. (2022). Energy efficient fault tolerance techniques in green cloud computing: A systematic survey and taxonomy. *In sustainable energy technologies and assessments* 53, 102613. doi: 10.1016/j.seta.2022.102613

Bharany, S., Sharma, S., Badotra, S., Khalaf, O. I., Alotaibi, Y., Alghamdi, S., et al. (2021). Energy-efficient clustering scheme for flying ad-hoc networks using an optimized LEACH protocol. *Energies* 14, 6016. doi: 10.3390/en14196016

Bharany, S., Sharma, S., Bhatia, S., Rahmani, M. K. I., Shuaib, M., and Lashari, S. A. (2022). Energy efficient clustering protocol for fanets using moth flame optimization. *Sustainability* 14 (10), 6159. doi: 10.3390/su14106159

Bharany, S., Sharma, S., Frnda, J., Shuaib, M., Khalid, M. I., Hussain, S., et al. (2022). Wildfire monitoring based on energy efficient clustering approach for FANETS. *Drones* 6, 193. doi: 10.3390/drones6080193

Bharany, S., Sharma, S., Khalaf, O. I., Abdulsahib, G. M., Al Humaimeedy, A. S., Aldhyani, T. H. H., et al. (2022). A systematic survey on energy-efficient techniques in sustainable cloud computing. *Sustainability* 14 (10), 6256. doi: 10.3390/su14106256

Bhatti, D., Saeed, N., and Nam, H. (2016). Fuzzy c-means clustering and energy efficient cluster head selection for cooperative sensor network. *Sensors* 16 (9), 1459. doi: 10.3390/s16091459

Durrani, M. Y., Tariq, R., Aadil, F., Maqsood, M., Nam, Y., and Muhammad, K. (2019). Adaptive Node Clustering Technique for Smart Ocean Under Water Sensor Network (SOSNET). *Sensors* 19 (5), 1145. doi: 10.3390/s19051145

Fattah, S., Gani, A., Ahmedy, I., Idris, M. Y. I., and Targio Hashem, I. A. (2020). A survey on underwater wireless sensor networks: Requirements, taxonomy, recent advances, and open research challenges. *Sensors* 20 (18), 5393. doi: 10.3390/s20185393

He, S., Belacel, N., Hamam, H., and Bouslimani, Y. (2016). A hybrid artificial fish swarm simulated annealing optimization algorithm for automatic identification of clusters. *Int. J. Inf. Technol. Decision Mak* 15 (5), 949–974. doi: 10.1142/S0219622016500267

Heinzelman, W. R., Chandrakasan, A., and Balakrishnan, H. (2000). Energy-efficient communication protocol for wireless microsensor networks In: *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences. HICSS33: Hawaii International Conference on System Sciences*. IEEE Comput. Soc. doi: 10.1109/hicss.2000.926982

Heinzelman, W. B., Chandrakasan, A. P., and Balakrishnan, H. (2002). An application-specific protocol architecture for wireless microsensor networks, IEEE trans. *Wireless Commun.* 1 (4), 660–670. doi: 10.1109/TWC.2002.804190

Javadpour, A., Adelpour, N., Wang, G., and Peng, T. (2018). "Combing fuzzy clustering and PSO algorithms to optimize energy consumption in WSN networks" in *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart CityInnovation* (IEEE). doi: 10.1109/smartworld.2018.00238

Kazmi, H. S. Z., Javaid, N., Imran, M., and Outay, F. (2019). "Congestion control in wireless sensor networks based on support vector machine, grey wolf optimization and differential evolution," in *2019 wireless days (WD)* (IEEE). doi: 10.1109/wd.2019.8734265

Khan, A., Aftab, F., and Zhang, Z. (2019). Self-organization-based clustering scheme for FANETs using glowworm swarm optimization. *Phy. Com* 36, 100769. doi: 10.1016/j.phycom.2019.100769

Khan, M. F., Bibi, M., Aadil, F., and Lee, J.-W. (2021).Adaptive node clustering for underwater sensor networks. *Sensors* 21, 4514. doi: 10.3390/s21134514

Krishnaswamy, V., and Manvi, S. S. (2019). Fuzzy and PSO based clustering scheme in underwater acoustic sensor networks using energy and distance parameters. *Wireless Pers. Commun*, vol. 108 (3), 1529–1546. Springer Science and Business Media LLC. doi: 10.1007/s11277-019-06483-y

Kumari, S., Mishra, P. K., and Anand, V. (2019). Fault resilient routing based on moth flame optimization scheme for underwater wireless sensor networks. *Wireless networks* 26, 1417–1431. doi: 10.1007/s11276-019-02209-x

Lipare, A., Edla, D. R., Cheruku, R., and Tripathi, D. (2020). "GWO-GA based load balanced and energy efficient clustering approach for WSN," in *Smart trends in computing and communications* (Berlin/Heidelberg, Germany: Springer), 287–295.

Li, P., Wang, S., Zhang, H., and Zhang, E. (2017). "Improved particle swarm optimization algorithm of clustering in underwater acoustic sensor networks," in *Proc. OCEANS*(Aberdeen, U.K.), PP. 1_7.

Mazhar, M. S., Saleem, Y., Almogren, A., Arshad, J., Jaffery, M. H., Rehman, A. U., et al. (2022). Forensic analysis on Internet of things (IoT) device using machine-to-Machine (M2M) framework. *Electronics* 11, 1126. doi: 10.3390/electronics11071126

Mehmood, G., Khan, M. S., Waheed, A., Zareei, M., Fayaz, M., Sadad, T., et al. (2021). An Efficient and Secure Session Key Management Scheme in Wireless Sensor Network. In D. Volchenkov Ed. *Complexity* 1–10. doi: 10.1155/2021/6577492

Mehmood, G., Khan, M. Z., Waheed, A., Zareei, M., and Mohamed, E. M. (2020). "A trust-based energy-efficient and reliable communication scheme (Trust-based ERCS) for remote patient monitoring in wireless body area networks," in *IEEE*, vol. Vol. 8. (Institute of Electrical and Electronics Engineers (IEEE), 131397–131413. doi: 10.1109/access.2020.3007405

Mittal, N. (2018). "Moth flame optimization based energy efficient stable clustered routing approach for wireless sensor networks," in *Wireless personal communications*, vol. vol. 104, no. 2. (Springer Science and Business Media LLC), 677–694. doi: 10.1007/s11277-018-6043-4

Mohamed, A. W., Hadi, A. A., and Mohamed, A. K. (2019). Gaining-sharing knowledge based algorithm for solving optimization problems: a novel nature-inspired algorithm. *Int J Mach Learn Cybern* 11, 1501–1529. doi: 10.1007/s13042-019-01053-

Mohamed, A. W., and Mohamed, A. K. (2017). Adaptive guided differential evolution algorithm with novel mutation for numerical optimization. *Int. J. Mach. Learn. Cybernet.* 10, 253–277. doi: 10.1007/s13042-017-0711-7

Mohamed, A. W., and Suganthan, P. N. (2017). "Real-parameter unconstrained optimization based on enhanced fitness-adaptive differential evolution algorithm with novel mutation," in *Soft computing*, vol. Vol. 22. (Springer Science and Business Media LLC), 3215–3235. doi: 10.1007/s00500-017-2777-2

Namasudra, S., and Roy, P. (2018). PpBAC. *J. Organizat. End User Comput.* 30, 14–31. doi: 10.4018/joeuc.2018100102

Nguyen, N.-T., Le, T. T. T., Nguyen, H.-H., and Voznak, M. (2021). Energy-efficient clustering multi-hop routing protocol in a UWSN. *Sensors* 21, 627. doi: 10.3390/s21020627

Okoth, M. A., Shang, R., Jiao, L., Arshad, J., Rehman, A. U., and Hamam, H. (2022). A Large scale evolutionary algorithm based on determinantal point processes for Large scale multi-objective optimization problems. *Electronics* 11, 3317. doi: 10.3390/electronics1120

Patel, H., Singh Rajput, D., Thippa Reddy, G., Iwendi, C., Kashif Bashir, A., and Jo, O. (2020). A review on classification of imbalanced data for wireless sensor networks. *Int. J. Distributed Sensor Networks* 16 (4), 1550147720916404. doi: 10.1177/1550147720916404

Potthuri, S., Shankar, T., and Rajesh, A. (2018). Lifetime improvement in wireless sensor networks using hybrid differential evolution and simulated annealing (DESA). *Ain Shams Eng. J.* 9, 655–663. doi: 10.1016/j.asej.2016.03.004

Rajakumar, R., Amudhavel, J., Dhavachelvan, P., and Vengattaraman, T. (2017). GWO-LPWSN: Grey wolf optimization algorithm for node localization problem in

wireless sensor networks. *J. Comput. Networks Commun.* 2017, 1–10. doi: 10.1155/2017/7348141

Rao, P. C. S., Jana, P. K., and Banka, H. (2017). *A particle swarm optimization-based energy effcient cluster head selection algorithm for wireless sensor networks* Vol. vol. 23 (Wireless Netw.), 2005_2020.

Reddy, G. T., and Khare, N. (2017). An efficient system for heart disease prediction using hybrid OFBAT with rule-based fuzzy logic model. *J. Circuits Syst. Comput.* 26 (04), 1750061. doi: 10.1142/S021812661750061X

Saleem, F., Majeed, M. N., Iqbal, J., Waheed, A., Rauf, A., Zareei, M., et al. (2021). "Ant lion optimizer based clustering algorithm for wireless body area networks in livestock industry," in *IEEE*, vol. Vol. 9. (Institute of Electrical and Electronics Engineers (IEEE), 114495–114513. doi: 10.1109/access.2021.3104643

Sharma, R., Vashisht, V., and Singh, U. (2019). "EEFCM-DE: energy-efficient clustering based on fuzzy c means and differential evolution algorithm in WSNs," in *IET communications*, vol. vol. 13, no. 8. (Institution of Engineering and Technology (IET), 996–1007. doi: 10.1049/iet-com.2018.5546

Sharma, R., Vashisht, V., and Singh, U. (2020). eeTMFO/GA: a secure and energy efficient cluster head selection in wireless sensor networks. *Telecommun. Syst.* 74 (3), 253–268. doi: 10.1007/s11235-020-00654-0

Shokouhifar, M., and Jalali, A. (2017). Optimized sugeno fuzzy clustering algorithm for wireless sensor networks. *Eng. applications of artificial intelligence* 60, 16–25. doi: 10.1016/j.engappai.2017.01.007

Singh, B., and Lobiyal, D. K. (2012). A novel energy-aware cluster head selection based on particle swarm optimization for wireless sensor networks. *Human-centric computing and information sci.* 2 (1). doi: 10.1186/2192-1962-2-13.\

Srivastava, R., Bhardwaj, V. P., Othman, M. T. B., Pushkarna, M., Anushree, A., Mangla, A., et al. (2022). "Match-level fusion of finger-knuckle print and iris for human identity validation using neuro-fuzzy classifier," in *Sensors*, vol. 22. , 3620. doi: 10.3390/s22103620

Wang, J., Gao, Y., Liu, W., Sangaiah, A., and Kim, H.-J. (2019). An improved routing schema with special clustering using PSO algorithm for heterogeneous wireless sensor network. *Sensors* 19 (3), 671. doi: 10.3390/s19030671

Wan, Z., Liu, S., Ni, W., and Xu, Z. (2018). An energy-efficient multi-level adaptive clustering routing algorithm for underwater wireless sensor networks. *Cluster computing* 22 (S6), 14651–14660. doi: 10.1007/s10586-018-2376-8

Zeb, A., Wakeel, S., Rahman, T., Khan, I., Uddin, M. I., and Niazi, B. (2022). Energy-efficient cluster formation in IoT-enabled wireless body area network. *Computational intelligence and neuroscience* 2022, 1–11. doi: 10.1155/2022/2558590

Zhang, H., Wang, S. L., and Sun, H. X. (2017). A robust clustering architecture for medium scale Underwater acoustic sensor networks. *J. Xiamen Univ. Nat. Sci.* 56 (1), 129–136.

Zhao, X., Wei, Z., Cong, Y., and Yin, B. (2018). "A balances energy consumption clustering routing protocol for a wireless sensor network," in *2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC). 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC)* (IEEE). doi: 10.1109/itoec.2018.8740385

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 10 March 2023
DOI 10.3389/fmars.2023.1135058

Check for updates

OPEN ACCESS

EDITED BY
Hong Song,
Zhejiang University, China

REVIEWED BY
Yuan Lu,
Ocean University of China, China
Tianlong Zhang,
Northwest University, China
Aofei Mao,
University of Nebraska-Lincoln,
United States

*CORRESPONDENCE
Yang Zhu
zhuyang@jlu.edu.cn
Xiangqian Quan
quanxq@idsse.ac.cn

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 31 December 2022
ACCEPTED 10 February 2023
PUBLISHED 10 March 2023

CITATION
Yang G, Chen G, Cai Z, Quan X and Zhu Y (2023) Laser-induced breakdown spectroscopy instrument and spectral analysis for deep-ocean Fe-Mn crusts. *Front. Mar. Sci.* 10:1135058. doi: 10.3389/fmars.2023.1135058

COPYRIGHT
© 2023 Yang, Chen, Cai, Quan and Zhu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Laser-induced breakdown spectroscopy instrument and spectral analysis for deep-ocean Fe-Mn crusts

Guang Yang[1,2], Guanyu Chen[1,2], Zixuan Cai[1,2], Xiangqian Quan[3]* and Yang Zhu[1,4]*

[1]College of Instrumentation and Electrical Engineering, Jilin University, Changchun, China, [2]Jilin Province Key Laboratory of Trace Analysis Technology and Instruments, Changchun, China, [3]Institute of Deep-ocean Science and Engineering, Chinese Academy of Science, Sanya, China, [4]National Engineering Research Center of Geophysics Exploration Instruments, Jilin University, Changchun, China

Introduction: Marine ferromanganese crusts are potentially important metal resources.The deep-ocean research and survey ships often need to carry out rapid chemical element component analysis of mineral resources, so as to plan for the geological resource exploration mission.

Methods: The laser-induced breakdown spectroscopy can obtain the spectrum of chemical elements by the plasma excited by high-energy laser irradiation on the surface of the sample. A laser induced breakdown spectroscopy optical system for the detection of deepocean ferromanganese crusts is designed and built, which can meet the requirements of near-insitu chemical component detection of deep-ocean mineral resources on ocean-going survey ships.

Results: Hyperspectral data of Fe-Mn crusts are carried out by the Laser-induced breakdown spectroscopy (LIBS) system during a deep-sea exploration mission at a depth of 2,490 m in the South China Sea. The experimental parameters of laser energy and spectral acquisition delay are optimized to improve the spectral measurement accuracy. Based on the calibration-free method, the significant spectral features of the chemical elements Fe and Mn were obtained through proper alignment with the National Institute of Standards and Technology (NIST) chemical element spectral library.

Discussion: The LIBS instrument can be placed on board long-range survey vessels in the future to provide a fast, convenient, accurate, and economical detection method for deep-ocean resource exploration.

KEYWORDS

deep ocean, marine spectral analysis, marine resource exploration, laser-induced breakdown spectroscopy, marine Fe-Mn crusts, parameter optimization

1 Introduction

The resource detection and exploration of deep-ocean minerals have been studied for decades (Hein et al., 2013). Marine ferromanganese oxide crusts (Fe-Mn crusts) are potentially important metal resources formed on the seafloor by precipitation of dissolved and colloidal components from ambient seawater onto rocky surfaces (Lusty and Murton, 2018; Lusty et al., 2018). Ferromanganese crusts may provide a future resource for a large variety of metals, including emerging high-tech and green technologies applications. Fe-Mn crusts have a high potential economic value and is a hotspot of marine mineral resources research. Fe-Mn crusts are found typically at cold, ambient ocean depths of 400– 7,000 m onto the surface of seamounts, ridges, and plateaus, with the thickest and most metal-rich crusts occurring at depths of approximately 800–2,500 m (Corliss et al., 1979; Hein and Koschinsky, 2014). In addition, based on the assumption that the chemical composition of the ferromanganese crusts is not affected by late diagenesis, the chemical information of the crust can be used to reverse the ancient marine evolution (Chen et al., 2013; Jiang et al., 2021).

The distribution of Fe-Mn crusts can be preliminarily determined by deep-ocean optical imaging, while the finer classification of Fe-Mn crusts should be obtained by deep-ocean geological sampling and chemical analysis in the laboratory (Huang et al., 2021; Xue et al., 2021; Bell et al., 2022; Zhou et al., 2022). At present, there are two main methods to analyze the components of deep-ocean crust samples: laser spectroscopic detection and inductively coupled plasma optical emission spectrometer (ICP-OES) in the laboratory. The deep-ocean research and survey ships often need to carry out rapid chemical element component analysis of mineral resources (Lai et al., 2017), so as to plan for geological resource exploration missions. ICP-OES can achieve detection limits of 1–10 ppb for most elements and a linear range of four to six orders of magnitude on the standard curve (Becker-Ross et al., 2002; Piskunov and Valenti, 2002; Xie et al., 2009; Lavigne et al., 2010; Xu et al., 2011). However, it generally requires pretreatment such as microwave digestion, evaporation, ashing, and the individual preparation of the solutions for the samples to be tested and the standards. A large amount of high-purity argon gas is consumed in the ICP-OES analysis process. In addition, ICP-OES generally relies on the standard curve method for quantitative elemental analysis, with relatively long sample analysis cycles and complex equipment, which is unsuitable for rapid near-insitu testing in oceanic regions. In contrast, this paper proposes the use of the laser-induced breakdown spectroscopy (LIBS) technique, where high-energy laser pulses are focused onto the surface of a sample and excited to produce a high-brightness, high-heat plasma. The qualitative and quantitative analysis method of the excitation spectra of complex matrix samples in the field through the calibration-free method is presented, without complex sample pretreatment and high-purity gas environments (Kiefer et al., 2015; Yang et al., 2020). It also effectively prevents the interference of the spectrum of seawater saline substances to the spectrum of seabed resources, which is more conducive to rapid chemical component detection near-insitu in the deep ocean. The LIBS instrument could be placed on board long-range survey vessels in the future to provide a fast, convenient, accurate, and economical detection method for deep-ocean resource exploration.

In this study, a laser-induced breakdown spectroscopy optical system for the detection of deep-ocean ferromanganese-rich ores is designed and built, which can meet the requirements of the near-insitu chemical component detection of deep-ocean mineral resources on ocean-going survey ships. The 1,064-nm-wavelength and 90-mJ-energy laser is chosen as the excitation light source, realizing spectral information detection with a spectral resolution of 0.05 nm and completing the integration and testing of the principal prototype. Hyperspectral detection of ores collected from a seafloor exploration mission to the South China Sea area at a depth of 2,490 m below the seafloor was carried out using LIBS, and hyperspectral data were obtained for deep-ocean ferromanganese-rich ores. The experimental parameters of laser energy and spectral acquisition delay are optimized for the spectral data, and the stability of laser energy fluctuations as well as acquisition angle fluctuations was discussed. Finally, based on the calibration-free method, the significant spectral features of the chemical elements Fe and Mn were obtained through proper alignment with the National Institute of Standards and Technology (NIST) chemical element spectral library, and the chemical element contents of Fe and Mn are obtained by calculating the plasma temperature, distribution function, and instrument parameters. The principal prototype enables the detection of deep-ocean ferromanganese-rich ores, and through experimental optimization, the impact on the detection stability of the instrument due to offshore work is reduced and provides a basis for the engineering of near-insitu component detection of seabed ores on board ocean resources exploration in the future.

2 Methods

2.1 LIBS instruments and components

As shown in Figure 1, the laser emits a pulsed laser beam that is split by a beam splitting mirror and a portion of the laser is monitored in real time by an energy meter; as the beam splitting ratio of the beam splitting mirror is fixed, the energy ratio of the two splitting beams is also fixed and the actual laser energy of the optical path can be calculated from the monitored laser energy. The actual optical path is focused onto the surface of the sample through a focusing mirror to create a plasma. The elemental composition of the material is then analyzed by collecting the plasma spectra.

For the deep-ocean ferromanganese detection instrument, the wavelength of laser is 1,064 nm, the maximum laser energy is 100 mJ, and the standard deviation of laser energy stability is 0.433. The polarization attenuator is set at the laser outlet port. For linearly polarized beams, a tunable attenuator can be constructed by a half-wave plate and a polarization beam splitter. The ratio of S light and P light can be changed continuously by the attenuator to achieve the purpose of continuous attenuation. The polarization angle can be adjusted by rotating the knob at the laser output port, thereby adjusting the offset angle of the laser's output energy, so that the

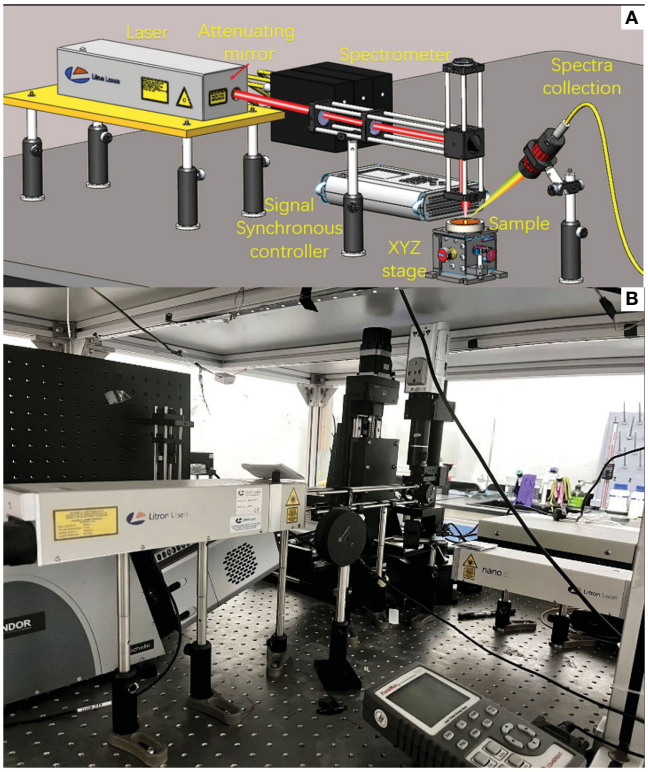


FIGURE 1
The experimental setup of laser-induced breakdown spectroscopy: **(A)** the schematic diagram of experimental setup; **(B)** the inside components of the LIBS system.

laser energy can be continuously adjusted from 0 mJ to 100 mJ and provide a hardware foundation for the subsequent optimization of the laser energy parameters of the LIBS system.

The overall spectral response range of the hyperspectral detector is 200–950 nm. The broad spectrum is divided into three channels: 200–320 nm, 320–420 nm, and 420–950 nm, with a detection accuracy of 0.05 nm, 0.05 nm, and 0.1 nm, respectively. The sensitivity of the detector is 310,000 (in counts/μW per millisecond integration time), with a sampling rate of 1.1 ms each and a minimum integration time of 3 ms, which can meet the requirements for spectral collection of subsea ferromanganese crust samples.

As illustrated in Figure 1, the LIBS optical path is built using a high-precision cage system, which ensures both stability and convenience of adjustment. The three-dimensional translation platform control system consists of a high-precision three-dimensional translation platform and an electronically controlled 3D translation platform control box, which mainly functions to control the coordinate translational transformation of the laser focus position on the sample surface, to control the excitation position of the sample, and to assist in focusing and spectral surface array scanning, to name a few. The high-precision three-dimensional translation platform is equipped with precision-grade ball screws, in combination with linear slider guides, high-quality couplings, and other components to ensure that it can be used for high-precision control over a stroke range of 100 mm and a resolution of 20 μm. The electronically controlled three-dimensional translation platform control box allows for delicate control of the three-dimensional movement direction, movement speed, and movement acceleration of the three-dimensional translation platform.

The delay pulse generator is a self-developed component of the laboratory, which allows precise control for the time sequence pulses of the various functional components by receiving and sending precise time sequence pulse signals. The delay accuracy of 10 ns provides hardware support for the subsequent optimization and analysis of the spectral acquisition delay parameters. The self-developed 3D printed slope simulation component was independently designed by the laboratory of Jilin University to simulate the surface roughness of the sample. The slope can be applied to represent the cut surface of the rough sample and then used to verify the angular size of the sample, as well as the effect of the roughness size on the spectral data acquisition.

2.2 Deep-ocean sample acquisition

The *Hai Yang 6 Hao* ocean survey ship navigated the South China Sea to explore the deep-ocean geological and seabed minerals. The deep-ocean mining vehicle developed by the Institute of Deep-ocean Science and Engineering (Chinese Academy of Science) has been placed on the seabed and collected samples of deep-ocean rocks at a water depth of 2,490 m at 115°06′ E, 18°17′ N, as illustrated in Figure 2A. The LIBS measurement

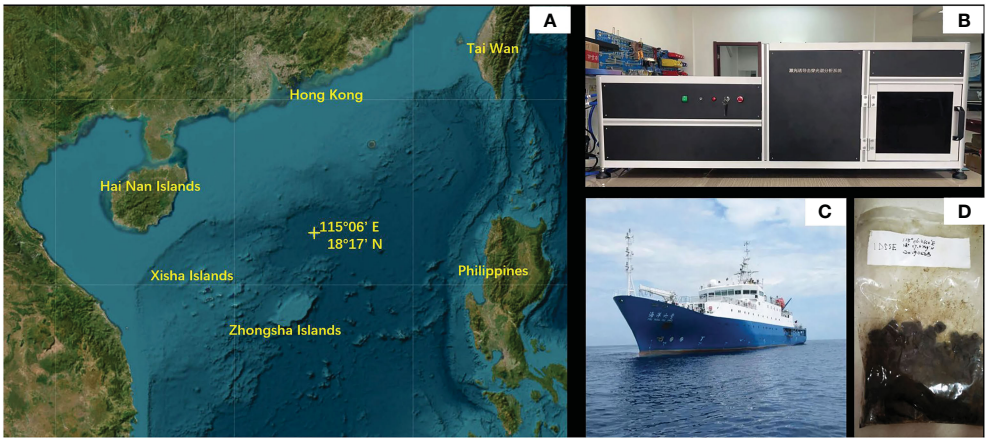


FIGURE 2
(A) The deep-ocean geological survey region of the South China Sea; **(B)** the LIBS measuring system with the implementation of packaging and integration; **(C)** the *Hai Yang 6 Hao* ocean survey ship; **(D)** the seabed mineral sample mined by a deep-ocean mining vehicle at a water depth of 2,490 m.

system has completed the development of the principal prototype with the implementation of packaging and integration (Figure 2B). The *Hai Yang 6 Hao* ocean survey ship is shown in Figure 2C. The seabed mineral sample mined by the deep-ocean mining vehicle is shown in Figure 2D. After a simple surface leveling treatment, the spectrum of the sample is measured by the LIBS system in the laboratory, and there are plans to implement near-insitu measurement on ocean survey ships in the future.

3 Optimization of experimental parameters

Before the exploration of deep-ocean ferromanganese crusts by LIBS, we should first run an experimental simulation. Based on the simulation experiment platform, the experimental conditions are optimized to simulate the marine exploration environment and avoid the measurement errors attributed to the measurement conditions of the marine exploration environment.

3.1 Optimization of laser energy

Based on the LIBS system, we have conducted the laser excitation spectroscopy measurements. The results are shown in Figure 3; the laser-induced breakdown spectra obtained significantly change as the laser energy is gradually increased. The number and intensity of the characteristic peaks also tend to rise. Several strong peaks in the diagram are caused by metals such as Na, K, and Ca in seawater or carbonates in stone, which have been processed by filtering to remove the influence on the Fe-Mn crust spectrum. The characteristic spectral lines of typical mineral elements of Fe and Mn are extracted at laser energies of 30 mJ, 40 mJ, 60 mJ, and 90 mJ, respectively (A–I spectral lines).

Due to the low energy compared to the high energy in the excitation of the sample, the chemical bonds internal to the sample cannot absorb enough energy to break, thus failing energy level transition. In addition, because of the small amount of sample ablation, as the internal structure of the sample is not sufficiently homogeneous and dense at this point, without a complex pretreatment, the ablated sample is not guaranteed to contain the measured elements completely, resulting in the characteristic lines of the measured elements not being fully excited.

As shown in Table 1, by analyzing the data from actual measurements, when the laser energy is 30 mJ, only the G (Mn 369.391 nm) and H (Fe 437.450 nm) spectral lines could be excited properly. At a laser energy of 40 mJ, A (Fe 225.977 nm), C (Fe 285.208 nm), and E (Fe 301.991 nm) are not properly excited, while the other A–I spectral lines can be excited normally. As the laser energy increases to 90 mJ, all the individual A–I spectral lines are excited. Once the laser energy is further increased to 100 mJ, part of the spectrum appears saturated due to the limited intensity of the spectral lines collected by the spectrometer, which will reduce the analytical accuracy of typical elements; thus, we finally selected 90 mJ of laser energy as the experimental parameter for the next step.

3.2 Optimization of spectral acquisition delay

During the initial stages of plasma formation, the strong signal generation is accompanied by a high level of background noise due to the interference of the bremsstrahlung (Fu et al., 2020; Nosrati et al., 2020). As the signal decays at a much lower rate than the background noise, time-delayed collection is required. When the acquisition is performed after the background noise has been almost exhausted, the signal with a high signal-to-noise ratio (SNR) can be obtained.

As presented in Figure 4, when the delay is −150 μs to +18 μs, the SNR tends to rise slowly as the signal decays at a rate much lower than the background decay. The SNR increases abruptly and rapidly during the delay phase of 18–18.5 μs and decreases fast

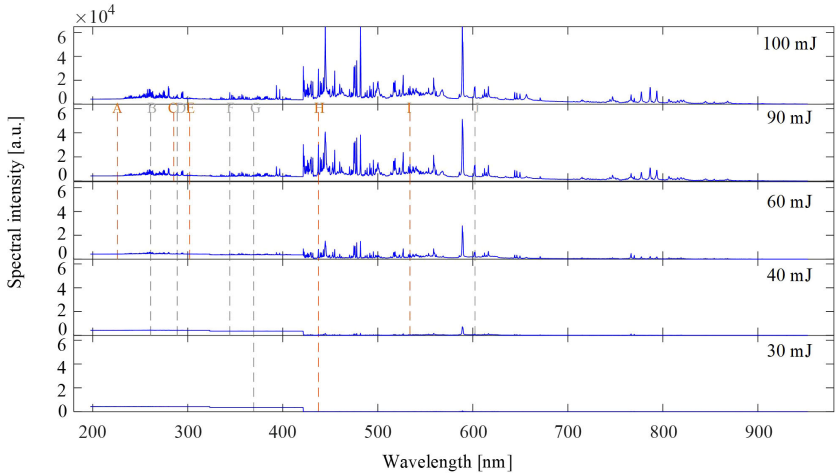


FIGURE 3
The spectra of the sample at a laser energy of 30–100 mJ.

TABLE 1 Characteristic peaks of Fe, Mn, and Co at a laser energy of 30–90 mJ.

Number	Characteristic peak	90 mJ	60 mJ	40 mJ	30 mJ
A	Fe 225.977 nm	√	√	×	×
B	Mn 260.983 nm	√	√	√	×
C	Fe 285.208 nm	√	×	×	×
D	Mn 288.896 nm	√	√	√	×
E	Fe 301.991 nm	√	√	×	×
F	Mn 344.195 nm	√	√	√	×
G	Mn 369.391 nm	√	√	√	√
H	Fe 437.450 nm	√	√	√	√
I	Fe 533.865 nm	√	√	√	×
J	Mn 602.295 nm	√	√	√	×

during the phase of 18.5–19 μs. With a delay time of 18.5–30 μs, the SNR shows a decreasing trend, which could be due to the background noise nearly decaying out and the signal continuing to decay. The signal can be extracted to a highly SNR peak in a short time, which puts a demand on the delay time resolution of the spectral collection delay; a higher resolution of the delay time enables the near step change features in the vicinity of the peak to be captured, thus allowing the highest SNR transient excitation spectrum to be obtained. The reason for spectral acquisition delay optimization is that the high-energy laser focused on the sample surface causes charged particles to collide with atoms and suddenly slow down to produce bremsstrahlung.

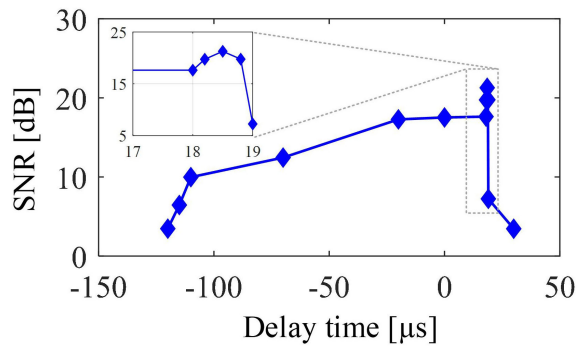


FIGURE 4
The figure of signal-to-noise ratio changing with spectral acquisition delay.

4 Experiment and results

4.1 Stability measurement and analysis of laser energy

During deep-ocean exploration missions, we tend to execute long-term ocean-going offshore field exploration surveys. Due to the power supply and distribution on ocean-going measurement vessels and the harsh environment at sea, the intensity, surge, and reflection ripple of the supply voltage to the LIBS equipment and its

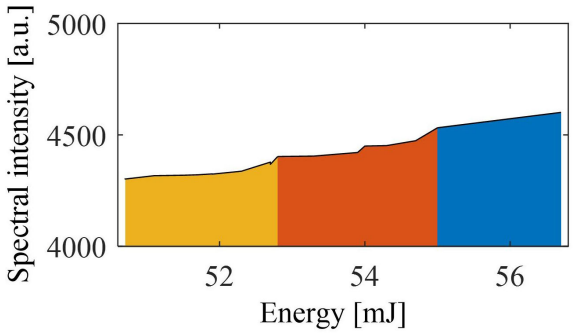


FIGURE 5
The figure of relationship between the fluctuation of laser energy and the change of spectral intensity.

lasers may change accordingly, resulting in the instability of the laser energy and a drift in the intensity of the excitation spectrum, which will finally reduce the accuracy of the spectral measurements. Hence, there is a need to measure and evaluate the effect of laser energy instability on the precision of spectral measurements.

The instrument utilizes a beam splitting mirror fitted to the laser outlet for laser energy splitting, with a beam splitting ratio of 1:1. The laser we use has an energy of approximately 100 mJ, and the energy of the laser beam after the beam splitting mirror is approximately 50 mJ. Under conditions of stable supply voltage, the standard deviation of the laser's own energy is 0.433 after actual measurements, with the small fluctuations in its energy having essentially no effect on the system's measurements.

We then simulate the undulating drift of the laser energy by adjusting the linear polarization angle of the polarizer. When the laser energy fluctuates in the interval of 50–56 mJ, the collected sample excitation characteristic peak spectrum is also obtained to fluctuate in the intensity range of 4,300–4,600 by the real measurements. As shown in Figure 5, for every 2-mJ fluctuation in laser energy, the characteristic spectral lines fluctuate by roughly 100. For the laser employed in this equipment, it is required to ensure that the laser energy fluctuates by less than 2 mJ if the intensity of the characteristic spectrum is to be fluctuated within 100.

4.2 Influence of sample surface roughness on spectral measurement

As the mineral samples we collected from the deep ocean was simply polished without more complex pretreatment, the surfaces of the samples are still undulating. In order to verify the effect of the extent of surface undulations on the spectra of deep-ocean ores, we microstimulated the laser focus location (i.e., the surface of the sample at the sample ablation location) by modeling the cut surface as a smooth bevel at a certain angle. The effect of surface roughness on the spectra of deep-ocean ores is illustrated by measuring the effect of different angle bevels at the sample surface on the spectrum.

During the changing of the slope of the sample with the use of the slope tool, if the height of the 3D translation platform is not adjusted, the focusing position will change accordingly and reduce the focusing quality of the laser spot. A Charge-coupled Device (CCD) sensor is placed coaxially with the focusing mirror to ensure consistency of focus. The accuracy of spot focusing can be judged according to the focusing image obtained by the CCD sensor. As shown in Figure 6, the spot image is clear in the best focusing position. The spot image is blurred under defocus conditions. The larger the defocus distance, the blurrier the image.

We have simulated the ore spectra generated from the same sample ablation point for inclination angles of 0°, 15°, 22°, 30°, 37°, and 45° of the slope. As shown in Figure 7, the spectrum varies insignificantly at slope angles of 0° and 15°, the total intensity of the spectrum decreases at a bevel angle of 22°, the decline is 1/5 of the overall, and at a bevel angle of 22° – 45°, the overall intensity of the spectrum follows a continuing downward trend. The results show that if the spectral signal needs to remain stable, the roughness of the sample surface should be less than 15° in the section of the sample ablative position.

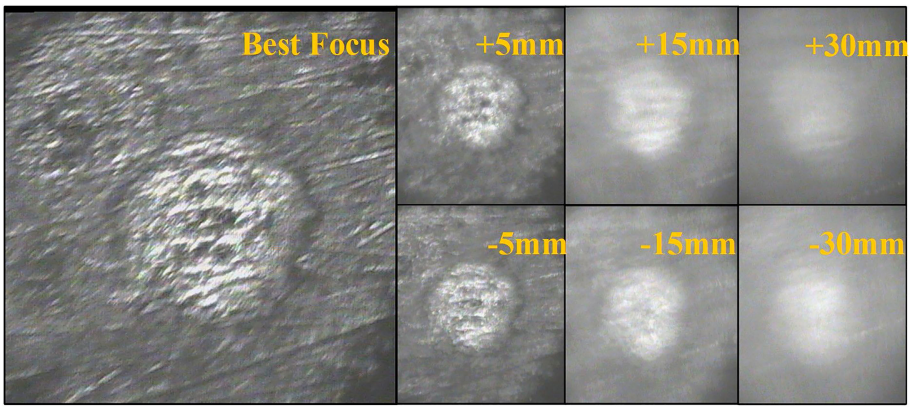


FIGURE 6
Focus photo and defocus photos by CCD.

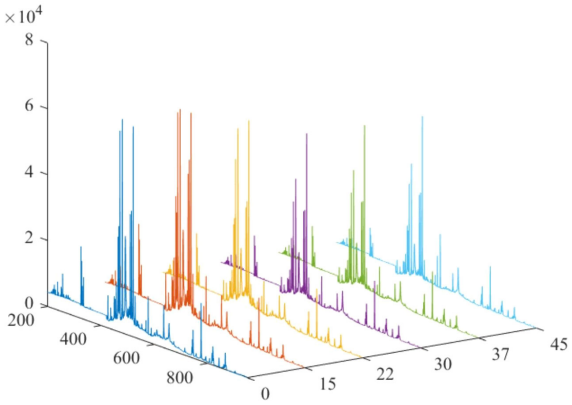


FIGURE 7
The variation of spectra of the sample ablative section angle from 0° to 45°.

5 Data analysis and calibration

The calibration-free model (CF-LIBS) (Ciucci et al., 1990; Dell'Aglio et al., 2018; Ferus et al., 2018; Umar et al., 2018) is implemented on the premise of three basic assumptions: (1) the elemental composition of the plasma produced by LIBS corresponds exactly to the elemental constitution of the sample analyzed before ablation, which is also known as chemometric ablation; (2) the plasma formed by LIBS is in the local thermodynamic equilibrium (LTE), judged by Eq. (1), which specifically means that the excitation rate of all the energy levels p and q that collide cannot be less than 10 times the rate of spontaneous radiation; and (3) the plasma from LIBS is in an optically thin state, meaning that no self-absorption is present in the plasma emission spectrum.

$$N_e \geq 1.6 \times 10^{12} T_e^{\frac{1}{2}} \Delta E^3 \tag{1}$$

where T_e is the electron temperature, N_e is the electron density, and ΔE is the maximum energy difference of the ions analyzed. Based on the three fundamental assumptions, the mathematical relationship between the spectral line intensities and the plasma parameters can be expressed by Eq. (2):

$$I_\lambda^{ij} = FC_s A_{ij} \frac{g_i}{U_s(T)} e^{\frac{E_i}{k_B T}} \tag{2}$$

In Eq. (2), I_λ^{ij} is the intensity of the integrated spectral line at the leap wavelength λ; i and j are the high and low energy levels of the electron leap corresponding to the leap wavelength λ; F is the experimental constant; C_s is the relative concentration of emitting species in the plasma; s is the particle species; A_{ij} is the probability of jumping for a given spectral line; g_i is the statistical weight of energy level i; T is the excitation temperature of the plasma; E_i is the energy of particle energy level i; k_B is the Boltzmann constant; and $U_s(T)$ is the partition function of the species s at temperature T, which can be calculated by Eq. (3):

$$U_s(T) = \sum g_i e^{\frac{-E_i}{k_B T}} \tag{3}$$

By taking the logarithm of Eq. (2), the equation can be transformed into a linear form:

$$y = ax + q_s \tag{4}$$

In Eq. (4), a is the slope; q_s is the intercept, where $q_s = \ln C_s F U_s(T)$, $a = -\frac{1}{k_B T}$, $y = \ln I_\lambda^{ij} g_i A_{ij}$, and $x = E_i$. The two-dimensional plane defined by the x and y coordinates is called the Boltzmann plane, and every characteristic spectral line obtained by the experiment can be plotted as a point on the Boltzmann plane. Therefore, each chemical element in the sample that needed to be measured can be formulated using a relationship similar to Eq. (4) and plotted on the Boltzmann line. From Eq. (4), the slope a is related to the plasma temperature, while q_s is proportional to the logarithm of the particle concentration. According to the properties of LIBS plasma, the temperature of each particle in the plasma is the same; therefore, the Boltzmann lines of different elemental particles in the sample have theoretically the same slope.

The plasma temperature obtained from the plotted Boltzmann diagram is utilized to calculate the particle distribution function $U_s(T)$; then, combining with the intercept q_s of each particle in the Boltzmann diagram, the concentration C_s of each particle can be obtained. F can be obtained by normalizing the elemental concentrations:

$$\sum_s C_s = 1, \quad F = \sum_s U_s(T) e^{q_s} \tag{5}$$

By substituting the calculated F into Eq. (6), the mass percentages of the various particles and analyzed elements can be acquired.

$$C_s = \frac{1}{F} U_s(T) e^{q_s} \tag{6}$$

$$C_s^{TOT} = C_s^I + C_s^{II} \tag{7}$$

where C_s^I is the concentration of the atom of the element being analyzed, C_s^{II} is the concentration of the ion of the element being analyzed, and C_s^{TOT} is the concentration of the element being analyzed.

As shown in Table 2, the final mass fractions of the elements Fe and Mn were calculated to be 26.2% and 15.3%, respectively, by using the calibration free method; in comparison with the ICP

TABLE 2 The relative measurement error analysis of ICP and CF-LIBS measurement results and.

Element	ICP (mass fraction)	CF-LIBS (mass fraction)	Relative error
Fe	22.9	26.2	14.4%
Mn	17.6	15.3	13.1%

elemental determination, the relative errors reached 14.4% and 13.1% (Xinglong et al., 2012; Dehua et al., 2016; Lei et al., 2019), respectively, and the measurement accuracy could meet the practical requirements for the offshore near-*in situ* chemical fraction analysis of deep-sea Fe-Mn crusts.

6 Conclusion

A laser-induced breakdown spectroscopy instrument has been designed and built to the spectrum of chemical elements of deep-ocean Fe-Mn crusts. The overall spectral response range of the hyperspectral detector is 200–950 nm. The broad spectrum of LIBS was divided into three channels: 200–320 nm, 320–420 nm, and 420–950 nm, with a detection accuracy of 0.05 nm, 0.05 nm, and 0.1 nm, respectively. Hyperspectral data of Fe-Mn crusts have been carried out by the LIBS system during a deep-sea exploration mission at a depth of 2,490 m in the South China Sea. Through measurement and parameter optimization, and with a laser energy of 90 mJ and a delay time of 18.5–19 μs, the highest spectral resolution and SNR were obtained. The significant spectral features of the chemical elements Fe and Mn were obtained through proper alignment with the NIST chemical element spectral library based on the free spectral calibration. As shown by the results of the experiment, the LIBS instrument and spectrum processing method provide a technical foundation for the near-insitu component detection in deep-ocean resource explorations in the future.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Author contributions

GY completed the design and construction of the equipment/instrument. GC conducted the experimental tests and data analysis method. ZC participated in manuscript preparation and data processing. XQ organized the deep-ocean exploration and supervised the project. YZ conceived the work and provided guidance in manuscript preparation. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Natural Science Foundation of China (62105119 and 62275099) and Youth Innovation Promotion Association CAS (No. 2020361).

Acknowledgments

The authors thank Ming Chen (Institute of Deep-sea Science and Engineering, Chinese Academy of Science) for the assistance during the experiment and data measurement.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Becker-Ross, H., Okruss, M., Florek, S., Heitmann, U., and Huang, M. D. (2002). Echelle-spectrograph as a tool for studies of structured background in flame atomic absorption spectrometry, spectrochim. *Acta Part B* 57, 1493–1504. doi: 10.1016/S0584-8547(02)00107-6

Bell, K. L.C., Chow, J. S., Hope, A., Quinzin, M. C., Cantner, K. A., Amon, D. J., et al. (2022). Low-cost, deep-ocean imaging and analysis tools for deep-ocean exploration: a collaborative design study. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.873700

Chen, T. Y., Ling, H. F., Hu, R., and Frank, M. (2013). Lead isotope provinciality of central north pacific deep water over the Cenozoic. *Geochemistry Geo-physics Geosystems* 14 (5), 1523–1537. doi: 10.1002/ggge.20100

Ciucci, A., Corsi, M., Palleschi, V., Rastelli, S., Salvetti, A., and Tognoni, E. (1990). New procedure for quantitative elemental analysis by laser-induced plasma spectroscopy. *Appl. Spectrosc.* 53, 960–964. doi: 10.1366/0003702991947612

Corliss, J. B., Dymond, J., Gordon, L. I., Edmond, J. M., Von Herzen, R. P., Ballard, R. D., et al. (1979). Submarine thermal springs on the galápagos rift. *Science* 203, 1073–1083. doi: 10.1126/science.203.4385.1073

Dehua, Z., Cao, Y., and Rong and C. Xiaojing, Z. (2016). Quantitative analysis of composition change in AZ31 magnesium alloy using CF-LIBS after laser material processing. *Plasma Sci. Technol.* 17, 909–913. doi: 10.1088/1009-0630/17/11/03

Dell'Aglio, M., Lopez-Claros, M., Laserna, J. J., Longo, S., and Giacomo, A. (2018). Stand-off laser induced breakdown spectroscopy on meteorites: calibration-free

approach, spectrochim. *Acta Part B At. Spectrosc* 147, 87–92. doi: 10.1016/j.sab.2018.05.024

Ferus, M., Koukal, J., Len˙za, L., Srba, J., Kubelík, P., Laitl, V., et al. (2018). Calibration-free quantitative elemental analysis of meteor plasma using reference laser-induced breakdown spectroscopy of meteorite samples. *Astron. Astrophys* 610, 1–12. doi: 10.1051/0004-6361/201629950

Fu, C., Wu, D., Wang, Q., Sun, L., and Wang and H. Ding, Y. (2020). Time-resolved study of bremsstrahlung emission and spectra at the early stage in a nanosecond laser ablated tungsten plasma. *J. Of Insturment* 15, C02022. doi: 10.1088/1748-0221/15/02/C02022

Hein, J. R., and Koschinsky, A. (2014). "Deep-ocean ferromanganese crusts and nodules," in *Treatise on geochemistry*, 273–291. doi: 10.1016/B978-0-08-095976-7.01111-6

Hein, J. R., Mizell, K., Koschinsky, A., and Conrad, T. A. (2013). Deep-ocean mineral deposits as a source of critical metals for high-and green-technology applications: Comparison with land-based resources. *Ore Geology Rev.* 51, 1–14. doi: 10.1016/j.oregeorev.2012.12.001

Huang, H., Sun, Z., Liu, S., Di, Y. N., Xu, J. Z., Liu, C. C., et al. (2021). Underwater hyperspectral imaging for *in situ* underwater microplastic detection. *Sci. Total Environ.* 776, 145960. doi: 10.1016/j.scitotenv.2021.145960

Jiang, X. D., Zhao, X., Zhao, X. Y., Chou, Y. M., Hein, J. R., Sun, X. M., et al. (2021). A magnetic approach to unravel-ling the paleoenvironmental significance of nanometer-sized fe hy-droxide in NW pacific ferromanganese deposits. *Earth Planet-ary Sci. Lett.* 565, 116945. doi: 10.1016/j.epsl.2021.116945

Kiefer, J., Zhou, B., Li, Z. S., and Alden, M. (2015). Impact of plasma dynamics on equivalence ratio measurements by laser-induced breakdown spectroscopy. *Appl. Opt* 54, 4221–4226. doi: 10.1364/AO.54.004221

Lai, C.-Z., De Grandpre, M. D., and Darlington, R. C. (2017). Autonomous optofluidic chemical analyzers for marine applications: insights from the submersible autonomous moored instruments (SAMI) for pH and pCO(2). *Front. Mar. Sci.* 4. doi: 10.3389/fmars.2017.00438

Lavigne, J.-F., Doucet, M., Wang, M., Lacoursière, J., Grill, M., Melchiorri, R., et al. (2010). Study of the image quality and stray light in the critical design phase of the compact echelle spectrograph for aeronomical research (CESAR). *Proc. SPIE* 7735, 773539. doi: 10.1117/12.857509

Lei, Z., Shuxia, Z., and Ying, S. (2019). Quantitative CF-LIBS analysis of alloys via comprehensive calibration of plasma temperature and spectral intensity. *Scientia Sinica-physica Mechanica Astronomica* 48. doi: 10.1360/SSPMA2017-00221

Lusty, P. A. J., Hein, J. R., and Josso, P. (2018). Formation and occurrence of ferromanganese crusts: earth's storehouse for critical metals. *Elements* 14 (5), 313–318. doi: 10.2138/gselements.14.5.313

Lusty, P. A. J., and Murton, B. J. (2018). Deep-ocean mineral deposits: metal resources and windows into earth processes. *Elements* 14, 301–306. doi: 10.2138/gselements.14.5.313

Nosrati, Y., Tavassoli, S. H., and Hassanimatin and A. Safi, M. M. (2020). Study of material ablation and plasma radiation in double-pulse laser induced breakdown spectroscopy at different delay times:Modeling and numerical simulation. *Phys. Plasmas* 27, 023301. doi: 10.1063/1.5132804

Piskunov, N. E., and Valenti, J. A. (2002). New algorithms for reducing cross-dispersed echelle spectra, astron. *Astrophys* 385, 1095–1106. doi: 10.1051/0004-6361:20020175

Umar, Z. A., Ahmed, N., Ahmed, R., Liaqat, U., and Baig, M. A. (2018). Elemental composition analysis of granite rocks using LIBS and LA-TOF-MS. *Appl. Opt* 57, 4985. doi: 10.1364/AO.57.004985

Xie, P., Ni, Z., Huang, Y., Zhang, D., and Zhang, Y. (2009). Application research progress in the echelle grating. *Laser J.* 30, 4–6. doi: 10.1109/JSTQE.2012.2213804

Xinglong, C., Fengzhong, D., and Qi, W. (2012). Quantitative analysis of slag by calibration-free laser-induced breakdown spectroscopy. *Spectrosc. Spectral Anal.* 31, 3289–3293. doi: 10.3964/j.issn.1000-0593(2011)12-3289-05

Xu, L., Davonport, M. A., Turner, M. A., Sun, T., and Kelly, K. F. (2011). Compressive echelle spectroscopy. *Proc. SPIE* 8165, 81650E. doi: 10.1117/12.894191

Xue, Q., Tian, Z., Bai, Y., Liang, J. S., Li, C., Wang, F. P., et al. (2021). Underwater hyperspectral imaging system using a prism-Grating-Prism structure. *Appl. Optics.* 60 (4), 894–900. doi: 10.1364/AO.415351

Yang, G., Liu, L., Wang, T., Fan, L. S., Huang, X., Tian, D., et al. (2020). Laser-induced breakdown spectroscopy of ammonia gas with resonant vibrational excitation. *Optics Express* 28, 1197–1205. doi: 10.1364/OE.382663

Zhou, J., Huang, H., Huang, S. H., Si, Y., Shi, K., Quan, X., et al. (2022). AUH, a new technology for ocean exploration. *Engineering* 1–10. doi: 10.1016/j.eng.2022.09.007

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 13 March 2023
DOI 10.3389/fmars.2023.1031869

Check for updates

OPEN ACCESS

EDITED BY
Andrew James Manning,
HR Wallingford, United Kingdom

REVIEWED BY
Jessica Lunt,
Dauphin Island Sea Lab, United States
Xiaoping Wang,
Zhejiang University, China

*CORRESPONDENCE
Hui Huang
huih@zju.edu.cn

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 30 August 2022
ACCEPTED 20 February 2023
PUBLISHED 13 March 2023

CITATION
Song H, Mehdi SR, Li Z, Wang M, Wu C, Venediktov VY and Huang H (2023) Investigating the rate of turbidity impact on underwater spectral reflectance detection. *Front. Mar. Sci.* 10:1031869. doi: 10.3389/fmars.2023.1031869

COPYRIGHT
© 2023 Song, Mehdi, Li, Wang, Wu, Venediktov and Huang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Investigating the rate of turbidity impact on underwater spectral reflectance detection

Hong Song [1], Syed Raza Mehdi [1], Zixin Li[1], Mengjie Wang[1], Chaopeng Wu[1], Vladimir Yu Venediktov[2] and Hui Huang[1,3]*

[1]Department of Marine Engineering, Ocean College, Zhejiang University, Zhoushan, Zhejiang, China, [2]Faculty of Physics, St.-Petersburg State Electrotechnical University "LETI", St.-Petersburg, Russia, [3]Hainan Institute, Zhejiang University, Sanya, Hainan, China

Spectral reflectance detection of the targeted object is considered a vital inherent optical property for its potential to provide abundant spectral information, which is crucial in underwater spectral imaging. However, the coarse condition of the underwater environment due to turbidity causes extreme distortions in spectral reflectance detection due to the high absorption and scattering of light. To cope with the effects of light degradation on underwater spectral reflectance detection accuracy, the rate of the impacts of turbidity on spectral reflectance should be examined thoroughly. Therefore, we utilize a stare-type underwater spectral imaging system based on a liquid crystal tunable filter (LCTF) to study the effects of turbidity in underwater spectral imaging of various colored bodies. To examine the accuracy of underwater spectral reflectance detection based on escalating turbidity, the paper models the rate of increase in scattering intensity of the water body. Results show that, based on the non-linear increase in the pixel response of the black and white board, the rapid upsurge in scattering intensity occurs between 400nm to 500nm at different turbidity levels. Additionally, the spectral reconstruction of color bodies relative to the black and white board shows the maximum absolute deviation of 5.3% in spectral reflectance detection accuracy under varying turbidity. While employing underwater spectral imaging, the above findings of optimal band selection can find significant applications to improve the quality of underwater object detection.

KEYWORDS

spectral reflectance, spectral imaging, turbidity, spectral detection, liquid crystal tunable filters, spectral features, underwater scattering

1 Introduction

Since the 1980s, optical-based marine observation has become a vital technology in various marine sciences research (Jian et al., 2021; Shahani et al., 2021). However, the techniques for marine observation are still very limited due to the minimal knowledge of underwater environment conditions. For instance, turbidity limits the light penetration capacity to the depths of any water body due to absorption and scattering by total suspended particles (Tananaev and Debolskiy, 2014). This leads to underwater light degradation which poses significant challenges in underwater observation through optical imaging (Zhang et al., 2019); resulting in a considerable amount of undiscoverable information.

Meanwhile, the necessity for adequate underwater imaging systems for marine observation is pushing scientific horizons on multiple fronts (Akiba and Kakui, 2000). In the past decade, the concept of underwater spectral imaging technology has been developed by combining two-dimensional spatial information with one-dimensional spectral data, revealing the spectral features hidden in the narrow band (Polerecky et al., 2009). Because of its advantages of high spectral and spatial resolution, it has shown great potential in remote sensing applications (Yasir et al., 2023), exploration and mapping of seabed minerals, investigation of seabed ecological environment (Johnsen et al., 2013; Liu et al., 2020), marine archaeology (Ødegård et al., 2018), and classification of marine species such as sponges and corals (Song et al. 2021b).

Currently, a variety of underwater spectral imagers have been applied to various underwater detection applications. In 2009, the Norwegian University of science and technology and Ecotone developed a push broom underwater spectral imager for spectral detection of corals, submarine pipelines, algae, and mining ships (Johnsen et al., 2016). In 2016, another underwater spectral imaging was developed based on the rotary spectral imager, the system used 31 filters with different transmission bands to acquire spectral images in various bands. The system has been tested within a 50 m underwater depth range and carried out short-range spectral imaging research of coral reefs (Guo et al., 2016). In 2018, Zhejiang University cooperated with the Helmholtz Institute of Oceanography in Germany to develop a multispectral imaging system to monitor the underwater ecosystem (Liu et al., 2018). More recently, a push broom underwater hyperspectral imager was developed, which demonstrated high stability and a good detection effect for large-scale targets (Xue et al., 2021).

Distinguishing wavelengths in different spectral bands lies at the core of spectral imaging technology. Spectroscopic technology directly affects the structural complexity, volume, weight, and performance of the spectral imaging system. According to spectroscopy principles, spectra generation methods can be divided into dispersion, interference, computed tomography, coded aperture, and filter types. A liquid crystal tunable filter (LCTF) can switch its central wavelength through electrical signal control without moving mechanical structure, simplify the system structure, and improve reliability, response speed, and automation. Based on a stare-type underwater spectral imager based on LCTF (Song et al., 2021a), this paper explores the influence of different water turbidity and lighting conditions on the accuracy of spectral reflectance detection methods.

Spectral reflectance is an inherent optical property of an object widely used for target identification. Using appropriate methods to reconstruct the spectral reflectance of the targeted objects is the premise of spectral analysis and spectral image processing. There are two ways to reconstruct the spectral reflectance of underwater spectral images: first is the spectral irradiation calibration of the imaging and lighting system to obtain the spectral irradiation energy distribution of incident light. The obtained absolute spectral radiation energy is then inverted according to the pixel response of the target position in the spectral image resulting in the spectral irradiation energy of reflected light at the target position. Finally, the ratio of reflected light spectral irradiation energy to incident light spectral irradiation energy is taken as the spectral reflectance (Ping et al., 2017). The second spectral reflectance reconstruction method is a comparative measurement method reliant upon the known spectral reflectance of the standard target (Xia et al., 2018). In this paper, a method of underwater spectral reflectance reconstruction of a black and white board based on a comparative measurement method is proposed: the spectral reflectance of the target is reconstructed by introducing standard reference objects (standard black and white board with known spectral reflectance). This method synchronously collects the spectral image data of a standard whiteboard, blackboard, and target, and reconstructs the spectral reflectance of the target object through post-processing.

2 The underwater spectral imaging system

Figure 1 shows the underwater spectral imaging system based on LCTF. The imaging system takes LCTF as the core element spectrum splitter element, and the internal optical structure design mainly adopts the form of “imaging lens + LCTF + image sensor”. Compared with the “filter + imaging lens + image sensor” system, the above-discussed structure makes the spectral imaging system more compact and smaller. The imaging lens adopts a large-diameter electric focusing lens, which can be focused by a stepping motor. Using LCTF as a spectral device, an electrical signal can control the transmission wavelength of LCTF to generate narrowband scanning in the range of visible light. A scientific-grade charged coupled device (CCD) is used as an image sensor to improve imaging sensitivity in the underwater environment. Later, the underwater spectral imaging system is sealed for waterproofing followed by the optical system installation and underwater pressure test analysis. To ensure the accuracy and reliability of the spectral imaging system, imaging parameters such as spectral resolution, spectral response sensitivity, and underwater detection sensitivity are tested and calibrated with a standard monochromator and spectrometer (Song et al. 2021a).

The underwater spectral imaging system isolates the water body through the compressive cabin glass. Since the optical medium on

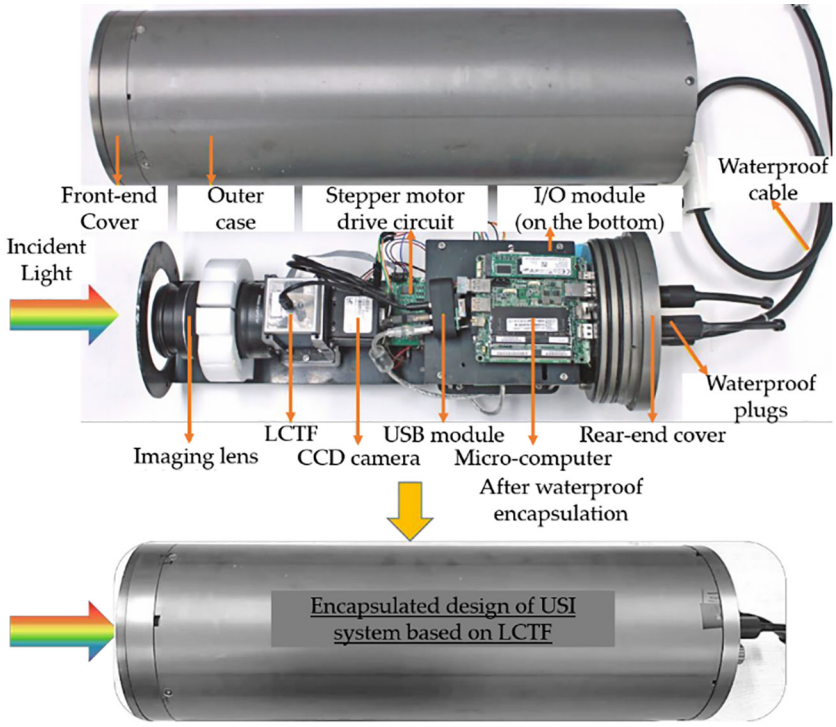


FIGURE 1
Encapsulated structure of the designed USI system. The hardware includes optical (lens, LCTF, and imaging sensor) and control (microcomputer, motor drive circuit, remote I/O circuitry, and voltage converter) modules installed in a waterproof mechanical structure (Song et al. 2021a).

the sides of compressive glass is water and air respectively, the refraction of the water body and the glass window must be considered in imaging. To simplify the model and ignore the influence of the illumination source, it is assumed that the self-luminous target is used for underwater spectral imaging, and the light emitted by the target is refracted twice by water-optical glass-air interface before incident into the imaging lens. As shown in Figure 2, the optical path of underwater spectral imaging can be equivalent to imaging the object plane (refracted virtual plane in case of underwater) in the air after removing the optical glass. Therefore, the image distance z_w in the imaging model can be defined as:

$$z_w = l_a + \frac{l_g}{n_g} + \frac{l_w}{n_w}$$

where l_a is the distance between the optical center of the lens and the inner surface of the optical glass, l_g is the thickness of the optical glass, l_w is the distance between the target plane and the outer surface of the optical glass, n_g is the refractive index of the optical glass (that is approximately equal to 1.49 for acrylic glass), and n_w is the refractive index of the water body (that is 1.33).

Defining the spatial position coordinate on the object plane of a point on the underwater target as $P(x_w, y_w)$, focal length of the camera f, and the distance between image plane and optical center of the lens z_w. The position coordinates of the conjugate point on the image plane of the image sensor are recorded as $P(x'_w, y'_w)$, the three conjugate coordinate are defined as:

$$x'_w = -\frac{f}{z_w - f} x_w$$

$$y'_w = -\frac{f}{z_w - f} y_w$$

$$z'_w = -\frac{f}{z_w - f} z_w$$

3 Experimental setup

The spectral reflectance curve information of the object surface can be used as important feature information of object recognition. In the air, the spectral reflectance reconstruction generally adopts the comparison measurement method: the spectral reflectance and spectral response value of the standard whiteboard are $R_w(\lambda_c)$ and I_w, respectively, and the spectral response value of the target is I_c, then the spectral reflectance of the target $R_c(\lambda_c)$ is as follow;

$$R_c(\lambda_c) = \frac{I_c}{I_w} R_w(\lambda_c)$$

In water, it is difficult to detect the spectral reflectance of the target. Especially in turbid water, suspended particles have a significant scattering effect on light, resulting in blurred imaging and a considerable loss of color and other characteristic information. To mitigate these distortions, this work proposes a

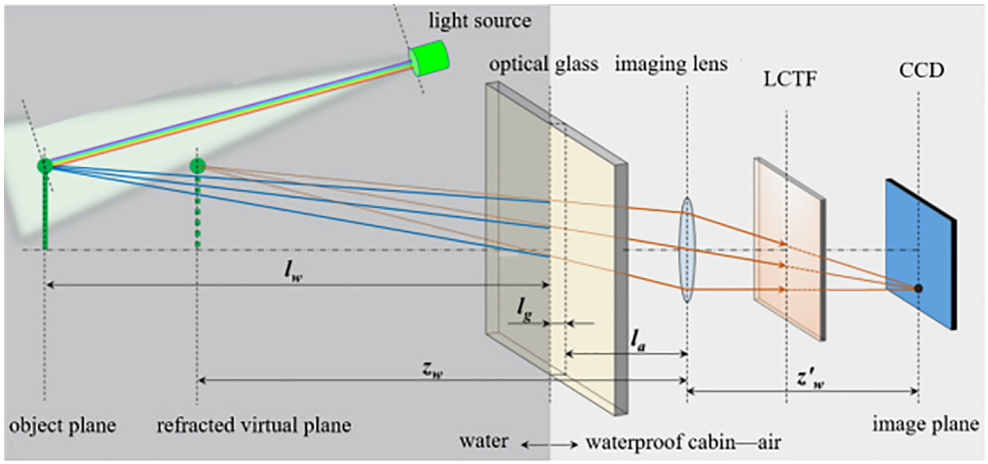


FIGURE 2
Schematics of underwater spectral imaging and spectral reflectance measurement by underwater spectral imaging system based on LCTF.

method of underwater spectral reflectance reconstruction of a black and white board based on a comparative measurement method. Standard black and white board is introduced as correction plates to reconstruct the spectral reflectance of targets in the water. The spectral reflectance and spectral response of the whiteboard are $R_w(\lambda_c)$ and I_w, respectively. The spectral reflectance and spectral response of the blackboard are $R_b(\lambda_c)$ and I_b, respectively, and the spectral response of the target is I_c,then the spectral reflectance of the underwater targeted object $R_c(\lambda_c)$ is as follows;

$$R_c(\lambda_c) = \frac{I_c - I_b}{I_w - I_b} R_w(\lambda_c) - \frac{I_w - I_c}{I_w - I_b} R_b(\lambda_c)$$

3.1 Spectral imaging setup

The bandwidth of the absorption peak or reflection peak caused by the spectral reflection of the underwater target is generally about 5nm to 50nm. In this work, the underwater spectral imager based on LCTF collects the spectral image data of the target in 2nm steps and extracts the spectral response curve of any pixel or combination of pixels in the image space.

Due to the dominant Rayleigh scattering of turbid water in the test tank, the optical thickness is 1.155, and the extinction coefficient is considered equivalent to seawater. Figure 3 shows the details of experimental device for measuring the spectral reflectance of

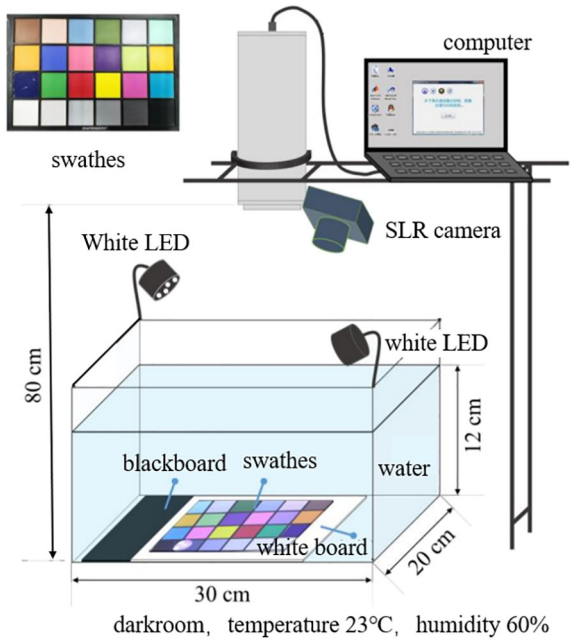


FIGURE 3
Schematic diagram of the experimental setup to measure the underwater spectral reflectance by LCTF-based spectral imaging system under varying conditions of turbidity and illumination.

underwater targets. The setup quantitatively analyze the influence of water turbidity and illumination conditions on the measurement of spectral reflectance and then identify the accuracy of underwater spectral reflectance reconstruction method. The execution of the experimental process in the darkroom environment is as follows: initially, the underwater target object (i.e. white, black and color board) is illuminated with two white white light emitting diaodes (LED) sources (refered as illumination A). After setting the focus and exposure time, the spectral images of target boards were collected in 151 wave bands ranging from 400 nm to 700 nm with 2 nm step difference. With the addition of certain amount of milk emulsion, the degree of turbidity of water body is varied from minimum to maximum of 51 NTU to simulate natural turbid water body. The results were calculated for 0 NTU, 11 NTU, 12 NTU, 21 NTU, 39 NTU, and 51 NTU. The experiment is repeated several times to collect spectral image data under varying water turbidity. Furthermore, the spectral data of target boards were also collected under the illumination of two blue LED light sources (refered as illumination B).

3.2 Data processing

For the spectral image collected at a particular central wavelength, the pixel response is related to the response bandwidth of the system and the spectral characteristics of each device within the bandwidth. Assuming that the spectral characteristics are evenly distributed within the bandwidth, the pixel response $I(x, y, \lambda_c)$ is:

$$I(x,\ y,\ \lambda_c) = K(\lambda_c)(R(\lambda_c)E_s(\lambda_c)e^{-c(\lambda_c)l_w} + E_t(\lambda_c))$$

where,(x,y) represents the pixel position in the image corresponding to a specific pixel,λ_c signifies the central wavelength of the band corresponding to the spectral image,$K(\lambda_c)$ characterizes the mapping relationship from the spectral energy to the pixel response,$R(\lambda_c)$ embodies the spectral reflectance of the target,$E_s(\lambda_c)$ represents the incident spectral radiation energy at the target position,$E_t(\lambda_c)$ denotes the water body or water surface stray light entering the image surface,$c(\lambda_c)$ shows the spectral attenuation coefficient of the water body, andl_w represents the water depth.

The data processing technique involves denoising all the spectral images of the whiteboard, blackboard, and color board. Seven areas of equally spaced array distribution are selected on the color board. The size of each area is 5 × 5 pixels. Median pixel response as the pixel response of the color board is considered. Similarly, pixel spectral responses of the white and blackboard are extracted. The pixel response of the pixel area in row i^{th} and column j^{th} of the color block can be recorded as I_c, the pixel response at the exact position of the whiteboard spectral image can be recorded as I_w, and the pixel response of the blackboard spectral image can be recorded as I_b. Since the pixel spectral response is in the same position, it is approximately considered that $K(\lambda_c)$, $E_s(\lambda_c)$, $c(\lambda_c)$, and $E_t(\lambda_c)$ remain unchanged, and the spectral reflectance corresponding to the pixel region is as the spectral reflectance of the underwater targeted object $R_c(\lambda_c)$.

4 Results and discussion

As shown in Figure 4, under illumination A, the ratio of spectral image pixel responses of the blackand whiteboard at the same position represents the spectral response of the system to the water body under different turbidity. When the turbidity of the water body is minimum or below 4 NTU, the black and whiteboard pixel response ratio is low and stable in the range of 450 nm to 700 nm, and the average value is 3.9%. When the water turbidity gradually increases, the concentration of emulsion dispersed particles in the water body increases so that the scattering effect of the water body is significantly enhanced. The proportion of the scattered light signal received by the CCD image plane is increased dramatically compared with the reflected signal of the target, resulting in the non-linear increase of the pixel response ratio of the black and whiteboard and the scattering effect of the water body in the range of 400 nm to 500 nm is enhanced rapidly.

As shown in Figure 5, under illumination A in clear water, the spectral image data of the color plate collected by the system are corrected by the black and white board to obtain the spectral reflectance data of different color blocks. The spectral reflectance of blocks of various hues, such as red, yellow, green, and blue, has varying spectral characteristics. It is also noticed that the color

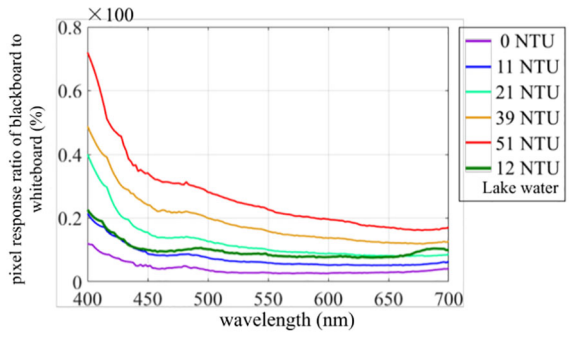


FIGURE 4
Pixel-based spectral response ratio of blackboard and whiteboard under varying turbidity conditions of the water body.

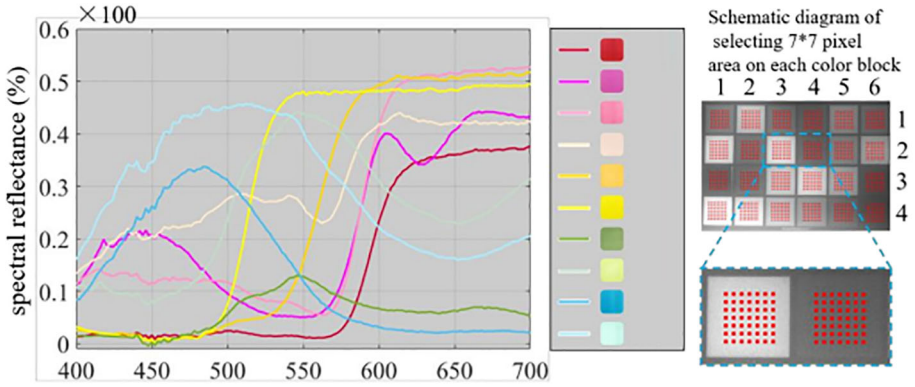


FIGURE 5
Pixel-based spectral reflectance detection of different color blocks in water at turbidity below 4NTU.

blocks with different saturation and lightness in the same hue also have different spectral characteristics. For example, dark red color blocks have high and stable spectral reflectance in the wavelength band of 630 nm to 700 nm, with an average of 36.6%. The spectral reflectance curve of the pink color block has three peak characteristics, and the peak wavelengths are 446 nm, 604 nm, and 672 nm respectively. The light pink color block has high and stable spectral reflectance in the wave band of 610 nm to 700 nm, with an average of 51.9%; The yellow color block has high and stable spectral reflectance in the 540 nm to 700 nm band, with an average of 48.2%.

Furthermore, the spectral images are collected under illumination A and B in varying turbidity of minimum turbidity as 0 NTU and 11 NTU, as in Figure 6, and the spectral reflectance of the color plate is calculated. Under the turbidity of 11 NTU, the maximum absolute deviation of the spectral reflectance of the color plate under illumination condition A is 1.9%, and the maximum relative deviation is 4.1%. The maximum absolute deviation of the spectral reflectance of the color plate under light condition B from the benchmark is 2.3%, and the maximum relative deviation is 4.9%. In the lake water (collected from Minghui Lake in Zhoushan campus of Zhejiang University, with a turbidity of 11 NTU), compared with the benchmark, the maximum absolute deviation and the spectral reflectance of the color plate under illumination B are 2.8% and 6.1% respectively.

As shown in Figure 7, under illumination A, the turbidity of the water body is gradually increased, and the color plate's spectral images are collected simultaneously. The spectral reflectance curves of different color blocks under varying turbidity conditions can be obtained through black and whiteboard correction. When the water turbidity is lower, the results of spectral reflectance of the same color block under different water turbidity are close to the same. However, when the water turbidity is high, spectral reflectance fluctuates, and the maximum absolute deviation of 5.3% is noted. It is also observed that the overall spectral characteristics of color plates have not changed. Through the black and whiteboard correction, underwater targets' spectral reflectance reduction results are close to the actual values, and the changes in lighting conditions and water turbidity have little impact on spectral reflectance.

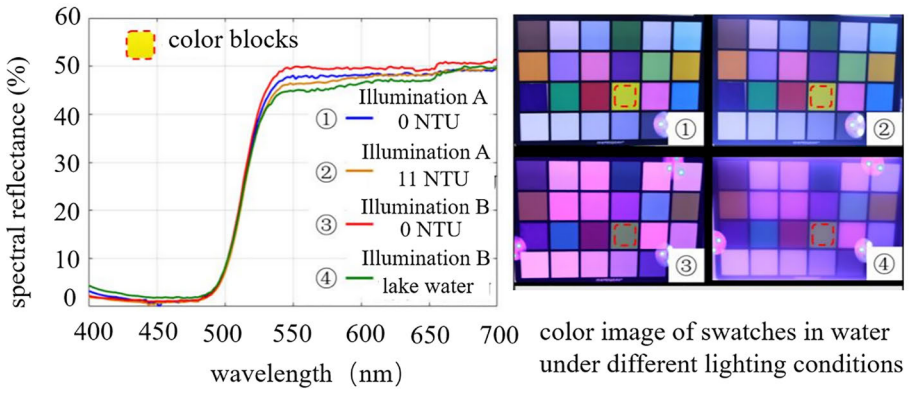


FIGURE 6
Explanation of spectral reflectance detection of specific color block; yellow patch under varying turbidity and illumination.

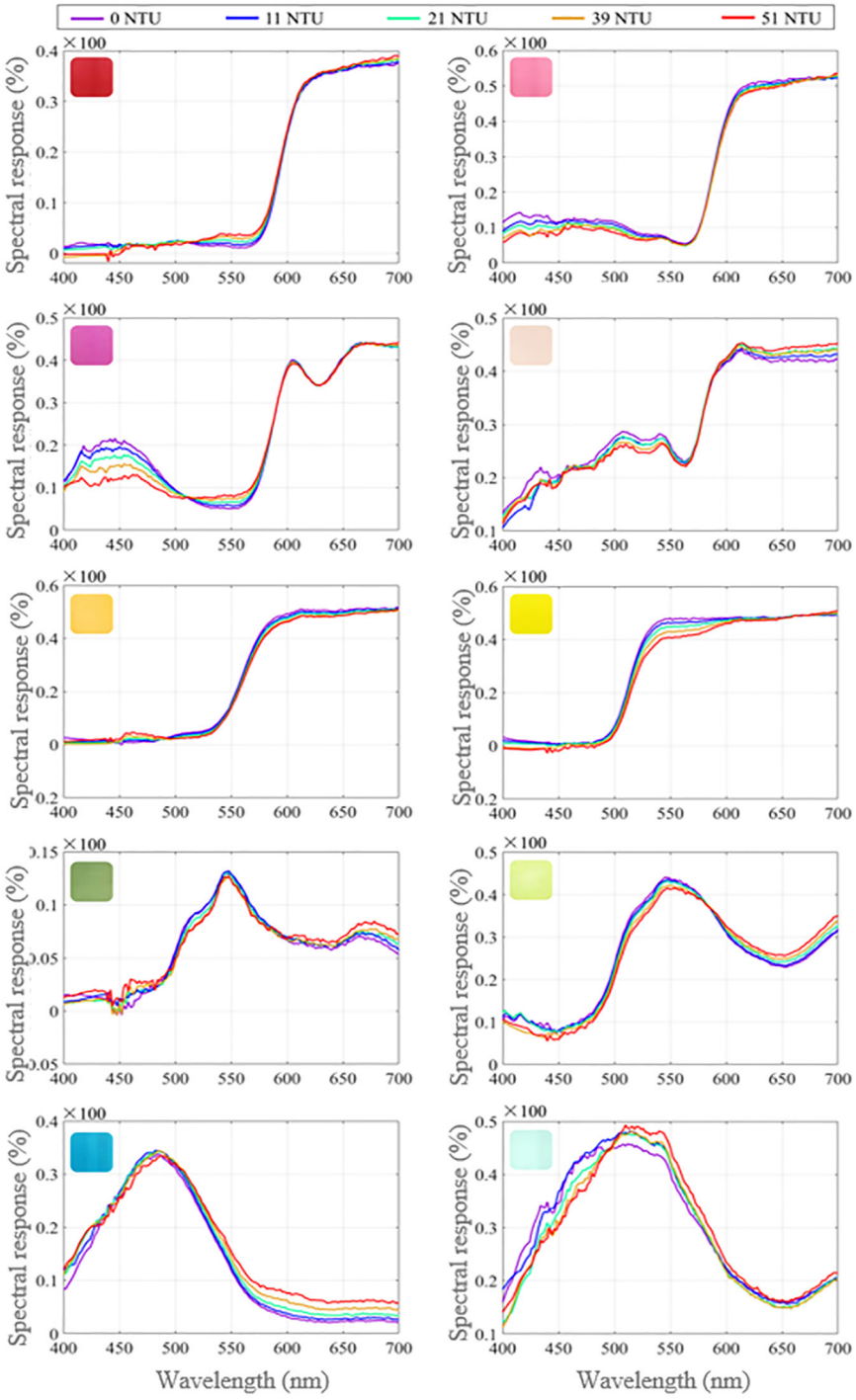


FIGURE 7
Spectral reflectance detection of color blocks after the spectral reconstruction process through black and white board correction technique.

5 Conclusion

In this work, underwater spectral imaging is observed under the impact of varying turbidity of water body. The paper also described the spectral reflectance reconstruction method of the underwater target by utilizing the underwater spectral imaging system based on LCTF. In order to overcome the influence of water turbidity causing light scattering on the measurement of spectral reflectance, a comparison-based black and white board correction method is introduced for spectral reconstruction. The experiments showed that by changing the water turbidity and using various illumination conditions, the spectral reflectance measurement method based on the black and white board correction method could be applied. The spectral response and reflectance curves of several targets were accurately obtained. The results also verified the peak spectral response of different underwater color bodies in the wave

spectrum in the varying turbidity of water body. The system obtained the maximum absolute deviation of 5.3% for various colors under varying turbidity conditions. The reliability and accuracy of the data obtained by the underwater spectral imaging system based on LCTF can be further utilized in the applications of underwater spectral imaging for target detection. The significance of this work lies in the fact that it can enable researchers to determine the possible spectral range needed for a specific degree of turbidity thereby enhancing the quality of underwater spectral reflectance detection.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

HS, SRM, and HH contributed to the conceptualization of the experiment. SRM, ZL, MW, and CW devised and conducted the experiment and organized the data processing. HH and SRM wrote the original manuscript. SRM and VV conducted the comprehensive revision. All authors contributed to the article and approved the submitted version.

Funding

This study was financially supported by the Key Research and Development Plan of Zhejiang Province, China (Grant number: 2020C03012, 2021C03181).

Acknowledgments

The authors present special gratitude to the Intelligent Underwater Optics laboratory team of Ocean College, Zhejiang University, for their assistance during the work.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Ødegård, Ø., Mogstad, A. A., Johnsen, G., Sørensen, A. J., and Ludvigsen, M. (2018). Underwater hyperspectral imaging: A new tool for marine archaeology. *Appl. Opt.* 57 (12), 3214–3223. doi: 10.1364/AO.57.003214

Akiba, T., and Kakui, Y. (2000). Design and testing of an underwater microscope and image processing system for the study of zooplankton distribution. *IEEE J. Ocean. Eng.* 25 (1), 97–104. doi: 10.1109/48.820741

Guo, Y., Song, H., Liu, H., Wei, H., Yang, P., Zhan, S., et al. (2016). Model-based restoration of underwater spectral images captured with narrowband filters. *Opt. Express* 24 (12), 13101–13120. doi: 10.1364/OE.24.013101

Jian, M., Liu, X., Luo, H., Lu, X., Yu, H., and Dong, J. (2021). Underwater image processing and analysis: A review. *Signal Process.: Image Commun.* 91, 116088. doi: 10.1016/j.image.2020.116088

Johnsen, G., Ludvigsen, M., Sørensen, A., and Aas, L. M. S. (2016). The use of underwater hyperspectral imaging deployed on remotely operated vehicles-methods and applications. *IFAC-PapersOnLine* 49 (23), 476–481. doi: 10.1016/j.ifacol.2016.10.451

Johnsen, G., Volent, Z., Dierssen, H., Pettersen, R., Van Ardelan, M., Søreide, F., et al. (2013). "Underwater hyperspectral imagery to create biogeochemical maps of seafloor properties," in *Subsea optics and imaging* (Woodhead Publishing), 508–540e.

Liu, B., Liu, Z., Men, S., Li, Y., Ding, Z., He, J., et al. (2020). Underwater hyperspectral imaging technology and its applications for detecting and mapping the seafloor: A review. *Sensors* 20 (17), 4962. doi: 10.3390/s20174962

Liu, H., Sticklus, J., Köser, K., Hoving, H. J. T., Song, H., Chen, Y., et al. (2018). TuLUMIS-a tunable LED-based underwater multispectral imaging system. *Opt. express* 26 (6), 7811–7828. doi: 10.1364/OE.26.007811

Ping, Y., Yilu, G., He, W., Dan, S., Hong, S., Yunfei, Z., et al. (2017). Method for spectral restoration of underwater images: Theory and application. *Infrared Laser Eng.* 46 (3), 323001–0323001. doi: 10.3788/IRLA201746.0323001

Polerecky, L., Bissett, A., Al-Najjar, M., Faerber, P., Osmers, H., Suci, P. A., et al. (2009). Modular spectral imaging system for discrimination of pigments in cells and microbial communities. *Appl. Environ. Microbiol.* 75 (3), 758–771. doi: 10.1128/AEM.00819-08

Shahani, K., Song, H., Mehdi, S. R., Sharma, A., Tunio, G., Qureshi, J., et al. (2021). Design and testing of an underwater microscope with variable objective lens for the study of benthic communities. *J. Mar. Sci. Appl.* 20 (1), 170–178. doi: 10.1007/s11804-020-00185-9

Song, H., Mehdi, S. R., Wu, C., Li, Z., Gong, H., Ali, A., et al. (2021a). Underwater spectral imaging system based on liquid crystal tunable filter. *J. Mar. Sci. Eng.* 9 (11), 1206. doi: 10.3390/jmse9111206

Song, H., Mehdi, S. R., Zhang, Y., Shentu, Y., Wan, Q., Wang, W., et al. (2021b). Development of coral investigation system based on semantic segmentation of single-channel images. *Sensors* 21 (5), 1848. doi: 10.3390/s21051848

Tananaev, N. I., and Debolskiy, M. V. (2014). Turbidity observation in sediment flux studies: Example from Russian rivers in cold environments. *Geomorphology* 218, 63–71. doi: 10.1016/j.geomorph.2013.09.031

Xia, W., Ning-fang, L., Ya-sheng, L., Hao-bo, C., Bin, C., Wen-ming, Y., et al. (2018). The research on reconstruction of spectral reflectance in LCTF imaging system based on comparative measurement. *Spectrosc. Spectral Anal.* 38 (1), 290–295.

Xue, Q., Tian, Z., Yang, B., Liang, J., Li, C., Wang, F., et al. (2021). Underwater hyperspectral imaging system using a prism–grating–prism structure. *Appl. Opt.* 60 (4), 894–900. doi: 10.1364/AO.415351

Yasir, M., Shanwei, L., Mingming, X., Hui, S., Hossain, M. S., Colak, A. T. I., et al. (2023). Multi-scale ship target detection using SAR images based on improved Yolov5. *Front. Mar. Sci.* 9, 1086140. doi: 10.3389/fmars.2022.1086140

Zhang, W., Dong, L., Pan, X., Zou, P., Qin, L., and Xu, W. (2019). A survey of restoration and enhancement for underwater images. *IEEE Access* 7, 182259–182279. doi: 10.1109/ACCESS.2019.2959560

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 17 March 2023
DOI 10.3389/fmars.2023.1135356

Check for updates

OPEN ACCESS

EDITED BY
Hong Song,
Zhejiang University, China

REVIEWED BY
Renliang Huang,
Tianjin University, China
Bing Chen,
Memorial University of Newfoundland, Canada
Jianliang Xue,
Shandong University of Science and Technology, China
Yuanzhi Zhang,
The Chinese University of Hong Kong, China

*CORRESPONDENCE
Junfang Yang
yangjunfang@upc.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 31 December 2022
ACCEPTED 06 March 2023
PUBLISHED 17 March 2023

CITATION
Yang J, Hu Y, Zhang J, Ma Y, Li Z and Jiang Z (2023) Identification of marine oil spill pollution using hyperspectral combined with thermal infrared remote sensing. *Front. Mar. Sci.* 10:1135356. doi: 10.3389/fmars.2023.1135356

COPYRIGHT
© 2023 Yang, Hu, Zhang, Ma, Li and Jiang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Identification of marine oil spill pollution using hyperspectral combined with thermal infrared remote sensing

Junfang Yang[1,2,3*†], Yabin Hu[4†], Jie Zhang[1,2,3,4], Yi Ma[2,4], Zhongwei Li[1] and Zongchen Jiang[5]

[1]College of Oceanography and Space Informatics, China University of Petroleum, Qingdao, China, [2]Technology Innovation Center for Ocean Telemetry, Ministry of Natural Resources, Qingdao, China, [3]Technology Innovation Center for Maritime Silk Road Marine Resources and Environment Networked Observation, Ministry of Natural Resources, Qingdao, China, [4]First Institute of Oceanography, Ministry of Natural Resources, Qingdao, China, [5]School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, China

The types of marine oil spill pollution are closely related to source tracing and pollution disposal, which is an important basis for oil spill pollution punishment. The types of marine oil spill pollution generally include different types of oil products as well as crude oil and its emulsions in different states. This paper designed and implemented two outdoor oil spill simulation experiments, obtained the hyperspectral and thermal infrared remote sensing data of different oil spill pollution types, constructed a hyperspectral recognition algorithm of oil spill pollution type based on classical machine learning, ensemble learning and deep learning models, and explored to improve the identification ability of hyperspectral oil spill pollution type by adding thermal infrared features. The research shows that hyperspectral combined with thermal infrared remote sensing can effectively improve the recognition accuracy of different oils, but thermal infrared remote sensing cannot be used to distinguish crude oil and high concentration water-in-oil emulsion. On this basis, the recognition ability of hyperspectral combined with thermal infrared for different oil film thicknesses is also discussed. The combination of hyperspectral and thermal infrared remote sensing can provide important technical support for emergency response to maritime emergencies and oil spill monitoring business of relevant departments.

KEYWORDS

marine oil spills, hyperspectral remote sensing, thermal infrared remote sensing, oil pollution types identification, deep learning

1 Introduction

Marine oil spill is the most typical and serious environmental pollution accident in the process of marine development. Various oil spills pollution types on the sea not only damage the marine environment and coastline ecology, pollute fishery resources, endanger marine food security, affect tourism, but also threaten human health, and even hinder the

healthy development of the marine economy (Washburn et al., 2018; Silva et al., 2022; Wang et al., 2022). It is difficult to eliminate negative impacts of major marine oil spills on the marine ecological environment in a short time.

In general, the types of marine oil spill pollution include different types of oil products as well as crude oil and its emulsions in different states. There are five typical oil products: 1) Crude oil, such as the oil product leaked in the Deepwater Horizon platform oil spill accident in the Gulf of Mexico in 2010 (Leifer et al., 2012) and the Penglai 19-3C platform oil spill accident in 2010 (Yang et al., 2019); 2) Fuel oil, that is, the fuel used for large ship engines, such as the oil product leaked in the oil tanker grounding accident in southeast Mauritius in 2020 (Rajendran et al., 2021); 3) Condensate, also known as natural gasoline, is similar to gasoline, such as the oil product leaked from the East China Sea oil tanker collision accident in 2018 (Lu et al., 2019a); 4) Vegetable oil, such as palm oil with the largest production, consumption and international trade volume in the world (Yang et al., 2021); 5) Diesel oil, that is, the fuel used for high-speed diesel engine for small ships. Crude oil and its emulsions in different states include non-emulsified crude oil, water-in-oil emulsion and oil-in-water emulsion (Lu et al., 2019b; Lu et al., 2020). Once the spilled oil on the sea surface is not removed in time, a series of complex physical and chemical changes, such as diffusion, drift, emulsification, evaporation, dissolution, adsorption precipitation, photooxidation and biodegradation, will occur under the combined action of wind, wave, current and other environmental dynamics, forming water-in-oil (WO) and oil-in-water (OW) emulsions of different concentrations (Lu et al., 2013a).

The types of marine oil spill pollution are closely related to source tracing and pollution disposal, which is an important basis for oil spill pollution punishment. Different types of oil products, crude oils and its emulsions in different states need to adopt different emergency treatment strategies, such as combustion elimination, oil absorption felt adsorption, dispersant spraying, skimmer recovery, etc (Zhong and You, 2011; French-McCay et al., 2022). Timely identification of different types of marine oil spill pollution is of great significance for marine oil spill monitoring and emergency response.

Optical remote sensing and microwave remote sensing are important means for marine oil spill monitoring (Fingas and Brown, 2014; Lu et al., 2016a; Pärt et al., 2021). Synthetic Aperture Radar (SAR) is the most commonly used technology for marine oil spill monitoring in microwave remote sensing, which has the advantages of all-day, all-weather, and can also successfully obtain images under cloudy and rainy conditions. Therefore, SAR has become the primary means for relevant business departments to monitor marine oil spills (Velotto et al., 2011; Marghany, 2014; Alpers et al., 2017; Mdakane and Kleynhans, 2020; Ma et al., 2021; Zhu et al., 2021). However, SAR also has a higher false alarm rate in detecting marine oil spills, and it is difficult to identify oil spill pollution types and estimate the oil film thickness. The oil spill information contained by optical remote sensing is more abundant than SAR, and has been studied and applied widely in oil spill monitoring (Shi et al., 2018; Shen et al., 2020; Hu et al., 2021; Yang et al., 2022b). Among them, hyperspectral remote sensing has higher spectral resolution, and can extract rich subtle features among different oil spill pollution types. Therefore, it has been widely concerned about oil spill monitoring using hypersectral remote sensing (Lu et al., 2013b; Cui et al., 2017; Zhu et al., 2019; Zhu et al., 2019; Li et al., 2020; Wang B. et al., 2021), especially oil spill pollution type identification (Wettle et al., 2009; Yang et al., 2020; Jiang et al., 2021; Lai et al., 2021; Li et al., 2021; Yang et al., 2022a), but is vulnerable to the impact of sun flare in complex marine environment, which lead to the oil film boundary is not clearly identified (Sun and Hu, 2016; Lu et al., 2016b; Duan et al., 2020). The infrared emissivity of seawater and oil film is different, and the oil-water boundary is well identified. Thermal infrared remote sensing is almost not affected by the change of light, which can realize the detection of marine oil spill and the inversion of oil film thickness, but it is difficult to distinguish the type of oil spill emulsion (Jing et al., 2011; Wang and Hu, 2015; Lu et al., 2016c; Guo et al., 2020; Jiao et al., 2021; Li et al., 2022). The comparison of marine oil spill monitoring capabilities of various remote sensing technologies is listed in Table 1.

Most of the optical remote sensing monitoring of marine oil spill is carried out based on a single sensor type, which has certain limitations and cannot meet the needs of accurate monitoring. Recently, more and more scholars pay attention to combining the advantages of different sensors to effectively improve the remote sensing monitoring ability of marine oil spill, and have made certain achievements (Mohammadi et al., 2021; Rajendran et al., 2021). Lu et al., (2019a) used China's GF-3 SAR data to delineate suspected oil spill areas in the East China Sea "SANCHI" collision oil spill event in January 2018. The multispectral data of Sentinel-2 satellite is used to carry out optical remote sensing detection of marine oil spill, and further carried out optical remote sensing identification and classification of oil spill pollution types based on the analysis of spectral response characteristics of oil spill simulation experiments, which is mutually verified with the suspected oil spill monitoring results of GF-3 SAR. Wang L. F. et al. (2021) proposed a new method to determine the oil film area using the fusion of visible light and thermal infrared images. This method integrates the advantages of visible light and thermal infrared images, and can accurately determine the oil film area under different lighting conditions, with an average error of 2.78%. Wang and Gao (2020) used SAR and laser fluorescence sensors to carry out airborne platform oil spill monitoring. First, they used SAR for long-distance and large-scale detection. Once the suspected oil spill area was detected, they used laser fluorescence sensor for close detection. The combination of SAR and laser fluorescence sensor greatly improved the detection effect and recognition ability of marine pollutants in large areas.

To sum up, this paper designed and implemented two outdoor oil spill simulation experiments, carried out the extraction and analysis of hyperspectral and thermal infrared optical features of different oil spill pollution types, introduced classification algorithms such as classical machine learning, ensemble learning and deep learning, studied and constructed hyperspectral oil spill pollution type recognition algorithms, and explored how to improve the ability of hyperspectral oil spill pollution type identification by adding thermal infrared features.

TABLE 1 Comparison of oil spill monitoring capability of various remote sensing technologies.

Remote sensing technologies	Advantages	Disadvantages
SAR	all-day, all-weather and not covered by cloud	high false alarm rate, unable to identify the type of oil spill pollution and estimate the oil film thickness
Multispectral remote sensing	wide space coverage, low cost, able to distinguish heavy oils from light oils and estimate the approximate oil film thickness	low spectral resolution, vulnerable to sunglint interference
Hyperspectral remote sensing	high spectral resolution, able to identify different oil products and crude oil emulsions, and reverse oil film thickness	not applicable to large-scale marine oil spill monitoring, vulnerable to sunglint interference
Ultraviolet remote sensing	sensitive to thiner oil film	low spatial resolution, vulnerable to interference from sunlight and bio-oil film, unable to identify non-emulsified oil and oil-water emulsions
Thermal infrared remote sensing	all-day, sensitive to thicker oil film, not disturbed by the sunglint	unable to identify non-emulsified oil and oil-water emulsions, vulnerable to interference from targets with similar thermal properties to oil film
Laser radar	active remote sensing, able to identify different oil products, and reverse oil film thickness	unable to detect very thick oil film, vulnerable, and expensive

2 Materials and methods

2.1 Experiments and materials

This part mainly introduces two outdoor oil spill simulation experiments. One is the experimental scenario (Figures 1A, B) for different types of oil products (Figure 2A), namely crude oil, fuel oil, palm oil, diesel oil and gasoline. In September 2020, the experiment was conducted in the pool (45 m× 40 m) of Qingdao Nanjiang Wharf, five kinds of oil products are distributed in the enclosure made of PVC boards, and the seawater depth is 1.2 m. The second is the experimental scenario (Figures 1C, D) for crude oil and its emulsions in different states (Figure 2B), namely, non-emulsified crude oil, WO emulsion and OW emulsion, which was carried out in Aoshanwei experimental base in Qingdao in May 2022. High speed disperser and emulsifier were used to prepare WO emulsion and OW emulsion with different volume concentrations. The preparation method was referred to (Lu et al., 2019). Crude oil with different thickness and crude oil emulsions with different volume concentration are distributed in the enclosure made of black PVC boards. The oil products information used in the experiment is shown in Table 2.

2.2 Data and preprocessing

2.2.1 Remote sensing data

For the experimental scenarios of different types of oil products, the Cubert S185 airborne hyperspectral imager and DJI "Yu" 2 airborne thermal infrared camera were used to obtain approximately synchronous hyperspectral images and thermal infrared images of different types of oil products (Figures 3A, B). For the experimental scenes of crude oil and its emulsions in different states, the Cubert S185 airborne hyperspectral imager

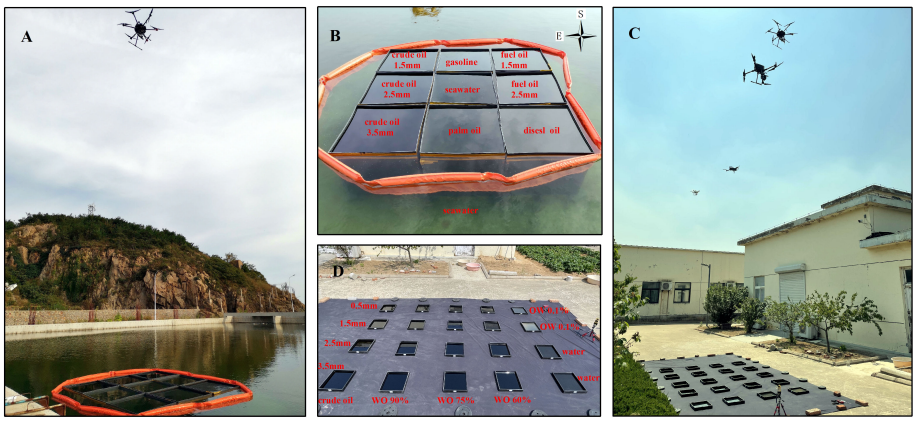


FIGURE 1
Oil spill observation experimental scenarios and oil distribution using hyperspectral combined thermal infrared remote sensing: **(A)** experimental scenarios of different types of oil products in September 2020; **(B)** distribution of different types of oil products; **(C)** experimental scenario of crude oil and its emulsions in different states in May 2022; **(D)** distribution of crude oil and its emulsions in different states.

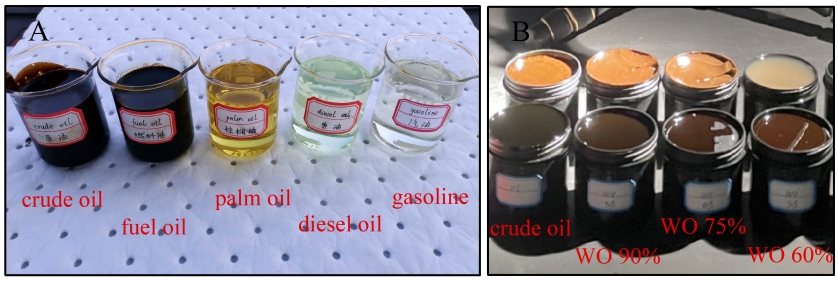


FIGURE 2
Experimental oils: **(A)** different types of oil products; **(B)** crude oil and its emulsions in different states.

and Zemmuse H20T airborne thermal infrared imager were used to obtain approximately synchronous hyperspectral and thermal infrared images of crude oil and its emulsions in different states (Figures 3C, D). The main parameters of the Cubert S185 airborne hyperspectral imager and Zemmuse H20T airborne thermal infrared imager are shown in Table 3. Except for the camera focus and stability, the main parameters of Zemmuse H20T airborne thermal infrared imager and DJI "Yu" 2 airborne thermal infrared camera are the same.

The data acquired by the airborne hyperspectral imager include the original measured spectral data, reference plate and dark current calibration data. The data to be converted was imported to the Cubert Utils Touch software. The spectral reflectance data of different oil spill pollution types was exported. Import the thermal infrared image acquired by the airborne thermal infrared imager into the DJI Thermal Analysis Tool to export the brightness temperature data of different oil spill pollution types. Due to the different coverage of airborne hyperspectral image and airborne thermal infrared image, two images need to be registered and cropped. The image size of different oil products as well as crude oil and its emulsions in different states used for identification experiments is 638×630 and 551×765 respectively.

2.2.2 Field data

Field data includes Analytical Spectral Devices (ASD) data and site photos used to assist in the production of ground truth image. The ASD FieldSpec4 spectrometer (350-2500 nm) was used to measure Lambertian standard plate, different oil spill pollution types, seawater and skylight to obtain their radiance, and then convert the radiance into the spectral reflectance of different oil spill pollution types and seawater according to the formula in literature (Yang et al., 2020). The detailed index parameters of ASD FieldSpec4 spectrometer are shown in literature (Yang et al., 2020). The ground truth image (Figure 4) obtained through human-computer interaction interpretation is used as the benchmark for evaluating the classification and recognition results, and the performance of the classification and recognition model is tested at the same time.

TABLE 2 The properties and description of experimental oils.

Oil spill pollution types	Oil products	Density g/mL	Apparent color	Description
Different types of oil products	crude oil	0.882	black	produced in the Shengli oil field of China with added anticoagulant
	fuel oil	0.853	dark cyan	fuel for large-scale marine engines, which is a residual heavy oil after crude oil extraction of gasoline and diesel oil
	palm oil	0.836	palm yellow	the largest vegetable oil product in the world in terms of production, consumption, and international trade.
	0# diesel oil	0.835	light cyan	high-speed diesel engine fuel for small ships
	95# gasoline	0.737	yellowish transparent	strong volatility, properties are similar to condensate oil that leaked from the East China Sea oil tanker collision accident in 2018
Crude oil and its emulsions in different states	crude oil	0.935	black	produced in the Shengli oil field of China, (dehydrated), asphaltene 1.57%, resin 18.77%, wax 7.56%
	WO emulsion		dark brown	there are many small water droplets in the oil
	OW emulsion		light yellow	there are many small oil droplets in the water

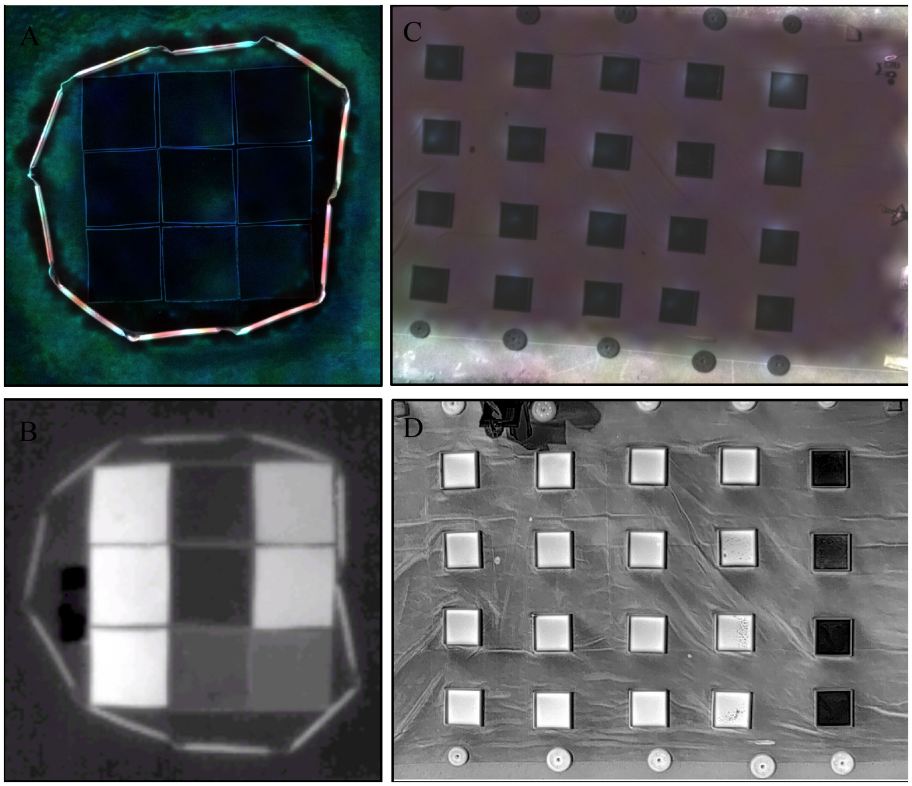


FIGURE 3
Hyperspectral and thermal infrared images of different oil spill pollution types: **(A)** airborne hyperspectral image of different oil products; **(B)** airborne thermal infrared images of different oil products; **(C)** airborne hyperspectral images of crude oil and its different emulsions; **(D)** airborne thermal infrared images of crude oil and its different emulsions.

2.3 Methods

In this paper, the Support Vector Machine model in classical machine learning, the Random Forest model in ensemble learning and the Convolution Neural Network model in deep learning are selected to identify the types of oil spill pollution. For the experiment of different types of oil products, the number of labeled samples is 375606, of which about 3% are used for training and about 3% for validation. For the experiment of crude oil and its emulsions in different states, the number of labeled samples is 135695, of which about 3% are used for training and about 3% for validation. The number of training samples, validation samples and test samples for each oil spill pollution type in the experiment is listed in Table 4.

TABLE 3 The main parameters of Cubert S185 hyperspectral imager and Zenmuse H20T thermal infrared imager.

Cubert S185 hyperspectral imager		Zenmuse H20T thermal infrared imager	
parameter	index	parameter	index
spectral range/nm	450~950	spectral range/μm	8~14
spectral resolution/nm	8	resolution	640×512
spatial resolution/cm	3.6@10000	sensitivity(NETD)/mK	≤50 @ f/1.0
field of view angle/°	23	field of view angle/°	40.6
specification/px^2	1000×1000	measuring range/°C	-40~ +550
sampling interval/nm	4	pixel spacing/μm	12
number of channels	125	aperture	f/1.0
focal length/mm	16	focal length/mm	13.5

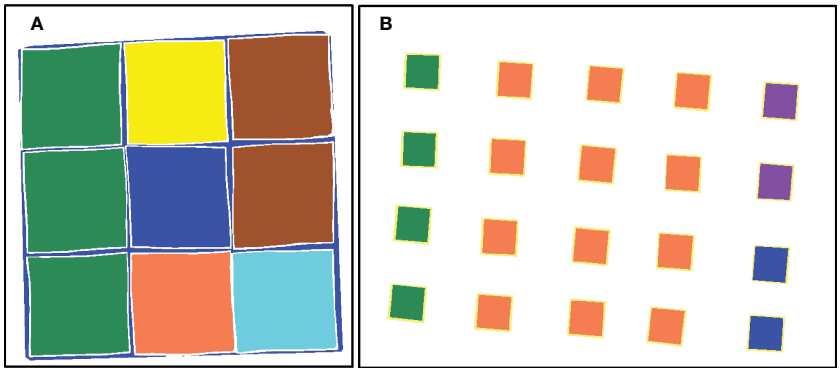


FIGURE 4
Ground truth image: **(A)** different oil products; **(B)** crude oil and its emulsions in different states.

2.3.1 Support vector machine

SVM is a traditional machine learning method based on statistical learning theory. It can automatically find the support vector that has a greater ability to distinguish classification and then construct the classifier, which can maximize the interval between classes to achieve good statistics when the number of samples is small. This method has high convergence efficiency, training speed, and classification accuracy, and has been widely used in many fields of research in recent years (Wang et al., 2011; Hu et al., 2019a; Hu et al., 2019b). The kernel function selected in this paper is the radial basis function.

2.3.2 Random forest

Random Forest is an algorithm that effectively uses multiple decision trees to train and predict samples and optimize decisions through the idea of ensemble Learning. The basic unit is the decision tree. The randomness of RF is mainly reflected in two aspects: one is the random selection of data, and the other is the random selection of features to be selected. This can make the decision trees in the RF different from each other, and further enhance the generalization ability of the model.

The RF algorithm uses multiple CART decision trees as weak classifiers. In CART tree, the criterion of impure measure used to select variables is Gini coefficient. The minimum Gini coefficient criterion is used for feature selection. Here, the number of decision trees in the RF is 100.

2.3.3 Convolutional neural network

CNN is a significant achievement in the field of deep learning, and has been widely used in hyperspectral remote sensing classification in recent years. CNN has two main characteristics. One is local receptive fields, the other is weight sharing, which effectively reduces the number of parameters in the network and makes CNN have displacement invariance.

The CNN model structure used in this paper consists of seven information layers, including one input layer, two convolutional

TABLE 4 Number of training samples, validation samples and test samples in the experiment of oil spill pollution type identification.

	Data size	Class	Number of training samples	Number of validation samples	Number of test samples
Different oil products	638×630	crude oil	3000	3000	101897
		fuel oil	2000	2000	70036
		gasoline	1000	1000	34624
		diesel oil	1000	1000	34173
		palm oil	1000	1000	34624
		seawater	1600	1600	56286
		PVC board	700	700	23366
Crude oil and its emulsions in different states	551×765	crude oil	400	400	13082
		WO emulsion	1200	1200	40750
		OW emulsion	200	200	6776
		seawater	200	200	6773
		PVC board	1800	1800	60714

layers, two pooling layers, one full connection layer, and one output layer. The number of convolutional kernel in the first convolution layer is 10, and the size of convolutional kernel is 5×5. The number of convolutional kernel in the second convolution layer is 8, and the size of convolutional kernel is 3×3. The size of the subsampling filter in the first pooling layer and the second pooling layer is 1×1 and 1×1 respectively. The maximum pooling is adopted for the pooling layer. The number of batch training is 2, the number of iterations is 80, and the learning rate is 0.7.

3 Results

3.1 Identification results of different oil spill pollution types

In view of the hyperspectral image (HSI) and multidimensional image of hyperspectral combined thermal infrared (HTI), Convolution Neural Network (2D-CNN) model, Support Vector Machine (SVM) algorithm and Random Forest (RF) algorithm are used to carry out classification and identification of different oil spill pollution types, namely, identification of different types of oil products as well as crude oil and its emulsions in different states.

3.1.1 Identification results of different types of oil products

According to the classification results of different types of oil products (Figure 5), it can be seen that 2D-CNN, SVM and RF classifiers have different abilities in data feature mining, and the performance of classification is also different. Intuitively, SVM has relatively poor oil identification effect based on the HSI. The mixing of different oils, especially light oils (palm oil, diesel oil, gasoline) is serious. The recognition effect of heavy oils (crude oil, fuel oil) is good. The classification results are quite different from the ground truth image (Figure 4A). The recognition results of RF model and 2D-CNN model (Figures 5A–C) can better maintain the continuity of oil film, which is consistent with the ground truth image, and shows the powerful data mining ability and feature extraction ability of ensemble learning model and deep learning model.

For the HTI, the recognition results (Figures 5D–F) of the three algorithms are obviously better than those based on the HSI. With the addition of thermal infrared features, the oil type recognition results can better maintain the continuity of oil spill on the sea, with clearer boundaries, and the mixing phenomenon between light oils is greatly reduced, which is consistent with the ground truth image. However, gasoline and seawater still have a misclassification phenomenon, which is due to the volatility of gasoline, resulting in that the selected gasoline training samples are not completely pure gasoline pixels.

3.1.2 Identification results of crude oil and its emulsions in different states

It can be seen from the recognition results of crude oil and its emulsions in different states (Figure 6) that, for the HSI, the recognition results of 2D-CNN model on WO emulsion, OW emulsion and seawater (Figure 6C) are good, which are consistent with the ground truth image (Figure 4B), but the crude oil is partially divided into WO emulsion, showing the strong data mining ability and feature extraction ability of the deep learning model. The recognition results of SVM model and RF model for crude oil and its emulsions in different states (Figures 6A, B) are similar. In addition to the fact that crude oil is partially divided into

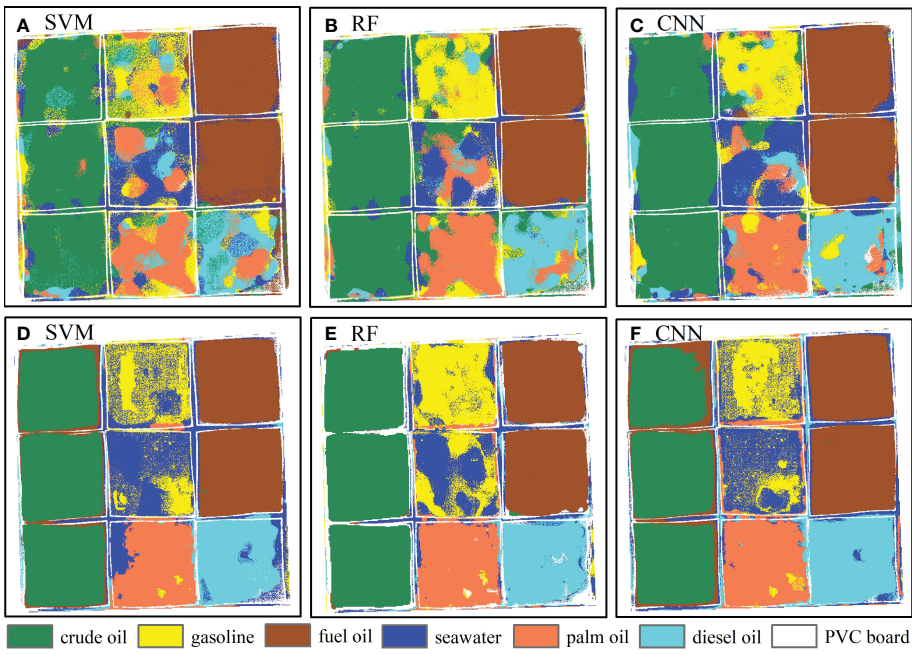


FIGURE 5
Classification results of oil products based on different dimensional images: **(A–C)** HSI; **(D–F)** HTI.

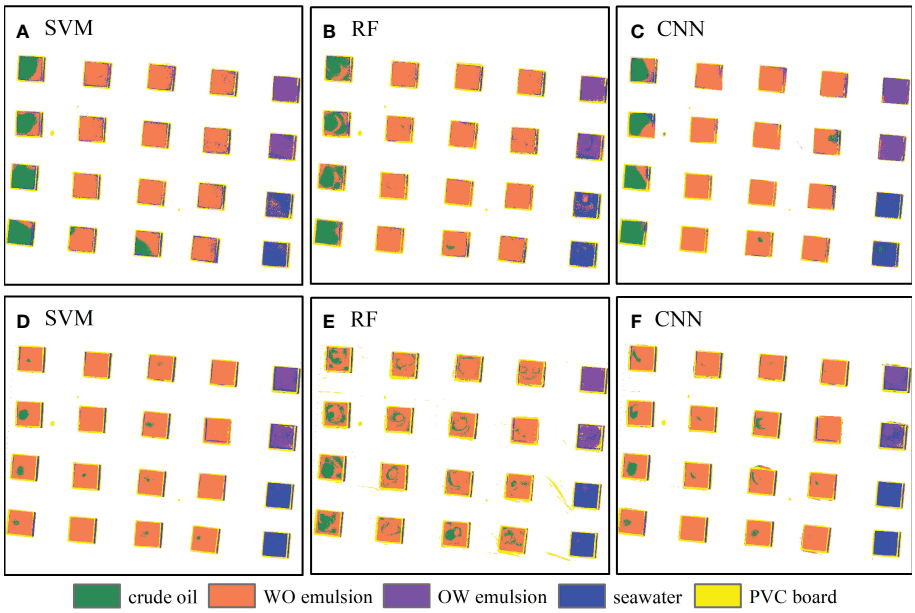


FIGURE 6
Classification results of crude oil and its different emulsions based on different dimensional images: **(A–C)** HSI; **(D–F)** HTI.

WO emulsion, and seawater is also partially divided into WO emulsion.

Different from the recognition results of different oil products in Section 3.1.1, the recognition results of the three algorithms for crude oil and its emulsions in different states based on the HSI are significantly better than those of the HTI. The addition of thermal infrared features enhances the recognition effect between seawater and WO emulsion, but aggravates the mixing phenomenon between crude oil and WO emulsion, making most crude oil be wrongly divided into WO emulsion (Figures 6D–F).

3.2 Identification accuracy evaluation of oil spill pollution types

The same training samples are used to train the 2D-CNN model, SVM and RF model to carry out oil spill pollution type recognition, and the recognition accuracy of oil spill pollution type is evaluated based on the ground truth image (Figure 4).

3.2.1 Overall identification accuracy evaluation

The overall accuracy and Kappa coefficient in the confusion matrix are used in this paper to evaluate the overall identification accuracy of oil spill pollution types based on HSI and HTI, as shown in Table 5. For different oils, the overall recognition accuracy of SVM, RF and 2D-CNN models based on the HSI is 72.66%, 78.82% and 80.09% respectively. After adding thermal infrared features, the overall recognition accuracy of SVM, RF and 2D-CNN models reached 83.24%, 85.08% and 82.9% respectively, and Kappa coefficients were 0.80, 0.82 and 0.80 respectively, indicating that the prediction results of the HTI were in good agreement with the actual results. For crude oil and its emulsions in different states, the overall recognition accuracy of SVM, RF and 2D-CNN models based on the HSI reaches 94.52%, 94.62% and 95.2% respectively. After adding thermal infrared features, the overall recognition accuracy of SVM, RF and 2D-CNN models is 92.06%, 91.67% and 91.45% respectively.

3.2.2 Identification accuracy evaluation of single oil spill pollution type

The F1 score is used to evaluate the identification accuracy of single oil spill pollution type based on HSI and HTI, as shown in Figure 7. In general, for different oils, the recognition accuracy of the three algorithms based on HTI has been improved to varying degrees compared with that based on HSI. Taking the recognition results of RF model as an example, the F1 scores of gasoline, palm oil and diesel oil increased by 0.07, 0.21 and 0.20 respectively. The F1 score of crude oil and fuel oil increased by 0.11 and 0.03 respectively. The F1 score of seawater increased by 0.09. The experimental results show that the thermal infrared features can increase the identification ability of hyperspectral for different oils, especially light oils, and can effectively improve the oil identification accuracy. For crude oil and its emulsions in different states, the recognition accuracy of the three algorithms based on HTI is lower than that based on HSI to varying degrees, but the recognition accuracy of seawater is improved. Taking the recognition results of RF model as an example, the F1 scores of crude oil, WO emulsion and OW emulsion decreased by 0.37, 0.09 and 0.04 respectively, while the F1 scores of seawater increased by 0.10.

4 Discussion

4.1 Spectral response of different oil spill pollution types

Remote sensing reflectivity of five typical oil products, crude oil and its emulsions in different states, and seawater were obtained by

TABLE 5 The accuracies for oil pollution type identification of three algorithms based on HSI and HTI.

Methods \ Evaluation criterion			Overall Accuracy (%)	Kappa Coefficient
Different oil products	SVM	HSI	72.66	0.68
		HTI	83.24	0.80
	RF	HSI	78.82	0.75
		HTI	85.08	0.82
	2D-CNN	HSI	80.09	0.76
		HTI	82.93	0.80
Crude oil and its emulsions in different states	SVM	HSI	94.52	0.90
		HTI	92.06	0.85
	RF	HSI	94.62	0.90
		HTI	91.67	0.85
	2D-CNN	HSI	95.20	0.91
		HTI	91.45	0.84

ASD, of which the area around 1.4 μm and 1.9 μm is affected by the strong absorption zone of water, so the reflectivity within this range is abnormal. At the same time, considering the edge effect of the photosensitive devices used, irregular oscillations will occur at the end of the sensing spectrum of the spectrometer. Therefore, this paper studied the spectral response of oil spill in the spectral range of 360~1340 nm, 1440~1800 nm and 1980~2400 nm.

4.1.1 Spectral characteristics of different oils

In general, the reflectivity of the five oil products and seawater is very small (Figure 8), which is related to the absorption properties of oil products and seawater. In the visible light band, the reflectance spectra of light oils such as diesel, gasoline and palm oil are generally consistent with those of seawater, and there is an obvious reflection peak at about 480 nm; The reflectivity of heavy oil such as crude oil and fuel oil are obviously lower than those of light oils and seawater. In the near-infrared and shortwave infrared bands, the reflectivity of the five oils are higher than that of seawater. This is because the pure natural water is nearly a "black body" in the near-infrared band. Therefore, in the spectral range of 850~2500 nm, the reflectivity of the purer natural water is very low, almost zero.

Within the spectral range (450~950 nm) of the airborne hyperspectral image, the spectral curves of the three light oils and seawater are very similar. Therefore, based on the single dimensional characteristics of hyperspectral, the recognition results of the three algorithms for light oils (palm oil, diesel oil, gasoline) and seawater are poor, and the mixing phenomenon is serious. At the same time, the spectra of heavy oils and light oils are quite different, so except at the edge of the enclosure, the recognition accuracy of heavy oils such as crude oil and fuel oil based on the HSI is good, and there is little mixing with light oils and seawater.

4.1.2 Spectral characteristics of crude oil and its emulsion in different states

In general, the spectral response of crude oil and its emulsions in different states and seawater in different spectral ranges is different (Figure 9), which is related to its absorption and scattering properties. WO emulsion contains seawater droplets. When incident light enters seawater droplets from oil film, it will produce strong backscattering. In the near infrared and short wave infrared spectra, the WO emulsion has higher reflectivity. In addition, due to the absorption of -C-H in WO emulsion, obvious reflection valleys are formed at ~1725 nm, ~1760 nm and ~2170 nm. In the OW emulsion, dispersed small oil droplets exist in continuous water, which makes it have high reflectivity in the spectral range of 360~1400 nm. At the same time, due to the absorption of -O-H, obvious reflection valleys are formed at ~975 nm and ~1200 nm. This is very close to the results obtained from the indoor experiment carried out by Lu et al. (2019). Due to the

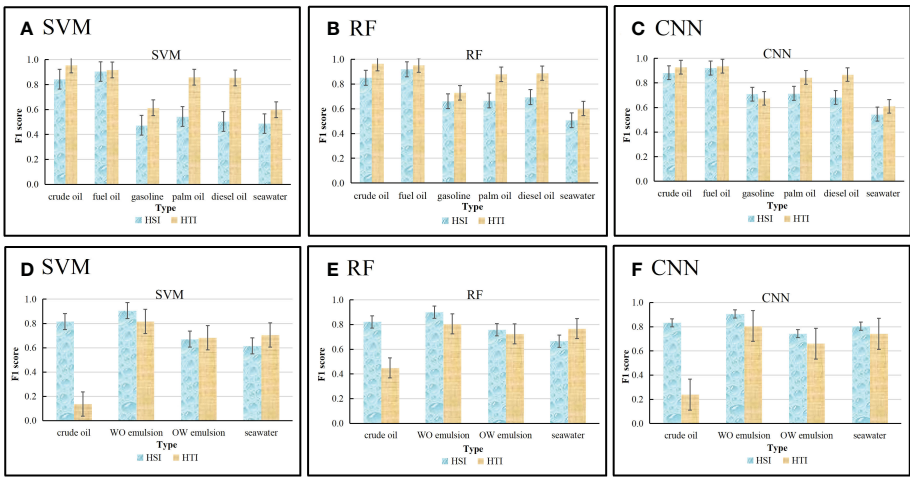


FIGURE 7
Identification accuracies of single oil spill pollution type based on HSI and HTI: **(A–C)** different oil products; **(D–F)** crude oil and its emulsions in different states.

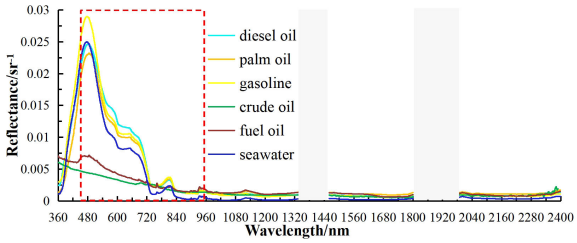


FIGURE 8
Mean spectral reflectance of typical oil products (The gray area is a non-atmospheric window, the red rectangular area represents the spectral range of airborne hyperspectral image).

strong absorption and low reflection of incident light, the spectral reflectance of crude oil is low, especially in the near infrared band, the spectral reflectivity of crude oil is nearly zero.

In the spectral range of airborne hyperspectral images (450~950 nm), there are large differences in the spectral response of crude oil and its emulsion different states and seawater. Therefore, based on the HSI, three algorithms are used to identify WO emulsion, WO emulsion and seawater with good results. However, due to the high oil concentration of the stable WO emulsion, the spectral curve of the WO emulsion is very similar to that of the crude oil. Therefore, the WO emulsion is mixed with the crude oil in the classification and recognition results based on the HSI.

4.2 Thermal response of different oil spill pollution types

4.2.1 Thermal infrared intensity characteristics of different oils

Because the heat capacity and infrared reflectance of oil film and seawater are different, there exists temperature difference between oil film and background seawater, which is the basis using infrared image to detect marine oil spill. It can be seen from Figure 3B that the heavy oils (crude oil and fuel oil) show a “bright” hue on the thermal infrared image, which is the most obvious difference from the seawater background, in which the crude oil shows “brightest”, and the fuel oil shows “brighter”. Light oils (palm oil, diesel oil and gasoline) is displayed as “dark” hue on the thermal infrared image. Diesel oil is “darker”, palm oil is “dark”, and gasoline is “darkest”. This is mainly related to the characteristics of the oil products themselves, the ability to absorption and radiation and other factors. Heavy oils can absorb more solar radiation and emit it in the form of thermal radiation.

There is a certain conversion relationship between intensity of thermal infrared image and brightness temperature. Here, intensity is used to replace brightness temperature to analyze the thermal response characteristics of different oils. It can be seen from Figure 10 that the thermal infrared radiation intensity of the five oil products and seawater from high to low is crude oil, fuel oil, diesel oil, palm oil, seawater and gasoline. The thermal infrared intensity of gasoline is lower than that of seawater, because gasoline volatilization will cause the surface temperature to drop. The thermal infrared intensity of crude oil and fuel oil are obviously higher than those of diesel oil, palm oil and gasoline. The thermal infrared intensity of the three kinds of light oils is quite different. It is precisely because of this difference that the mixing phenomenon between light oils in the recognition results of the HTI is greatly weakened. However, there is still a certain mixing phenomenon between gasoline and seawater, which is related to the strong volatility of gasoline.

4.2.2 Brightness temperature characteristics of crude oil and its emulsions in different states

It can be seen from Figure 3D that the crude oil and WO emulsions show “bright” hue on the thermal infrared image, which is the most obvious difference from the seawater background. The OW emulsion and seawater show a “dark” hue on the thermal infrared image, among which, the oil in water emulsion is relatively “dark”, and the seawater is the “darkest”, which is mainly related to

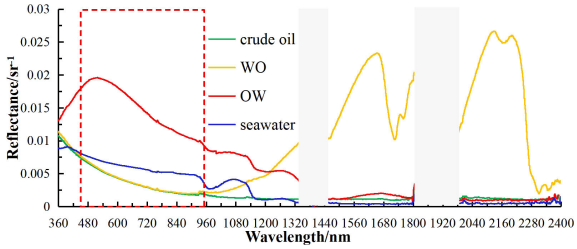


FIGURE 9
Mean spectral reflectance of crude oil and its emulsions in different states (The gray area is a non-atmospheric window, the red rectangular area represents the spectral range of airborne hyperspectral image).

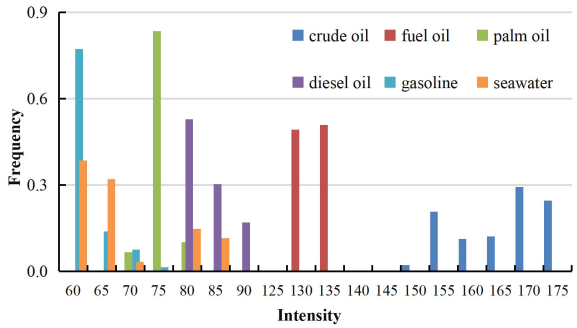


FIGURE 10
Thermal infrared intensity distribution of different oil products and seawater.

the concentration of oil. The higher the oil concentration, the more solar radiation absorbed and emitted in the form of thermal radiation.

The thermal infrared image of crude oil and its emulsion in different states was obtained using the Zemmuse H20T airborne thermal infrared imager. The brightness temperature analysis was carried out by selecting pixels of the same size for crude oil, WO emulsion, OW emulsion and seawater from the thermal infrared image. It can be seen from Figure 11 that the brightness temperature of crude oil and its emulsions in different states and seawater from high to low are crude oil, 90% WO emulsion, 75% WO emulsion, 60% WO emulsion, 0.1% OW emulsion and seawater. The brightness temperatures of crude oil and WO emulsions with different concentrations are close. The brightness temperature of OW emulsion and seawater is similar, but the brightness temperature of crude oil and WO emulsion is significantly different from that of OW emulsion and seawater, with an average difference of 20 °C. It is precisely because the brightness temperature difference between crude oil and WO emulsions of different concentrations is not obvious, which makes the mixing phenomenon of crude oil and WO emulsions enhanced in the recognition results of the HTI, so the vast majority of crude oil is wrongly divided into WO emulsions. This shows that thermal infrared remote sensing technology cannot be used to distinguish crude oil and WO emulsion, which is consistent with the conclusion reached by Jiao et al. (2021).

4.3 Application of hyperspectral combined thermal infrared features in oil film monitoring with different thickness

4.3.1 Spectral and brightness temperature characteristics of oil film with different thickness

In order to explain the identification results of different oil spill pollution types using the HTI, the spectral and brightness temperature characteristics of different oils, crude oil and its emulsions in different states have been analyzed and discussed in the previous section. Here we further analyze and discuss the spectral and brightness temperature characteristics of oil films with different thicknesses, which can be used to explain the possible results of oil film thickness classification using the HTI.

It can be seen from Figure 12 that the spectral curves of oil films with different thicknesses are generally the same. With the increase of wavelength, the spectral reflectance of crude oil decreases gradually. Similarly, with the increase of oil film thickness, the spectral reflectance of crude oil decreases gradually. In the spectral range of airborne hyperspectral images (450~950 nm), the spectral

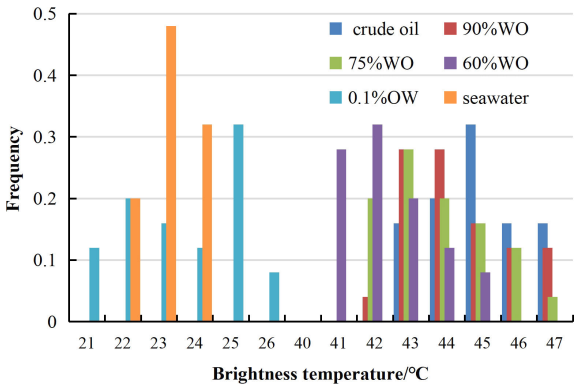


FIGURE 11
Thermal infrared brightness and temperature distribution of crude oil and its different emulsions and seawater.

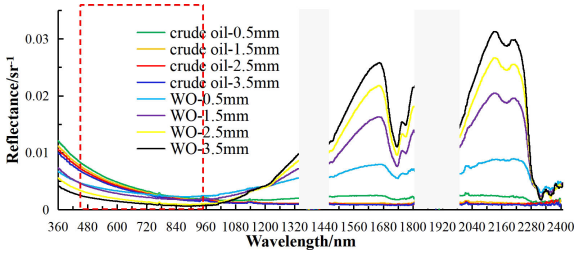


FIGURE 12
Spectral reflectance of crude oil film with different thickness and WO emulsion with different thickness at the same concentration (The gray area is a non-atmospheric window, the red rectangular area represents the spectral range of airborne hyperspectral image).

responses of oil films with different thicknesses exist difference. Therefore, theoretically, the classification results of oil films with different thicknesses should be good based on the HSI. It can be seen from Figure 13 that the brightness temperature of oil film with different thickness is different, which increases with the increase of oil film thickness, and the brightness temperature difference is greater than 1°C. In theory, the addition of thermal infrared features will enhance the recognition ability of oil films with different thicknesses. The combination of hyperspectral and thermal infrared remote sensing can be used to classify and identify oil film with different thickness on the sea surface.

4.3.2 Spectral and brightness temperature characteristics of WO emulsions with different thickness at the same concentration

Here, in order to explain the possible classification results of WO with different thicknesses at the same concentration using hyperspectral and thermal infrared remote sensing, the spectral and brightness temperature characteristics of WO emulsions with different thicknesses at the same concentration are analyzed and discussed. It can be seen from Figure 12 that, taking WO emulsions with a concentration of 60% as an example, the spectral curves of WO emulsions with different thickness at the same concentration are generally consistent. In the spectral range before 1200 nm, the spectral reflectance of WO emulsions decreases gradually with the increase of oil film thickness; In the near-infrared range after 1200 nm, the spectral reflectance of WO emulsions increases gradually with the increase of oil film thickness.

In the spectral range of airborne hyperspectral images (450~950 nm), the spectral responses of WO emulsions with different thicknesses at the same concentration are different. Therefore, theoretically, the classification results of oil films with different thicknesses should be good based on the HSI. It can be seen from Figure 13 that the brightness temperature of WO emulsions with different thicknesses at the same concentration is different, which increases with the increase of oil film thickness. However, the brightness temperature difference is not all greater than 1°C, and there may be mixing between oil films with different thicknesses. In conclusion, theoretically, the addition of thermal infrared features may enhance the recognition ability of WO emulsions with different thicknesses at the same concentration.

5 Conclusion

In this paper, we designed and implemented two outdoor oil spill simulation experiment, carried out the extraction and analysis of hyperspectral and thermal infrared multidimensional optical features of different oil spill pollution types, constructed hyperspectral oil spill pollution type recognition algorithm based on traditional machine learning, ensemble learning, and deep learning model, and explored how to improve the ability of hyperspectral oil spill pollution type

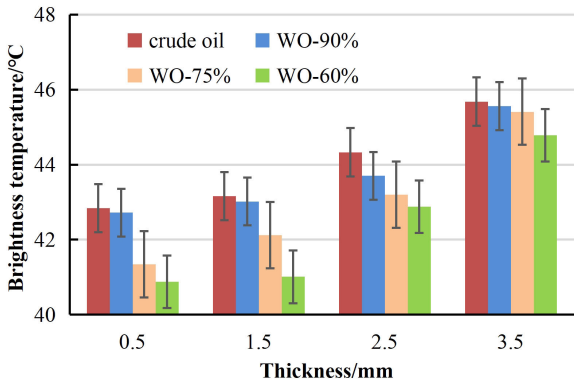


FIGURE 13
Thermal infrared brightness and temperature of crude oil film with different thickness and WO emulsion with different thickness at the same concentration.

identification by adding thermal infrared features. At the same time, the classification ability of hyperspectral combined with thermal infrared remote sensing for different oil film thicknesses and WO emulsions with different thicknesses at the same concentration is also discussed. The main conclusions can be drawn as follows: (1) The addition of thermal infrared features can effectively improve the hyperspectral recognition ability of different oils (crude oil, fuel oil, palm oil, diesel oil, gasoline), especially light oils. Among them, the F1 scores of gasoline, palm oil and diesel oil are increased by 0.07, 0.21 and 0.20 respectively, the F1 scores of crude oil and fuel oil are increased by 0.11 and 0.03 respectively, and the F1 scores of seawater are increased by 0.09. (2) Thermal infrared remote sensing technology cannot be used to distinguish crude oil and high concentration WO emulsions. The addition of thermal infrared features will reduce the recognition effect of hyperspectral on crude oil and WO emulsions; (3) Through the analysis of the spectral characteristics and brightness temperature characteristics of oil films with different thicknesses and WO emulsions with different thicknesses at the same concentration, it is theoretically shown that the thermal infrared characteristics can enhance the ability of hyperspectral classification of oil films with different thicknesses and WO emulsions with different thicknesses at the same concentration.

This paper analyzes the spectral characteristics and brightness temperature characteristics of oil films with different thicknesses and WO emulsions with different thicknesses at the same concentration, and obtains corresponding theoretical conclusions. Next, in order to verify the relevant theories and obtain reliable conclusions, we will use airborne hyperspectral images and airborne thermal infrared images to identify oil films with different thicknesses and WO emulsions with different thicknesses at the same concentration.

Affected by the weather, complex marine environment and lighting conditions, the data obtained by optical remote sensing during the oil spill accident is limited. In the case of few samples, sample self-learning and expansion is a problem worthy of attention, and machine learning models in the case of limited samples should also be concerned. At the same time, facing the needs of oil spill remote sensing business monitoring, transfer learning model based on historical oil spill remote sensing data should be explored to achieve marine oil spill monitoring without samples.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

JY and ZJ designed outdoor oil spill simulation experiments, JY and YH designed the method with experiments, and JY wrote the manuscript. JY and YH revised the manuscript according to the comments of reviewers. The manuscript was supervised by YH, JZ, YM and ZL. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by National Natural Science Foundation of China (No. U1906217, No. 42206177, No. 61890964), Fund of Technology Innovation Center for Ocean Telemetry, Ministry of Natural Resources (No. 2022004), Shandong Provincial Natural Science Foundation (No. ZR2022QD075), Qingdao Postdoctoral Application Research Project (No. qdyy20210082, No. QDBSH202105) and the Fundamental Research Funds for the Central Universities (No. 21CX06057A).

Acknowledgments

We would like to express our sincere appreciation to the editor and reviewers who provided valuable comments to help improve this paper.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Alpers, W., Holt, B., and Zeng, K. (2017). Oil spill detection by imaging radars: Challenges and pitfalls. *Remote Sens. Environ.* 201, 133–147. doi: 10.1016/j.rse.2017.09.002

Cui, C., Li, Y., Liu, B. X., Li, G. N., Salehi, B., and Kainz, W. (2017). A new endmember preprocessing method for the hyperspectral unmixing of imagery containing marine oil spills. *ISPRS Int. J. Geo-Inf* 69, 286. doi: 10.3390/ijgi6090286

Duan, P. H., Lai, J. B., Kang, J., Kang, X. D., Ghamisi, P., and Li, S. T. (2020). Texture-aware total variation-based removal of sun glint in hyperspectral images. *ISPRS J. Photogrammetry Remote Sens.* 166, 359–372. doi: 10.1016/j.isprsjprs.2020.06.009

Fingas, M., and Brown, C. (2014). Review of oil spill remote sensing. *Mar. pollut. Bull.* 831, 9–23. doi: 10.1016/j.marpolbul.2014.03.059

French-McCay, D. P., Frediani, M., and Gloekler, M. D. (2022). Modeling emulsification influence on oil properties and fate to inform effective spill response. *Front. Environ. Sci.* 10. doi: 10.3389/fenvs.2022.90898

Guo, G., Liu, B. X., and Liu, C. Y. (2020). Thermal infrared spectral characteristics of bunker fuel oil to determine oil-film thickness and API. *J. Mar. Sci. Eng.* 82, 135. doi: 10.3390/jmse8020135

Hu, C. M., Lu, Y. C., Sun, S. J., and Liu, Y. X. (2021). Optical remote sensing of oil spills in the ocean: what is really possible? *J. Remote Sens.* 2021, 9141902. doi: 10.34133/2021/9141902

Hu, Y. B., Zhang, J., Ma, Y., An, J. B., Ren, G. B., and Li, X. M. (2019a). Hyperspectral coastal wetland classification based on a multi-object convolutional neural network model and decision fusion. *IEEE Geosci. Remote Sens. Lett.* 167, 1110–1114. doi: 10.1109/LGRS.2018.2890421

Hu, Y. B., Zhang, J., Ma, Y., Li, X. M., Sun, Q. P., and An, J. B. (2019b). Deep learning classification of coastal wetland hyperspectral image combined spectra and texture features: a case study of yellow river estuary wetland. *Acta Oceanol Sin.* 385, 142–150. doi: 10.1007/s13131-019-1445-z

Jiang, Z. C., Ma, Y., and Yang, J. F. (2020). Inversion of the thickness of crude oil film based on an OG-CNN model. *J. Mar. Sci. Eng.* 89, 653. doi: 10.3390/jmse8090653

Jiang, Z. C., Zhang, J., Ma, Y., and Mao, X. P. (2021). Hyperspectral remote sensing detection of marine oil spills using an adaptive long-term moment estimation optimizer. *Remote Sens.* 141, 157. doi: 10.3390/RS14010157

Jiao, J. N., Lu, Y. C., Hu, C. M., Shi, J., Sun, S. J., and Liu, Y. X. (2021). Quantifying ocean surface oil thickness using thermal remote sensing. *Remote Sens. Environ.* 2611, 112513. doi: 10.1016/J.RSE.2021.112513

Jing, Y., An, J. B., and Liu, Z. X. (2011). A novel edge detection algorithm based on global minimization active contour model for oil slick infrared aerial image. *IEEE Trans. Geosci. Remote Sens.* 496, 2005–2013. doi: 10.1109/TGRS.2010.2103671

Lai, Q. Z., Xie, Y. M., Wang, C. G., Wang, M., and Tan, J. Y. (2021). Multiband directional reflectance properties of oil-in-water emulsion: application for identification of oil spill types. *Appl. optics* 6023, 6902–6909. doi: 10.1364/AO.427978

Leifer, I., Lehr, W. J., Simecek-Beatty, D., Bradley, E., Clark, R., Dennison, P., et al. (2012). State of the art satellite and airborne marine oil spill remote sensing: Application to the BP deepwater horizon oil spill. *Remote Sens. Environ.* 1249, 185–209. doi: 10.1016/j.rse.2012.03.024

Li, Y., Lu, H. M., Zhang, Z. D., and Liu, P. (2020). A novel nonlinear hyperspectral unmixing approach for images of oil spills at sea. *Int. J. Remote Sens.* 4112, 4684–4701. doi: 10.1080/01431161.2020.1723179

Li, Y., Yu, Q., Xie, M., Zhang, Z. D., Ma, Z. J., and Cao, K. (2021). Identifying oil spill types based on remotely sensed reflectance spectra and multiple machine learning algorithms. *IEEE J. Sel Topics Appl. Earth Observ Remote Sens.* 14, 9071–9078. doi: 10.1109/JSTARS.2021.3109951

Li, K., Yu, H. L., Xu, Y. Q., and Luo, X. Q. (2022). Detection of oil spills based on gray level co-occurrence matrix and support vector machine. *Front. Environ. Sci.* 10. doi: 10.3389/fenvs.2022.1049880

Lu, Y. C., Hu, C. M., Sun, S. J., Zhang, M. W., Zhou, Y., Shi, J., et al. (2016a). Overview of optical remote sensing of marine oil spills and hydrocarbon seepage. *J. Remote Sens.* 205, 1259–1269. doi: 10.11834/jrs.20166122

Lu, Y. C., Li, X., Tian, Q., Zheng, G., Sun, S. J., Liu, Y. X., et al. (2013a). Progress in marine oil spill optical remote sensing: detected targets, spectral response characteristics, and theories. *Mar. Geodesy* 363, 334–346. doi: 10.1080/01490419.2013.793633

Lu, Y. C., Liu, J. Q., Ding, J., Shi, J., Chen, J., and Ye, X. (2019a). Optical remote identification of spilled oils from the SANCHI oil tanker collision in the East China Sea (in Chinese). *China Sci. Bull.* 6431, 3213–3222. doi: 10.1360/N972019-00094

Lu, Y. C., Shi, J., Wen, Y. S., Hu, C. M., and Liu, Y. X. (2019b). Optical interpretation of oil emulsions in the ocean-part I: Laboratory measurements and proof-of-concept with AVIRIS observations. *Remote Sens. Environ.* 230, 111183. doi: 10.1016/j.rse.2019.05.002

Lu, Y. C., Shi, J., Wen, Y. S., Zhang, M. W., Sun, S. J., and Liu, Y. X. (2020). Optical interpretation of oil emulsions in the ocean-part II: Applications to multi-band coarse-resolution imagery. *Remote Sens. Environ.* 242, 111778. doi: 10.1016/j.rse.2020.111778

Lu, Y. C., Sun, S. J., Zhang, M. W., Murch, B., and Hu, C. M. (2016b). Refinement of the critical angle calculation for the contrast reversal of oil slicks under sunglint. *J. Geophys Res: Oceans* 1211, 148–161. doi: 10.1002/2015JC011001

Lu, Y. C., Tian, Q. J., Wang, X. Y., Zheng, G., and Li, X. (2013b). Determining oil slick thickness using hyperspectral remote sensing in the bohai Sea of China. *Int. J. Digital Earth* 61, 76–93. doi: 10.1080/17538947.2012.695404

Lu, Y. C., Zhan, W., and Hu, C. M. (2016c). Detecting and quantifying oil slick thickness by thermal remote sensing: A ground-based experiment. *Remote Sens. Environ.* 181, 207–217. doi: 10.1016/j.rse.2016.04.007

Ma, X. S., Xu, J. G., Wu, P. H., and Kong, P. (2021). Oil spill detection based on deep convolutional neural networks using polarimetric scattering information from sentinel-1 SAR images. *IEEE Trans. Geosci. Remote Sens.* 60, 4204713. doi: 10.1109/TGRS.2021.3126175

Marghany, M. (2014). Utilization of a genetic algorithm for the automatic detection of oil spill from RADARSAT-2 SAR satellite data. *Mar. pollut. Bull.* 891, 20–29. doi: 10.1016/j.marpolbul.2014.10.041

Mdakane, L. W., and Kleynhans, W. (2020). Feature selection and classification of oil spill from vessels using sentinel-1 wide-swath synthetic aperture radar data. *IEEE Geosci. Remote Sens. Lett.* 19, 4002505. doi: 10.1109/LGRS.2020.3025641

Mohammadi, M., Sharifi, A., Hosseingholizadeh, M., and Tariq, A. (2021). Detection of oil pollution using SAR and optical remote sensing imagery: a case study of the Persian gulf. *J. Indian Soc. Remote Sens.* 4910, 2377–2385. doi: 10.1007/S12524-021-01399-2

Pärt, S., Kankaanpää, H., Björkqvist, J. V., and Uiboupin, R. (2021). Oil spill detection using fluorometric sensors: Laboratory validation and implementation to a ferrybox and a moored smartbuoy. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.778136

Rajendran, S., Vethamony, P., Sadooni, F., Al-Saad, A. H., Jassim, A., Vashist, S., et al. (2021). Detection of wakashio oil spill off Mauritius using sentinel-1 and 2 data: Capability of sensors, image transformation methods and mapping. *Environ. Pollut.* 274, 116618. doi: 10.1016/J.ENVPOL.2021.116618

Shen, Y. F., Liu, J. Q., Ding, J., Jiao, J. N., Sun, S. J., and Lu, Y. C. (2020). HY-1 CCOCTS and CZI observation of marine oil spills in the south China Sea. *J. Remote Sens.* 248, 933–944. doi: 10.11834/jrs.20209475

Shi, J., Jiao, J. N., Lu, Y. C., Zhang, M. W., Mao, Z. H., and Liu, Y. X. (2018). Determining spectral groups to distinguish oil emulsions from sargassum over the gulf of Mexico using an airborne imaging spectrometer. *ISPRS J. Photogr Remote Sens.* 146, 251–259. doi: 10.1016/j.isprsjprs.2018.09.017

Silva, F. R., Schiavetti, A., Malhado, A. C. M., Ferreira, B., Sousa, C. V. P., Vieira, F. P., et al. (2022). Oil spill and socioeconomic vulnerability in marine protected areas. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.85969

Song, M. P., Cai, L. F., Lin, B., An, J. B., and Chang, C. (2016). "Hyperspectral oil spill image segmentation using improved region-based active contour model," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2016).* (Beijing, China: IEEE), 6352–6355. doi: 10.1109/IGARSS.2016.7730660

Sun, S. J., and Hu, C. M. (2016). Sun glint requirement for the remote detection of surface oil films. *Geophys Res. Lett.* 431, 309–316. doi: 10.1002/2015GL066884

Velotto, D., Migliaccio, M., Nunziata, F., and Lehner, S. (2011). Dual-polarized TerraSAR-X data for oil-spill observation. *IEEE Trans. Geosci. Remote Sens.* 4912, 4751–4762. doi: 10.1109/TGRS.2011.2162960

Wang, Y. B., Du, P. P., Liu, J. Q., and Chen, C. T. (2022). Spatial variation of coastal wetland vulnerability to oil spill stress in the bohai Sea. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.107

Wang, X. L., Du, P. J., and Tan, K. (2011). Research on training sample selection in SVM classification of hyperspectral remote sensing image. *Sci. Surveying Mapp.* 03, 127–129.

Wang, G. C., and Gao, C. (2020). Airborne ocean oil spill detection mission system. *Ind. Technol. Innovation* 73, 97–101. doi: 10.14103/j.issn.2095-8412.2020.03.018

Wang, M. Q., and Hu, C. M. (2015). Extracting oil slick features from VIIRS nighttime imagery using a gaussian filter and morphological constraints. *IEEE Geosci. Remote Sens. Lett.* 1210, 2051–2055. doi: 10.1109/LGRS.2015.2444871

Wang, B., Shao, Q. F., Song, D. M., Li, Z. W., Tang, Y. H., Yang, C. L., et al. (2021). A spectral-spatial features integrated network for hyperspectral detection of marine oil spill. *Remote Sens.* 138, 1568. doi: 10.3390/rs13081568

Wang, L. F., Xin, L. P., Yu, B., Ju, L., and Wei, L. (2021). A novel method for determination of the oil slick area based on visible and thermal infrared image fusion. *Infrared Phys. Technol.* 119, 103915. doi: 10.1016/J.INFRARED.2021.103915

Washburn, T. W., Yoskowitz, D. W., and Montagna, P. A. (2018). Valuing nature waste removal in the offshore environment following the deepwater horizon oil spill. *Front. Mar. Sci.* 5. doi: 10.3389/fmars.2018.0047

Wettle, M., Daniel, P. J., Logan, G. A., and Thankappan, M. (2009). Assessing the effect of hydrocarbon oil type and thickness on a remote sensing signal: A sensitivity study based on the optical properties of two different oil types and the HYMAP and quickbird sensors. *Remote Sens. Environ.* 113, 2000–2010. doi: 10.1016/j.rse.2009.05.010

Yang, J. F., Hu, Y. B., Ma, Y., Li, Z. W., and Zhang, J. (2022a). "Research on oil spill pollution type identification using RPnet deep learning model and airborne hyperspectral image," in *2022 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2022).* (Kuala Lumpur, Malaysia: IEEE), 807–810. doi: 10.1109/IGARSS46834.2022.9883530

Yang, J. F., Ma, Y., Hu, Y. B., Jiang, Z. C., Zhang, J., Wan, J. H., et al. (2022b). Decision fusion of deep learning and shallow learning for marine oil spill detection. *Remote Sens.* 143, 666. doi: 10.3390/RS14030666

Yang, J. F., Wan, J. H., Ma, Y., Jiang, Z. C., and Hu, Y. B. (2021). Accuracy assessments of hyperspectral characteristic waveband for common marine oil spill types identification. *Mar. Sci.* 454, 97–105. doi: 10.11759/hykx20200720003

Yang, J. F., Wan, J. H., Ma, Y., Zhang, J., and Hu, Y. B. (2020). Characterization analysis and identification of common marine oil spill types using hyperspectral remote sensing. *Int. J. Remote Sens.* 4118, 7163–7185. doi: 10.1080/01431161.2020.1754496

Yang, J. F., Wan, J. H., Ma, Y., Zhang, J., and Jiang, Z. C. (2019). Oil spill hyperspectral remote sensing detection based on DCNN with multi-scale features. *J. Coast. Res.* 90 (sp1), 332–339. doi: 10.2112/SI90-042.1

Zhong, Z. X., and You, F. Q. (2011). Oil spill response planning with consideration of physicochemical evolution of the oil slick: a multiobjective optimization approach. *Comput. Chem. Eng.* 358, 1614–1630. doi: 10.1016/j.compchemeng.2011.01.009

Zhu, X. Y., Li, Y., Zhang, Q., and Liu, B. X. (2019). Oil film classification using deep learning-based hyperspectral remote sensing technology. *ISPRS Int. J. Geo-Inf* 84, 181. doi: 10.3390/ijgi8040181

Zhu, Q. Q., Zhang, Y. N., Li, Z. Q., Yan, X. R., Guan, Q. F., Zhong, Y. F., et al. (2021). Oil spill contextual and boundary-supervised detection network based on marine SAR images. *IEEE Trans. Geosci. Remote Sens.* 60, 5213910. doi: 10.1109/TGRS.2021.3115492

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 24 March 2023
DOI 10.3389/fmars.2023.1124185

Check for updates

OPEN ACCESS

EDITED BY
Ran Liao,
Tsinghua University, China

REVIEWED BY
Wei Cai,
Southern University of Science and Technology, China
Tianlei Ma,
Zhengzhou University, China

*CORRESPONDENCE
Shijun Wu
bluewater@zju.edu.cn

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 14 December 2022
ACCEPTED 13 March 2023
PUBLISHED 24 March 2023

CITATION
Wang X, Cao Y, Wu S and Yang C (2023) Real-time detection of deep-sea hydrothermal plume based on machine vision and deep learning.
Front. Mar. Sci. 10:1124185.
doi: 10.3389/fmars.2023.1124185

COPYRIGHT
© 2023 Wang, Cao, Wu and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Real-time detection of deep-sea hydrothermal plume based on machine vision and deep learning

Xun Wang, Yanpeng Cao, Shijun Wu* and Canjun Yang

State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, Hangzhou, China

Recent years have witnessed an increase in applications of artificial intelligence (AI) in the detection of oceanic features with the tremendous success of deep learning. Given the unique biological ecosystems and mineral-rich deposits, the exploration of hydrothermal fields is both scientifically and commercially important. To achieve autonomous and intelligent sampling of the hydrothermal plume by using AUV, this paper proposes an innovative method for real-time plume detection based on the YOLOv5n deep learning algorithm designed with a light-weight neural network architecture to meet the requirements of embedded platforms. Ground truth labeler app LabelImg was used to generate the ground truth data from the plume dataset created by ourselves. To accurately and efficiently detect hydrothermal plumes using an embedded system, we improved the original structure of YOLOv5n in two aspects. First, SiLU activation functions in the model were replaced by ReLU activations at shallow layers and Hard-SiLU activations at deep layers to reduce the number of calculations. Second, an attention module termed Coordinate Attention (CA) was integrated into the model to improve its sensitivity to both channel and spatial features. In addition, a transfer learning training method was adopted to further improve the model's accuracy and generalizability. Finally, we successfully deployed the proposed model in a low-cost embedded device (NVIDIA Jetson TX2 NX) by using the TensorRT inference engine. We then installed the Jetson TX2 NX into a hovering-type AUV as its vision processing unit and conducted a plume detection test in the water tank. The water tank experimental results demonstrated that the proposed method can achieve real-time onboard hydrothermal plume detection.

KEYWORDS

hydrothermal plume, deep-sea, real-time, object detection, deep learning, transfer learning, YOLOv5

1 Introduction

Since its first discovery at the mid-ocean ridge in the eastern Pacific by the research submersible ALVIN in 1977 (Corliss et al., 1979), seafloor hydrothermal activity has become a popular topic of research in many academic fields owing to its unique mineralization system, ecosystem, and its contribution to the heat and material of the ocean (Luther et al., 2001; Petersen et al., 2011; Tao et al., 2020). Collecting hydrothermal fluids with suitable samplers and then analyzing their chemical composition in the laboratory has always been an effective means of studying the hydrothermal fluids. In the past few decades, fluids have been collected from seafloor hydrothermal systems with a variety of samplers that can be deployed using remotely operated vehicles (ROVs) and human-occupied vehicles (HOVs) (Seewald et al., 2002; Chen et al., 2007; Wang et al., 2020). However, the operation and maintenance of ROV or HOV are expensive, time-consuming, and might cause security concerns for the vehicle operator. By contrast, using autonomous underwater vehicles (AUVs) to explore hydrothermal deposits is more cost-effective and safer.

Obviously, unmanned and intelligent ocean exploration is the future development trend. In the past few years, AUV has played an essential role in the exploration of hydrothermal fields (German et al., 2008; Kumagai et al., 2010; Minami and Ohara, 2020). To obtain detailed visual data about seafloor hydrothermal activity more efficiently, Okamoto et al. (2019) developed a hovering-type AUV named Hobalin to perform visual observations with submillimeter image resolution using still cameras. However, the shutters of the still cameras were triggered every 4 seconds and the images captured were then analyzed by experts after the recovery of the AUV. In another word, Hobalin AUV was not capable of detecting the hydrothermal plume in real-time. To achieve autonomous sampling, AUV must first be endowed with the ability of intelligent detection of hydrothermal plumes. Several decades of exploration have collected a large number of images and videos of the deep-sea hydrothermal plume, making it feasible to build and train a deep learning model for plume detection based on these data.

In recent years, deep learning, part of a broader family of AI methods, has become a hot research method in various fields, providing intelligent solutions to some complex problems that previously required human expertise. Deep learning is a data-driven technique, learning from a large amount of labeled data to build a model. The model is then used to analyze unlabeled data and make predictions. With the continuous advancement of ocean observation technology, the amount and dimensions of ocean data have risen sharply. Applying deep learning methods to marine exploration has attracted more and more attention. For example, Ditria et al. (2020) utilized the Mask R-CNN model to analyze fish abundance automatically; Xu et al. (2021) applied three deep learning schemes to oceanic eddy detection; Li et al. (2021) proposed an improved YOLOV3 model with densely connected structures to achieve *in situ* zooplankton detection; Kandimalla et al. (2022) developed a fish passage observation platform to monitor fish with deep learning methods; Liao et al. (2022) used the MobileNet-SSD model and key-frame extraction detection method to detect the damage of far-sea underwater cage. As far as we know, there is limited research work on deep-sea hydrothermal plume detection *via* deep learning techniques.

This paper proposes a real-time object detection model for deep-sea hydrothermal plumes based on machine vision and deep learning. Firstly, a large number of hydrothermal plume images and videos were collected and then labeled manually to build a dataset for model training. Secondly, several state-of-the-art models, such as Faster R-CNN, SSD, and YOLO series were custom trained on the hydrothermal plume dataset. By comparing the plume detection performance, we selected the YOLOv5n algorithm as the baseline model in this paper. In addition, according to the characteristics of the hydrothermal plume and the real-time detection requirement, the original YOLOv5n network was optimized and improved, which not only maintains a good performance in inference speed but also achieves a tremendous increase in detection accuracy. Finally, the proposed plume detector was deployed on an edge AI computing device NVIDIA Jetson TX2 NX using TensorRT deep learning accelerator, as a vision processing unit on the AUV.

The main contributions and innovations of this paper are summarized below:

- We proposed a novel deep-sea hydrothermal plume dataset, which consists of 1589 images of the hydrothermal plume and 756 images of the seafloor background.
- We proposed a lightweight model based on the YOLOv5n algorithm to perform real-time plume detection. The original structure of YOLOv5n was improved in two aspects, including (1) replacing the SiLU activation functions with ReLU functions at shallow layers and Hard-SiLU functions at deep layers to increase the inference efficiency on the embedded system and (2) insetting an attention module termed Coordinate Attention (CA) into the backbone network to improve the overall accuracy.
- We adopted a transfer learning approach to utilize the knowledge learned from a fire and smoke dataset to better train the deep-sea hydrothermal plume detection model and enhance the model's robustness and generalization.
- We deployed the proposed plume detection model on an embedded AI device NVIDIA Jetson TX2 NX using the TensorRT inference framework and achieved the real-time on-board high-accuracy plume detection.

2 Materials and methods

2.1 Object detection deep learning models

Deep learning in the machine vision field includes image classification, object detection, and instance segmentation. Object detection refers to the technique for recognizing and locating some

specific objects in an image. As one of the primary tasks in machine vision, it has been extensively studied in the past few decades. In recent years, convolutional neural networks (CNNs) have gradually been the mainstream of object detection with the tremendous successes of deep learning in image classification (Liu et al., 2020). Generally, CNN-based object detection algorithms could be further classified into two-stage detectors and one-stage detectors (Wu et al., 2020). The former methods firstly generate the region proposals from input images and then do object classification and refinement for each region proposal. The latter innovatively reframe the object detection as a simple regression problem and perform the localization and classification in one network. Thus, one-stage methods normally have a better real-time performance.

Typical two-stage object detection methods are the R-CNN series, which include R-CNN (Girshick et al., 2014), Fast R-CNN (Girshick, 2015), and Faster R-CNN (Ren et al., 2015), etc. Representative one-stage object detection models are SSD (Liu et al., 2016) and YOLO series (including YOLO (Redmon et al., 2016), YOLOv2 (Redmon and Farhadi, 2017), YOLOv3 (Redmon and Farhadi, 2018), YOLOv4 (Bochkovskiy et al., 2020), and YOLOv5[1]). YOLOv5 is the latest version of the YOLO series and its structure remains close to YOLOv4. The major difference between YOLOv4 and YOLOv5 is that they are developed in different deep learning frameworks. They are comparable in terms of accuracy, but YOLOv5 outperforms YOLOv4 in the aspect of speed and ease of model deployment. Unlike previous versions of YOLO developed in the Darknet[2], YOLOv5 is built in the PyTorch[3] framework. Darknet is an open-source neural network framework written in C and CUDA. PyTorch, written in python, is much more easily configurable than Darknet, making YOLOv5 much more production ready.

Like EfficientNet (Tan and Le, 2019), YOLOv5 adopts a simple yet effective compound scaling method that uniformly scales the network width and depth through a set of fixed scaling coefficients, called depth_multiple and width_multiple respectively. Based on differences in network depth and width, there are five models in the official release of YOLOv5-v6.1, named YOLOv5l, YOLOv5m, YOLOv5s, YOLOv5n, and YOLOv5x. Among them, YOLOv5l is the baseline model with both depth_multiple and width_multiple equal to 1, and the other four models are acquired at different scaling scales. Figure 1 shows the overall structure diagram of the YOLOv5-v6.1 series models. The CBS module is the basic module in YOLOv5-v6.1. It is composed of a 2D convolution (conv2d) layer, a batch normalization (BN) layer, and a sigmoid-weighted linear unit (SiLU) activation layer. YOLOv5 adopts CSPDarknet53 as the backbone and uses the same head as YOLOv3. Different from YOLOv4, YOLOv5 introduces the CSP network into PAN and then uses it as the neck. There are two kinds of CSP modules: the CSP in the backbone (noted as CSP1) and the CSP in the neck (noted as CSP2). The former consists of three CBS modules and several Bottleneck modules, while the latter only replaces a Bottleneck module with two CBS modules based on the former. It should be pointed out that the number of the Bottleneck modules is determined by the depth_multiple. In addition, YOLOv5 implements a new cascaded SPP module (called SPPF) that produces mathematically identical results to SPP with faster speeds. Finally, YOLOv5 has some small but useful tricks, such as the auto-learning bounding box anchors generating mechanism for custom datasets and the maximum batch-size auto-computing mechanism.

2.2 Deep-sea hydrothermal plume dataset

Deep Learning models are data-hungry for creating the best model or system with high performance. Finding or creating a quality dataset is a fundamental requirement for developing any real-world AI application. To our knowledge, there are currently no publicly available deep-sea hydrothermal plume datasets. As mentioned before, we have collected a lot of videos and images of the hydrothermal plume during the sampling process using ROV or HOV in the past few years. According to these private data and other small parts of images acquired from the internet, we built the first deep-sea hydrothermal plume dataset, consisting of 1589 images of the hydrothermal plume and 756 images of the deep-sea background. The plume dataset serves as a valuable resource for deep learning-based object detection model. Specifically, the background images are instrumental in training models to differentiate between areas with and without objects, which can reduce false positives and improve the accuracy of plume detection. Figure 2 shows some example images of the plume dataset.

2.2.1 Image extraction and labelling

The initial plume dataset contains about 60 videos filmed by ROV or HOV in different hydrothermal fields. In total there are approximately 10 h of video. We extracted 68 clips with hydrothermal plume target. Object detection models require still images with precisely labelled bounding boxes for training. However, the original video clips do not match the requirements. To meet the demands, we further extracted frames from these video clips at intervals of 2 seconds and then manually labelled the images using the LabelImg[4] application. We saved these label files in the YOLO format, needed for YOLO models' training. In addition, we wrote a python script to convert the YOLO format labels to the COCO format automatically. The COCO format labels are used to train and evaluate Faster-RCNN and SSD models developed in MMDetection open source object detection toolbox (Chen et al., 2019).

2.2.1 Dataset partitioning

All images and their corresponding labels were divided into 3 subsets, namely Train, Val, and Test. The former two subsets (also

1 https://github.com/ultralytics/yolov5

2 https://pjreddie.com/darknet/

3 https://pytorch.org/

4 https://github.com/heartexlabs/labelImg

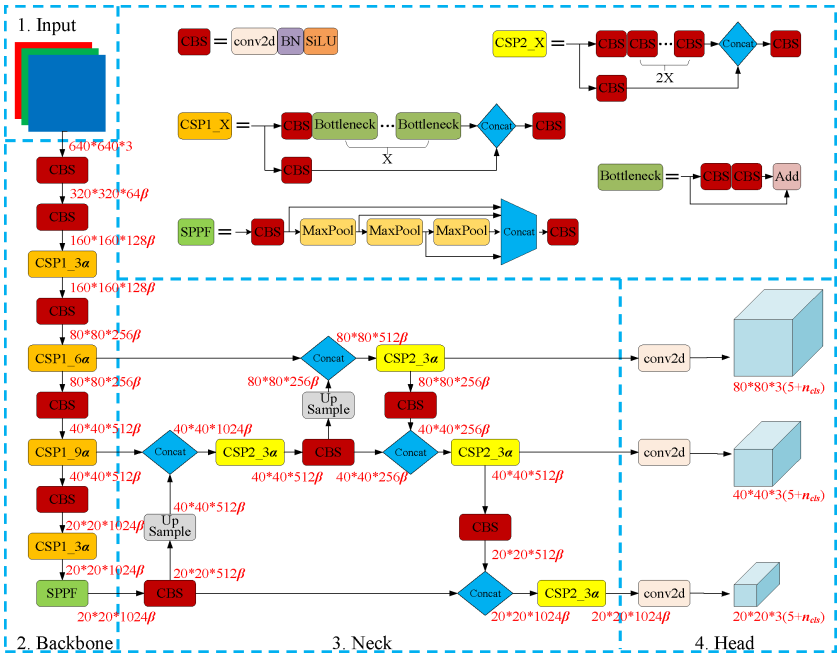


FIGURE 1
YOLOv5-v6.1 network structure diagram. α is the depth_multiple, β is the width_multiple, and $n_{cls}$is the number of object classes. 640*640*3, 320*320*64β etc. indicate the height*width*channels of the feature maps.

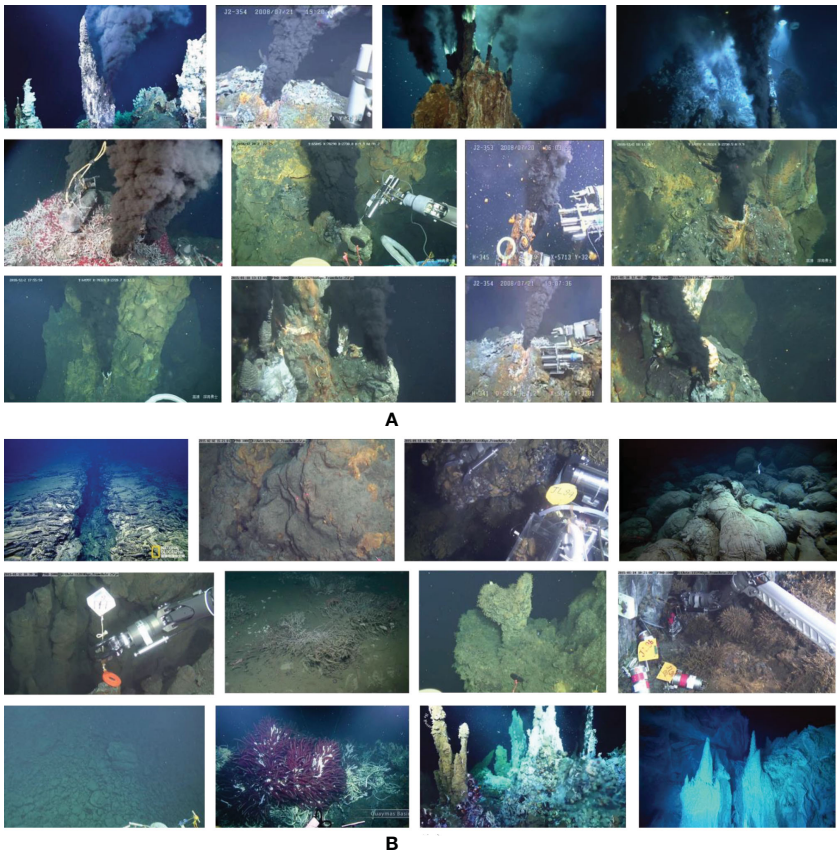


FIGURE 2
Example images from the deep-sea hydrothermal plume dataset: **(A)** images of hydrothermal plume; **(B)** images of deep-sea background.

can be seen together as TrainVal set) are used for model training, while the latter one subset is used for final model performance evaluation. To better assess the trained model's real performance, the images in the Test set and the images in the TrainVal set are from different videos. The numbers of images in different subsets of the plume dataset are shown in Table 1. We randomly picked 80% samples of the annotated images in the TrainVal set for model learning, with the remaining 20% used to form a validation set for best learning result assessment. Overfitting, which refers to a phenomenon that the model tries to fit the training data entirely and ends up performing poor in the case of unseen data scenarios, is a common problem in deep learning. We minimized overfitting by using the early-stopping technique. In our case, this was achieved by assessing the AP@0.5 on the Val set at intervals of one epoch and stopping the training when there were at least 50 epochs without improvement.

2.2.3 Data augmentations

Deep convolutional neural networks have significantly improved the state of the arts on object detection machine vision tasks. However, their impressive performances are heavily reliant on massive amounts of labelled data for supervised learning. To improve the performance, we use the data augmentation technique to increase both the size and the diversity of the labelled plume dataset by leveraging label preserving transformations.

Common augmentations can be divided into two categories: geometric transformations and photometric transformations. Geometric transformations alter the geometry of the image to make the CNN insensitive to changes in position and orientation. Example transformations include flipping, cropping, scaling, translating, and rotating; As for photometric transformations, it adjusts the color channels or adds some noise to make the CNN insensitive to changes in illumination and color (Taylor and Nitschke, 2018). YOLOv5 not only contains these generic augmentations, but also has some unique augmentations such as Mosaic, MixUp, and Copy-Paste. However, it does not mean that the more data augmentation methods are used, the better the performance of the detector will be. For a specific dataset, it is necessary to choose some appropriate data augmentations. Obviously, flapping up-down is not practical in plume detection. In the proposed plume detector, we used a combination of Mosaic and some suitable traditional augmentation methods. To make the plume detector insensitive to changes in illumination and color, the enhancement coefficient of hue (H), saturation (S), and lightness (V) were set to 0.1, 0.8, and 0.5, respectively. Other coefficients were kept the same as those in YOLOv5.

TABLE 1 Numbers of images of different subsets in the deep-sea hydrothermal plume dataset.

	Train +Val	Train (80%)	Val (20%)	Test	Train+Val +Test
Plume	1205	963	242	384	1589
Background	666	532	134	90	756
Sum	1871	1495	376	474	2345

2.3 Deep-sea hydrothermal plume detector

2.3.1 Model architecture

Prior studies have shown that network depth and width are both important for models' expressive power. Normally, deeper CNN networks can capture richer and more complex features, and wider networks are capable of capturing more fine-grained features and are easier to train (Zagoruyko and Komodakis, 2016; Raghu et al., 2017). However, with the network getting deeper and wider, the number of weights increases and the inference speed goes slower. For a specific object detection task, choosing suitable depth and width is necessary for the trade-off between accuracy and efficiency. Through a comprehensive comparative experiment (related results are detailed in section 3.1), we finally chose to build the real-time deep-sea hydrothermal plume detector based on YOLOv5n. Furthermore, to accurately and efficiently detect hydrothermal plumes by using an embedded system, the original structure of YOLOv5n was improved in two aspects. First, SiLU activation functions in the model were replaced by ReLU at shallow layers and Hard-SiLU at deep layers to reduce the number of calculations (see section 2.3.1.1 for more details). Second, a CA attention module was integrated into the backbone network to improve the model's sensitivity to channel and spatial features (see section 2.3.1.2 for more details). The architecture of the proposed hydrothermal plume detector is shown in Figure 3.

2.3.1.1 Activation functions

The choice of activation functions in deep neural networks has a significant effect on the training dynamics and task performance. Currently, the most successful and widely-used activation function is the Rectified Linear Unit (ReLU) (Nair and Hinton, 2010; Krizhevsky et al., 2017), defined as:

$$\mathrm{ReLU}(x) = \begin{cases} 0, & x \le 0 \\ x, & x > 0 \end{cases} \tag{1}$$

Thanks to its simplicity and effectiveness, ReLU has become the default activation function used across the deep learning community. In YOLOv5, a nonlinearity called SiLU (Elfwing et al., 2018), also called Swish (Ramachandran et al., 2017), was employed as the activation function, which significantly improves the accuracy of neural networks. The activation of the SiLU is computed by the sigmoid function multiplied by its input. This nonlinearity is defined as:

$$\mathrm{SiLU}(x) = x \cdot \sigma(x) = x \cdot \frac{1}{1+e^{-x}} \tag{2}$$

where the $\sigma(x) = \frac{1}{1+e^{-x}}$ denotes the sigmoid function.

While this nonlinearity improves accuracy, it comes with a higher cost in embedded environments as the sigmoid function is much more expensive to compute than ReLU. Inspired by

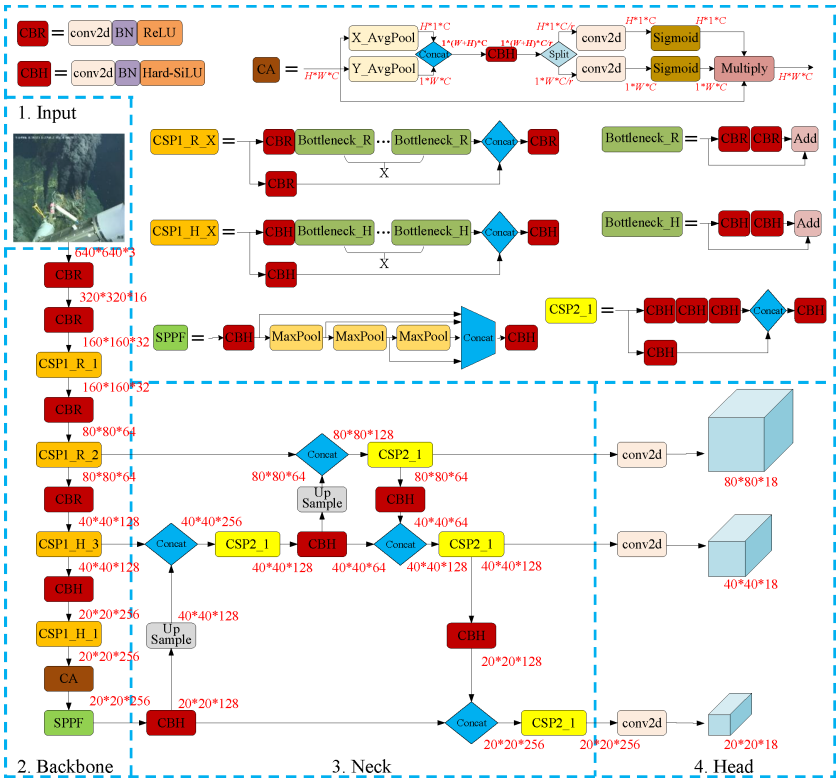


FIGURE 3
Network structure diagram of the proposed hydrothermal plume detector.

MobileNetV3 (Howard et al., 2019), we handled this problem by replacing the sigmoid function with its piece-wise linear hard analog $\frac{\text{ReLU6}(x+3)}{6}$, where

$$\text{ReLU6}(x) = \begin{cases} 0, & x \le 0 \\ x, & 0 < x \le 6 \\ 6, & x > 6 \end{cases} \tag{3}$$

ReLU6 is a modification of the ReLU with a maximum size of 6. Similarly, the hard version of SiLU becomes:

$$\text{Hard-SiLU}\ (x) = x \cdot \frac{\text{ReLU6}(x+3)}{6} = \begin{cases} 0, & x \le -3 \\ \frac{x^2+3x}{6}, & -3 < x \le 3 \\ x, & x > 3 \end{cases} \tag{4}$$

The values of constants in Hard-SiLU was determined by being a good match to the original smooth version. Replacing SiLU with Hard-SiLU has no discernible difference in accuracy, but the piece-wise implementation of Hard-SiLU can reduce the number of memory accesses, improving the training and inference efficiency substantially. In addition, since most of the benefits of SiLU are obtained by using them only in the deeper layers, Hard-SiLU was only used in the deep layers in our modified model architecture. We replaced SiLU activations in the shallow layers with more simple ReLU activations to further reduce the calculation.

2.3.1.2 Attention modules

Attention modules, inspired by the attention mechanisms in the human brain, have been widely used for boosting the performance of modern deep neural networks, thanks to their ability to provide additional information on "where" and "what" to focus on. However, their application for lightweight networks deployed in the embedded systems with limited computing power significantly lags behind that for large networks running in the workstation with powerful and expensive GPU graphics cards. This is mainly because the computational overhead brought by most attention modules is not affordable for embedded devices. Considering the restricted computation competence of embedded systems, to date, the most popular attention mechanisms for lightweight neural networks are still the Squeeze-and-Excitation (SE) (Hu et al., 2018), Convolutional Block Attention Module (CBAM) (Woo et al., 2018), and Coordinate Attention (CA) (Hou et al., 2021). Among them, the SE attention only considers encoding inter-channel information but neglects the position information. The CBAM network exploits positional information by reducing the channel dimension of the input tensor and then computing spatial attention using convolutions. However, convolutions can only capture local relations but fail in modeling long-range dependencies. The CA module embeds positional information into channel attention to attend over large regions. It can capture not only cross-channel but also

position-sensitive information and generates coordinate-aware attention maps while avoiding increasing significant computation overhead. A previous study demonstrates that the CA module outperforms the SE and CBAM in object detection tasks (Hou et al., 2021). Thus, we inserted a CA module in the plume detector's backbone network to lift its performance.

The block diagram of the CA module is shown at the top of Figure 3. Unlike channel attention that converts a feature tensor to a single feature vector *via* 2D global pooling, the CA module employs two 1D global pooling operations to respectively aggregate the input features map X (dimension is $H*W*C$) along the vertical and horizontal directions into two separate direction-aware feature maps $X1$ (dimension is $H*1*C$) and $X2$ (dimension is $1*W*C$). These two feature maps embedded with direction-specific information are then concatenated to form feature map $X3$ (dimension of $1*(H+W)*C$) which is subsequently passed through a CBH module. The CBH module reduces the channels from C to C/r based on a specified reduction ratio r (in our implementation, r was set to 32). After that, the feature map acquired is split into two tensors which are then individually passed through two 2D convolution kernels to increase the channels back to C from C/r. Finally, two sigmoid activations are separately applied on the resultant two tensors to form the two attention maps, each of which captures long-range dependencies of the input feature map along one spatial direction. Therefore, the positional information can be preserved in the generated attention maps. The two attention maps are then sequentially element-wise multiplied with the original input feature map to emphasize the representations of interest. The dimension of the feature map does not change after it was processed by the CA module. Thus, it is convenient to plug the CA module into any classic deep neural network.

2.3.2 Model training and deployment

After preparing the plume dataset and defining the model's structure, we trained the plume detector in the Pytorch deep learning framework. A weights file was obtained after the training. For model deployment on Jetson TX2 NX, we firstly rebuilt the plume detector by the NVIDIA TensorRT C++ API and then generated the plume detection TensorRT engine file with the weights file obtained in the model training process. The detailed flow chart of the proposed plume detection method including model training and deployment process has been shown in Figure 4.

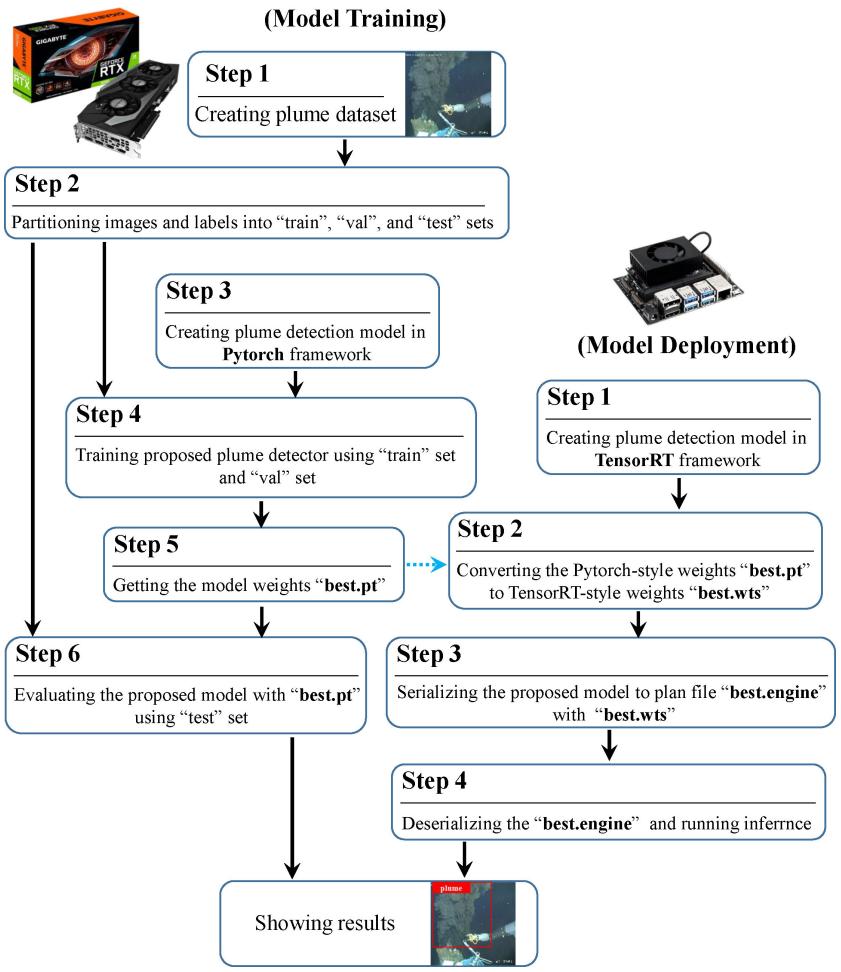


FIGURE 4
Flow chart of the proposed method, including model training (left) and model deployment (right) two parts.

2.3.2.1 Model training platform

A workstation running the Ubuntu 20.04 operating system was used as the model training platform. The Pytorch framework and the proposed plume detection algorithm were built in the Anaconda3 environment. The program was written in Python 3.7, and the CUDA version was 11.1. For hardware, the processor was an AMD 64-Core Ryzen Threadripper Pro 3995WX with 3.5 GHz main frequency, the memory was 126G, and the graphics card was a GeForce RTX 3090 24G.

2.3.2.2 Training with transfer learning

Transfer learning, inspired by human beings' capabilities to transfer knowledge across domains, is an effective model training method that can leverage knowledge from a different but related source domain to improve the performance of target learners in a specific target domain where labelled data is scarce (Zhuang et al., 2020). As we described before, there are no open-source deep-sea hydrothermal plume datasets. The plume dataset created by ourselves only contains 2345 images in total. For training of deep learning models, it is not large enough. Although the data augmentation technique can rich the plume dataset to a certain extent, its improvement is still limited. Thus, we further utilized the transfer learning method to boost the plume detector's performance. As can be seen in Figure 5, the deep-sea hydrothermal plume is very similar in shape, texture, and color to the smoke on land. After searching, a fire and smoke dataset[5], consisting of 23.7k images, was found in the Kaggle community. We treated the fire and smoke dataset as the source domain in our transfer learning experiment.

The transfer learning method commonly used in the field of deep learning is to copy the weights of the base network trained on the source dataset to the target network, and then train the target network on the target dataset. For the transferred weights, there are two methods to deal with: fine-tuning or freezing. The former method backpropagates the errors from the new task into the copied features to fine-tune them to the new task. The latter method leaves the transferred feature layers frozen in place, meaning that the copied features do not update during training on the new task. How to handle the copied features of the target network is determined by the size of the target dataset and the number of parameters of the network. On the one hand, when the number of parameters is large and the target dataset is small, an overfitting phenomenon may occur if only fine-tune the transferred features, so some of the weights (first n layers) can be left frozen. On the other hand, when the target dataset is large or the number of parameters is small, there are enough data for parameter update to make that overfitting do not happen, so all the copied features can be fine-tuned to the new task to boost performance. Of course, if the target dataset is very large, there would be no need to utilize the transfer learning method since the labelled images in the dataset are rich enough for model training from scratch (Yosinski et al., 2014). In our case, we conducted some comparative experiments to do a better choice. An overview of the experimental treatments and controls of the transfer learning method is shown in Figure 6. Through experiments, we found that the best detection performance was obtained by only fine-tuning the target network with no need for freezing the first n layers. Although the plume dataset is not very large, our proposed plume detector is a lightweight network with few parameters, so this result is reasonable. The detailed experimental results can be seen in section 3.4.

2.3.2.3 Model deployment platform

NVIDIA Jetson TX2 NX[6] is a powerful but compact embedded product. It provides up to 1.33 TFLOPs AI performance, which is 2.5 times the performance of Jetson Nano. Jetson TX2 NX shares form-factor and pin compatibility with Jetson Nano. It consists of an NVIDIA Pascal architecture GPU with 256 CUDA cores, a CPU complex with Denver 2 (Dual-Core) processor and ARM Cortex-A57 MPcore (Quad-Core) processor, and a 4GB-128bit-LPDDR4 memory. Thanks to its high performance, low power consumption, and small form-factor, Jetson TX2 NX is a preferred hardware platform to deploy a hydrothermal plume detection model for AUV with limited space and constrained battery. The trained plume detection model has been transplanted into the Jetson TX2 NX using the TensorRT[7] inference engine, a C++ library for high-performance inference on NVIDIA GPUs.

3 Results

3.1 Performances of different scale models of YOLOv5

To find a suitable baseline model for real-time plume detection, besides the officially released four models (YOLOv5 l, m, s, n), we still trained and tested three custom models (named YOLOv5 s1, n1, n2) on the plume dataset. The tested seven models have a similar structure with only differences in depth (layers) and width (channels). In this experiment, the number of training epochs was set to 600, and the optimal batch-size, which refers to the number of training examples utilized in one iteration, was auto-computed for the 90% utilization of GPU memory. All the hyper-parameters used were the default values in the *hyp.scratch-low.yaml*[8] file.

The parameters and plume detection performances of different scale models of YOLOv5 are detailed in Table 2. From Table 2, the following conclusions can be drawn:

- As the network gets deeper and wider, the required inference time increases dramatically.

5 https://www.kaggle.com/datasets/hhhhhhdoge/fire-smoke-dataset

6 https://developer.nvidia.com/embedded/jetson-tx2-nx

7 https://github.com/NVIDIA/TensorRT

8 https://github.com/ultralytics/yolov5/blob/master/data/hyps/hyp.scratch-low.yaml

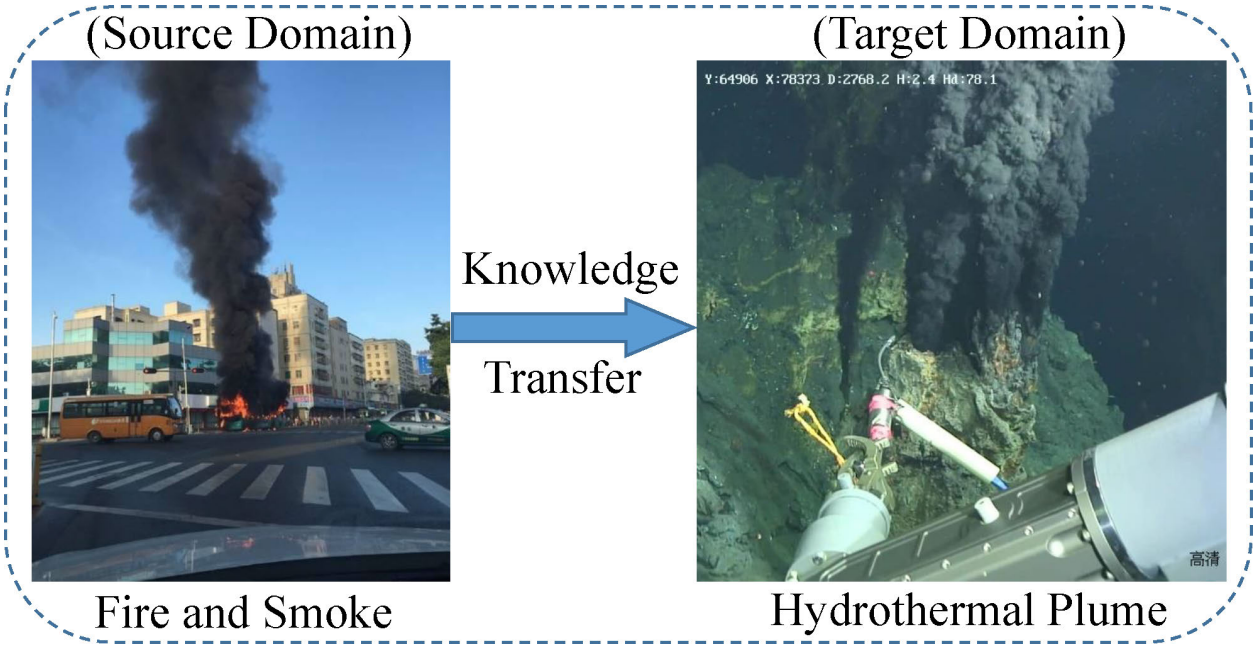


FIGURE 5
Schematic diagram of the knowledge transfer.

- In general, large models have better accuracy performance in plume detection than small models, but the difference is not as big as that of inference time.
- Reducing the computation of model, the maximum batch-size that GPU can support increases.
- To balance accuracy and efficiency, YOLOv5n is a better choice for building the plume detector. Compared with YOLOv5m, it only has a slight loss of accuracy (0.069 AP decrease, i.e., from 0.741 to 0.672) but almost 4 times increase of inference speed (i.e., from 3.7 ms/image to 0.9 ms/image). Compared with YOLOv5n2, it has comparable speed but much higher accuracy performance.

3.2 The effectiveness of CA module

For the qualitative analysis of the CA module's impact on the performance of the plume detector, we apply the Grad-CAM as our visualization tool to visualize the class activation maps of models with and without attention mechanism. Grad-CAM is a recently proposed visualization method that uses gradients to calculate the importance of the spatial locations in convolutional layers (Selvaraju et al., 2017). The visualization results of Grad-CAM can clearly illustrate the attended regions of models. Figure 7A is the original input image with ground truth marked, Figure 7B shows the Grad-CAM heat-map and predicted plume bounding box

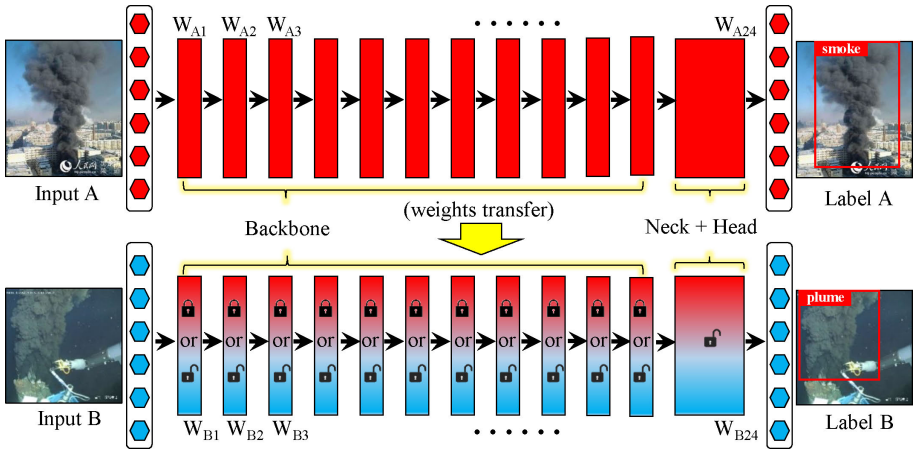


FIGURE 6
Overview of experimental treatments and controls of the transfer learning.

TABLE 2 Plume detection performance comparisons between different scale models of YOLOv5-v6.1.

	l	m	s1	s	n1	n	n2
depth_multiple	1	0.67	0.33	0.33	0.33	0.33	0.33
width_multiple	1	0.75	0.625	0.50	0.375	0.25	0.125
layers	468	369	270	270	270	270	270
channels	64	48	40	32	24	16	8
batch-size (training)	40	65	98	118	149	206	392
weights (MB)	92.9	42.23	22.3	14.5	8.3	3.9	1.3
GFLOPs	107.9	48	24.5	15.8	9	4.2	1.2
AP@0.5_test	0.73	0.741	0.704	0.712	0.626	0.672	0.578
inference time (ms/image)	5.1	3.7	2.5	1.7	1.4	0.9	0.7

The number of layers is counted before layer fusing. The channels denotes the output channels of the first CBS module. The inference time is tested at shape (16, 640, 640, 3) on NVIDIA RTX 3090 24G graphics card (16 is the batch-size, (640, 640, 3) is the dimension of image). The batch-size denotes the maximum number of images for 90% utilization of GPU memory during model training.

produced by the baseline model (YOLOv5n), and Figure 7C displays the visualization results of the model with the CA module integrated (YOLOv5n+CA). It is evident that the Grad-CAM mask of the CA-integrated model can more precisely cover the target plume region than that of the model without the CA module, which fully reveals the CA module's effectiveness in promoting the plume detector to exploit information in target object regions. Consequently, the CA module improves the prediction of the plume bounding box. For the quantitative analysis, the CA-integrated model outperforms the baseline model by 3.1% AP (from 0.672 to 0.703) with only a bit of parameters increase (from 1.76M to 1.77M).

3.3 The effectiveness of activations

Table 3 shows the ablation study of YOLOv5n with and without activations changed. From Table 3, one can find that replacing SiLU with Hard-SiLU can guarantee a larger batch-size under the same GPU memory condition and less inference time, which confirms that the Hard-SiLU activation can offer a model with reduced memory utilization and computation. In addition, there is even a slight increase in the AP performance, which indicates that replacing SiLU with Hard-SiLU has no discernible difference in accuracy.

TABLE 3 The ablation study of YOLOv5n with and without activations changed.

	YOLOv5n	YOLOv5n with activations changed
batch-size (training)	206	225
AP@0.5_test	0.672	0.681
inference time (ms/image)	0.9	0.8

The inference time is calculated at shape (16, 640, 640, 3) on NVIDIA RTX 3090 24G graphics card.

3.4 The effectiveness of transfer learning

In the transfer learning experiments, we first trained the improved YOLOv5n model (i.e., our proposed model) on the fire and smoke dataset for 600 epochs. After about 12 hours of training, a weights file of about 3.9 MB in size was obtained. Subsequently, we trained the improved YOLOv5n model on the custom-made deep-sea hydrothermal plume dataset with loading the fire and smoke pre-trained weights. We choose to freeze the first n (from 0 to 10) modules of the model and then see the performance differences in different situations. Figure 8 shows the results of this transfer learning experiment. From Figure 8, one can see that the best AP performance can be obtained by only fine-tuning the target network without freezing. With transfer learning, we achieved the best AP of 0.765.

3.5 Comparisons with state-of-the-arts (other alternatives)

Table 4 shows the performance comparisons of our proposed plume detector with the other three state-of-the-art methods. One

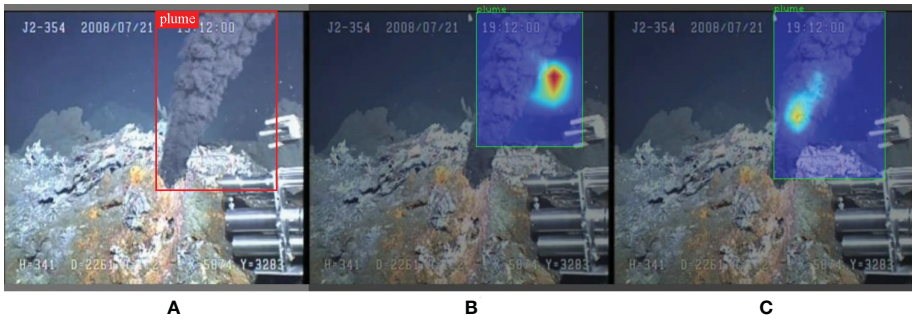


FIGURE 7
Grad-CAM visualization results. **(A)** The input image with ground truth marked. **(B)** The visualization result of baseline model (YOLOv5n). **(C)** The visualization result of CA-integrated model (YOLOv5n+CA).

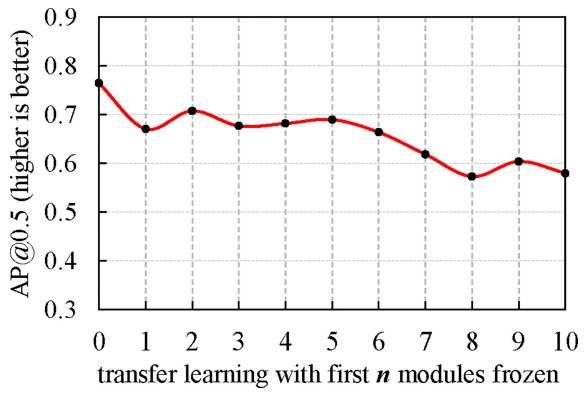


FIGURE 8
Results of the transfer learning experiment.

TABLE 4 The comparison results with other state-of-the-art methods.

	Faster-RCNN	SSD-512	YOLOv5n	ours
weights (MB)	330.3	195.2	3.9	3.9
parameter size	41.12M	24.39M	1.76M	1.77M
AP@0.5_test	0.598	0.661	0.672	0.765
inference time (ms/image)	482.9	463.0	4.9	4.9

The inference time is tested on NVIDIA RTX 3090 24G graphics card with batch-size equals to 1. The parameter size refers to model complexity.

can see that our proposed plume detector outperforms Faster-RCNN and SSD-512 in both accuracy and inference speed. Compared to the baseline model YOLOv5n, our model can make a 9.3% AP improvement (i.e., from 0.672 to 0.765) and requires the same inference time, which fully reveals that our improved plume detector can not only maintain a good performance of efficiency but also achieves a tremendous increase in detection accuracy.

3.6 Test results on Jetson TX2 NX

We tested the deployed plume detector as a stand-alone application in Jetson TX2 NX to evaluate its final application performance. As shown in Figure 9, a USB camera and a monitor are connected to the Jetson TX2 NX. The camera was exposed to another computer that was playing a video with the target of the deep-sea hydrothermal plume. The monitor was used to present the detection results. we recorded various parameters when the proposed plume detector was running in the Jetson TX2 NX. The resolution of the USB camera is 640x480. Although this camera only supports a 30 fps frame rate at this resolution, the video processing speed of the plume detector calculated was about 37 fps. It should be noted that we take not only the inference time but also the pre-processing and post-processing into account to calculate the fps. The pre-processing includes converting BGR to RGB and resizing the image to 640x640. The post-processing includes non-maximum suppression (NMS) processing of the predicted bounding box and putting the final bounding box on the image. The real-time on-board detection capability of the proposed plume detector is demonstrated. During the test, the power consumption and GPU temperature of the Jetson TX2 NX were also measured using jtop[9], a system monitoring utility. The power consumption was 2.58 W on average and the temperature was 40.5°C when the plume detector was off. While our method was executed, the power consumption was 4.02 W and the temperature was 49.5°C. Thus, the power consumption of our proposed plume detector is only 1.44 W, which benefits a lot for application in battery-powered devices.

3.7 Plume detection test in water tank with AUV

We then installed the Jetson TX2 NX into a hovering-type AUV as its vision processing unit. A water tank test was conducted in the laboratory. As shown in Figure 10, we simulated a dynamic hydrothermal plume in the water tank, and then used the hovering-type AUV to explore it. The experimental results show that our proposed hydrothermal plume detector can endow the AUV with the ability of real-time and intelligent detection of the hydrothermal plume. Based on the detection result, we also designed an image-based visual servo (IBVS) algorithm to help the AUV approach the target automatically and did some tests in the water tank. Since that is out of the scope of this paper, we will present the related results in another paper.

4 Conclusion and discussion

This paper proposed a real-time and embedded implementation of the deep-sea hydrothermal plume detection technique that can give AUV the competence of intelligent hydrothermal plume detection. The proposed solution achieved more promising results for accuracy and efficiency performance compared to other state-of-the-art methods. The preliminary water tank test verified the feasibility of the proposed method and laid the foundation for the accurate detection or sampling of deep-sea hydrothermal plumes by using AUV. However, only the static spatial features are considered in our proposed plume detector now, as only the still images are used for training. In the future, we will further improve our algorithm by taking the dynamic and temporal characteristics of the hydrothermal plume into account. One possible method is to combine CNNs and LSTM (long short-term memory) neural networks in a consecutive way so that the

9 https://github.com/rbonghi/jetson_stats

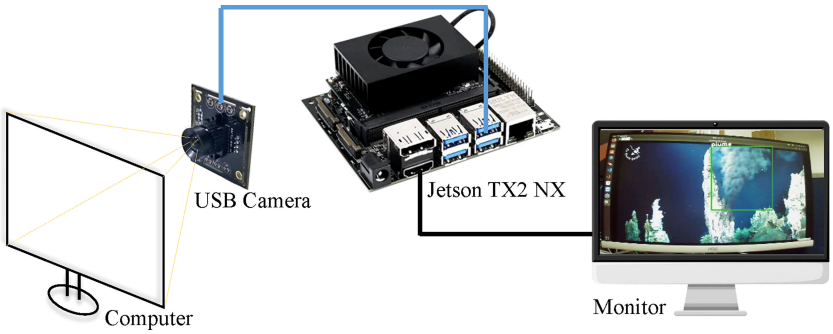


FIGURE 9
Testing the proposed plume detector on Jetson TX2 NX.

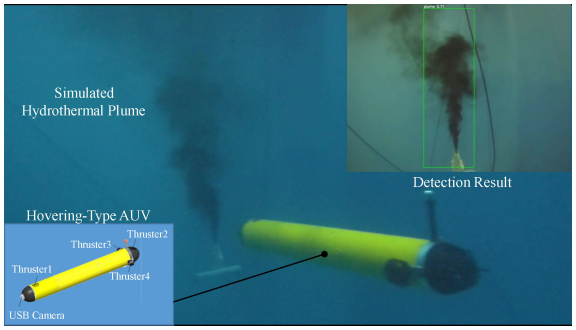


FIGURE 10
Water tank test of a hovering-type AUV configured with the proposed plume detector.

model can learn the dynamic and temporal features of the hydrothermal plume. However, the LSTM usually causes high computational costs. How to make this method be practically used on AUV needs to be further studied. Finally, we will extend our research to deploy the proposed plume detection system on an AUV for achieving automatic plume sampling by providing real-time visual status and feedback on the hydrothermal plumes in actual sea trials.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

SW and CY designed the study. XW created the plume dataset and performed the experiments and drafted the manuscript. YC and SW proofread the manuscript and gave valuable advice. All authors contributed to the article and approved the submitted version.

Funding

This work is supported by the National Science Foundation of Zhejiang Province, China (Grant No. LR21E090001) and the National Natural Science Foundation of China (Grant No. 51879232).

Acknowledgments

Most of the videos in our hydrothermal plume dataset are captured during the past research cruises we have attended, which include the cruise of KNOX18RR with the R/V Roger Revelle and the ROV Jason, the cruise of DY35 with R/V Xiang Yang Hong 09 and HOV Jiaolong, and the cruise of TS 10 with R/V Tan Suo Yi Hao and HOV Shen Hai Yong Shi, etc. We are grateful to all the crews for their support during these cruises.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1124185/full#supplementary-material

References

Bochkovskiy, A., Wang, C. Y., and Liao, H. Y. M. (2020). "Yolov4: Optimal speed and accuracy of object detection," in *arXiv*. doi: 10.48550/arXiv.2004.10934

Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., et al. (2019). "MMDetection: Open mmlab detection toolbox and benchmark," in *arXiv*. doi: 10.48550/arXiv.1906.07155

Chen, Y., Wu, S., Xie, Y., Yang, C., and Zhang, J. (2007). A novel mechanical gas-tight sampler for hydrothermal fluids. *IEEE J. Oceanic Eng.* 32 (3), 603–608. doi: 10.1109/joe.2007.891887

Corliss, J. B., Dymond, J., Gordon, L. I., Edmond, J. M., von Herzen, R. P., Ballard, R.D., et al. (1979). Submarine thermal springs on the Galapagos rift. *Sci.* 203 (4385), 1073–1083. doi: 10.1126/science.205.4409.856.d

Ditria, E. M., Lopez-Marcano, S., Sievers, M., Jinks, E. L., Brown, C. J., and Connolly, R. M. (2020). Automating the analysis of fish abundance using object detection: Optimizing animal ecology with deep learning. *Front. Mar. Sci.* 429. doi: 10.3389/fmars.2020.00429

Elfwing, S., Uchibe, E., and Doya, K. (2018). Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks* 107, 3–11. doi: 10.1016/j.neunet.2017.12.012

German, C. R., Yoerger, D. R., Jakuba, M., Shank, T. M., Langmuir, C. H., and Nakamura, K. I. (2008). Hydrothermal exploration with the autonomous benthic explorer. *Deep Sea Res. Part I* 55 (2), 203–219. doi: 10.1016/j.dsr.2007.11.004

Girshick, R. (2015). "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision* (Santiago: IEEE), 1440–1448. doi: 10.1109/iccv.2015.169

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Columbus: IEEE), 580–587. doi: 10.1109/cvpr.2014.81

Hou, Q., Zhou, D., and Feng, J. (2021). "Coordinate attention for efficient mobile network design," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (Virtual: IEEE), 13713–13722. doi: 10.1109/cvpr46437.2021.01350

Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., et al. (2019). "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF international conference on computer vision* (Seoul: IEEE), 1314–1324. doi: 10.1109/iccv.2019.00140

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks," in. *Proc. IEEE Conf. Comput. Vision Pattern recognition* (Salt Lake City: IEEE), 7132-7141. doi: 10.1109/cvpr.2018.00745

Kandimalla, V., Richard, M., Smith, F., Quirion, J., Torgo, L., and Whidden, C. (2022). Automated detection, classification and counting of fish in fish passages with deep learning. *Front. Mar. Sci.* 2049. doi: 10.3389/fmars.2021.823173

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60 (6), 84-90. doi: 10.1145/3065386

Kumagai, H., Tsukioka, S., Yamamoto, H., Tsuji, T., Shitashima, K., Asada, M., et al. (2010). Hydrothermal plumes imaged by high-resolution side-scan sonar on a cruising AUV, urashima. *Geochem., Geophys. Geosyst.* 11 (12). doi: 10.1029/2010gc003337

Li, Y., Guo, J., Guo, X., Zhao, J., Yang, Y., Hu, Z., et al. (2021). Toward *in situ* zooplankton detection with a densely connected YOLOV3 model. *Appl. Ocean Res.* 114, 102783. doi: 10.1016/j.apor.2021.102783

Liao, W., Zhang, S., Wu, Y., An, D., and Wei, Y. (2022). Research on intelligent damage detection of far-sea cage based on machine vision and deep learning. *Aquac. Eng.* 96, 102219. doi: 10.1016/j.aquaeng.2021.102219

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., et al. (2016). "Ssd: Single shot multibox detector," in *European Conference on computer vision* (Amsterdam: Springer), 21–37. doi: 10.1007/978-3-319-46448-0_2

Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., et al. (2020). Deep learning for generic object detection: A survey. *Int. J. Comput. Vision* 128 (2), 261–318. doi: 10.1007/s11263-019-01247-4

Luther, G. W., Rozan, T. F., Taillefert, M., Nuzzio, D. B., Di Meo, C., Shank, T. M., et al. (2001). Chemical speciation drives hydrothermal vent ecology. *Nat.* 410 (6830), 813–816. doi: 10.1038/35071069

Minami, H., and Ohara, Y. (2020). Tectonic, volcanic and hydrothermal features of a nascent rift graben in the southern Okinawa trough. *Mar. Geo.* 430, 106348. doi: 10.1038/s41467-020-15062-w

Nair, V., and Hinton, G. E. (2010). "Rectified linear units improve restricted boltzmann machines," in *Icml* (Haifa: Omnipress), 807-814.

Okamoto, A., Seta, T., Sasano, M., Inoue, S., and Ura, T. (2019). Visual and autonomous survey of hydrothermal vents using a hovering-type AUV: Launching hobalin into the Western offshore of kumejima island. *Geochem. Geophys.* 20 (12), 6234–6243. doi: 10.1029/2019gc008406

Petersen, J. M., Zielinski, F. U., Pape, T., Seifert, R., Moraru, C., Amann, R., et al. (2011). Hydrogen is an energy source for hydrothermal vent symbioses. *Nat.* 476 (7359), 176–180. doi: 10.1038/nature10325

Raghu, M., Poole, B., Kleinberg, J., Ganguli, S., and Sohl-Dickstein, J. (2017). "On the expressive power of deep neural networks," in *International conference on machine learning*, 2847–2854.

Ramachandran, P., Zoph, B., and Le, Q. V. (2017). "Searching for activation functions," in *arXiv*. doi: 10.48550/arXiv.1710.05941

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Las Vegas: IEEE), 779–788. doi: 10.1109/cvpr.2016.91

Redmon, J., and Farhadi, A. (2017). "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition* (Honolulu: IEEE), 7263–7271. doi: 10.1109/cvpr.2017.690

Redmon, J., and Farhadi, A. (2018). "Yolov3: An incremental improvement," in *arXiv*. doi: 10.48550/arXiv.1804.02767

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* 28. doi: 10.1109/tpami.2016.2577031

Seewald, J. S., Doherty, K. W., Hammar, T. R., and Liberatore, S. P. (2002). A new gas-tight isobaric sampler for hydrothermal fluids. *Deep Sea Res. Part I* 49 (1), 189–196. doi: 10.1016/s0967-0637(01)00046-2

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision* (Venice: IEEE), 618–626. doi: 10.1109/iccv.2017.74

Tan, M., and Le, Q. (2019). "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning* (Long Beach: PMLR), 6105–6114.

Tao, C., Seyfried, W. E.Jr., Lowell, R. P., Liu, Y., Liang, J., Guo, Z., et al. (2020). Deep high-temperature hydrothermal circulation in a detachment faulting system on the ultra-slow spreading ridge. *Nat. com.* 11 (1), 1300. doi: 10.1038/s41467-020-15062-w

Taylor, L., and Nitschke, G. (2018). "Improving deep learning with generic data augmentation," in *2018 IEEE symposium series on computational intelligence (SSCI)* (Bangalore: IEEE), 1542–1547. doi: 10.1109/ssci.2018.8628742

Wang, X., Wu, S. J., Fang, Z. F., Yang, C. J., and Wang, S. (2020). A pressure-tight sampler with flexible titanium bag for deep-Sea hydrothermal fluid samples. *J. Atmos. Oceanic Technol.* 37 (11), 2065–2073. doi: 10.1175/jtech-d-20-0017.1

Woo, S., Park, J., Lee, J. Y., and Kweon, I. S. (2018). "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)* (Munich: Springer), 3–19. doi: 10.1007/978-3-030-01234-2_1

Wu, X., Sahoo, D., and Hoi, S. C. (2020). Recent advances in deep learning for object detection. *Neurocomputing* 396, 39–64. doi: 10.1016/j.neucom.2020.01.085

Xu, G., Xie, W., Dong, C., and Gao, X. (2021). Application of three deep learning schemes into oceanic eddy detection. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.672334

Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). "How transferable are features in deep neural networks?," in *Advances in neural information processing systems*, vol. 27. .

Zagoruyko, S., and Komodakis, N. (2016). "Wide residual networks," in *arXiv*. doi: 10.48550/arXiv.1605.07146

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., et al. (2020). A comprehensive survey on transfer learning. *Proc. IEEE* 109 (1), 43–76. doi: 10.1109/JPROC.2020.3004555

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 20 April 2023
DOI 10.3389/fmars.2023.1138013

Check for updates

OPEN ACCESS

EDITED BY
Rizwan Ali Naqvi,
Sejong University, Republic of Korea

REVIEWED BY
Chao Chen,
Zhejiang Ocean University, China
Ning Wang,
Dalian Maritime University, China
Ahmad Naeem,
University of Management and Technology, Lahore, Pakistan

*CORRESPONDENCE
Tao Liu
liutao08@zju.edu.cn
Shuangyan He
hesy@zju.edu.cn

SPECIALTY SECTION
This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 05 January 2023
ACCEPTED 30 March 2023
PUBLISHED 20 April 2023

CITATION
Deng X, Liu T, He S, Xiao X, Li P and Gu Y (2023) An underwater image enhancement model for domain adaptation. *Front. Mar. Sci.* 10:1138013. doi: 10.3389/fmars.2023.1138013

COPYRIGHT
© 2023 Deng, Liu, He, Xiao, Li and Gu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

An underwater image enhancement model for domain adaptation

Xiwen Deng[1,2], Tao Liu[1*], Shuangyan He[1,2*], Xinyao Xiao[1,2], Peiliang Li[1,2] and Yanzhen Gu[1,2]

[1]Ocean College, Zhejiang University, Zhoushan, China, [2]Hainan Institute, Zhejiang University, Sanya, China

Underwater imaging has been suffering from color imbalance, low contrast, and low-light environment due to strong spectral attenuation of light in the water. Owing to its complex physical imaging mechanism, enhancing the underwater imaging quality based on the deep learning method has been well-developed recently. However, individual studies use different underwater image datasets, leading to low generalization ability in other water conditions. To solve this domain adaptation problem, this paper proposes an underwater image enhancement scheme that combines individually degraded images and publicly available datasets for domain adaptation. Firstly, an underwater dataset fitting model (UDFM) is proposed to merge the individual localized and publicly available degraded datasets into a combined degraded one. Then an underwater image enhancement model (UIEM) is developed base on the combined degraded and open available clear image pairs dataset. The experiment proves that clear images can be recovered by only collecting the degraded images at some specific sea area. Thus, by use of the scheme in this study, the domain adaptation problem could be solved with the increase of underwater images collected at various sea areas. Also, the generalization ability of the underwater image enhancement model is supposed to become more robust. The code is available at https://github.com/fanren5599/UIEM.

KEYWORDS

underwater image, image enhancement, underwater dataset, domain adaptation, deep learning

1 Introduction

High-quality underwater photography and videography benefit ocean resource development and sustainable utilization (Mariani et al., 2018; Liu et al., 2020). However, underwater image degradation, including blurred edges and reduced visibility, is unavoidable due to the strong absorption and scattering of light in the water (Xie et al., 2022). The degree of degradation varies with different water constituents in various water areas (Fu et al., 2022). The enhanced visibility can make scenes and objects more

highlighted (Wang et al., 2023), although enhancing these poor-quality underwater images has been a challenging prerequisite for underwater object detection, monocular depth estimation, and underwater object tracking (Gao et al., 2019).

As shown in Figure 1, the underwater image quality mainly depends on two aspects (Rizzi et al., 2002; Foster, 2011; Ancuti et al., 2018). The first is brightness consistency. The reflected light of the underwater object is absorbed and scattered by the particles in the medium, and the camera usually captures low-contrast images. The second is color consistency. Light at long waves is more easily absorbed than short ones (Liu et al., 2020), thus underwater images generally appear bluish or greenish (Galdran et al., 2015). Accordingly, underwater image enhancement (UIE) algorithms basically aim at contrast enhancement and color cast correction (van de Weijer et al., 2007). In recent years, a variety of methods have been proposed for underwater image enhancement. Generally, the UIE methods can be divided into three categories: traditional methods without explicit model-building, physical prior-based methods, and data-driven deep learning methods.

Several traditional method scan better deal with noise, such as (Buchsbaum, 1980; Singh and Kapoor, 2014; Singh et al., 2015), lack of explicit modeling, which usually leads to problems of low contrast and color deviation. Several physical prior-based methods, such as (Kaiming He et al., 2009; Drews et al., 2016), can make the restored image conform to physical concepts, but the disadvantage is that there is often a lack of prior knowledge of experts and hand-crafted features (Lan et al., 2023). Several deep learning-based methods, such as (Li et al., 2017; Cao et al., 2018; Hou et al., 2018), have shown good enhancement and restoration performance. However, these studies are based on the data-driven approach, and thus their performance is limited by their specific dataset in two aspects. Firstly it is the domain adaptation problem (Huang and Belongie, 2017). The dataset used by different studies were collected in different water environments, and their optically active constituents vary with time and location (Devlin et al., 2009; Peterson et al., 2020). Besides, solar illumination also differs, and it also makes a significant difference in underwater objects' appearance and water styles. Limited training sample scan not sufficiently cover changeably complex underwater environments (Chen et al., 2021). Therefore, even though one model recovers the underwater image of a specific sea area well, it is usually difficult to apply with good generalization abilities in other sea areas. Secondly, it is difficult for the marine engineering community to obtain a large dataset of degraded-clear pairwise underwater images (Wang et al., 2022a). The performance of the deep learning-based models based on pairwise datasets is supposed to be better than those using unpaired ones. In natural water such as rivers, lakes, and oceans, the degraded underwater images can be continuously collected, while it is usually impossible to obtain pairwise clear ones simultaneously.

As a result, in order to solve the above domain adaptation problem, we propose an underwater attenuation fitting network in this study. By combining utilizing individual degraded images and publicly available pairwise underwater data sets, it can merge the individual localized and publicly available degraded datasets into a combined degraded one. Thus, the combined degraded images and open available clear images compose a pairwise dataset. An underwater image enhancement method based on this dataset is also provided in this study. Using our method, even if only degraded images of a certain sea area are collected, an underwater enhancement model in the sea area can be built and applied well. In this study, we used Underwater Image Enhancement Benchmark (UIEB) (Li et al., 2020b), which contains 950 real Underwater images. Among them, 890 have corresponding reference images. We collected three datasets from different sea areas using the submarine online observation system for training and testing. We evaluated the quality of the generated images and the image quality of our dataset using the evaluation quality index of no reference images (NIQE) (Mittal et al., 2013). Then we propose a new image-to-image underwater image enhancement model to verify the validity of our dataset. The enhancement effect of the underwater image enhancement algorithms on our dataset is significantly higher than that on the public dataset.

This paper is organized as follows. Chapter two reviews the existing UIE algorithms and public datasets. In chapter three, the structure of our proposed UDFM and UIEM are described in detail. In the fourth chapter, the performance of experiments on different

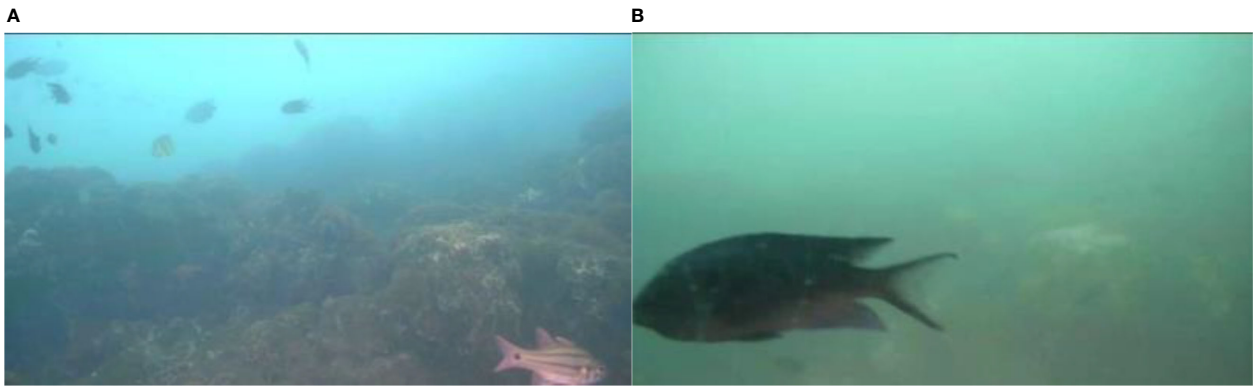


FIGURE 1
Challenges and problems in natural marine ranch environments: **(A)** a bluishimage with color distortion; **(B)** a low contrast and texture blurred image.

datasets is evaluated and compared with other image enhancement methods. The fifth part concludes and discusses this paper.

2 Relation work

2.1 Underwater image enhancement methods

Spatial domain methods such as histogram equalization (Singh and Kapoor, 2014)and gray world hypothesis algorithm (Buchsbaum, 1980), and frequency domain methods such as Fourier transform (Prabhakar and Praveen Kumar, 2011) and wavelet transform (Singh et al., 2015) are traditional methods. These methods can handle noises better but often bring problems of low contrast, loss of details, and color deviation.

Some researchers also built underwater imaging models to characterize the physical formation process of the underwater images and then estimated the depth-related transmission coefficients (Drews et al., 2016). Thus, they can reverse the degradation process to restore a clear underwater image, such as the Kai-ming DCP method (Kaiming He et al., 2009). The advantage of the method is that it can make the image enhancement process conform to physical concepts. The disadvantage is that experts' prior domain knowledge needs to be introduced into the model (Cheng et al., 2019), which does not always hold. These priors are often sensitive to the water environment where underwater images are collected, which causes the model may lack generalization. Besides, the imaging processing could be very complicated in various natural waters, and it is difficult to establish an appropriate general model.

With the improvement of the computing ability of hardware platforms, the introduction of excellent networks such as AlexNet (Krizhevsky et al., 2017) in 2012, and an ever-expanding database of digital images (Wang et al., 2022b), deep learning methods have been widely used in image processing (Li et al., 2017; Cao et al., 2018; Hou et al., 2018), natural language analysis (dos Santos and Gatti, 2014), and speech processing (Bollepalli et al., 2017). Data-driven deep learning-based methods are end-to-end methods, and they can solve the above problems easily by making the model learn basic parameters directly from the input. Deep convolutional network CNNs on large-scale datasets have shown excellent performance in many computer vision tasks (LeCun et al., 2015; He et al., 2016), which has motivated the development of data-driven UIE methods. At the same time, the proper availability of patterns for the training set affects the accuracy of test results (Naeem et al., 2022). However, real paired image datasets that meet the training objectives are often scarce. Thus, UIE methods usually use synthetic degraded images and high-quality counterpart images for training. Some UIE methods take a different approach and consider training with unpaired images.

As a pioneering work, Li et al. (Li et al., 2017) proposed the WaterGAN, which utilizes a generative adversarial network (GAN) and an image formation model to synthesize degraded/clear image pairs for unsupervised learning. To avoid the requirement of paired training data, Li et al. (Li et al., 2018) proposed a weakly supervised underwater color transformation model based on cycle-consistent adversarial networks, which alleviated the need for paired underwater images for training. However, the nature of the multiple possible outputs tends to produce unrealistic results in some cases. Li et al. (Li et al., 2020a) proposed to simulate real underwater images according to different water types and underwater imaging physical models. They first synthesized ten underwater images based on a revised underwater imaging model (Chiang and Chen, 2012). Then they use synthetic images to train the corresponding ten underwater image enhancement (UWCNN) models. This method can obtain a stable and good output. But facing the input underwater images, how to choose the appropriate UWCNN model is a challenge. Recently, Li et al. (Li et al., 2020b) collected a real-world paired underwater image dataset UIEBD to train a deep network. UIEBD can provide high-quality paired training data for depth models and a good evaluation of various underwater image enhancement methods.

2.2 Underwater image enhancement datasets

There are public underwater datasets, such as Fish4Knowledge (Boom et al., 2012), LifeCLEF2014 (Salman et al., 2016), LifeCLEF2015 (Salman et al., 2016), Sea-THRU (Salman et al., 2016), Haze-Line (Akkaynak and Treibitz, 2019), and UIEB(Devlin et al., 2009). These datasets are mainly used for object detection tasks. The Fish4Knowledge contains about 700,000 10-minute video clips of coral reefs spanned a time period of five years, including videos taken from sunrise to sunset. Note that these images were collected with varying water attenuation coefficients and solar illumination conditions and cannot be directly used in underwater image enhancement algorithms. The Sea-THRU dataset includes 1100 underwater images and range maps (Akkaynak and Treibitz, 2019), and the Haze-Line dataset has complete content, providing original images and camera calibration files (Berman et al., 2021). The UIEB (Li et al., 2020b) dataset contains 890 pairs of sharp and degraded pictures. In general, these existing datasets are individually collected in some specific sea areas in rather clear waters, so their representativeness is limited for real underwater images. Therefore, Li et al. (Li et al., 2020a) proposed to simulate real underwater images in different waters using underwater imaging physical models. They used the synthetic images to train ten underwater image enhancement (UWCNN) models. However, the underwater environment is complex, and ten models are not enough to characterize the underwater environment.

3 Method

The underwater imaging model (Chavez, 1988) can be expressed as:

$$I(x) = J(x)*T(x) + A*(1 - T(x)) \quad (1)$$

Where I(x) is the degraded underwater image; J(x) is the image to be recovered; A is background light; T(x) is the underwater medium transmission map depending on the water quality conditions, such as scattering and absorption coefficients of the water constituents. Therefore, the water quality conditions of other sea areas can be fused into the UIEB degraded images by extracting the medium transmission map.

3.1 Underwater dataset fitting model

Our underwater dataset fitting network is shown in Figure 2A. The input I are the degraded images of UIEB, and input II are the underwater images collected in other certain sea areas. In this study, the images were collected by an online seabed observation system in China Wei-zhou island. We first calculate the underwater medium transmission map of image II, and then splice the RGB image I and underwater medium transmission map of image II into the four-channel image III. Next, the output of a new degraded underwater image is generated by inputting Image III into the encoder and decoder. Note that the output image is generated from the degraded image I of UIEB, and it is further blurred with the underwater medium transmission map of image II by a neural network.

The neural network structure adopts the Encoder-Decoder structure. The structure of the Decoder is shown in Figure 2B. Encoder1 and Encoder2 use the first 31 layers of the VGG model. But the convolution kernel of the first layer is modified to 4×4 in Encoder1.

In this study, the underwater medium transmission map of T(x) is calculated based on the generalized dark channel prior algorithm (Peng et al., 2018), in which we improve the depth estimation method as shown in Equation (2).

$$D(x) = \min_{c,y \in \Theta(x)} (1 - w_c + w_c \times (|s_c - I^c(y)|) \tag{2}$$

In the formula, C represents the R, G, B three channels. s_c and w_c are determined by a_c. The a_c represents the proportional relation of C(R, G, or B channel) with the change of depth of field. Under different water quality conditions, the proportional relationship between light intensity and depth of field is different. A gradient map is first computed as the rough depth image on R, G, and B channel by using Sobel operators. The proportional relation can be fitted by the linear least square method under depth image. The a_c is the slope of the linear regression equation. If a_c is greater than 0, it means that the object is further away from the lens, the intensity of the corresponding channel is higher, and if a_c is less than 0, it means that the object is closer to the lens, and the intensity of the corresponding channel is higher. The relationship between a_c and s_c and w_c is shown as follows:

$$s_c = \begin{cases} 1 & if a_c > 0 \\ 0 & if a_c < 0 \end{cases} \tag{3}$$

$$w_c = tanh(k|a_c|) \tag{4}$$

The w_c is the weight factor of each channel, and k is constant, which is generally taken as 5. The largest top 0.1% pixel values from the depth map are averaged as P. Then the Equation (5) can be used to estimate the media transmission map:

$$T(x) = \max_{c,y \in \Theta(x)} \left(\frac{|P - I^c(x)|}{max(P, \quad 1 - P)} \right) \tag{5}$$

In the neural network structure, our loss function is divided into two parts: the content loss function and the ray attenuation loss function. The content loss function controls that the content of the generated image and the input I are the same object. The image attenuation of UIEB's degraded image I is controlled by the ray attenuation loss function. The ray attenuation is carried out according to the degraded degree of the underwater image II in a certain sea area. The content loss function is shown in Equation (6):

$$\mathcal{L}_{content} = \frac{\sum_{i=1}^{n} (Enc(x1)_i - Enc(x2)_i)^2}{n} \tag{6}$$

Where $x1$ is the output image, $x2$ is the input image I, Enc is the first 31 layers of VGG model (VGG31), n is the pixel number of one image. The ray attenuation loss function uses the L1 loss function,

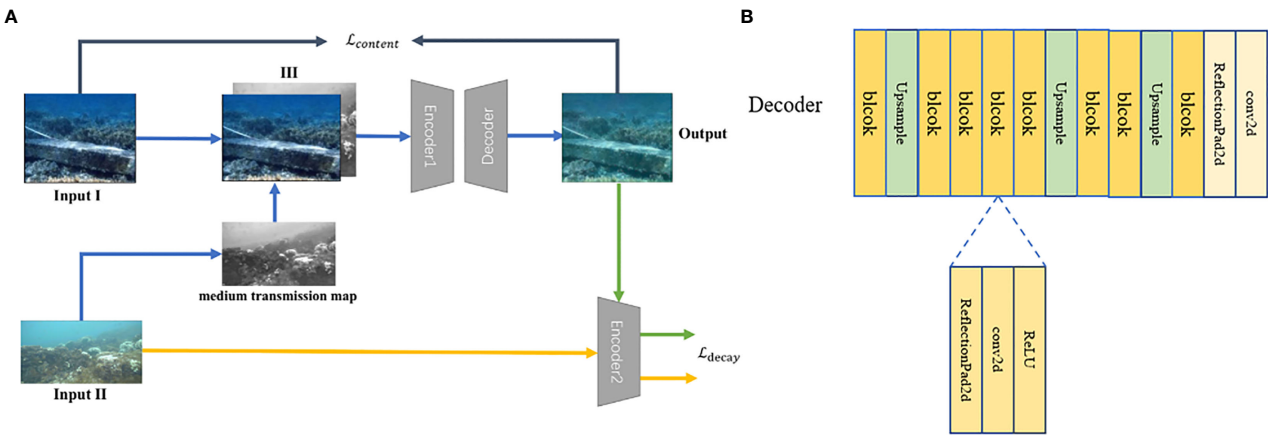


FIGURE 2
The framework of the UDFM. **(A)** the overall framework the underwater datasets fitting newtork; **(B)** the structure of the decoder.

and the specific equation is shown in Equation (7):

$$\mathcal{L}_{decay} = \frac{\sum_i^n |Mean(Relu_i(\mathrm{Enc}(x1))) - Mean(Relu_i(Enc(x1)))|^2}{n} \quad (7)$$

Where x1 is the output image, x2 is the input image II, and $Relu_i$ refers to the RELU layer network of VGG31. In specific, it refers to the last four RELU layer networks. Then the output difference of each layer is averaged and calculated.

3.2 Underwater image enhancement model

In the underwater image enhancement task, more feature information about each pixel needs to be used than in other tasks, such as object detection and object classification. For example, each pixel uses global information to estimate the relative depth of the current data, to correct the color deviation. In addition, for every single pixel, the feature information of its nearby pixels is also needed to restore its texture information. To retain a large amount of existing information, we use the fifteen residual dense block (RDB) modules as the backbone network to further improve the enhanced image quality. The RDB module is composed of four dense blocks connected by residual structure, as shown in Figure 3. The bypass keeps the input information retained in the dense block (Huang et al., 2018), and the network layer learns and generates new information. Through the above design, our model is suitable for the image enhancement task. Moreover, we add residual structure between dense blocks to make training and convergence easier.

In the UIEM, our loss function is divided into the content loss function and the pix-to-pix loss function. The content loss function assures that the content of the generated image and the input are the same object. The U-net network (Ronneberger et al., 2015) implements the content loss function. The content loss function is shown in Equation (8):

$$\mathcal{L}_{content} = \frac{\sum_{i=1}^n (Unet(output)_i - Unet(gt)_i)^2}{n} \quad (8)$$

Where output is the output image, gt is the clear image in UIEB, Unet is the U-net model, n is the pixel number of an image. The pix-to-pix loss function uses the L1 loss function, and the specific equation is shown in Equation (9):

$$\mathcal{L}_{pixtopix} = \frac{\sum_{i=1}^n (output_{i,j} - gt_{i,j})^2}{n} \quad (9)$$

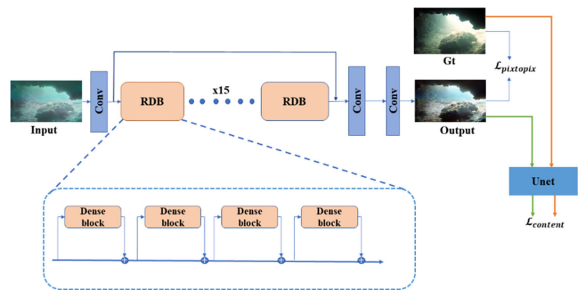


FIGURE 3
The network architecture of the underwater image enhancement model (UIEM).

4 Experiments and results

The underwater image enhancement has gradually formed its indicators, divided into two parts. One part is the comparison of result images, which we have put in Figure 4. The other part is the comparison of quantitative indicators. PSNR(peak signal-to-noise ratio) and SSIM(structural similarity) are most commonly used. However, it is found that the above indicators are not well consistent with human perception. And then, some scholars proposed an improved indicator named NIQE(no-reference image evaluation index). Therefore, we adopt NIQE in this paper. The results are compared in Tables 1, Table 2.

4.1 Underwater dataset fitting model experiments

Implementation details. The training and reference of the algorithm were carried out on a server configured with Intel(R) Xeon(R) Gold 6342 CPU ×2and NVIDIA A100 40 GB × 4 in this experiment. The PyTorch was used to build the backbone network and run on the Ubuntu 18.04 LTS operating system. We used Adam to optimize the overall objective. The learning rate was initialized as 1×10^{-4} and then decayed to 1×10^{-6} in the last 20 epochs. We trained with a batch size of 16 (on 4 GPUS) for 300K epochs.

Firstly, the performance of the data fitting model was verified. The loss function curve is shown in Figure 5 in the training stage, and it can be seen that the neural network began to converge at the 300K epoch, so we tested the model trained tillthe300K epoch.

In addition, the proportions of the content and ray loss functions were tested. The $\mathcal{L}_{content}/\mathcal{L}_{decay}$ was tested at 1:1, 5:1, 10:1 and 20:1, respectively. The convergence function curves are shown in Figure 6.

It has been proved by a lot of experiments that, as the data index of the experiment, higher peak signal-to-noise ratio (PSNR) or structural similarity (SSIM) does not always represent a better reconstruction effect, and the reconstructed texture often fails to meet the expectation of human eyes. Thus, the no-reference image evaluation index (NIQE) is used in this study. NIQE (Mittal et al., 2013) is based on 'quality perception' features. Usually, the features of simple and highly regular natural landscapes are extracted and used to fit a multivariate Gaussian model. The index quantifies the difference in the multivariate distribution of the image. In this study, the NIQE is used to quantify the difference between a set of generated images and an input set of images. The NIQE average is calculated in Equation (10).I are the UIEB degraded datasets, II are the other nature degraded datasets. III are the output images of the underwater dataset fitting model.

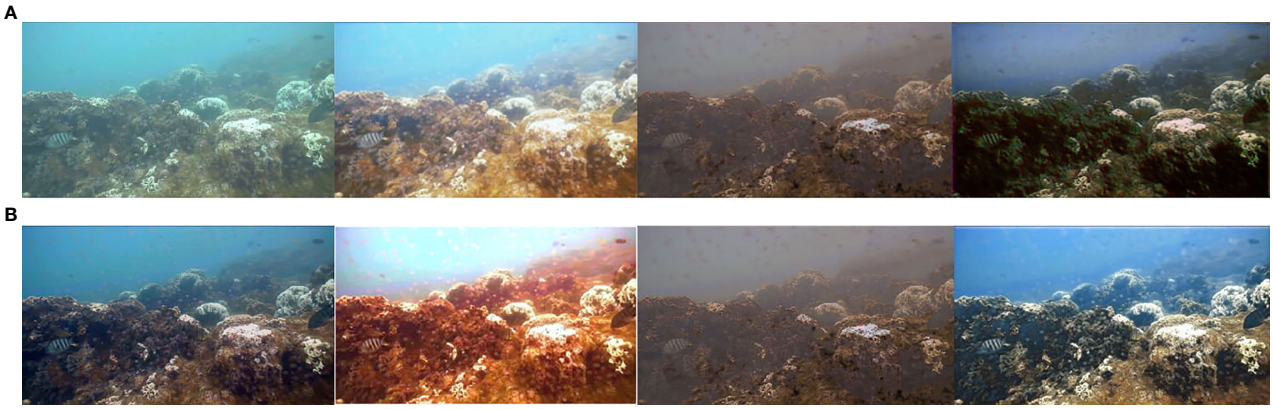


FIGURE 4
The enhancement result with different methods. **(A)** Origin UIEB (Li et al., 2020b) UWCNN-typeI (Li et al., 2020a) UWGAN (Wang et al., 2021), **(B)** Water-net (Li et al., 2017) GDCP (Peng et al., 2018) UWCNN-typeIII (Li et al., 2020a) Ours.

TABLE 1 Comparison of results before and after using the underwater dataset fitting model.

Input I	Input II	NIQE average for the original dataset	NIQE average of the dataset after fitting
UIEB degraded datasets	North Sea dataset	3.88	0.73
UIEB degraded datasets	East Pole Island dataset	5.32	0.92

$$NIQE_{avg} = \frac{\sum_{i=1}^{n}(NIQE(I,\ II) - NIQE(III,\ II))}{n} \quad (10)$$

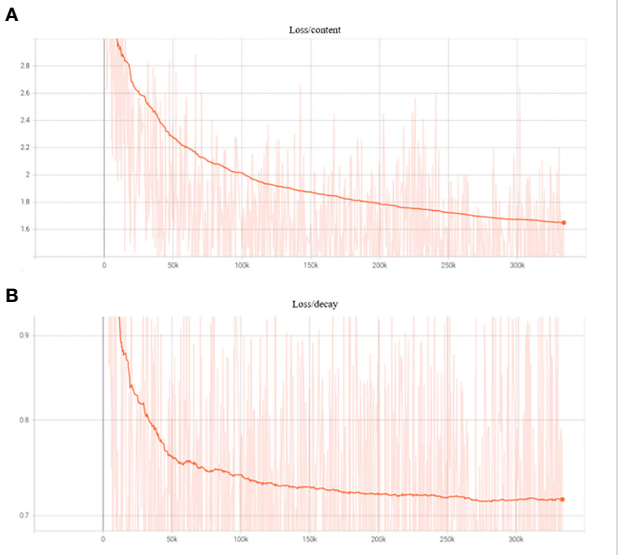


FIGURE 5
Convergence diagrams of the content loss function. **(A)** is the content loss function. **(B)** is the ray attenuation loss function.

TABLE 2 The Comparison of results from different encoders.

Model	$\mathcal{L}_{content}/\mathcal{L}_{decay}$	NIQE average
VGG31	1.61/0.73	0.73
DenseNet121	1.9/1.6	0.95
ResNet50	1.21/0.32	0.70
ResNet101	0.96/0.28	0.65

We tested two nature datasets, and the results are shown in Table 1. The *NIQE* average has all dropped below 1.

Besides that, the output results using different encoders are also given in Table 2. The neural network can encode more apparent information with the complex network structure, thus improving performance. As can be seen from the table, Resnet101 has the best effect as an encoder. The NIQE average is 0.65.

From the visual perspective, we demonstrate some images in Figure 7. Class (a) is the images of the publicly available dataset, class (b) is the underwater images of a sea area, and class (c) is the generated images.

4.2 Underwater image enhancement experiments

To evaluate the image restoration performance in a real underwater dataset, we compared UIEM with six deep learning methods, including UWCNN, UWGAN, Water-net, GDCP (Chavez, 1988), and UIEB. The network's input is the same underwater image without additional input, and the reasoning parameters of each network are the default parameters provided in the open-source network. It can be seen that when the input image quality is poor, the enhancement effects of UWCNN and UWGAN are not obvious and even appear more blurred. Water-net, GDCP, and UIEB can enhance the image to different degrees. However, Water-Net loses the picture's original color and enhances the contrast, making the whole image dark. UIEB also appears to strengthen excessively brightness,

and the details and texture of the picture after GDCP enhancement are unclear. The UIEM enhancement is the best. All the above image enhancement models use the generated dataset as the training set. The closer part of the results obtained by the proposed method is more precise than the UWCNN-typeI and UWCNN-typeIII, but the farther part is blurrier. In this study, we need to calculate the underwater medium transmission map, which represents the attenuation coefficient of underwater images. The distance of the contents in the image limits its accuracy. Its accuracy for the results of the near part is much greater than that of the far part, so it can be seen that the recovery of the near part is better than that of the far part. However, UWCNN-typeI and UWCNN-typeIII perform better in the restoration of the farther part of the image because the UWCNN method designs a variety of water types and then selects the algorithm model according to the water type. Therefore, The algorithm tends to restore the far water rather than the near target.

5 Discussion and conclusion

In this paper, we proposed a dataset fitting model UDFM to solve a challenging problem in underwater image enhancement. There are few high-quality images, not to mention corresponding clear-enough reference images for a specific sea area. Although the UIEB dataset has contributed clear paired datasets of underwater images for training, the results are unsatisfactory to test in a specific sea area. Our model has good generalization. The features of underwater images in specific sea areas are extracted and given to real images in UIEB to obtain synthetic images. Specific sea areas characterize these synthetic images. We use NIQE of the reference images (Mittal et al., 2013) to ensure that the quality of the generated images is close to our sea areas. Finally, image-to-image enhancement training can be performed using synthetic images and UIEB reference images. In the testing phase, we adopted the weights obtained after the training to the real blurred pictures. Experiments verify that our dataset augmentation method can be applied to underwater images under different sea area conditions. Just put the image of any sea area into the encoder to simulate the underwater environment. Experiments also verify that these image-to-image underwater image enhancement models achieve better results on our dataset than on the UIEB dataset. Our method can provide a baseline for the dataset synthesis method so that the training of image-to-image underwater image enhancement algorithms is no longer limited to the dataset. Our method provides new ideas for the study of domain adaptation.

In addition, we propose a deep learning-based image-to-image underwater image enhancement model UIEM. Our algorithm is based on dense blocks that can obtain more structural texture information, and we add residual structure. It is beneficial to the image restoration stage. This method can eliminate the effects of degradation and scatter on underwater images. Finally, we apply the MSE loss and introduce the Content loss (Wang et al., 2021) to train the network for further control over the content information. In addition, the results of our model are compared with existing algorithms. The effects of other algorithms are reddish, dark, and blurred. The UIEM has a leading position in underwater datasets.

The shortcoming of this model is that the model needs to be retrained when the data sets of other sea areas are collected, and the fitting ability of the model can only become more robust with the expansion of the data sets. Secondly, for some pictures, there will be color bands. As shown in Figure 8, blue color bands exist in the upper part of the processed picture, which is caused by the fact that the data fitting model is not optimal. Our future research direction is to combine these two models and solve the phenomenon of color bands in the model.

In future work, we will try to collect data from different sea areas to improve the generalization ability of our underwater image enhancement model. And we will further optimize our data fitting model to solve the phenomenon of color bands. We will also try to combine our method with algorithms such as object detection for better progress in more advanced vision tasks.

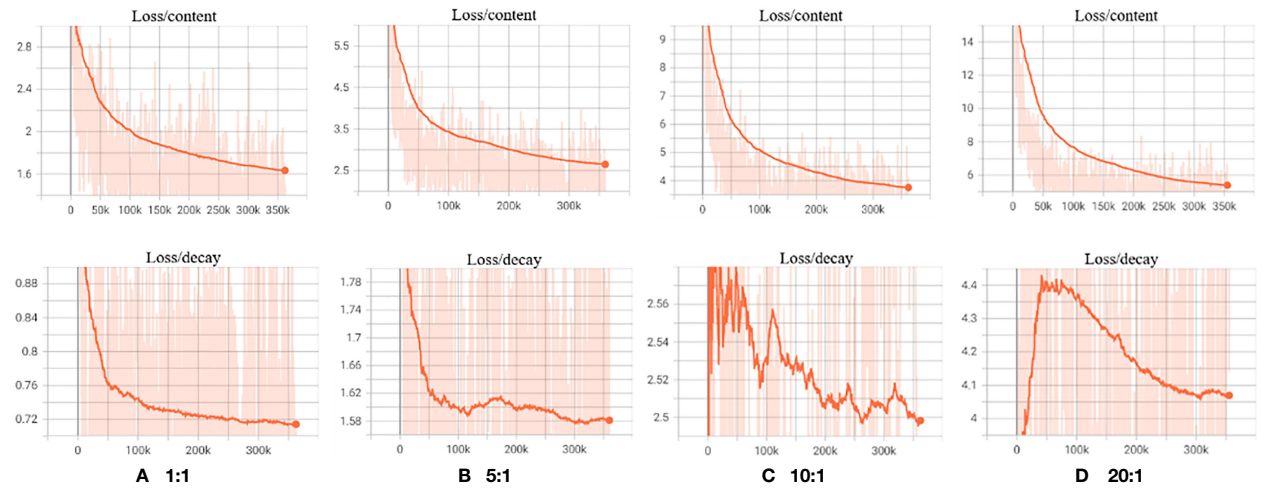


FIGURE 6
The loss function traincurve. **(A)** $\mathcal{L}_{content}$: $\mathcal{L}_{decay}$ 1:1, **(B)** $\mathcal{L}_{content}$: $\mathcal{L}_{decay}$ is 5:1, **(C)** $\mathcal{L}_{content}$: $\mathcal{L}_{decay}$ is 10:1, **(D)** $\mathcal{L}_{content}$: $\mathcal{L}_{decay}$ is 20:1.

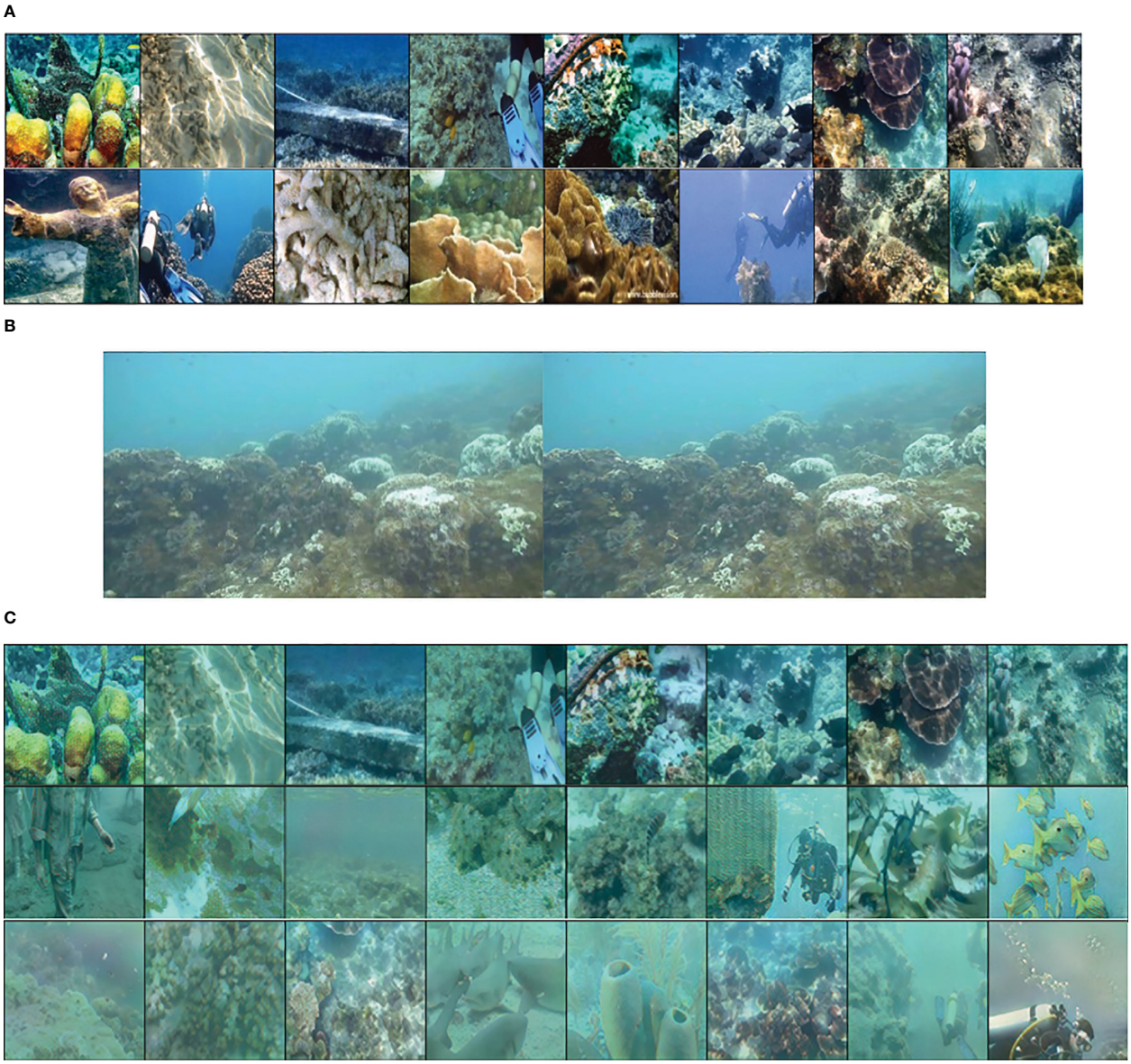


FIGURE 7
(A) The images of the public data set, **(B)** the underwater images of a sea area, and **(C)** the generated images.

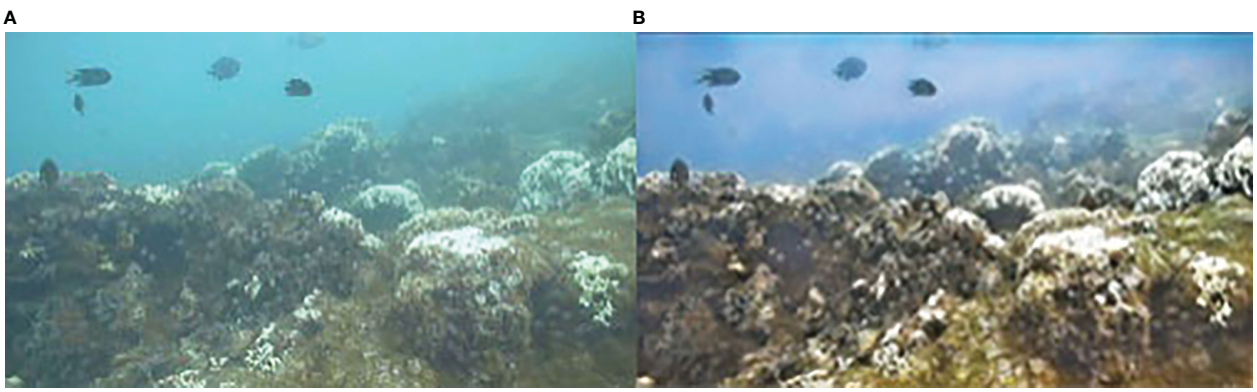


FIGURE 8
The enhancement result from the phenomenon of color bands. **(A)** The origin picture. **(B)** The enhancement picture with color bands.

Data availability statement

The data is provided by Hainan Observation and Research Station of Ecological Environment and Fishery Resource in Yazhou Bay. The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Author contributions

The algorithm studied by XD and TL, and the experimental results organized by SH and XX. Paper by PL and YG. All authors contributed to the article and approved the submitted version.

Funding

The paper was sponsored by [Hainan Provincial Joint Project ofSanya Yazhou Bay Science and Technology City] with grant number [No. 120LH001], [Hainan Provincial Joint Project of Sanya Yazhou Bay Science and Technology City] with grant number [No. 2021CXLH0020], [National Key R&D Program of China] with grant number [No. 2022YFC3103402].

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Akkaynak, D., and Treibitz, T. (2019). "Sea-Thru: A method for removing water from underwater images," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA. 1682–1691 (IEEE). doi: 10.1109/CVPR.2019.00178

Ancuti, C. O., Ancuti, C., De Vleeschouwer, C., and Bekaert, P. (2018). Color balance and fusion for underwater image enhancement. *IEEE Trans. Image Process.* 27, 379–393. doi: 10.1109/TIP.2017.2759252

Berman, D., Levy, D., Avidan, S., and Treibitz, T. (2021). Underwater single image color restoration using haze-lines and a new quantitative dataset. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 2822–2837. doi: 10.1109/TPAMI.2020.2977624

Bollepalli, B., Juvela, L., and Alku, P. (2017). "Generative adversarial network-based glottal waveform model for statistical parametric speech synthesis," in *Interspeech 2017*. 3394–3398. doi: 10.21437/Interspeech.2017-1288

Boom, B. J., Huang, P. X., He, J., and Fisher, R. B. (2012). "Supporting ground-truth annotation of image datasets using clustering," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. 1542–1545.

Buchsbaum, G. (1980). A spatial processor model for object colour perception. *J. Franklin Institute* 310, 1–26. doi: 10.1016/0016-0032(80)90058-7

Cao, K., Peng, Y.-T., and Cosman, P. C. (2018). "Underwater image restoration using deep networks to estimate background light and scene depth," in *2018 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*, Las Vegas, NV. 1–4 (IEEE). doi: 10.1109/SSIAI.2018.8470347

Chavez, P. S. (1988). An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sens. Environ.* 24, 459–479. doi: 10.1016/0034-4257(88)90019-3

Chen, T., Wang, N., Wang, R., Zhao, H., and Zhang, G. (2021). One-stage CNN detector-based benthonic organisms detection with limited training dataset. *Neural Networks* 144, 247–259. doi: 10.1016/j.neunet.2021.08.014

Cheng, S., Zhao, K., and Zhang, D. (2019). Abnormal water quality monitoring based on visual sensing of three-dimensional motion behavior of fish. *Symmetry* 11, 1179. doi: 10.3390/sym11091179

Chiang, J. Y., and Chen, Y.-C. (2012). Underwater image enhancement by wavelength compensation and dehazing. *IEEE Trans. Image Process.* 21, 1756–1769. doi: 10.1109/TIP.2011.2179666

Devlin, M. J., Barry, J., Mills, D. K., Gowen, R. J., Foden, J., Sivyer, D., et al. (2009). Estimating the diffuse attenuation coefficient from optically active constituents in UK marine waters. *Estuarine Coast. Shelf Sci.* 82, 73–83. doi: 10.1016/j.ecss.2008.12.015

dos Santos, C. N., and Gatti, M. (2014). Deep convolutional neural networks for sentiment analysis. *COLING* 10, 69–78. Available at: https://aclanthology.org/C14-1008

Drews, P. L. J., Nascimento, E. R., Botelho, S. S. C., and Montenegro Campos, M. F. (2016). Underwater depth estimation and image restoration based on single images. *IEEE Comput. Grap. Appl.* 36, 24–35. doi: 10.1109/MCG.2016.26

Foster, D. H. (2011). Color constancy. *Vision Res.* 51, 674–700. doi: 10.1016/j.visres.2010.09.006

Fu, Z., Lin, X., Wang, W., Huang, Y., and Ding, X. (2022)Underwater image enhancement via learning water type desensitized representations. In: *arXiv:2102.00676*. Available at: http://arxiv.org/abs/2102.00676 (Accessed April 8, 2022).

Galdran, A., Pardo, D., Picón, A., and Alvarez-Gila, A. (2015). Automatic red-channel underwater image restoration. *J. Visual Communication Image Representation* 26, 132–145. doi: 10.1016/j.jvcir.2014.11.006

Gao, S.-B., Zhang, M., Zhao, Q., Zhang, X.-S., and Li, Y.-J. (2019). Underwater image enhancement using adaptive retinal mechanisms. *IEEE Trans. Image Process.* 28, 5580–5595. doi: 10.1109/TIP.2019.2919947

He, K., Sun, J., and Tang, X. (2009). "Single image haze removal using dark channel prior," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL. 1956–1963 (IEEE). doi: 10.1109/CVPR.2009.5206515

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA. 770–778 (IEEE). doi: 10.1109/CVPR.2016.90

Hou, M., Liu, R., Fan, X., and Luo, Z. (2018). "Joint residual learning for underwater image enhancement," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, Athens. 4043–4047 (IEEE). doi: 10.1109/ICIP.2018.8451209

Huang, X., and Belongie, S. (2017)Arbitrary style transfer in real-time with adaptive instance normalization. In: *arXiv:1703.06868*. Available at: http://arxiv.org/abs/1703.06868 (Accessed April 8, 2022).

Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q. (2018) *Densely connected convolutional networks*. Available at: http://arxiv.org/abs/1608.06993 (Accessed September 6, 2022).

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Commun. of the ACM* 60, 84-90. doi: 10.1145/3065386.

Lan, Z., Zhou, B., Zhao, W., and Wang, S. (2023). An optimized GAN method based on the que-attn and contrastive learning for underwater image enhancement. *PloS One* 18, e0279945. doi: 10.1371/journal.pone.0279945

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Li, C., Anwar, S., and Porikli, F. (2020a). Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognition* 98, 107038. doi: 10.1016/j.patcog.2019.107038

Li, C., Guo, J., and Guo, C. (2018). Emerging from water: Underwater image color correction based on weakly supervised color transfer. *IEEE Signal Process. Lett.* 25, 323–327. doi: 10.1109/LSP.2018.2792050

Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., et al. (2020b). An underwater image enhancement benchmark dataset and beyond. *IEEE Trans. Image Process.* 29, 4376–4389. doi: 10.1109/TIP.2019.2955241

Li, J., Skinner, K. A., Eustice, R. M., and Johnson-Roberson, M. (2017). WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images. *IEEE Robot. Autom. Lett.*, 1. doi: 10.1109/LRA.2017.2730363

Liu, R., Fan, X., Zhu, M., Hou, M., and Luo, Z. (2020). Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. *IEEE Trans. Circuits Syst. Video Technol.* 30, 4861–4875. doi: 10.1109/TCSVT.2019.2963772

Mariani, P., Quincoces, I., Haugholt, K., Chardard, Y., Visser, A., Yates, C., et al. (2018). Range-gated imaging system for underwater monitoring in ocean environment. *Sustainability* 11, 162. doi: 10.3390/su11010162

Mittal, A., Soundararajan, R., and Bovik, A. C. (2013). Making a "Completely blind" image quality analyzer. *IEEE Signal Process. Lett.* 20, 209–212. doi: 10.1109/LSP.2012.2227726

Naeem, A., Anees, T., Ahmed, K. T., Naqvi, R. A., Ahmad, S., and Whangbo, T. (2022). Deep learned vectors' formation using auto-correlation, scaling, and derivations with CNN for complex and huge image retrieval. *Complex Intell. Syst.* doi: 10.1007/s40747-022-00866-8

Peng, Y.-T., Cao, K., and Cosman, P. C. (2018). Generalization of the dark channel prior for single image restoration. *IEEE Trans. Image Process.* 27, 2856–2868. doi: 10.1109/TIP.2018.2813092

Peterson, K. T., Sagan, V., and Sloan, J. J. (2020). Deep learning-based water quality estimation and anomaly detection using landsat-8/Sentinel-2 virtual constellation and cloud computing. *GIScience Remote Sens.* 57, 510–525. doi: 10.1080/15481603.2020.1738061

Prabhakar, C. J., and Praveen Kumar, P. U. (2011). An image based technique for enhancement of underwater images. *Int. J. Mach. Intell.* 3, 217–224. doi: 10.48550/arXiv.1212.0291

Rizzi, A., Gatta, C., and Marini, D. (2002). Color correction between gray world and white patch. *Human Vision and Electronic Imaging VII*. 4662, 367–375. doi: 10.1117/12.469534

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *Medical image computing and computer-assisted intervention, PT III*. 9351, 234-241. doi: 10.48550/arXiv.1505.04597

Salman, A., Jalal, A., Shafait, F., Mian, A., Shortis, M., Seager, J., et al. (2016). Fish species classification in unconstrained underwater environments based on deep learning: Fish classification based on deep learning. *Limnol. Oceanogr. Methods* 14, 570–585. doi: 10.1002/lom3.10113

Singh, G., Jaggi, N., Vasamsetti, S., Sardana, H. K., Kumar, S., and Mittal, N. (2015). "Underwater image/video enhancement using wavelet based color correction (WBCC) method," in *2015 IEEE Underwater Technology (UT)*, Chennai, India. 1–5 (IEEE). doi: 10.1109/UT.2015.7108303

Singh, K., and Kapoor, R. (2014). Image enhancement using exposure based Sub image histogram equalization. *Pattern Recognition Lett.* 36, 10–14. doi: 10.1016/j.patrec.2013.08.024

van de Weijer, J., Gevers, T., and Gijsenij, A. (2007). Edge-based color constancy. *IEEE Trans. Image Process.* 16, 2207–2214. doi: 10.1109/TIP.2007.901808

Wang, W., Jiao, P., Liu, H., Ma, X., and Shang, Z. (2022b). Two-stage content based image retrieval using sparse representation and feature fusion. *Multimed Tools Appl.* 81, 16621–16644. doi: 10.1007/s11042-022-12348-7

Wang, Z., Shen, L., Xu, M., Yu, M., Wang, K., and Lin, Y. (2023). Domain adaptation for underwater image enhancement. *IEEE Trans. Image Process.* 32, 1442–1457. doi: 10.1109/TIP.2023.3244647

Wang, N., Wang, Y., and Er, M. J. (2022a). Review on deep learning techniques for marine object recognition: Architectures and algorithms. *Control Eng. Pract.* 118, 104458. doi: 10.1016/j.conengprac.2020.104458

Wang, N., Zhou, Y., Han, F., Zhu, H., and Yao, J. (2021). UWGAN: Underwater GAN for real-world underwater color restoration and dehazing. *CoRR* doi: 10.48550/arXiv.1912.10269

Xie, Y., Yu, Z., Yu, X., and Zheng, B. (2022). Lighting the darkness in the sea: A deep learning model for underwater image enhancement. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.921492

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 15 June 2023
DOI 10.3389/fmars.2023.1167191

Check for updates

OPEN ACCESS

EDITED BY
Takafumi Hirata,
Hokkaido University, Japan

REVIEWED BY
Ronald Petie,
Wageningen University and Research, Netherlands
Liwei Wang,
Guangxi University, China

*CORRESPONDENCE
Tao Liu
liutao08@zju.edu.cn
Peiliang Li
lipeiliang@zju.edu.cn

RECEIVED 16 February 2023
ACCEPTED 31 May 2023
PUBLISHED 15 June 2023

CITATION
Li L, Liu T, Huang H, Song H, He S, Li P, Gu Y and Chen J (2023) An early warning model for starfish disaster based on multi-sensor fusion.
Front. Mar. Sci. 10:1167191.
doi: 10.3389/fmars.2023.1167191

COPYRIGHT
© 2023 Li, Liu, Huang, Song, He, Li, Gu and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

An early warning model for starfish disaster based on multi-sensor fusion

Longyu Li[1,2], Tao Liu[1]*, Hui Huang[1,2], Hong Song[1], Shuangyan He[1,2], Peiliang Li[1,2]*, Yanzhen Gu[1,2] and Jiawang Chen[1,2]

[1]Ocean College, Zhejiang University, Zhoushan, China, [2]Hainan Institute, Zhejiang University, Sanya, China

Starfish have a wide range of feeding habits, including starfish, sea urchins, sea cucumbers, corals, abalones, scallops, and many other marine organisms with economic or ecological value. The starfish outbreak in coastal areas will lead to severe economic losses in aquaculture and damage the ecological environment. However, the current monitoring methods are still artificial, time-consuming, and laborious. This study used an underwater observation platform with multiple sensors to observe the starfish outbreak in Weihai, Shandong Province. The platform could collect the temperature, salinity, depth, dissolved oxygen, conductivity, other water quality data, and underwater video data. Based on these data, the paper proposed an early warning model for starfish prevalence (EWSP) based on multi-sensor fusion. A deep learning-based object detection method extracts time-series information on the number of starfish from underwater video data. For the extracted starfish quantity information, the model uses the k-means clustering algorithm to divide the starfish prevalence level into four levels: no prevalence, mild prevalence, medium prevalence, and high prevalence. Correlation analysis concluded that the water quality factors most closely related to the starfish prevalence level are temperature and salinity. Therefore, the selected water quality factor and the number of historical starfish are inputted. The future starfish prevalence level of the starfish outbreak is used as an output to train the BP (back propagation) neural network to build EWSP based on multi-sensor fusion. Experiments show that the accuracy rate of this model is 97.26%, whose precision meets the needs of early warning for starfish outbreaks and has specific application feasibility.

KEYWORDS

starfish disaster, multi-sensor fusion, early-warning model, self-supervised model, feature selection

1 Introduction

Starfish are one of the most common carnivorous echinoderms in coastal areas. It has robust reproductive and regeneration capabilities and uses shellfish and corals as food sources. The characteristics and feeding habits of starfish enable them to reproduce rapidly in a suitable environment, forming a large-scale starfish outbreak disaster and causing significant damage to shellfish farming and coral ecosystems (Babcock et al., 2020; Wang et al., 2021). In coral ecosystems on the Great Barrier Reef in Australia, *Acanthaster planci* (Crown-of-Thorn Starfish, COTS) outbreaks are one of the most prominent factors in coral degradation (Pratchett et al., 2019; Westcott et al., 2020). In marine ranching, predatory starfishes such as Japanese common starfish and *Asterias rubens* (sugar starfish) are the main predators of farmed shellfish (Miyoshi et al., 2019; Agüera et al., 2021). The main food of adult *Asterias amurensis* and *Patiria pectinifera* are bivalve mollusks. A large-scale outbreak of the population will have a huge impact on the population of shellfish, especially the *Ruditapes philippinarum*. It is of great significance to protect coral ecosystems and marine pastures by monitoring the scale of starfish aggregation and assessing the possibility of starfish prevalence to take management measures to reduce the losses caused by starfish outbreaks. In 2012, Kayal et al. developed the SCUBA-tow technology based on Manta-tow and SCUBA to track the distribution and density of starfish aggregations (Kayal et al., 2012). Suzuki et al. used a DNA barcoding approach to describe the distribution of COTS larvae in Sekisei Lagoon, Ryukyu Archipelago, Japan (Suzuki et al., 2016). Saponari et al. used snorkeling at shallow depths (0–5 m) and scuba diving at greater depths (5–30 m) to monitor and assess starfish outbreaks in Ari Atoll, Republic of Maldives, over 2 years (Saponari et al., 2018). Dumas et al. adopted the method of citizen science to monitor starfish outbreaks, effectively increasing the observation range and detection quantity (Dumas et al., 2020). Uthicke et al. developed an eDNA method to detect the density of crown-of-thorns starfish early in the outbreak (Uthicke et al., 2022). Rogers et al. used an empirically tuned, individual-based simulation model to investigate how density and aggregation influence COTS reproductive success, and quantified a threshold level of density and aggregation above which reproductive success will increase dramatically (Rogers et al., 2017). Ecological analysis methods also informed management targets for the culling of starfish to prevent coral decline (Plagányi et al., 2020). Besides ecology-based methods, image-processing techniques are also used for starfish detection. Gesú et al. proposed three feature indicators for identifying starfish from starfish video sequences based on the unique shape of starfish (Gesú et al., 2003). Clement et al. developed a local binarization method based on texture features to automatically detect starfish images (Clement et al., 2005). Smith and Dunbabin developed a novel shape recognition algorithm to classify the Northern Pacific Sea Star autonomously (Smith & Dunbabin, 2007). Gobi demonstrated using local invariant features, specifically SIFT features, in the visual recognition and identification of starfish (Gobi, 2010). Dayoub et al. proposed a method based on a random forest classifier (RFC) to train images to assist underwater robotic systems in the detection and tracking of crown-of-thorns starfish (COTS) (Dayoub et al., 2015).

However, the methods, including SCUBA-tow technology and DNA barcoding technology, are all implemented in the form of on-the-spot surveys, which have problems of poor timeliness and high cost and have high requirements for human resources and material resources. Methods based on traditional image processing technology have the problem of poor generalization, and it is easy to reduce processing precision due to scene transformation. To overcome the above difficulties, image-processing techniques based on deep learning have been gradually applied to starfish detection in recent years. Liu et al. released a large-scale, annotated underwater image dataset, called the CSIRO dataset, from a COTS outbreak area on the GBR, to encourage research on Machine Learning and AI-driven technologies to improve the detection, monitoring, and management of COTS populations at reef scale (Liu et al., 2021). Based on the CSIRO dataset, Truong proposed a deep learning model based YOLOv5 (YOU ONLY LOOK ONCE version 5) algorithm to automatically detect the COTS to prevent the outbreak and minimize coral mortality in the Reef (Truong, 2022). Nguyen applied advanced data augmentation methods for enhancing the quality and quantity of the CSIRO dataset to train and evaluate the COTS detection model with the YOLOv5 algorithm, which is used for embedded systems and mobile devices (Nguyen, 2022). Heenaye-Mamode Khan et al. proposed a novel approach for the automatic detection of COTS-based Convolutional Neural Network (CNN) with an enhanced attention module (Heenaye-Mamode Khan et al., 2023). Sheth and Prajapati detected COTS using various deep learning models and compared their accuracies to find the best model based on its performance (Sheth & Prajapati, 2022).

It is plausible that environmental conditions affecting starfish larvae's survival and development act similarly on each species of the same type (Yamaguchi, 1973). Temperature and salinity are very important for the survival of COTS larvae. Larvae die quickly at temperatures above 32°C and slowly at lower temperatures. Therefore, COTS larvae are widely distributed in the waters of the Great Barrier Reef, where the maximum water temperature is 27–28°C. The study also showed that larvae developed faster in low-salinity conditions (Lucas, 1973). Kashenko researched the lower limit of seawater desalination tolerance range of *A. amurensis* and *P. pectinifera* (blue bat star). The results showed that the *P. pectinifera* has a higher tolerance to salinity, under 18‰ still alive, while the *A. amurensis* is 22‰ (Kashenko, 2003). Caballes et al. studied the environmental critical points of each link in the development of the COTS. The results showed that gametes, fertilization, and embryos have a wide range of adaptability to temperature, salinity, and pH. Then, they proposed that compared with the early developmental stages, the scope of environmental adaptation may be narrowed during ontogeny in the later stage (Caballes et al., 2017). Novia Arinda Pradisty et al. evaluated more than 10 environmental parameters and determined the environmental parameters with a more significant correlation with starfish survival through redundancy analysis (RDA), including salinity, turbidity, and dissolved inorganic nitrogen (Pradisty et al., 2020). Chen et al. found that environmental (nutrient concentration) and oceanic

factors (ocean currents) played an essential role in the genetic feature and larval dispersal of starfish populations (Chen et al., 2021b). In recent years, multi-sensor fusion methods that integrate multiple data sources have been widely used. Manzione and Castrignanò have used a multi-source data fusion method based on multivariate geostatistics to predict the depth of groundwater levels (Manzione & Castrignanò, 2019). Nti et al. fused data from different sensors to predict the stock market (Nti et al., 2021). Hou et al. proposed a multi-source spatiotemporal data fusion model for sea surface temperature prediction (Hou et al., 2022). Jiang et al. proposed a deep learning method based on multi-sensor fusion to predict the water quality of urban sewage pipe networks (Jiang et al., 2021). To promote the monitoring and prediction of harmful algal blooms, a multi-element fusion prediction (MEFP) method for cyanobacteria bloom was proposed (X. Chen et al., 2022). Tang et al. created a multisource hybrid dataset for deep learning model training to predict harmful algal bloom events in Lake Okeechobee (Tang et al., 2022).

In the past, the impact of environmental factors on starfish was mainly the result of laboratory experiments or long-term scattered collection results. However, no intensive collection of environmental factors was taken. The data source of environmental factors used in this paper comes from the *in situ* intensive collection method of the multi-parameter water quality meter. The amount of data is larger, which is significant for accurately identifying environmental factors related to starfish outbreaks and early warning research. The data used in this paper come from the underwater observation platform arranged in the marine ranch in Weihai, Shandong Province, which mainly breeds sea cucumbers. Limited by the location of the platform, the research object of this paper is *P. pectinifera*. Although it poses no threat to coral or shellfish farming, it is an important predator of sea cucumbers (Lambert, 1997; Popov et al., 2014); *P. pectinifera* has been observed by the platform. Based on the platform and machine learning algorithm, the starfish disaster situation in the target sea area was subdivided. A starfish disaster warning was established—EWSP. The following describes the overall framework of the model first and then describes the implementation steps of each sub-module. Finally, the performance of the model is evaluated.

2 Materials and methods

The data used to study the early warning model for starfish outbreak disasters (EWSP) include starfish number, temperature, salinity, depth, and dissolved oxygen. Among them, the number of starfish was calculated according to the maximum value per hour, and the temperature, salinity, depth, and dissolved oxygen were calculated according to the average value per hour. Figure 1 shows the entire process from raw data acquisition to the final EWSP. The process was mainly divided into four parts.

2.1 The data preprocessing section

This section described the method of processing video data into starfish numbers and the preprocessing operation of water quality parameters.

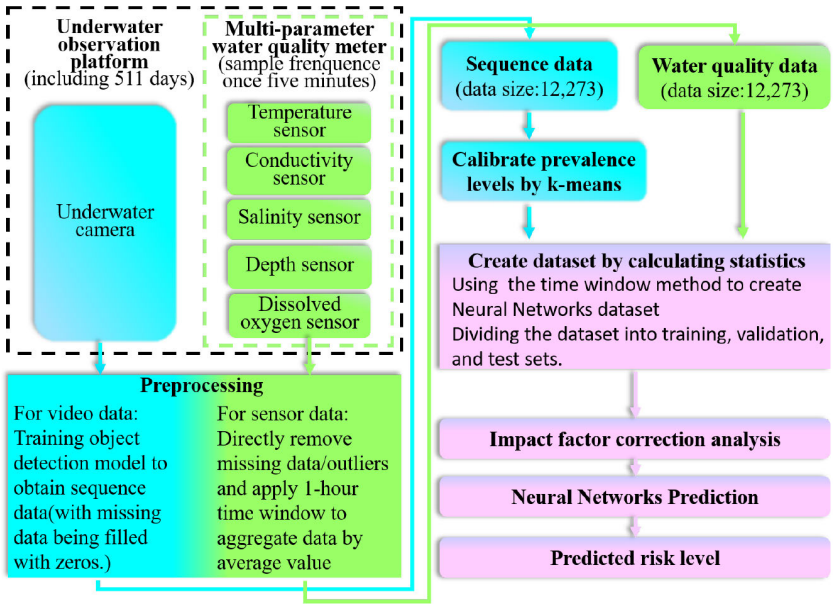


FIGURE 1
Schematic diagram of the EWSP.

2.2 The calibration section of the starfish prevalence level

This section used a clustering algorithm to calibrate the starfish prevalence level.

2.3 Parameter selection section

This part combined the correlation analysis between the water quality parameters and the starfish prevalence level to obtain the influence of the water quality parameters. It selected the data corresponding to the parameters with a more significant impactful than the training data.

2.4 Neural networks training and prediction section

The parameters with significant correlation and the historical number of starfish were the input, and the future starfish prevalence level was the output. A starfish early warning model was established using a BP neural network based on deep learning.

2.5 Data source

The data were collected from the underwater observation platform located in the marine ranching in the northern part of Weihai, China, as shown in Figure 2. The platform was placed 13 m underwater and measures 2 m long, 2 m wide, and 1 m high. The platform is equipped with various sensors, including underwater cameras and multi-parameter water quality sensors, to provide underwater video data and various water quality parameters such as temperature, salinity, depth, dissolved oxygen, and conductivity. The underwater camera used in this platform is a ROS-C 600 marine high-definition color camera with 30× optical zoom and 80° ultra-wide angle. The multi-parameter water quality sensor configured on the platform is the UMI-OT series, which can simultaneously observe multiple water quality parameters. The sensor parameters are shown in Table 1. The period of video and water quality parameter data used in this study is from May 2020 to October 2021. Eight kinds of marine life can be detected in the video, such as *P. pectinifera*, *Gymnocorymbus ternetzi*, *Stichopus japonicus*, and *Scophthalmus maximus*, and obtained 278 days of data between May 2020 and October 2021; the missing part was concentrated in December 2020 to June 2021 and October 2020, the main reason for which is that the water body is completely turbid, leading to video that cannot be used or video that cannot be saved owing to hard disk failure.

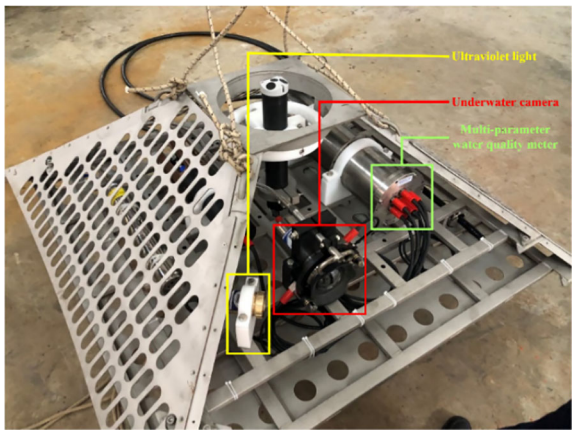


FIGURE 2
Underwater observation platform.

2.6 Data preprocessing

2.6.1 Preprocessing of video

In this study, 800,000 images were extracted at fixed intervals from the video data from May 2020 to October 2021, and 5,349 images were selected as a dataset and labeled. The selection principles were as follows:

2.6.1.1 Principle 1

The number of pictures selected each month is as close as possible. Because the turbidity of the water body varies slightly in each month, this can prevent the data distribution from shifting and the object detection model can only be used in certain months and is not robust.

2.6.1.2 Principle 2

Do not select pictures that are close in time because the *P. pectinifera*, as a slow-moving echinoderm, will not have a big difference in the distribution in a short time.

2.6.1.3 Principle 3

Pick images with better lighting conditions that will be more conducive to labeling.

The selected pictures were divided into the training set, testing set, and validation set according to the commonly used ratio of 6:2:2. In a total of 5,349 images, 3,209 images are selected as the training data, 1,070 images are chosen as the testing data randomly, and the remaining 1,070 images are used as the testing data. The training set was used to train the detection model. The validation set

TABLE 1 Sensor parameters.

Sensor Type	Range	Resolution
Salinity sensor	0–85 mS/cm	± 0.008 mS/cm
Temperature sensor	−5–45°C	± 0.003°C
Dissolved oxygen sensor	0–500 µmol/L	± 0.1 µmol/L
Conductivity sensor	0–90 mS/cm	± 0.003 mS/cm
Depth sensor	0–100 m	± 0.03%

was used to reflect the training process. The testing set was used to test the generalization ability of the detection model.

The detection model used in this study was the yolov5 algorithm based on deep learning. The detection model was constructed through the target detection algorithm, the target in the image was recognized, and the number of the same category is counted. Using 5,349 pictures to obtain a trained detection model, the detection model had an accuracy rate of 95% for the primary research target starfish. The detection model was applied to all videos from May 2020 to October 2021. The sequence data of the number of starfish over time resulted from preprocessing of the video. The sequence data were divided by taking the maximum value every hour, with missing data being filled with zeros. This method resulted in 12,273 data samples corresponding to 12,273 h of data.

2.6.2 Water quality data preprocessing

Water quality data of the multi-parameter water quality meter used in this study include five parameters: temperature, conductivity, depth, dissolved oxygen, and salinity. Sensors that provide raw data may experience various malfunctions that result in missing values or outliers that are significantly different from other values. For such data, we adopt the method of direct discarding to remove them from our analysis. The sensor sampling frequency was once every 5 min. To align the water quality data with the sequence data in time, the following operations are performed: firstly, the missing values and outliers are directly deleted, and then the moving average method with a window size of 1 h is used to calculate each average value over the hour. The water quality data were divided into 12,273 subsets at 1-h intervals.

2.6.3 Data normalization

Data normalization is an essential operation in data preprocessing. In this study's data analysis process, the units and magnitudes of various data are vastly different because the data come from different sensors. The data normalization method is used to scale the original data and eliminate the effect of the unit. It is calculated as follows:

$$w_{ij} = \frac{w_{ij} - \min(w_{ij})}{\max(w_{ij}) - \min(w_{ij})} \quad (1)$$

where w_{ij} is the value of the ith water quality parameter at the jth moment.

2.7 Calibration of starfish prevalence level

The starfish prevalence level is often determined according to the number of starfish. Owing to starfish's different survival and development statuses in different sea areas, it was difficult to define a unified standard (Hughes et al., 2014). When people classify starfish prevalence levels, they usually draw a rough classification standard based on the damage caused by starfish outbreaks to the local ecological environment. However, the description of the boundary between different levels was often vague, and a large area of uncertainty is prone to appear. In this study, we used the self-supervised method. The K-means clustering algorithm, which often was used for level classification, was used to classify starfish outbreak disasters, which can calibrate the number of different starfish to as many levels as possible to maximize the interval between samples (Hartigan & Wong, 1979; Li et al., 2016; Rifa et al., 2020).

2.7.1 Initial point selection

The interval between different calibration results is determined according to the data distribution, avoiding human subjectivity. The results of the k-means algorithm largely depend on the position of the initial point, which means that the results have great randomness (Erisoglu et al., 2011). Because of the random selection of the initial point, the calculation results will be different each time, and this error cannot be avoided. There are specific errors in any clustering algorithm. To reduce the error caused by randomness, this study will select multiple random initial points, calculate the similarity between classification results, and select the one with the highest similarity with other classification results. The classification result is the final level calibration result. The calculation of the similarity refers to Eq. 2 in *Section 2.8.1*.

2.7.2 Hyperparameter selection

To begin with, the sequence data are first aggregated using a time window of size s and a step size of 1 day. This process results in a three-dimensional feature point for each aggregation part, including data statistics within the calculation window. The three statistics included in the feature point are the average value, maximum value, and standard deviation. Choose k value as the number of clusters in the k-means algorithm and the number of classifications of the starfish prevalence level. The specific meanings of k and s are shown in Table 2.

TABLE 2 Specific meanings of k and s.

Parameters	Name	Meanings
k	Number of classifications of the starfish prevalence level	The physical interpretation of this parameter pertains to the number of classifications of the starfish prevalence level, which is expressed as the number of clusters in the clustering algorithm in the algorithm.
s	Time window size	The physical interpretation of this parameter is the number of how long to assess the starfish prevalence level, which is expressed as the amount of data contained in the prevalence level label in the algorithm.

Therefore, two hyperparameters, k and s, need to be selected in the steps of using this algorithm. The specific selection method is as follows: give a set of commonly used k values and a set of s values, combine them, and give each feature point a feature label through the k-means algorithm. In the complete dataset, according to the time series, select the feature label to which the nearest feature center in time belongs. The new feature label sequence obtained by combining each kind of k and s is calculated according to the correlation between the corresponding time and the original data sequence of the number of starfish. Then, the k and s values are determined by comprehensively considering the actual engineering situation and the correlation. Table 3 shows the classification of marine biological prevalence levels by countries around the world. It can be seen from the table that the classification of marine animal prevalence level in practice is mainly divided into two to five levels, with three levels and four levels being the most common. Therefore, a set of k = [2,3,4,5] is selected.

At the same time, according to some China official documents, such as "the notice on establishing and improving the marine ecological early warning and monitoring system by the ministry of natural resources", "the emergency response plan for storm surge, waves, tsunamis and sea ice", and "the provision on national marine warning report consultation", weekly and monthly periods are common reporting cycles for marine disaster warning reports. Therefore, a set of s = 7–30 was selected.

2.8 Parameter selection

It is necessary to select parameters to reduce the complexity of the EWSP model, reduce the prevalence of overfitting, and enhance the generalization ability.

2.8.1 Linear correlation analysis

To describe the qualitative relationship between the factors affecting the starfish outbreak disaster and the starfish prevalence level and to screen out water quality parameters that are more relevant to the starfish prevalence level, it is necessary to calculate the correlation between each parameter and the starfish prevalence level (Chen et al., 2021a). We use the Pearson correlation coefficient during the feature selection process (Li et al., 2016; Jayaweera & Aziz, 2018). The Pearson correlation coefficient is usually used to measure the degree of linear correlation between two variables. The larger the absolute value, the higher the correlation, and positive or negative represents the positive or negative correlation, respectively. It is calculated as follows:

$$r_i = \frac{\sum_{j=1}^{m}(w_{ij} - \bar{w}_i)(l_j - \bar{l})}{\sqrt{\sum_{j=1}^{n}(w_{ij} - \bar{w}_i)^2}\sqrt{\sum_{j=1}^{n}(l_j - \bar{l})^2}} \tag{2}$$

where r_i is the Pearson correlation coefficient between the ith water quality parameter and the starfish prevalence level ($-1 \leq r_i \leq 1$), $\bar{w}_i$ is the average of the numerical sequence corresponding to the ith water quality parameter, l_j is the calibration value of the starfish prevalence level at the jth moment, and j is the average value of the calibration value of the starfish prevalence level.

2.8.2 Nonlinear correlation analysis

There are some disadvantages to using only the Pearson correlation coefficient analysis. The Pearson correlation coefficient can well characterize the linear relationship between two variables, but when there is a nonlinear relationship between them, this ability to represent it will be significantly reduced. Therefore, this paper uses the Spearman correlation coefficient to measure the nonlinear correlation between parameters and quantities (Paul et al., 2017; Cai et al., 2019). Here, we briefly explain the Spearman correlation coefficient. The Spearman correlation coefficient is the Pearson correlation coefficient between grade variables. Its original calculation formula is similar to the Pearson correlation coefficient calculation formula. The difference is that the original data are converted into grade data, and the original data are based on their average. The descending position is given a corresponding rank. The absolute value of Spearman's coefficient indicates the degree of dependence. The larger the absolute value, the higher the monotonic correlation. Its calculation formula can be simplified as:

$$\rho_i = 1 - \frac{6\sum_{j=1}^{n} d_j^2}{n(n^2 - 1)} \tag{3}$$

$$d_j = R_{(w_{ij})} - R_{l_j} \tag{4}$$

TABLE 3 Marine biological prevalence management file.

Filename	Country	Prevalence Level Number
Emergency plan for aquatic animal diseases in Shandong province	China	4
Aquatic nuisance species task force strategic plan	United States	4
Australian pest animal strategy	Australia	4
Invasive alien species act	Japan	3
Harmonia+	Belgium	2
German–Austrian black list information system	Germany–Austria	3
Great Britain non-native species risk assessment	England	4
European non-native species in aquaculture risk analysis scheme	European Union	5

ep_i is the Spearman correlation coefficient between the ith water quality parameter and the starfish prevalence level ($-1 \le p_i \le 1$), $R(w_{ij})$ and R_{ij} respectively represent the ranking of the value of the ith water quality parameter at the jth moment in its numerical sequence in descending order and the ranking of the starfish prevalence level calibration value at the jth moment in its numerical sequence in descending order, and n is the sample size of the numeric sequence.

2.8.3 Correlation parameter selection

The Pearson correlation coefficient is mainly used to measure the linear correlation between two variables, so it is susceptible to outliers. Compared with the Pearson correlation coefficient, the Spearman correlation coefficient is a correlation coefficient based on the rank of the data. Since the amount of data directly manipulated is rank, the coefficient is robust to outliers and can reflect nonlinear relationships. Therefore, the selection of high-correlation parameters in this paper should consider both. The calculation method of the parameter selection evaluation index p_i is as follows (Cai et al., 2019):

$$p_i = \frac{r_i - \min(r_i)}{\max(r_i) - \min(r_i)} + \frac{\rho_i - \min(\rho_i)}{\max(\rho_i) - \min(\rho_i)} \quad (5)$$

where p_i is the selection evaluation index of the ith water quality parameter ($0 \le p_i \le 2$). According to its calculation formula and range, the category of parameters defined here with a p_i greater than 1.5 will be selected as BP neural network model's input features.

2.9 Training based on BP neural network model

2.9.1 Structure of the BP neural network model

A multilayer perceptron BP neural network model is established as the early warning model for starfish outbreak disasters. The output is the average of the starfish prevalence level in the next period. The structure of the BP neural network model is shown in Figure 3, which includes an input layer, three hidden layers, and an output layer. The input layer receives a tuple, and after the activation of three hidden layers, a probability tuple is an output in the output layer.

2.9.2 Dataset of the BP neural network model

To construct the BP neural network dataset, the time window method was used with a step size of 1 h. Specifically, the dataset feature part consists of 336 consecutive hours of the water quality data and sequence data statistics, including mean, maximum, and standard deviation values. The dataset label part consists of the corresponding maximum prevalence level occurring in the next 72 h after the 336 h. The dataset contains 11,864 data samples, which are divided into three parts: 7,118 for training, 2,374 for validation, and 2,372 for testing, in a ratio of 6:2:2.

3 Results and discussion

3.1 Prevalence level calibration

This study used the K-means algorithm to divide the starfish prevalence level. The division method is based on the actual number of starfish. To divide results more relevant and robust to the number of starfish, a set of commonly used k = [2,3,4,5] was set, and according to the observations widely used in actual engineering interval $s = 7 - 30$. Among them, k was the number of classification of starfish prevalence levels in this method, and s was the length of the sub-dataset after division. According to the ecological prevalence level classification standard commonly used in marine engineering practice and the requirement that the calibration results be robust to the number of oceans, we set k to 4 and s to 14 as the model hyperparameters among the parameter combination results. After choosing the hyperparameters, 18 random selections are made to determine the initial point, and the correlation between each clustering result and other results is calculated and summed. The result with the highest correlation sum is used as the final initial point. The dataset used in this model contains 511 days of data. After data aggregation with s = 14, 482 feature points are obtained for clustering. The clustering results are interpreted as follows: Each feature point is represented by three statistics (average value,

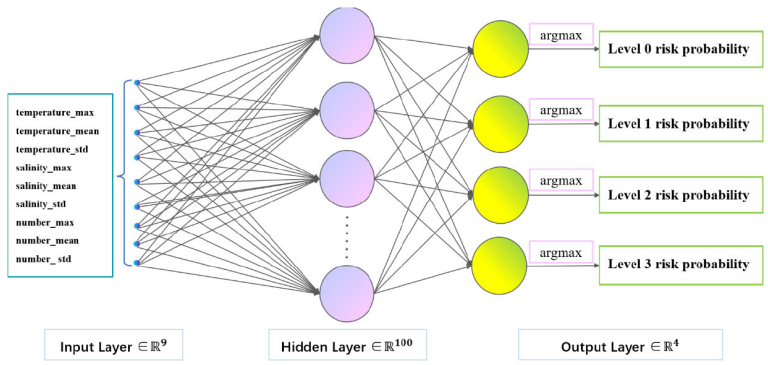


FIGURE 3
The BP neural network structure.

maximum value, and standard deviation) as the coordinate axes, and the resulting clustering is visualized in Figure 4A. Figure 4B shows the results of calibrating the clustering results to each sequence data according to the date, including 12,273 data.

The specific meaning of Figure 4 is as follows: Level 0 feature points are mostly concentrated in areas where the average, maximum, and standard deviation are relatively small, indicating that in this prevalence level, the target waters are likely to maintain a stable and low number of starfish. Feature points with level 1 have the characteristics of a small mean, small maximum, and large standard deviation, indicating that the number of starfish in the target waters at this prevalence level will fluctuate greatly and may return to a stable state of low numbers, but there is also a possibility of a sudden increase in number. Prevalence level 2 feature points have the characteristics of a small mean, large maximum, and large standard deviation, indicating that the number of starfish at this prevalence level is likely to reach a level of excessive aggregation that poses a threat to the ecology of the target waters. In prevalence level 3, the maximum, mean, and standard deviation are all very large, and the number is in a range that needs urgent control and is likely to continue to aggregate in the future. For the clustering results under s = 14,k = 4, we can give the description of the prevalence level division in Table 4.

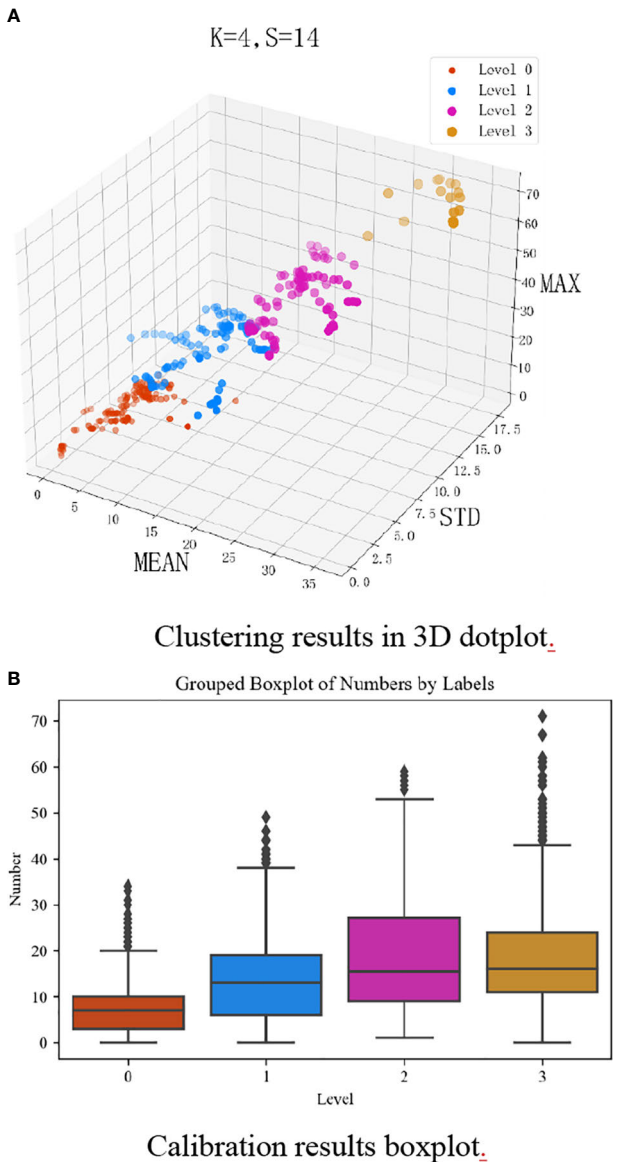


FIGURE 4
(A) Visualization of clustering results for different prevalence levels when k = 4 and s = 14. **(B)** The dotplot illustrates the clustering of data points, while the annotated boxplot presents the corresponding results.

3.2 Impactful factor selection

This study used the Pearson correlation coefficient and the Spearman correlation coefficient to calculate the parameters related to the starfish habitat collected by the sensors. Firstly, the Pearson correlation coefficient was used to calculate the correlation between the temperature, depth, salinity, electrical conductivity, dissolved oxygen, and the starfish prevalence level. Among them, depth and dissolved oxygen are negatively correlated with the starfish prevalence level, and other indicators are positively correlated with the starfish prevalence level. Secondly, the Spearman correlation coefficient is used to calculate the correlation between the temperature, depth, salinity, electrical conductivity, dissolved oxygen, and the starfish prevalence level. Among them, temperature and electrical conductivity have a more significant correlation with the starfish prevalence level, indicating that they have a more impactful effect on the occurrence of disasters.

After obtaining the calculation results of Pearson's and Spearman's correlation coefficients, we use the parameter selection evaluation index to evaluate each impactful factor. The results are shown in Table 5. It can be seen from the table that salinity and temperature are the two most impactful factors according to the calculation method of this evaluation index.

TABLE 4 Different prevalence level specific meanings and description.

Prevalence level	Meaning	Description
0	Very low prevalence	The number of starfish is within the acceptable range, and the number fluctuates slightly. Almost no impact on sea cucumber farming.
1	Low prevalence	The number of starfish is within the acceptable range, and there may be unacceptable aggregation. Little impact on sea cucumber farming.
2	Medium prevalence	The number of starfish has been in the unacceptable range for a long time and has a large change range. It is recommended to remove it regularly. Relatively great impact on sea cucumber farming.
3	High prevalence	The number of starfish has been in an unacceptable range for a long time and has changed greatly. It is recommended to take immediate measures. Extremely great impact on sea cucumber farming.

TABLE 5 Calculation results of parameter selection evaluation index p_i.

Parameters	Pearson r_i	Spearman p_i	Evaluation index p_i
Temperature	0.2798	0.1773	*1.6976*
Salinity	0.3520	0.5314	*1.7831*
Conductivity	0.0296	0.1537	0.8950
Dissolved oxygen	−0.0043	−0.0881	1.3420
Depth	−0.0919	−0.0927	0.5356

Therefore, the input feature group of the starfish outbreak disaster-grade BP neural network model should include temperature and salinity.

3.3 Parameters of the BP neural network

The BP neural network used in this paper is a three-layer BP neural network model, with 9 nodes in the input layer, 1,000 nodes in the hidden layer, and 4 nodes in the output layer. The input vector includes the average, standard deviation, and maximum value of historical temperature, salinity, and the number of starfish. The output results represent the predicted probabilities of the four prevalence levels. The activation function adopts the relu function. The essence of predicting the prevalence level is a multi-classification problem. The loss function uses cross-entropy loss. Figure 5 shows the number of starfish and their predictions over time.

3.4 Testing of early warning model effect

Use 20% of the data in the total dataset as a test set to evaluate the effect of the EWSP. A total of 2,372 data samples were tested, and the confusion matrix is shown in Figure 6.

For the starfish outbreak disaster early warning model, the historical temperature, salinity, and the number of starfish are selected as input tuples. The predicted results were compared with the actual results. Levels 0–3 represent four starfish prevalence levels: very low prevalence, low prevalence, medium prevalence, and high prevalence. Among the 2,372 prevalence data, 2,307 were correctly predicted, and 65 were wrongly predicted; for low prevalence, 401 were correctly predicted, 11 were wrongly predicted as low prevalence, 11 were predicted as very low prevalence, but no case was wrongly predicted as high prevalence; 73 out of 94 medium prevalence are correctly predicted; 284 out of 298 high prevalence are correctly predicted. The accuracy rate and the precision rate of four kinds of starfish prevalence level are 0.99, 0.95, 0.78, and 0.95, and the recall rates are 0.99, 0.94, 0.75, and 0.99, respectively; The F1-scores are 0.99, 0.94, 0.76, and 0.97 respectively. From the results of accuracy, the EWSP is good at predicting the prevalence level of starfish, especially for the early warning of lower prevalence levels. The early warning model has a very high accuracy rate.

The main errors in analyzing the early warning model come from many aspects. One is due to the error in the data itself, that is, the data collection process of the water quality sensors and the underwater camera. For example, the water quality sensor is affected by underwater attachments, or the water body is severely turbid, causing the camera to lose sight. This paper uses a simple method for missing values and outliers in raw data. The method of direct discarding makes the dataset lose a lot of original information; second, the biological aggregation characteristics of starfish in the sea area led to unbalanced sample categories. For example, in July, the concentration of starfish in the entire month was significantly more than in other months. In addition, the water quality sensor parameters used in this study only include five types, and after selection, only two parameters are used as input

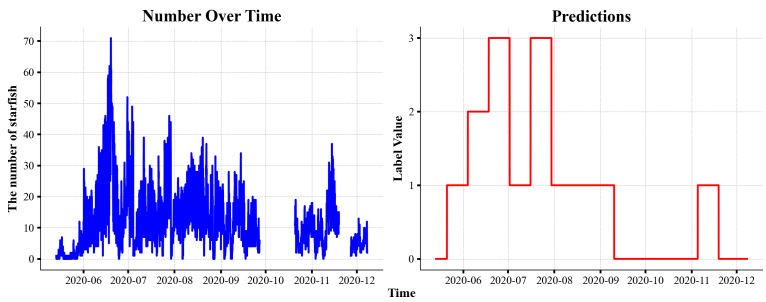


FIGURE 5
The number of starfish and its predictions over time.

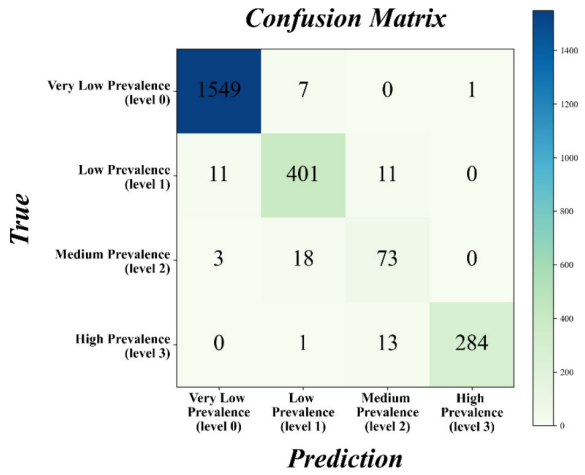


FIGURE 6
Confusion matrix of early warning results.

components, which belong to a small number of categories. This research assumes that this situation is also the cause of the error in the early warning model. The final result is that the accuracy of the EWSP is 0.9726.

4 Conclusions

This study used a deep learning-based object detection model and clustering algorithm to preprocess the dataset. The linear and nonlinear correlation coefficients are used to calculate the correlation of each water quality participating in the starfish prevalence level. Combining the number of historical starfish, water quality parameters, and starfish prevalence levels, select the highly correlated parameters and extract the data corresponding to the impactful factors and the number of starfish. The number of starfish, critical parameters, and starfish prevalence level are used as the training data of the neural network model. Finally, the early warning model for starfish outbreak disasters is obtained. In the Table 6, Judging from the prediction results of the four starfish prevalence levels, the accuracy rates of the 0, 1, 2, and 3 starfish prevalence levels predicted by the model are 0.99, 0.95, 0.74, and 0.95, respectively. The sample data are all from the natural underwater environment of marine ranching. The actual underwater environment parameter data ensure that the predictions are closer to the nature scene. At the same time, the training data selection can be adjusted, and the model architecture is simple and flexible. The neural network has a strong nonlinear fitting ability and exhibits excellent predictive performance in the case of sufficient data. At the same time, the research can be further expanded. In the future, if the natural enemies or food of starfish appear in the field of view of the underwater camera, the types of input data or the number of water quality parameters can be added to the model. For other starfishes related to shellfish farming or coral protection, such as COTS, underwater images containing COTS (such as the CSIRO dataset) can be added to the object detection model. The experimental results show that the EWSP in marine ranching has high precision, which has a specific application value for the regular operation of marine ranching, and EWSP can help to reduce the damage of starfish outbreak disaster time to marine ranching.

TABLE 6 Metrics of early warning results.

	Very Low (Level 0)	Low (Level 1)	Medium (Level 2)	High (Level 3)
Accuracy	0.99486191	0.94799054	0.77659574	0.95302013
Precision	0.99486191	0.94799054	0.77659574	0.95302013
Recall	0.99104287	0.93911007	0.75257732	0.99649123
F_1-score	0.99294872	0.94352941	0.76439791	0.97427101

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

All authors wrote the parts of various sections of the manuscript and contributed to the manuscript revision, read, and approved the submitted version.

Funding

This study was funded by Finance science and technology project of Hainan province (Grant No: ZDKJ202019), Scientific and technological projects of Zhoushan (Grant No: 2022C81010), and Hainan Provincial Joint Project of Sanya Yazhou Bay Science and Technology City (Grant No: 2021CXLH0020).

Acknowledgments

The authors are grateful for support from the science and technology projects in the “seed breeding, deep-sea and aerospace” industries, the Key Research and Development Plan of Zhejiang, the Major Science and Technology Project of Sanya, the High-level Personnel of Special Support Program of Zhejiang Province, and the Hainan Provincial Joint Project of Sanya Yazhou Bay Science and Technology. Thanks for data provided by Hainan Provincial Observatory of Ecological Environment and Fishery Resource in Yazhou Bay.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Agüera, A., Saurel, C., Møller, L. F., Fitridge, I., and Petersen, J. K. (2021). Bioenergetics of the common seastar asterias rubens: a keystone predator and pest for European bivalve culture. *Mar. Biol.* 168 (4), 48. doi: 10.1007/s00227-021-03854-7

Babcock, R. C., Plagányi, É.E., Condie, S. A., Westcott, D. A., Fletcher, C. S., Bonin, M. C., et al. (2020). Suppressing the next crown-of-thorns outbreak on the great barrier reef. *Coral Reefs* 39 (5), 1233–1244. doi: 10.1007/s00338-020-01978-8

Caballes, C., Pratchett, M., Raymundo, M., and Rivera-Posada, J. (2017). Environmental tipping points for sperm motility, fertilization, and embryonic development in the crown-of-Thorns starfish. *Diversity* 9 (1), 10. doi: 10.3390/d9010010

Cai, J., Xiao, D., Lv, L., and Ye, Y. (2019). An early warning model for vegetable pests based on multidimensional data. *Comput. Electron. Agric.* 156, 217–226. doi: 10.1016/j.compag.2018.11.019

Chen, X., Fu, Y., and Zhou, H. (2022). An approach of multi-element fusion method for harmful algal blooms prediction. *Environ. Sci. Pollut. Res.* 30 (11), 32083–32094. doi: 10.1007/s11356-022-23944-3

Chen, P., Li, F., and Wu, C. (2021a). Research on intrusion detection method based on Pearson correlation coefficient feature selection algorithm. *J. Physics: Conf. Ser.* 1757 (1), 12054. doi: 10.1088/1742-6596/1757/1/012054

Chen, B., Yu, K., Yao, Q., Liao, Z., Qin, Z., Yu, X., et al. (2021b). Insights into the environmental impact on genetic structure and larval dispersal of crown-of-Thorns starfish in the south China Sea. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.728349

Clement, R., Dunbabin, M., and Wyeth, G. (2005). Toward robust image detection of crown-of-thorns starfish for autonomous population monitoring. *Proceedings of the 2005 Australasian conference on robotics and automation* (Australian Robotics and Automation Association), 1–8.

Dayoub, F., Dunbabin, M., and Corke, P. (2015). "Robotic detection and tracking of crown-of-Thorns starfish," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE), 1921–1928. doi: 10.1109/IROS.2015.7353629

Dumas, P., Fiat, S., Durbano, A., Peignon, C., Mou-Tham, G., Ham, J., et al. (2020). Citizen science, a promising tool for detecting and monitoring outbreaks of the crown-of-thorns starfish acanthaster spp. *Sci. Rep.* 10 (1), 291. doi: 10.1038/s41598-019-57251-8

Erisoglu, M., Calis, N., and Sakallioglu, S. (2011). A new algorithm for initial cluster centers in k-means algorithm. *Pattern Recognition Lett.* 32 (14), 1701–1705. doi: 10.1016/j.patrec.2011.07.011

Gesú, V. D., Tegolo, D., Isgró, F., and Trucco, E. (2003). "A Low Level Image Analysis Approach to Starfish Detection" in *Image: e-Learning, Understanding, Information Retrieval, Medical*, 132–139.

Gobi, A. F. (2010). "Towards generalized benthic species recognition and quantification using computer vision," in *2010 Fourth Pacific-Rim Symposium on Image and Video Technology* (IEEE), 94–100. doi: 10.1109/PSIVT.2010.23

Hartigan, J. A., and Wong, M. A. (1979). Algorithm AS 136: a K-means clustering algorithm. *Appl. Stat* 28 (1), 100. doi: 10.2307/2346830

Heenaye-Mamode Khan, M., Makoonlall, A., Nazurally, N., and Mungloo-Dilmohamud, Z. (2023). Identification of crown of thorns starfish (COTS) using convolutional neural network (CNN) and attention model. *PloS One* 18 (4), e0283121. doi: 10.1371/journal.pone.0283121

Hou, S., Li, W., Liu, T., Zhou, S., Guan, J., Qin, R., et al. (2022). MUST: a multi-source spatio-temporal data fusion model for short-term sea surface temperature prediction. *Ocean Eng.* 259, 111932. doi: 10.1016/j.oceaneng.2022.111932

Hughes, R. N., Hughes, D. J., and Smith, I. P. (2014). Limits to understanding and managing outbreaks of crown-of-thorns starfish (Acanthaster spp.). *Oceanogr. Mar. Biol.: Ann. Rev.* 52, 133–200.

Jayaweera, C., and Aziz, N. (2018). "Reliability of principal component analysis and Pearson correlation coefficient, for application in artificial neural network model development, for water treatment plants," in *IOP Conference Series: Materials Science and Engineering*, (IOP Publishing), Vol. 458. 012076. doi: 10.1088/1757-899X/458/1/012076

Jiang, Y., Li, C., Sun, L., Guo, D., Zhang, Y., and Wang, W. (2021). A deep learning algorithm for multi-source data fusion to predict water quality of urban sewer networks. *J. Cleaner Production* 318, 128533. doi: 10.1016/j.jclepro.2021.128533

Kashenko, S. D. (2003). The reaction of the starfish Asterias amurensis and Patiria pectinifera (Asteroidea) from Vostok Bay (Sea of Japan) to a salinity decrease. *Russ. J. Mar. Biol.* 29, 110–114. doi: 10.1023/A:1023952524166

Kayal, M., Vercelloni, J., Lison de Loma, T., Bosserelle, P., Chancerelle, Y., Geoffroy, S., et al. (2012). Predator crown-of-Thorns starfish (Acanthaster planci) outbreak, mass mortality of corals, and cascading effects on reef fish and benthic communities. *PloS One* 7 (10), e47363. doi: 10.1371/journal.pone.0047363

Lambert, P. (1997). *Sea Cucumbers of British Columbia, southeast Alaska and puget sound* The Royal British Columbia museum handbook.

Li, C., Sun, L., Jia, J., Cai, Y., and Wang, X. (2016). Risk assessment of water pollution sources based on an integrated k-means clustering and set pair analysis method in the region of shiyan, China. *Sci. Total Environ.* 557–558, 307–316. doi: 10.1016/j.scitotenv.2016.03.069

Liu, J., Kusy, B., Marchant, R., Do, B., Merz, T., Crosswell, J., et al. (2021) *The CSIRO crown-of-Thorn starfish detection dataset*. Available at: http://arxiv.org/abs/2111.14311.

Lucas, J. S. (1973). Reproductive and larval biology of acanthaster planci (L.) in great barrier reef waters. *Micronesica* 9 (2), 197–203.

Manzione, R. L., and Castrignanò, A. (2019). A geostatistical approach for multi-source data fusion to predict water table depth. *Sci. Total Environ.* 696, 133763. doi: 10.1016/j.scitotenv.2019.133763

Miyoshi, K., Kuwahara, Y., and Chiba, S. (2019). Interactions between predatory sea stars (*Asterias amurensis* and *Distolasterias nipon*) and Japanese scallops (*Mizuhopecten yessoensis*) and implications for scallop seeding in mariculture. *Aquacult Res.* 50 (9), 2419–2428. doi: 10.1111/are.14195

Nguyen, Q. T. (2022). Detrimental starfish detection on embedded system: a case study of YOLOv5 deep learning algorithm and TensorFlow lite framework. *J. Comput. Sci. Institute* 23, 105–111. doi: 10.35784/jcsi.2896

Nti, I. K., Adekoya, A. F., and Weyori, B. A. (2021). A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction. *J. Big Data* 8 (1), 17. doi: 10.1186/s40537-020-00400-y

Paul, D., Su, R., Romain, M., Sébastien, V., Pierre, V., and Isabelle, G. (2017). Feature selection for outcome prediction in oesophageal cancer using genetic algorithm and random forest classifier. *Comput Med Imaging Graph* 60, 42–49. doi: 10.1016/j.compmedimag.2016.12.002

Plagányi, É.E., Babcock, R. C., Rogers, J., Bonin, M., and Morello, E. B. (2020). Ecological analyses to inform management targets for the culling of crown-of-thorns starfish to prevent coral decline. *Coral Reefs* 39 (5), 1483–1499. doi: 10.1007/s00338-020-01981-z

Popov, R. S., Avilov, S. A., Silchenko, A. S., Kalinovsky, A. I., Dmitrenok, P. S., Grebnev, B. B., et al. (2014). Cucumariosides F1 and F2, two new triterpene glycosides from the sea cucumber eupentacta fraudatrix and their LC-ESI MS/MS identification in the starfish patiria pectinifera, a predator of the sea cucumber. *Biochem. Systematics Ecol.* 57, 191–197. doi: 10.1016/j.bse.2014.08.009

Pradisty, N. A., Ampou, E. E., and Hanintyo., R. (2020). Water quality assessment in the occurrence of acanthaster spp. (Crown-of-Thorns starfish, CoTS) on coral reefs in menjangan island, Bali, Indonesia. *Makara J. Sci.* 24 (3), 8. doi: 10.7454/mss.v24i3.1016

Pratchett, M. S., Lang, B. J., and Matthews, S. (2019). Culling crown-of-thorns starfish (Acanthaster cf. solaris) on australia's great barrier reef: rationale and effectiveness. *Aust. Zoologist* 40 (1), 13–24. doi: 10.7882/AZ.2018.021

Rifa, I. H., Pratiwi, H., and Respatiwulan, R. (2020). CLUSTERING OF EARTHQUAKE RISK IN INDONESIA USING K-MEDOIDS AND K-MEANS ALGORITHMS. *Media Statistika* 13 (2), 194–205. doi: 10.14710/medstat.13.2.194-205

Rogers, J., Pláganyi, É., and Babcock, R. (2017). Aggregation, allee effects and critical thresholds for the management of the crown-of-thorns starfish acanthaster planci. *Mar. Ecol. Prog. Ser.* 578, 99–114. doi: 10.3354/meps12252

Saponari, L., Montalbetti, E., Galli, P., Strona, G., Seveso, D., Dehnert, I., et al. (2018). Monitoring and assessing a 2-year outbreak of the corallivorous seastar acanthaster planci in ari atoll, republic of Maldives. *Environ. Monit. Assess.* 190 (6), 344. doi: 10.1007/s10661-018-6661-z

Sheth, S., and Prajapati, D. J. (2022). "Recognition of underwater starfishes using deep learning," in *2022 Second International Conference on Next Generation Intelligent Systems (ICNGIS)*. (IEEE), 1–5. doi: 10.1109/ICNGIS54955.2022.10079811

Smith, D., and Dunbabin, M. (2007). "Automated counting of the northern pacific Sea star in the Derwent using shape recognition," in *9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications (DICTA 2007)*. (IEEE), 500–507. doi: 10.1109/DICTA.2007.4426838

Suzuki, G., Yasuda, N., Ikehara, K., Fukuoka, K., Kameda, T., Kai, S., et al. (2016). Detection of a high-density brachiolaria-stage larval population of crown-of-Thorns Sea star (Acanthaster planci) in sekisei lagoon (Okinawa, Japan). *Diversity* 8 (4), 9. doi: 10.3390/d8020009

Tang, Y., Feng, Y., Fung, S., Xomchuk, V. R., Jiang, M., Moore, T., et al. (2022). Spatiotemporal deep-Learning-Based algal bloom prediction for lake Okeechobee using multisource data fusion. *IEEE J. Selected Topics Appl. Earth Observations Remote Sens.* 15, 8318–8331. doi: 10.1109/JSTARS.2022.3208620

Truong, P. (2022). *Crown-of-Thorns starfish detection by state-of-the-art YOLOv5*.

Uthicke, S., Robson, B., Doyle, J. R., Logan, M., Pratchett, M. S., and Lamare, M. (2022). Developing an effective marine eDNA monitoring: EDNA detection at pre-outbreak densities of corallivorous seastar (Acanthaster cf. solaris). *Sci. Total Environ.* 851, 158143. doi: 10.1016/j.scitotenv.2022.158143

Wang, G., Guan, X. X., and Shi, Y. H. (2021). SIMULATION STUDY ON THE ARTIFICIAL ECOSYSTEM OF MARINE RANCHING AT DALIAN ZHANGZI ISLAND. *Appl. Ecol. Environ. Res.* 19 (1), 525–548. doi: 10.15666/aeer/1901_525548

Westcott, D. A., Fletcher, C. S., Kroon, F. J., Babcock, R. C., Plagányi, E. E., Pratchett, M. S., et al. (2020). Relative efficacy of three approaches to mitigate crown-of-Thorns starfish outbreaks on australia's great barrier reef. *Sci. Rep.* 10 (1), 1–12. doi: 10.1038/s41598-020-69466-1

Yamaguchi, M. (1973). Early life histories of coral reef asteroids, with special reference to Acanthaster planci (L.). *Biol. Geol. Coral Reefs* 2, 480. doi: 10.1016/B978-0-12-395526-5.50019-5

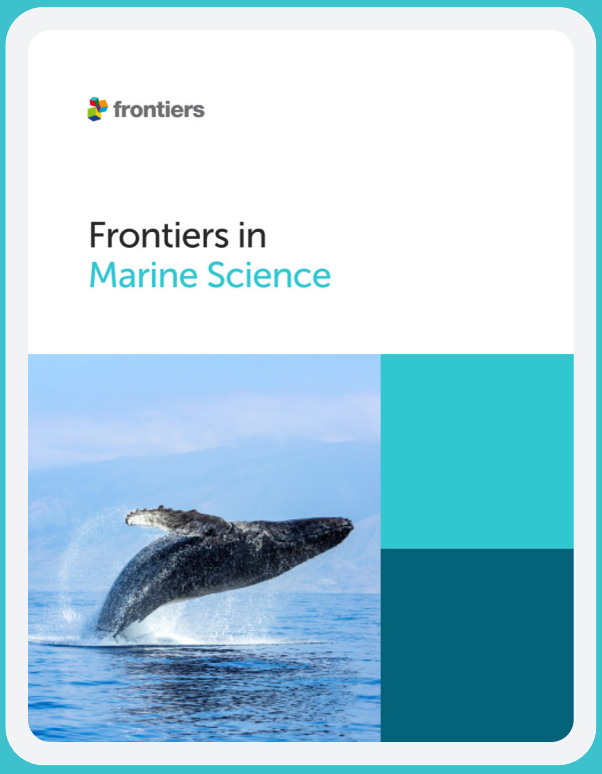
Frontiers in Marine Science

Explores ocean-based solutions for emerging global challenges

The third most-cited marine and freshwater biology journal, advancing our understanding of marine systems and addressing global challenges including overfishing, pollution, and climate change.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact