frontiers | Research Topics

Remote sensing applications in oceanography with deep learning

Edited by
Muhammad Yasir, Shah Nazir, Chao Chen and Weimin Huang

Published in
Frontiers in Marine Science
Frontiers in Remote Sensing

ISSN 1664-8714
ISBN 978-2-8325-6999-3
DOI 10.3389/978-2-8325-6999-3

Generative AI statement
Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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

Remote sensing applications in oceanography with deep learning

Topic editors

Muhammad Yasir — China university of petroleum (East China), China
Shah Nazir — University of Swabi, Pakistan
Chao Chen — Suzhou University of Science and Technology, China
Weimin Huang — Memorial University of Newfoundland, Canada

Citation

Yasir, M., Nazir, S., Chen, C., Huang, W., eds. (2025). *Remote sensing applications in oceanography with deep learning*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-6999-3

Table of contents

05 Editorial: Remote sensing applications in oceanography with deep learning
Muhammad Yasir, Chao Chen, Shah Nazir and Weimin Huang

07 Optimizing underwater connectivity through multi-attribute decision-making for underwater IoT deployments using remote sensing technologies
Inam Ullah, Farhad Ali, Amin Sharafian, Ahmad Ali, H. M. Yasir Naeem and Xiaoshan Bai

24 Development of VIIRS-OLCI chlorophyll-a product for the coastal estuaries
Alexander Gilerson, Mateusz Malinowski, Jacopo Agagliate, Eder Herrera-Estrella, Maria Tzortziou, Michelle C. Tomlinson, Andrew Meredith, Richard P. Stumpf, Michael Ondrusek, Lide Jiang and Menghua Wang

43 AB-LSTM: a mesoscale eddy feature prediction method based on an improved Conv-LSTM model
Xiaodong Ma, Lei Zhang, Weishuai Xu and Maolin Li

59 Two-decade variability and trend of chlorophyll-a in the Arabian Sea and Persian Gulf based on reconstructed satellite data
Mengmeng Yang, Faisal Ahmed Khan, Hua Fang, Elígio de Raús Maúre, Joji Ishizaka, Dong Liu and Shengqiang Wang

73 Oriented ice eddy detection network based on the Sentinel-1 dual-polarization data
Jinqun Wu, Yiqin Zheng, Tingting Wang, Chunyong Ma and Ge Chen

91 Machine learning-based analysis of sea fog's spatial and temporal impact on near-miss ship collisions using remote sensing and AIS data
Dan Liu, Ling Ke, Zhe Zeng, Shuo Zhang and Shanwei Liu

109 Leveraging ResUnet, oceanic and atmospheric data for accurate chlorophyll-a estimations in the South China Sea
Weiwei Fang, Ao Li, Haoyu Jiang, Chan Shu and Peng Xiu

124 A novel edge-feature attention fusion framework for underwater image enhancement
Shuai Shen, Haoyi Wang, Weitao Chen, Pingkang Wang, Qianyong Liang and Xuwen Qin

141 Small object detection in side-scan sonar images based on SOCA-YOLO and image restoration
Xiaodong Cui, Jiale Zhang, Lingling Zhang, Qunfei Zhang and Jing Han

158 Sonar-based object detection for autonomous underwater vehicles in marine environments
Zhen Wang, Jianxin Guo, Shanwen Zhang and Yucheng Zhang

181 RipFinder: real-time rip current detection on mobile devices
Fahim Khan, Akila de Silva, Ashleigh Palinkas, Gregory Dusek, James Davis and Alex Pang

195 A deep learning-based data augmentation method for marine mammal call signals
Jiaming Jiang, Wanlu Cheng, Shengwen Gong and Jingjing Wang

204 Retrieval algorithm based on locally sensitive hash for ocean observation data
Meijuan Jia, Xiaodong Mao, Shuai Guo and Xin Li

220 Satellite remote sensing of algal blooms in seagoing river in Eastern China
Zili Zhang, Jinsong Liu, Ruru Deng, Zunying Hu and Shuping Pan

230 A marine ship detection method for super-resolution SAR images based on hierarchical multi-scale Mask R-CNN
Jiancong Fan, Miaoxin Guo, Lei Zhang, Jianjun Liu and Yang Li

244 Detecting small seamounts in multibeam data using convolutional neural networks
Tobias Ziolkowski, Colin W. Devey and Agnes Koschmider

262 A deep-learning framework to detect green tide from MODIS images
Weidong Zhu, Yuelin Xu, Lei Zhang, Zitao Liu, Shuai Liu and Yifei Li

frontiers | Frontiers in Marine Science

TYPE Editorial
PUBLISHED 26 September 2025
DOI 10.3389/fmars.2025.1701125

OPEN ACCESS

EDITED AND REVIEWED BY
Johannes Karstensen,
Helmholtz Association of German Research Centres (HZ), Germany

*CORRESPONDENCE
Muhammad Yasir
lb2116001@s.upc.edu.cn

RECEIVED 08 September 2025
ACCEPTED 15 September 2025
PUBLISHED 26 September 2025

CITATION
Yasir M, Chen C, Nazir S and Huang W (2025) Editorial: Remote sensing applications in oceanography with deep learning.
Front. Mar. Sci. 12:1701125.
doi: 10.3389/fmars.2025.1701125

Editorial: Remote sensing applications in oceanography with deep learning

Muhammad Yasir[1]*, Chao Chen[2], Shah Nazir[3] and Weimin Huang[4]

[1]College of Oceanography and Space Informatics, Qingdao, China, [2]University of Swabi, Swabi, Pakistan, [3]Suzhou University of Science and Technology, Suzhou, China, [4]Memorial University of Newfoundland, St. John's, NL, Canada

KEYWORDS

deep learning, remote sensing applications, marine science and technology, oceanography, machine learning

Editorial on the Research Topic
Remote sensing applications in oceanography with deep learning

Deep learning and remote sensing for the ocean: from concept to operational impact

Deep learning (DL) and remote sensing (RS) are transforming how we observe and manage the ocean. Modern algorithms, platforms, and multi-sensor data integration now deliver insights at scales and speeds that were impossible just a few years ago. This Research Topic gathers 17 contributions across seafloor geomorphology, Ship and hazard monitoring, water quality assessment, mesoscale dynamics, under-ice processes, sonar perception, and enabling methods—demonstrating a field that is both technically innovative and mission-driven.

Seafloor to shoreline

Ocean science relies on accurate mapping of the seabed. Automation can quickly analyse broad regions while collecting characteristics that satellite altimetry misses, as demonstrated by a CNN + U-Net pipeline for recognising tiny seamounts in multibeam data. RipFinder, a mobile machine learning system for real-time rip current identification that also functions as a citizen-science tool in places with restricted connection, exemplifies "AI to edge" at the land–sea interface.

Ships, safety, and hazards

For marine awareness, synthetic aperture radar (SAR), is still essential. While previous research use AIS data, sea fog, and remote sensing to evaluate collision risk, a super-

resolution Mask R-CNN architecture uses scale-aware fusion to improve ship detection in noisy SAR settings, providing evidence-based navigation management tools.

Ecosystems and water quality

Chlorophyll-a (Chl-a) variations and harmful blooms are important ecological markers. Green tide identification from MODIS images is enhanced by WaveNet (VGG16 + BiFPN + CBAM). The importance of physics-aware features is demonstrated by ResUNet models that relate ocean-atmosphere dynamics to Chl-a in the South China Sea. Long-term variability in the Persian Gulf and Arabian Sea is revealed by rebuilt MODIS datasets, and new techniques also yield transferable Chl-a products for estuaries. MarGEN, a GAN-based augmentation technique that enhances marine mammal call categorisation in situations where labelled audio is limited, is one example of an advancement in acoustic ecology.

Mesoscale and cryosphere dynamics

OIEDNet generates the first large-scale MIZ eddy catalogues by detecting under-ice eddies from Sentinel-1 dual-pol data, whereas Conv-LSTM GAN hybrids predict mesoscale eddy properties with high fidelity.

Perception underwater

Sonar and visual sensing are crucial for autonomous systems. Forward-looking sonar object detection is improved by MLFANet, side-scan sonar small-object recognition is improved by SOCA-YOLO, and underwater optical imaging is improved by CUG-UIEF using edge- and attention-based fusion.

Data, platforms, and decision support

New contributions also address scalable data management (LSH-based retrieval for ocean archives) and decision-making (multi-criteria approaches for underwater IoT and AUV deployments), underscoring the need to co-design sensing, connectivity, and computation.

Cross-cutting lessons

Five themes emerge: (1) multi-scale architectures consistently boost detectability; (2) embedding physics-aware features enhances generalization; (3) translating models to edge-deployable tools enables real-world impact; (4) data efficiency strategies such as augmentation and self-supervision are critical in data-sparse regimes; and (5) benchmarking and openness will accelerate progress.

Outlook

This Research Topic highlights a decisive shift from proof-of-concept to operational potential in ocean AI. Future priorities include embedding physical priors, advancing generative/self-supervised methods for sparse data, and ensuring scalability, efficiency, and usability for real-world applications. Together, these works show how DL and RS can protect mariners, monitor ecosystems, and reveal ocean dynamics—bringing us closer to truly actionable ocean intelligence.

We thank all authors and reviewers for their contributions and the editors for their support. We hope that this Research Topic will serve as both a reference and a springboard for progress in observing and managing the blue planet.

Author contributions

MY: Writing – original draft, Writing – review & editing. CC: Writing – original draft, Writing – review & editing. SN: Writing – original draft, Writing – review & editing. WH: Writing – original draft, Writing – review & editing.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 19 September 2024
DOI 10.3389/fmars.2024.1468481

OPEN ACCESS

EDITED BY
Chao Chen,
Suzhou University of Science and Technology, China

REVIEWED BY
Prince Waqas Khan,
West Virginia University, United States
Chunjiong Zhang,
Ajou University, Republic of Korea
Abdur Rasool,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE
Xiaoshan Bai
baixiaoshan@szu.edu.cn

RECEIVED 22 July 2024
ACCEPTED 30 August 2024
PUBLISHED 19 September 2024

CITATION
Ullah I, Ali F, Sharafian A, Ali A, Naeem HMY and Bai X (2024) Optimizing underwater connectivity through multi-attribute decision-making for underwater IoT deployments using remote sensing technologies.
Front. Mar. Sci. 11:1468481.
doi: 10.3389/fmars.2024.1468481

Optimizing underwater connectivity through multi-attribute decision-making for underwater IoT deployments using remote sensing technologies

Inam Ullah[1,2], Farhad Ali[3], Amin Sharafian[1,2], Ahmad Ali[1,2], H. M. Yasir Naeem[1,2] and Xiaoshan Bai[1,4]*

[1]College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen, China, [2]College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China, [3]Department of Accounting and Information Systems, College of Business and Economics, Qatar University, Doha, Qatar, [4]National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University, Shenzhen, China

The underwater Internet of Things (UIoT) and remote sensing are significant for biodiversity preservation, environmental protection, national security, disaster assistance, and technological innovation. Assigning tasks to autonomous underwater vehicles (AUVs) is a fundamental challenge in underwater technology and exploration. Remote sensing and AUVs are vital for pollution detection, disaster prevention, marine observation, and ocean monitoring. This work presents an optimized network connectivity using a multi-attribute decision-making approach for underwater IoT deployment. A feature engineering approach highlights the significant characteristics of underwater things, incorporating remote sensing data, and a multi-objective optimization method is used to select optimal UIoT for effective task allocation in deep-sea environments. A balance between data transmission, energy economy, and operational performance is necessary for efficient task distribution. Effective communication algorithms and protocols are needed to maintain environmental sustainability, protect marine ecosystems, and improve underwater monitoring enhanced by remote sensing technologies. Multi-criteria decision-making (MCDM) is beneficial for addressing various challenges in underwater technology, considering factors such as mission objectives, energy efficiency, environmental conditions, vehicle performance, safety, and much more. The proposed criteria importance through intercriteria correlation (CRITIC) methodology will assess technical competencies like communication, resilience, navigation, and safety in an underwater environment, leveraging

remote sensing and aiding decision-makers in selecting appropriate undersea devices and vehicles for enhancing communication and transportation. This method prioritizes characteristics and aligns them with specific objectives, improving decision-making quality in the marine environment.

KEYWORDS

autonomous underwater vehicles, remote sensing, internet of underwater things, acoustics sensor networks, marine applications

1 Introduction

Emerging technologies such as the Internet of Things (IoT), machine learning, and big data analytics have revolutionized the lifestyles of common people. In essence, the term "IoT" solely pertains to the networking and communication stratum of the infrastructures in the Information Society, which establish connections between entities or devices and the Internet as well as among themselves. Through the linkage of entities that are ubiquitously present in our surroundings, the IoT has the potential to enhance our interactions with it (Jahanbakht et al., 2021; Gu et al., 2024). The term "underwater Internet of Things (UIoT)" describes an extensive worldwide network of networked underwater items that use embedded sensors, remote sensing technologies, tracking technologies, and the Internet to sense, understand, and react to their environment. Moreover, these gadgets can connect submerged and aboveground objects, including phones. Every underwater object has a fully functional virtual counterpart and is available to the public. Devices are connected to the Internet via the IoT, and underwater things are digitally identified via the underwater IoT (Domingo, 2012; Mariani et al., 2021). However, the lack of advanced sensors limits underwater surveillance technologies and sensor use. Low-power sensors, accompanied by remote sensing, can help address this issue, while marine sensors are crucial for ecological and environmental sustainability and saving lives (Refulio-Coronado et al., 2021).

The collaboration between remote sensing and underwater sensor networks (USNs) represents a significant advancement in marine technology, facilitating more precise and effective monitoring of oceanic conditions (Chen et al., 2022). USNs are employed for oceanography, pollution detection, underwater target detection, offshore exploration, and disaster prevention. These networks utilize unmanned underwater vehicles (UUVs) equipped with sensors specifically designed for underwater environments (Sun and Boukerche, 2018; Zacchini et al., 2022). The exchange of configuration, location, and motion information among these devices is made possible through underwater wireless acoustic networking. The UASN comprises diverse sensors and vehicles collaborating to monitor tasks within a designated area (Akyildiz et al., 2005). The next generation of USNs should have key characteristics such as reliability, robustness, adaptability, security, evaluability, efficiency, scalability, and intelligence, as illustrated in Figure 1 (Luo et al., 2018). Remote sensing is essential in these marine ecosystem procedures because it enables the gathering of data from underwater regions that are difficult to reach. The figure below depicts the fundamental characteristics essential for the future iteration of USNs. These characteristics guarantee that the USNs can operate efficiently and dependably in submerged surroundings.

Underwater communications necessitate the continuous monitoring of oceanic regions utilizing pre-existing technologies. However, such monitoring can lead to data loss during an interruption before recovery. To address this issue, it is imperative to establish instantaneous communication between the underwater instruments and the central control devices. This task is achieved by creating a rudimentary underwater acoustic network, which entails establishing a two-way acoustic connection between various devices, including autonomous underwater vehicles (AUVs) and sensors (Zhou et al., 2023). Remote sensing is essential for supplementing these acoustic networks by offering supplementary techniques for collecting data. This network is

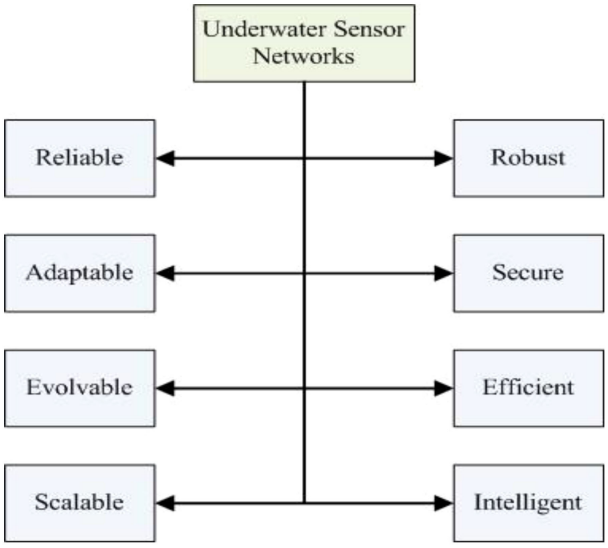


FIGURE 1
Performance metrics of UWSNs.

subsequently linked to a ground station, which can be connected to a host system via radio frequency (RF) communications, such as the Internet. Integrating the remote sensing data can augment the flexibility of such systems. Unlike terrestrial wireless sensor networks (WSNs), which trust radio waves for communication purposes, USNs use acoustic waves, which places a new research challenge in the scheme of MAC protocols. A comparison of various technologies for underwater communication and remote sensing is summarized in Table 1.

Underwater IoT (UIoT) applications utilize various network layers, often defined by the open systems interconnection (OSI) model and TCP/IP protocols. Remote sensing technologies can enhance these applications by offering additional sources of data. The data link layer uses a water channel for reliable transmission, while the physical layer uses specialized underwater communication technology. The OSI architecture uses a unique protocol considering depth, distance, and energy efficiency for packet routing (Luo et al., 2021). Remote sensing data can enhance and refine these techniques. It also handles reliability issues at the transport layer, improving latency and packet loss. The application layer analyzes data from underwater sensor platforms and devices to enable the implementation of IoT applications and services (Jiang, 2018). The use of remote sensing at this layer enables a more thorough analysis of data and the development of applications. Figure 2 shows the data transmission between various layers. The lower four layers of the OSI model comprise the main functionalities required for reliable transfer, which are divided into link and path levels. Link-level function objective is to lessen transmission errors caused by interference, noise, and frame collision between neighboring nodes. Path-level functions endeavor to guarantee end-to-end consistent transfer via network pathways, particularly by addressing packet losses.

AUVs utilized on underwater networks possess a promising capability to enhance their operational reach by transmitting control and data signals across extensive networks. AUVs and remote sensing can increase this capability by adding data and improving situational awareness. However, it should be noted that the capacity of shallow water acoustic channels is constrained, and numerous time-varying paths can result in significant symbol interference, as well as notable dispersion and Doppler shifts. To attain the necessary level of energy efficiency, underwater networks necessitate a hierarchical architecture (Sozer et al., 2000). Figure 3 shows the taxonomy of UASN and remote sensing. Underwater wireless satellites, called UWSNs, are crucial in coastal activities such as fish farm control, seabed mining, and water monitoring. Various factors influence underwater ecosystem instability, including temperature, lack of sensing capability, pressure, noise, and water density fluctuations (Abelson et al., 2020). Coping with several challenges, namely transmission delays, high probability of bit errors, limited bandwidth, and occasional loss of connectivity, poses significant problems in this domain (Garcia et al., 2011; Lloret et al., 2011). Moreover, RF signals are attenuated underwater, resulting in lower data rates and lesson remote sensing at very low frequencies. Alternatively, optical signals may not be useful due to light scattering in the underwater remote sensing environment. Acoustic modems fill a gap in existing technologies and must be energy efficient and economical due to the limited energy resources in the aquatic environment remote sensing. The hardware required to transmit audio signals is inexpensive, but transmission times are much slower than electromagnetic (EM) modems, at about 1500 m/s (Frampton, 2006; Farr et al., 2010).

The UIoT and remote sensing haven't received widespread attention due to their recent discovery and lack of scientific progress. Although 44% of the Earth's people live 150 kilometers or less from the ocean, 95% of the ocean's surface remains unexplored. Oceans cover 70% of the Earth's surface and provide habitat for nearly 500 million people. The development and use of

TABLE 1 Comparison of different technologies used for underwater communications and remote sensing.

Technology	Working Frequency	Modulation	Distance (m)	Data Rates (kbps)
EM Waves	2.4GHz	CCK	0.16m	11Mbps
	2.4GHz	QPSK	0.17m	2Mbps
	1KHz	BPSK	2m	1Kbps
	10KHz	BPSK	16m	1Kbps
	3KHz	–	40m	100bps
	5MHz	–	90m	500Kbps
Acoustic Waves	800KHz	BPSK	1m	80Kbps
	70KHz	ASK	70m	0.2Kbps
	24KHz	QPSK	2500m	30Kbps
	12KHz	MIMO-OFDM	–	24.36Kbps
Optical Waves	–	PPM	1.8m	100Kbps
	–	–	10m	10Mbps
	–	–	11m	9.69Kbps

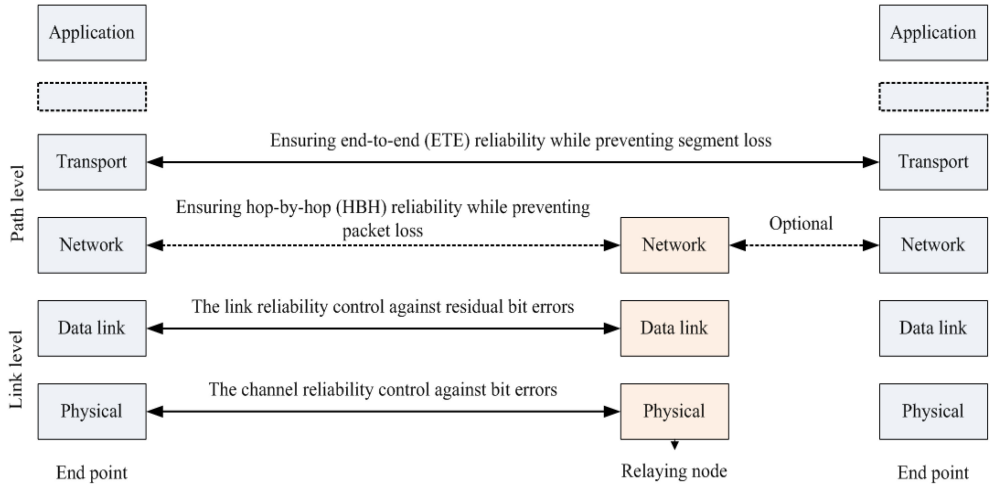


FIGURE 2
The data transmission between different layers (Zhou et al., 2023).

underwater exploration in UIoT can have a significant impact on people's lives. Due to the advancement in remote sensing and WSNs, the IoT has gained popularity for monitoring various applications, including volcanic activity, forest fire detection, air quality assessment, and home automation systems. However, underwater applications face challenges like sensor deployment and maintenance, energy acquisition, manufacturing costs, sensing issues, and signal propagation issues. Advanced wireless communication and sensing techniques are needed for underwater applications, which can be achieved through three-dimensional (3D space and algorithm placement). These networks offer opportunities for systematic examination of the underwater environment, including climate change impacts, deep-sea habitat research, sensing applications, coral reef population observations, ecological observation, military applications, mine exploration, water quality monitoring, disaster prevention systems, aquaculture supervision, and oceanic data collection and navigation (Sendra et al., 2015; Khan et al., 2023). Thus, UWSNs present a promising solution for various applications of remote sensing in the open sea. This work's key objectives are as follows:

- To expedite the use of modern technology for the AUV navigation and sensing system to enhance communication and networking.

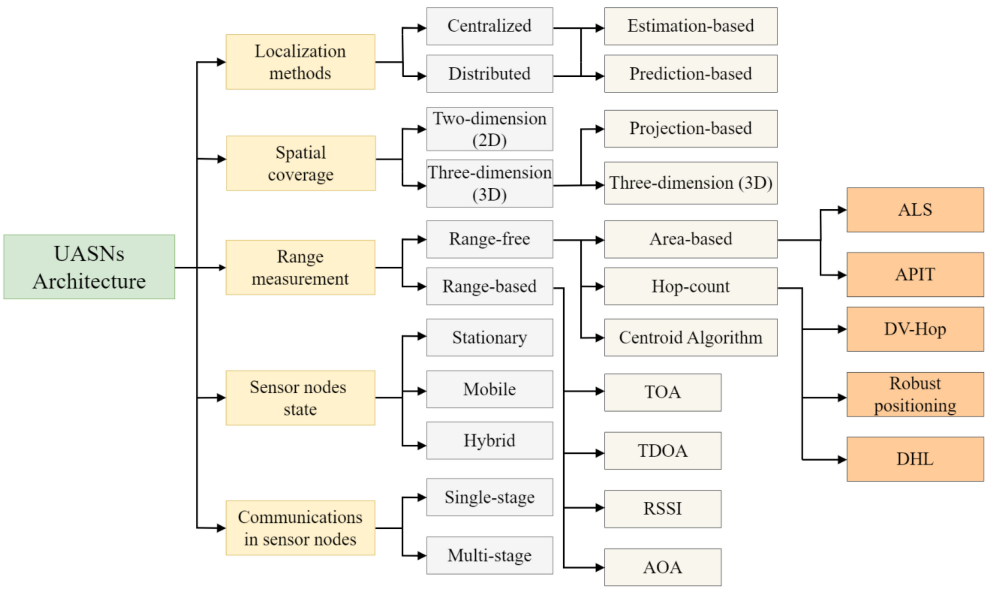


FIGURE 3
The taxonomy of underwater acoustics sensor networks (UASNs).

- To collect and retrieve various parameters, such as water quality, pressure, etc., that directly influence the behaviors of aquatic life and the UIoT.
- Using feature engineering strategy to highlight the significant characteristics of underwater things for efficient and effective sensing and tracking operations.
- To design a multifaceted criteria importance through intercriteria correlation (CRITIC)-based approach to prioritize and assess the essential attributes of UIoT for efficient task allocation and sensing in UIoT.

The rest of the article is organized as follows: Section II presents the overall literature review in the domain of underwater things, whereas Section III presents the methodology of the proposed model. Moreover, Section IV illustrates the results and discussions, while Section V concludes this work.

2 Literature review

Examining the extensive oceans, which cover two-thirds of the Earth's surface, requires UWSNs to understand this immense expanse fully. Future projections suggest that the market for AUVs is expected to grow substantially, with a compound annual growth rate (CAGR) of USD 1.638 billion by 2025. This represents a notable surge from the USD 638 million recorded in 2020. The applications of AUVs can be commercial, oceanographic, environmental, military, sensing, and more. Examples of commercial activities are surveying, port monitoring, and geophysical and archaeological research (Li et al., 2023; Wang et al., 2023). Scientific/oceanographic missions require seabed exploration, remote sensing, and water body exploration. Environmentally significant applications include habitat monitoring and water quality sampling. Anti-submarine warfare and border security are examples of military/defense activities. Shallow and medium water is the most typical deployments of AUVs in coastal waters (Brasier et al., 2020; Duan et al., 2020). Wireless charging in remote oceanic environments is expected to drive steady growth in the foreseeable future. The AUV Repository lists over 1,050 underwater platforms from over 350 universities, including standard components like battery modules, propulsion systems, sensing capacity, communication systems, navigation systems, and collision avoidance systems (Tian et al., 2023). However, the inertial navigation system's long-term accuracy is limited by accelerometer drift. To address this, ultra-short baseline or long baseline transponder systems can be used, or simultaneous localization and mapping (SLAM) can be employed (Cao et al., 2021; Hoeher et al., 2021).

However, current networks are hardware-centric, rigid, and need more resource-sharing capabilities. New models, such as software-defined technologies, have emerged to improve UWSNs by providing robustness, flexibility, adaptability, programmability, resource sharing, and easy administration (Sun and Boukerche, 2018). These technologies include software-defined networking (SDN), software-defined radio (SDR), network function virtualization (NFV), cognitive acoustic radio (CAR), underwater IoT sensing, and sensor clouds, turning network resources into software, improving resource efficiency and simplifying network management. These technologies could transform traditional UWSNs into next-generation, software-based, programmable, customizable, and service-oriented networks.

Many challenges arise when deploying IoT devices and networks in aquatic environments or underwater IoT for remote sensing purposes. These include challenges related to signal propagation and transmission in water, building and maintaining robust underwater positioning and navigation systems, optimizing energy efficiency for long-term operation, data processing and retrieval in low-bandwidth environments, and manufacturing waterproof and durable hardware (Arul et al., 2021; Wei et al., 2021). The data rates and usual bandwidth for underwater channels with different ranges are shown in Table 2 (Moradi et al., 2012). The underwater environment is particularly harsh and corrosive, making it difficult to build long-lasting sensors, secure networks, and keep sensors working as intended. Addressing these issues is crucial for the successful implementation of underwater IoT applications, from environmental monitoring to underwater robotics and exploration.

Autonomous underwater vehicles, or AUVs, are automated submersible platforms capable of operating at a maximum depth of three thousand meters. In 1957, the self-propelled underwater research vehicle (SPURV) became the inaugural AUV (Yang et al., 2021). Over the past six decades, AUV techniques have undergone significant advancements, enabling them to perform sensing tasks autonomously without human intervention (Bai et al., 2018). AUV navigation systems play a vital role in their operation by allowing computers and onboard sensors to govern and guide their movements. However, navigation and remote sensing can be exceedingly challenging due to the attenuation of GPS signals in submerged scenarios. Promising technologies, including cooperative navigation (CN) and SLAM, which can be swiftly implemented and adjusted with minimal infrastructure, are being proposed as potential solutions to this predicament. AUVs typically use batteries, but lithium batteries are now widely used due to their rechargeability and cost-effectiveness (Rymansaib et al., 2023). AUVs can serve as sensing platforms for various sensors, including echo sounders, underwater laser scanners, forward-looking sonars, and conductivity temperature depth sensors.

Ocean engineers are investigating over-actuated and under-actuated underwater vehicles. Over-actuated vehicles align with trajectories using surge, sway, and heave actuators, while under-actuated vehicles pitch and yaw. Tolerable thrust forces, damping

TABLE 2 Typical bandwidth and data rates for underwater channels with different ranges.

Span	Range (km)	Data rate (kbps)	Bandwidth (KHz)
Short Range	<1 km	20 kbps	20-50 KHz
Medium Range	1-10 km	10 kbps	10 KHz
Long Range	10-100 km	1 kbps	2-5 KHz
Basin-Scale	3000 km	10 bps	<1 KHz

limits, and inertia effects limit these models' attitude. Marine vehicles are managed remotely by input, status, and output barriers. A supplementary system and Doppler Indicator (DI) optimization are applied to track a fully-actuated underwater vehicle (He et al., 2022). In the event of a tracking error occurring within a confined area, the vehicle simultaneously moves in both the rightward and forward directions. Through simulations and experiments conducted under three different scenarios, the efficiency of the proposed strategy has been demonstrated with the tracking error confined to a narrow zone (Cao et al., 2022). A novel open-water path planning strategy called UP4O is designed for AUVs operating in challenging water conditions (Yang et al., 2022). The strategy uses an environmental encoder module to bind local obstacle data and combine it with relative position, velocity, and ocean currents, resulting in continuous operational decision-making using local dynamic information. The system has a diverse state space with at least 26 actions, ensuring motion accuracy and minimizing deviations from ocean current vectors (Yan et al., 2014). Experimental results support UP4O's ability to accelerate convergence and provide smoother paths in complex oceanic environments.

The domain of designing control systems for robotic arm systems and underwater vehicles is explained. The main focus is on the mathematical analysis of singular perturbation theory. Two control rules were proposed: one that is more straightforward and partially compensates for the sluggish subsystem and another that is a resilient nonlinear control not influenced by model parameters. The stability of both control rules is demonstrated using perturbation theory, and the performance of the suggested controller in a closed-loop system can be compared to that of a model-based correction (De Wit et al., 2000). Marine robotics has revolutionized the use of remotely operated underwater vehicles (ROVs) in science and industry, enabling humans to perform tasks like moving objects across long distances. The effectiveness of single- and multi-ROV systems depends on the right tracking controller. Issues related to individual ROV tracking include energy efficiency, Lyapunov-based model predictive control, feedback and linearization techniques, adaptive algorithms, proportional-derivative control, area tracking controllers, auto-tracking controller adjustment, multivariate control techniques, high-order adaptive sliding mode controllers, controllers based on models (Yan et al., 2019). Following the Deepwater Horizon disaster in 2010, public attention shifted to monitoring the subsurface sea environment (Vasilijević et al., 2017). Remote sensing technologies have proven effective for terrestrial disasters, but detecting and measuring underwater pollution requires field methods (Hao et al., 2022). A joint robotic system combining autonomous underwater and unmanned autonomous surface vehicles is proposed to rapidly detect/sense and quantify contaminants in the water column *in situ*. This system enables real-time contamination readings while minimizing human intervention and time commitments.

Autonomous underwater navigation relies on efficiency and autonomy, with dead reckoning techniques relying on proprioceptive data from compasses, Doppler Velocity Logs, and Inertial Navigation Systems. However, positioning errors tend to magnify over time, necessitating absolute georeferenced sources for precise positioning. Time-of-flight (ToF) acoustic positioning systems are the current method for correcting underwater positions and sensing (Li et al., 2018). As technology and hydroacoustic communication standards for AUVs continue to advance, CN may become a highly accurate method for locating multiple underwater vehicles. CN allows a group of AUVs to mutually estimate their current positions based on relative distance, velocity, and acceleration (Lyu et al., 2022). Figure 4 illustrates the USN and AUV architecture. The surface components like satellites, drones, ships, base stations, surface sinks, and servers aid communication and data management. AUVs communicate with stationary seabed sensors and other underwater mobile nodes. The Surface and underwater components communicate by two-way packet exchange and via wireless signals, safeguarding effective data collecting and network coordination in diverse underwater environments.

By utilizing ocean currents as control inputs during way-point tracking missions, the power consumption of the engine is reduced. The controller effectively considers multiple constraints, such as those related to the workplace, the vehicle's maximum speed, sensing capacity, the saturation of control inputs, and the presence of rare obstacles. The proposed technology accounts for all the vehicle dynamics, including ocean currents, enabling optimal thrust determination to minimize errors in waypoint tracking (Heshmati-Alamdari et al., 2019). Utilizing the ocean currents for control inputs during waypoint tracking missions reduces the power consumption of the engine. Analytical guarantees for stability and convergence are established for closed-loop systems. The presented work focuses on utilizing AUVs to monitor underwater pipelines and gather data from submarine networks (SNs) within the transmission range. This collected data is transmitted to a surface borehole using acoustic communication technology. This approach reduces power consumption, and the need for costly data re-transmissions is avoided. This architectural framework is well-suited for data applications that can tolerate latency and provide flexibility in implementing submarine networks. Various algorithms for AUV motion and remote sensing are put forth, and an investigation is conducted to determine the impact of the system on network performance. Furthermore, the system can be optimized by considering design parameters, such as SN density, distance, network reliability, medium access control (MAC), communication channel conditions, security measures, and quality of service (QoS) requirements (Jawhar et al., 2018).

Despite development in AUVs and underwater wireless sensor networks (UWSNs), there are still several limitations. The existing hardware-focused networks cannot exchange resources or familiarize themselves with software-defined solutions. Accelerometer drift disturbs navigation system accuracy, necessitating further research. Due to energy efficiency, localization, sensing, and navigation system resilience issues in demanding underwater environments, IoT device installation and conservation must be enhanced. Current challenges include creating long-lasting, water-resistant sensors and efficient data processing in low-network settings. Novel control and path planning systems like UP4O must also be authenticated in complex marine surroundings. The existing problems must be addressed to increase the performance, reliability, and expandability of AUVs and UWSNs. This will enable future advanced, flexible, and effective marine systems.

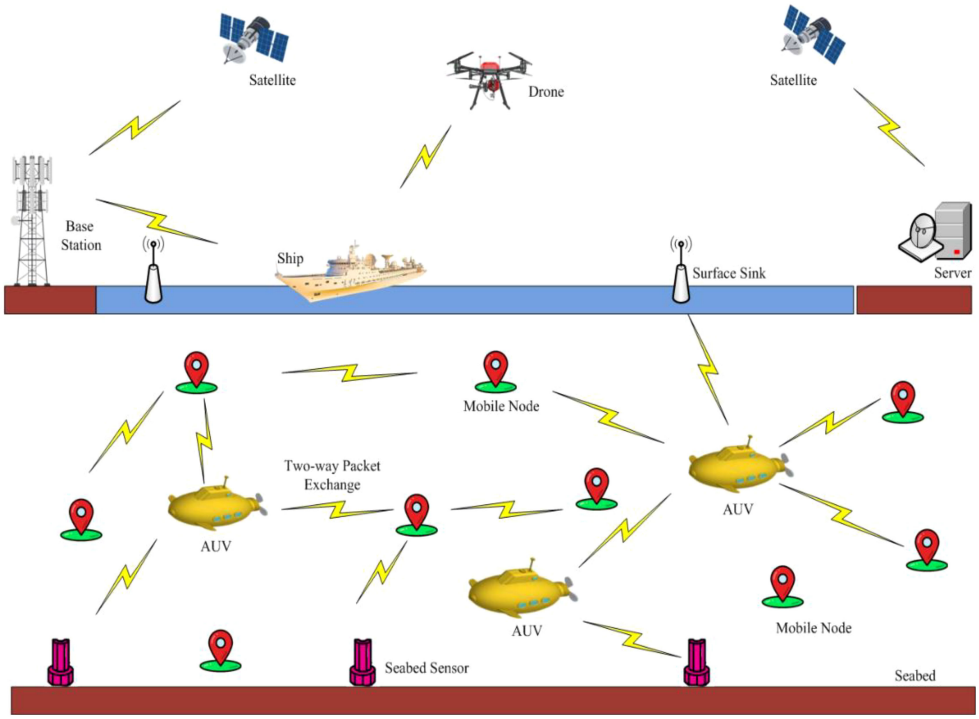


FIGURE 4
Remote sensing communication and navigation in deep-sea environment.

3 Methodology

Multi-criteria optimization is crucial for selecting the most efficient UIoT for underwater task assignment and sensing (Song, 2020; Fang et al., 2021). These devices must be chosen based on energy efficiency, communication range, remote sensing capacity, data transmission capability, durability, and adaptability to dynamic oceanic environments. The effectiveness of IoT devices depends on the specific objectives of underwater operations, such as accurate data collection in oceanography and robust and durable devices for marine remote sensing infrastructure repair work. Multi-criteria optimization helps decision-makers select IoT devices that align with operational goals and requirements. It also addresses trade-offs between features like energy efficiency and data transmission capabilities, ensuring the efficiency and effectiveness of IoT deployments tailored to specific underwater sensing applications. The optimal selection of vehicles based on multi-criteria is shown in Figure 5. Starting with the UIoT, digital library articles are assessed. These articles undergo feature engineering to extract relevant features. The selected features are evaluated for sufficiency. After recognizing significant attributes, ideal vehicles are selected using multi-criteria evaluation.

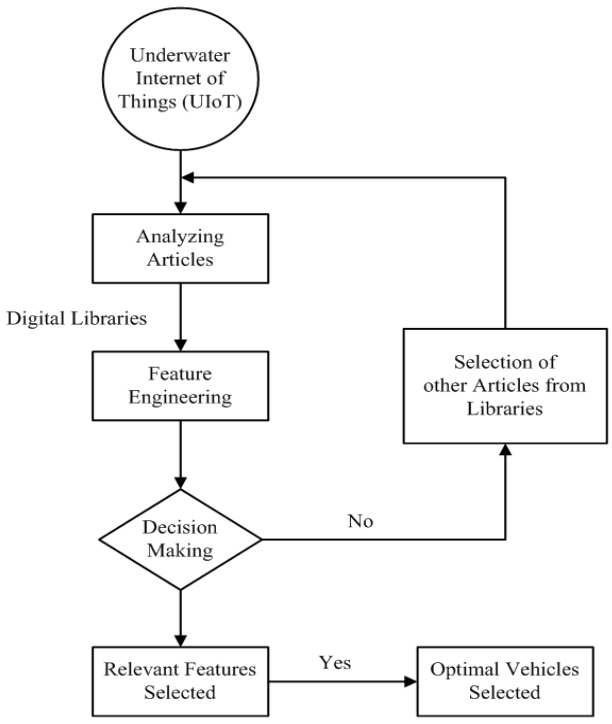


FIGURE 5
Optimal selection of vehicles based on multi-criteria.

3.1 Feature engineering

An important and challenging aspect of the UIoT is to identify the unique characteristics of the underwater sensing environment through an in-depth review of previous research (Zhu et al., 2023). These characteristics include energy efficiency, water pressure,

contact closure, resistance to corrosion, salinity, sensing issues, and temperature. Understanding and specifying these characteristics is important for effectively deploying IoT devices in an oceanic environment, as they directly affect network performance and reliability. Functional engineering is essential for optimizing IoT devices for specific underwater tasks such as oceanography, where data accuracy is critical, or maritime/oceanic infrastructure maintenance, where reliability and durability are priorities. As feature development is versatile, IoT devices may be customized to meet specific remote sensing needs. Functional engineering also promotes using and integrating cutting-edge scientific innovations and findings.

Engineers and researchers may uncover new possibilities and chances that extend the remote sensing potential of IoT devices by evaluating relevant literature and considering the latest developments. In addition to revolutionizing IoT technology, this iterative approach to feature development will produce creative solutions for applications emerging in underwater or oceanic environments. Feature engineering is essential to underwater IoT, enabling decision-makers to select, modify, and design IoT devices tailored to underwater remote sensing applications' needs and objectives. The UIoT is a network of submerged devices and systems that enable efficient communication, cooperation, and data transmission. These resilient and adaptable systems allow aquatic systems to withstand environmental pressures and recover from disturbances. Their flexibility allows them to respond to dynamic changes and gather oceanic research and exploration data.

Interoperability and integration are significant in the IoT, fostering collaboration and data exchange between devices. The dependability and efficiency of underwater technologies ensure consistent service provision, while their durability ensures functionality over time. Tolerance mechanisms and collision avoidance ensure secure sensing, navigation, and operations. Both bound and unbound deployment options allow flexibility in deployment. The visibility of the IoT enables real-time monitoring and visualization of underwater conditions, facilitating data analysis and seabed mapping for scientific research and exploration. These capabilities enable multitasking, responsiveness, and controllable sub-sea systems that provide essential services and data analytics, revolutionizing the way we explore and comprehend the vast depths of the ocean. The various key characteristics of underwater vehicles are illustrated in Figure 6.

3.2 Decision making

Ocean engineering is a rapidly developing field that relies on decision-making to guide undersea technologies. Researchers, producers, and institutions are constantly improving underwater vehicles to sense, navigate, and explore deep oceans (Rolland et al., 2023). These vehicles are designed for security, resource exploitation, remote sensing, ecological protection, and scientific investigation. Their creation ensures wise choices for the submerged environment, preserving marine ecosystems for future generations and allowing exploration of mysterious deep-sea enclaves (Borja et al., 2020). The ultimate goal is to provide creative and effective solutions for the vast, unexplored oceans.

3.3 Feature selection

The UIoT faces challenges in identifying unique underwater characteristics like power usage efficiency, water pressure, sensing issues, contact closure, corrosion resistance, salinity, and temperature (Khalil et al., 2020). These factors directly impact the remote sensing network's performance and reliability. Functional engineering promotes the use of state-of-the-art scientific discoveries and innovations, allowing engineers and researchers to assess the literature and consider recent advancements to uncover novel attributes and capabilities. This iterative approach revolutionizes IoT technology and provides innovative solutions for underwater/oceanic applications. Functional design is at the core of the UIoT, enabling decision-makers to choose, tailor, and design IoT devices to meet underwater remote sensing application requirements.

3.4 Multi-criterial decision making in UIoT

The Internet of vehicles (IoV) presents a challenge for policymakers and stakeholders in selecting optimal vehicle solutions to improve operations and productivity. Multi-criteria decision-making (MCDM) methodologies, such as the analytical hierarchy process (AHP) and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), help assess and prioritize vehicles across various dimensions, including fuel effectiveness, connectivity, environmental impact, safety attributes, and cost efficiency. These approaches provide a structured framework for informed choices, enabling stakeholders to navigate various alternatives, ultimately leading to efficiency, sustainability, and innovation in the transformative sector.

The UIoT aims to revolutionize underwater activities by combining interconnectivity, data sharing, and real-time monitoring. It enhances operational efficiency and predictability by enabling effective task control, route tracking, sensing, and location determination. UIoT's manoeuvrability, adaptability, multitasking, resource management, continuous integration, and efficient telemetry further enhance operational capabilities. It also provides durability, protection, safety, and resilience, promoting real-time monitoring and accountability in harsh environments. The responsiveness of IoT systems supports their reliability and interoperability, enabling seamless functioning and interaction. These capabilities can potentially revolutionize underwater remote sensing operations and usher in a new era of subsurface exploration and data collection. The overall methodology of this study is represented in Section 4, and the characteristics that are collected by properly analyzing existing approaches are presented in Table 3.

4 Results and discussions

The detailed methodology and the evaluation results of the proposed approach are presented in the below subsections.

FIGURE 6
Various key characteristics of underwater vehicles.

4.1 CRITIC approach

A technique utilized in MCDM to help with complex selection procedures where multiple factors require evaluation is called the CRITIC choice-making strategy. It works especially well when assessing and ranking choices or alternatives in situations where there is a lot of ambiguity and mutual dependence. Rather than considering each decision criterion separately, CRITIC focuses on their interrelatedness. Adapting to complicated and evolving decision situations, reducing subjectivity in weight distribution, and understanding hidden linkages between factors are only a few advantages of the CRITIC technique. It provides an organized and systematic manner to rank and incorporate multiple factors during the selection process, making it an extremely valuable tool across a wide range of industries, including sustainability management, engineering, finance, and healthcare. When everything is said and done, the CRITIC choice-making strategy offers a solid and methodical way to handle complicated problems that need a careful examination of criteria connections. These steps involved in the mechanism of the CRITIC calculation are portrayed in Figure 7.

The CRITIC technique assesses every criterion's significance by considering its relationships to other factors in addition to subjective perceptions. This implies that the method considers the possible consequences of altering one of the factors on others, resulting in a more accurate and comprehensive depiction of the decision problem. Decision-makers estimate every criterion's relative importance to all other factors by comparing them pairwise, creating an inter-criteria correlation matrix. After that, the ultimate weights for the criteria are obtained by subjecting this matrix to a number of mathematical operations, many of which include dynamic investigation.

Furthermore, all the chosen factors in the proposed work are beneficial. The chosen factors impact the alternatives more, which we can find after determining their weightage. The weights were assigned to every criterion based on their importance, according to expert opinion, using a scale ranging from one (1) to nine (9). The one value illustrates the equal significance of one factor over another. In contrast, the nine values state the extreme significance of one factor over another while comparing them against each other using the CRITIC approach. A 7 * 7 matrix has been constructed using Equation 1, and weights are distributed among criteria as per expert opinion. These factors have been properly comprehensively compared against each other to

TABLE 3 Multi-criteria-based optimal feature selection of underwater vehicles.

Alternatives		Criteria						
		C1	C2	C3	C4	C5	C 6	C7
Meth	1	Communication	Connectivity	Endurance	Sharing	Collaboration	Transmission	Propagation
Meth	2	Tasks Management	Path Monitoring	Surveillance	Real-time Monitoring	Predictability	Sensing	Data Collection
Meth	3	Positioning	Navigation	Localization	Positioning	Tracking	Detection	Tethered
Meth	4	Oceanography	Exploration	Virtualization	Seafloor mapping	Sampling	Visibility	Deployment
Meth	5	Service Provision	Forecasting	Multi-tasking	Resource Management	Integration	Telemetry	Controllability
Meth	6	Robustness	Protection	Resilience	Security	Tolerance	Durability	Shielding
Meth	7	Real-time Operation	Accountability	Responsiveness	Storage	Processing	Reliability	Interoperability
Meth	8	Adaptiveness	Recovery	Resistance	Collision Avoidance	Elasticity	Rejuvenation	Upgradation

determine the precise weightage of each criterion and determine their significance and impacts on the required alternatives. The evaluation matrix has been designed for eight alternatives based on specific factors. The maximum and minimum values have also been determined from every column, which states that every maximum value is the highest and the minimum is the lowest due to the beneficial nature of all the criteria. The alternatives have been set in rows, while the criteria have been set in columns, as depicted in Table 4.

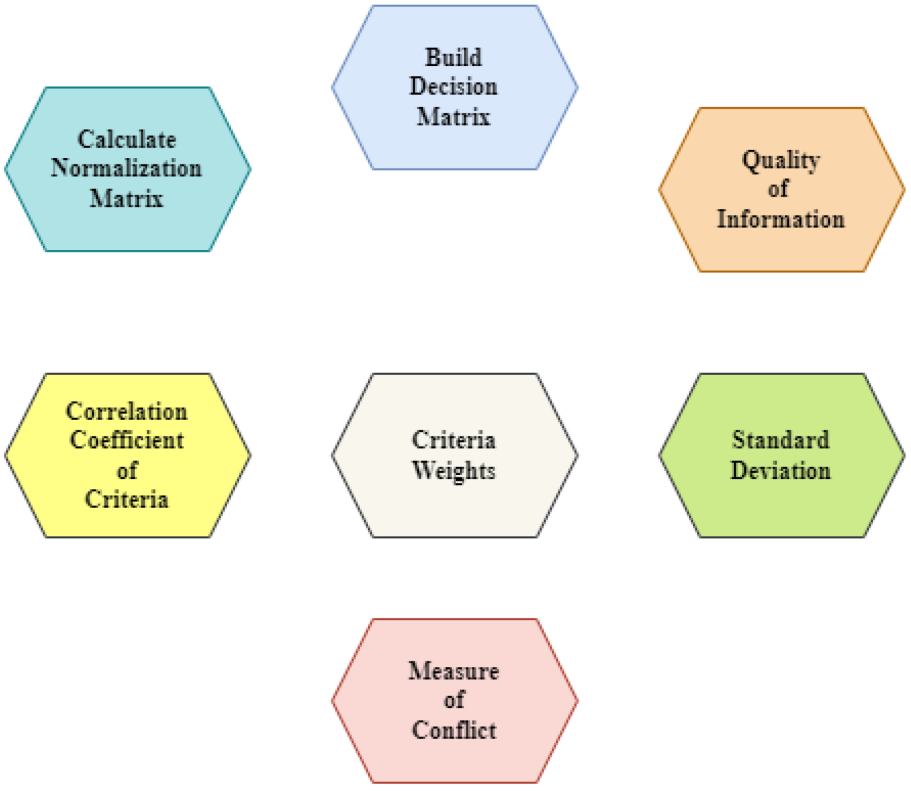


FIGURE 7
Representation of various (CRITIC approach) steps for underwater Things.

$$X = [Xij] \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \\ X_{31} & X_{32} & \dots & X_{3n} \\ X_{41} & X_{42} & \dots & X_{4n} \\ \dots & \dots & . & . \\ \dots & \dots & . & . \\ X_{m1} & X_{m2} & \dots & X_{mn} \end{bmatrix} \quad (1)$$

Here, the original matrix has been normalized through the utilization of Equation 2, as given below;

$$\bar{X}_{ij} = \frac{X_{ij} - \min(X_{ij})}{\max(X_{ij}) - \min(X_{ij})} \quad (2)$$

Where, $\bar{X}_{ij}$ indicates the normalized outputs, and it is essential to realize that normalization does not account for the kind of criterion.

The constructed matrix has been undergone to normalize original values by utilizing Equation 2 to reduce subjectivity and remove errors. The outcomes obtained from the entire calculation of the normalization process are listed in Table 5.

Equation 3 has been applied to the normalized matrix to obtain the standard deviation outputs. The entire calculation and obtained outputs are listed in Table 6.

$$\text{Standard Deviation } (\sigma j) = \sqrt{\frac{\sum_{i=1}^{n}(\chi_i - \overline{\chi})^2}{n-1}} \quad (3)$$

Figure 8 plots the calculated standard deviation values for every criterion. The correlation coefficient outputs have been obtained by comparing two pairs of criteria in the normalized matrix. The required values of the correlation coefficient between pairs of criteria have been identified, as listed in Table 7 and Figure 9.

$$\rho_{jk} = \frac{\sum_{i=1}^{m}(r_{ij} - \acute{r}_j)(r_{ik} - \acute{r}_k)}{\sqrt{(\sum_{i=1}^{m}(r_{ij} - \acute{r}_j)^2 \sum_{i=1}^{m}(r_{ik} - \acute{r}_k)^2}} \quad (4)$$

TABLE 4 Evaluation matrix.

Criteria / Alternatives	C1	C2	C3	C4	C5	C6	C7
Meth 1	3	7	2	5	6	8	4
Meth 2	9	4	6	2	8	3	5
Meth 3	2	5	3	7	4	6	3
Meth 4	5	2	7	3	9	4	6
Meth 5	7	6	2	4	5	2	4
Meth 6	4	3	5	6	2	7	5
Meth 7	6	4	8	2	3	5	2
Meth 8	8	2	4	3	7	6	3
MAX	9	7	8	7	9	8	6
MIN	2	2	2	2	2	2	2

The required values have been achieved through the use of Equation 5. Every correlation coefficient value mentioned above has been subtracted from one, and then these values are added in a row-wise manner to get the required values according to the equation. The calculated outputs of the measure of conflict are outlined in Table 8.

$$\text{Measure of conflict} = \left(\sum_{j'=1}^{n}(1 - r'_{jj})\right) \quad (5)$$

The calculated outcomes, known as the measure of conflict of every criterion, are plotted in Figure 10. The required outputs, known as the quantity of information, have been achieved through the application of (Equation 6). These values are obtained by multiplying the measure of conflict outputs with the standard deviation scores as per the quantity of information formula. The calculated outcomes of the quantity of information are listed in Table 9.

$$\text{Quality of information (Cj)} = \sigma j * \left(\sum_{j'=1}^{n}(\mathbf{1} - \mathbf{r}_{jj}')\right) \quad (6)$$

The calculated scores of the quantity of information have been plotted in Figure 11, which improves visibility and understanding of the calculated outcomes.

According to Equation 7, every single value of the quantity of information has been divided by the total of the values of the quantity of information in order to get the required weights of every criterion to determine the relative importance of the factors and identify their effects on the numerous essential alternatives. The calculated weightage of each criterion in the study is displayed in Table 10.

$$\textbf{Criteria weights } (W_j) = \frac{C_j}{\sum_{j=1}^{n} C_j} \quad (7)$$

The weightage of each criterion calculated by the CRITIC procedure is plotted in graphical form, as shown in Figure 12 to increase the readability and clarity for the user to easily understand the relative importance of numerous essential criteria chosen and evaluated in the study. A criterion with the highest weight indicates a greater significance and high effect on the chosen alternatives, as followed by the remaining criteria in a sequence.

The CRITIC technique will prioritize the primary characteristics of submersible vehicles, enabling decision-makers to optimize vessel deployment. The appropriate vehicles will be selected, and tasks will be assigned to them, leading to higher success rates, enhanced security measures, efficient resource allocation, and improved underwater operations accuracy. The technique is valuable for investigating subsea technologies and enhancing vessel deployment in the undersea IoT.

Comparing the proposed work with previous systems shows notable differences in important performance measures, such as the distribution of trust values, the time it takes for data to go from one end to another, the lifespan of individual nodes, and the time it takes for the system to reach a stable state, see Table 11. The proposed technique showcases the most minimal end-to-end latency of 50 ms, suggesting very efficient data transfer, whereas the other alternatives display the highest delay of 65 ms. The

TABLE 5 Normalized matrix.

	C1	C2	C3	C4	C5	C6	C7
Meth 1	0.142857143	1	0	0.600	0.571	1	0.500
Meth 2	1	0.4	0.66666667	0	0.857	0.16666667	0.75
Meth 3	0.000	0.600	0.16666667	1	0.28571429	0.667	0.250
Meth 4	0.428571429	0.000	0.83333333	0.2	1	0.333	1.000
Meth 5	0.714	0.800	0	0.400	0.429	0.000	0.5
Meth 6	0.286	0.200	0.5	0.800	0.000	0.833	0.750
Meth 7	0.571428571	0.400	1	0	0.143	0.500	0.000
Meth 8	0.857142857	0	0.33333333	0.200	0.714	0.667	0.250

TABLE 6 Calculation of standard deviation.

	C1	C2	C3	C4	C5	C6	C7
Meth 1	0.142857143	1	0	0.600	0.571	1	0.500
Meth 2	1	0.4	0.66666667	0	0.857	0.16666667	0.75
Meth 3	0.000	0.600	0.16666667	1	0.28571429	0.667	0.250
Meth 4	0.428571429	0.000	0.83333333	0.2	1	0.333	1.000
Meth 5	0.714	0.800	0	0.400	0.429	0.000	0.5
Meth 6	0.286	0.200	0.5	0.800	0.000	0.833	0.750
Meth 7	0.571428571	0.400	1	0	0.143	0.500	0.000
Meth 8	0.857142857	0	0.33333333	0.200	0.714	0.667	0.250
Std deviation	0.350	0.362	0.377	0.370	0.350	0.339	0.327

proposed work exhibits the highest trust value distribution (0.95), indicating a greater level of reliability among network nodes. In contrast, the other approach demonstrates the lowest trust value distribution (0.85), implying inferior trust management. The suggested work has the greatest node lifespan, lasting for 200 hours, compared to the other job with a shorter node lifetime of 175 hours. This emphasizes the energy efficiency of the proposed work. In addition, the suggested work demonstrates the fastest convergence time of 30 seconds, which indicates a rapid stabilization of the network. In contrast, the comparison work has

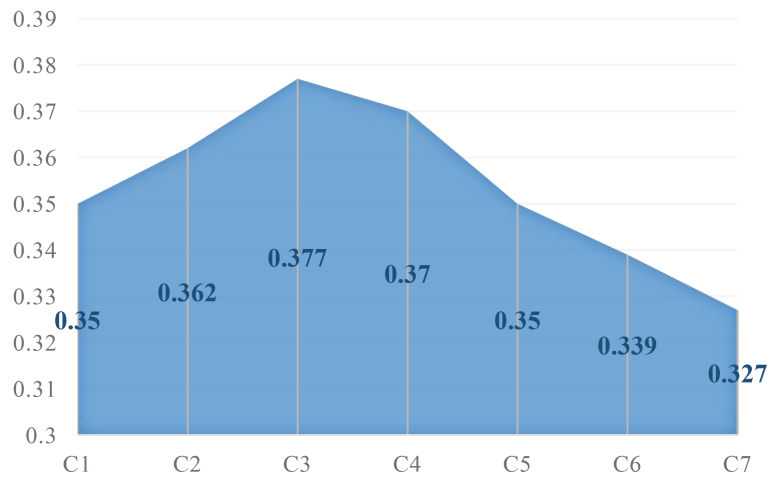


FIGURE 8
Standard deviation outputs.

10.3389/fmars.2024.1468481

TABLE 7 Calculation of correlation coefficient between criteria.

Criteria / Criteria	C1	C2	C3	C4	C5	C6	C7
C1	1.000	-0.339	0.296	-0.819	0.429	-0.646	0.045
C2	-0.339	1.000	-0.650	0.341	-0.242	0.073	-0.241
C3	0.296	-0.650	1.000	-0.613	0.090	-0.206	0.096
C4	-0.819	0.341	-0.613	1.000	-0.504	0.532	0.000
C5	0.429	-0.242	0.090	-0.504	1.000	-0.359	0.490
C6	-0.646	0.073	-0.206	0.532	-0.359	1.000	-0.215
C7	0.045	-0.241	0.096	0.000	0.490	-0.215	1.000

FIGURE 9
Correlation coefficients.

TABLE 8 Calculation of measure of conflict.

	C1	C2	C3	C4	C5	C6	C7	Measure of Conflict
C1	0.000	1.339	0.70373712	1.81892302	0.57142857	1.64609574	0.95545646	7.034
C2	1.339	0	1.650	0.65856832	1.24196696	0.92704422	1.24142866	7.058
C3	0.70373712	1.650159294	0	1.61343836	0.90983304	1.2058396	0.90360746	6.987
C4	1.818923025	0.65856832	1.61343836	0	1.50395263	0.46818398	1	7.063
C5	0.571428571	1.241966959	0.90983304	1.50395263	0	1.35894208	0.51002106	6.096
C6	1.646095738	0.927044217	1.2058396	0.46818398	1.35894208	0	1.21488612	6.821
C7	0.95545646	1.241428656	0.90360746	1	0.51002106	1.21488612	0	5.825

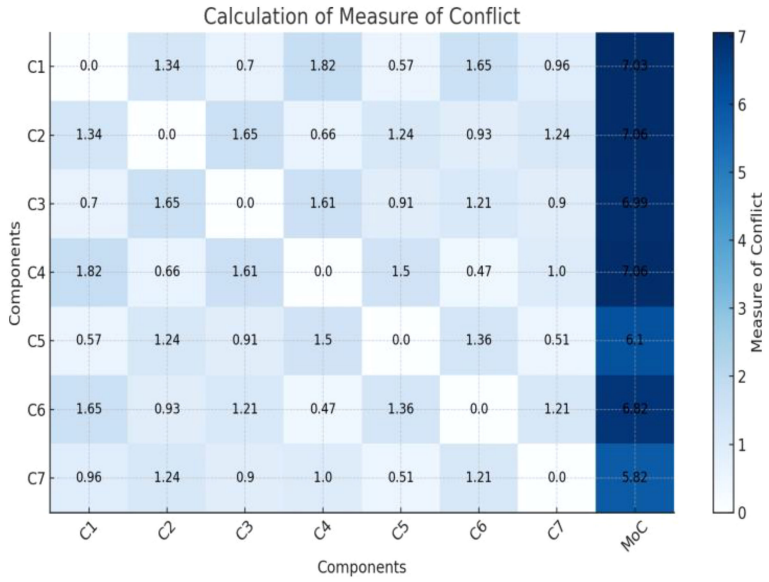


FIGURE 10
Measure of conflict outcomes.

TABLE 9 Calculation of quantity of information.

	Standard Deviation		Measure of Conflict	Quantity of Information (Cj)
C1	0.350		7.034	2.462
C2	0.362		7.058	2.552
C3	0.377		6.987	2.636
C4	0.370	×	7.063	2.616
C5	0.350		6.096	2.133
C6	0.339		6.821	2.309
C7	0.327		5.825	1.907

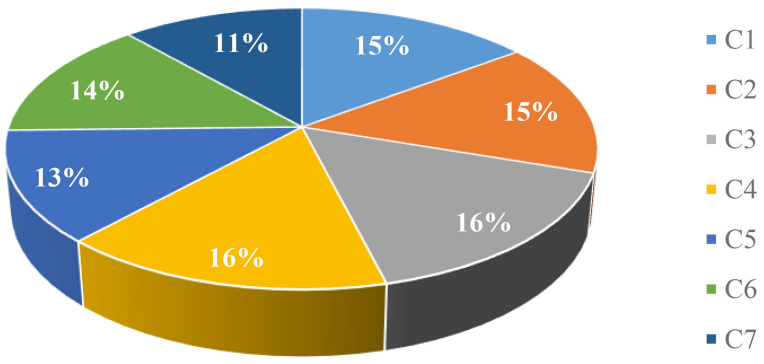


FIGURE 11
Quantity of information values.

TABLE 10 Calculation of weights.

	Quantity of Information	Weights	Weights in Percent (%)
C1	2.462	0.148	14.82%
C2	2.552	0.154	15.36%
C3	2.636	0.159	15.87%
C4	2.616	0.157	15.74%
C5	2.133	0.128	12.84%
C6	2.309	0.139	13.90%
C7	1.907	0.115	11.48%
Sum	16.614	1.000	100%

the slowest convergence time of 37 seconds, highlighting the superior performance of the proposed work. The data demonstrates that the suggested methodology surpasses other methods in all measurable dimensions, emphasizing reducing delay and maximizing trust, node lifetime, and speedy convergence.

5 Conclusion

The impact of 5G and 6G communication networks on underwater technology drives rapid growth in the IoT market. As a result, underwater automobiles, vessels, tracking devices, and surveillance devices, such as s, environmental sensitivity observation equipment and advanced aquatic study instruments, have emerged. These technologies have the potential to transform our understanding of the underwater oceanic environment and contribute to the long-term management of ocean resources. The Internet of underwater vehicles is an exciting breakthrough in the IoT area that is transforming sub-aquatic operations and communications. It has energy-efficient communication modules, quick data processing, flexible sensors, remote sensing capability, and better mobility. The Internet of underwater vehicles enhances sub-aquatic vehicle communication capabilities while also accelerating job completion, resulting in a more productive, automatic, and adaptive sub-aquatic network. These advancements in technology bring up new avenues for subaquatic applications, environmental surveillance, and marine exploration. In this study, the CRITIC technique is proposed to examine and evaluate appropriate UIoT characteristics such as localization, sensing,

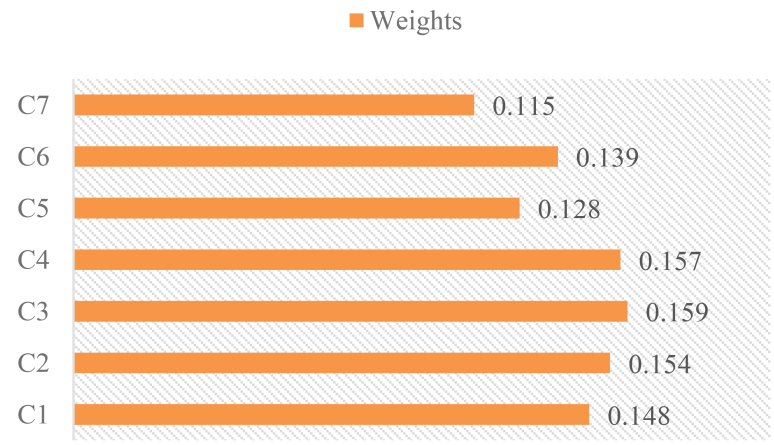


FIGURE 12
Illustration of criteria weights.

TABLE 11 Comparison with other approaches.

Work	End-to-End Delay (ms)	Trust Value Distribution	Node Lifetime (hrs)	Convergence Time (s)
(Jawhar et al., 2018)	62	0.89	195	36
(Song, 2020)	57	0.86	178	34
(Li et al., 2018)	60	0.90	190	32
(Lyu et al., 2022)	65	0.85	175	37
(Jahanbakht et al., 2021)	55	0.88	180	35
(Heshmati-Alamdari et al., 2019)	58	0.87	185	33
Proposed	50	0.95	200	30

positioning security, privacy, resource allocation, and optimization. These multi-characteristics will assist decision-makers in assessing correlations and interdependencies between various characteristics, which is critical for effective and well-informed decision-making in dynamic UIoT environments and for modifying Vehicles with multi-features to achieve specific objectives in undersea operations. Developing an integrated approach to submerged technology, ensuring the effectiveness and safety of underwater remote sensing systems, and developing energy-saving solutions to increase the lifespan of underwater vehicles are various domains that need further research and exploration to achieve the objectives of ocean and marine engineers.

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

IU: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Conceptualization. FA: Writing – review & editing, Writing – original draft, Software, Resources, Methodology. AS: Writing – review & editing, Validation, Resources, Investigation. AA: Writing – review & editing, Resources, Investigation, Formal analysis, Data curation. HN: Writing – review & editing, Visualization, Validation, Resources, Data curation. XB: Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Formal analysis.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported in part by the National Natural Science Foundation of China under Grants 62373255, in part by the Natural Science Foundation of Guangdong Province under Grant 2024A1515011204, in part by the Shenzhen Natural Science Fund through the Stable Support Plan Program under Grant 20220809175803001, and in part by the Open Fund of National Engineering Laboratory for Big Data System Computing Technology under Grant SZU-BDSC-OF2024-15.

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

References

Abelson, A., Reed, D. C., Edgar, G. J., Smith, C. S., Kendrick, G. A., Orth, R. J., et al. (2020). Challenges for restoration of coastal marine ecosystems in the anthropocene. *Front. Mar. Sci.* 7, 544105. doi: 10.3389/fmars.2020.544105

Akyildiz, I. F., Pompili, D., and Melodia, T. (2005). Underwater acoustic sensor networks: research challenges. *Ad. Hoc. Networks* 3, 257–279. doi: 10.1016/j.adhoc.2005.01.004

Arul, R., Alroobaea, R., Mechti, S., Rubaiee, S., Andejany, M., Tariq, U., et al. (2021). Intelligent data analytics in energy optimization for the internet of underwater things. *Soft. Computing.* 25, 12507–12519. doi: 10.1007/s00500-021-06002-x

Bai, X., Yan, W., Ge, S. S., and Cao, M. (2018). An integrated multi-population genetic algorithm for multi-vehicle task assignment in a drift field. *Inf. Sci.* 453, 227–238. doi: 10.1016/j.ins.2018.04.044

Borja, A., Andersen, J. H., Arvanitidis, C. D., Basset, A., Buhl-Mortensen, L., Carvalho, S., et al. (2020). *Past and future grand challenges in marine ecosystem ecology*, Vol. vol. 7. Ed. Frontiers Media SA, Frontiers in Marine Science, Switzerland, 362.

Brasier, M. J., McCormack, S., Bax, N., Caccavo, J. A., Cavan, E., Ericson, J. A., et al. (2020). Overcoming the obstacles faced by early career researchers in marine science: lessons from the marine ecosystem assessment for the Southern Ocean. *Front. Mar. Sci.* 7, 564335. doi: 10.3389/fmars.2020.00692

Cao, B., Li, M., Liu, X., Zhao, J., Cao, W., and Lv, Z. (2021). Many-objective deployment optimization for a drone-assisted camera network. *IEEE Trans. Network. Sci. Eng.* 8, 2756–2764. doi: 10.1109/TNSE.2021.3057915

Cao, X., Ren, L., and Sun, C. (2022). Dynamic target tracking control of autonomous underwater vehicle based on trajectory prediction. *IEEE Trans. Cybernetics.* 53, 1968–1981. doi: 10.1109/TCYB.2022.3189688

Chen, B., Hu, J., Zhao, Y., and Ghosh, B. K. (2022). Finite-time observer based tracking control of uncertain heterogeneous underwater vehicles using adaptive sliding mode approach. *Neurocomputing* 481, 322–332. doi: 10.1016/j.neucom.2022.01.038

De Wit, C. C., Diaz, O. O., and Perrier, M. (2000). Nonlinear control of an underwater vehicle/manipulator with composite dynamics. *IEEE Trans. Control. Syst. Technol.* 8, 948–960. doi: 10.1109/87.880599

Domingo, M. C. (2012). An overview of the internet of underwater things. *J. Network. Comput. Appl.* 35, 1879–1890. doi: 10.1016/j.jnca.2012.07.012

Duan, R., Du, J., Jiang, C., and Ren, Y. (2020). Value-based hierarchical information collection for AUV-enabled Internet of Underwater Things. *IEEE Internet Things. J.* 7, 9870–9883. doi: 10.1109/JIoT.6488907

Fang, Z., Wang, J., Du, J., Hou, X., Ren, Y., and Han, Z. (2021). Stochastic optimization-aided energy-efficient information collection in internet of underwater things networks. *IEEE Internet Things. J.* 9, 1775–1789. doi: 10.1109/JIOT.2021.3088279

Farr, N., Bowen, A., Ware, J., Pontbriand, C., and Tivey, M. (2010). "An integrated, underwater optical/acoustic communications system," in *OCEANS'10 IEEE SYDNEY*, Sydney, NSW, Australia, 2010, pp. 1–6.

Frampton, K. D. (2006). Acoustic self-localization in a distributed sensor network. *IEEE Sensors. J.* 6, 166–172. doi: 10.1109/JSEN.2005.860361

Garcia, M., Sendra, S., Lloret, G., and Lloret, J. (2011). Monitoring and control sensor system for fish feeding in marine fish farms. *IET. Commun.* 5, 1682–1690. doi: 10.1049/iet-com.2010.0654

Gu, Y., Hu, Z., Zhao, Y., Liao, J., and Zhang, W. (2024). MFGTN: A multi-modal fast gated transformer for identifying single trawl marine fishing vessel. *Ocean. Eng.* 303, 117711. doi: 10.1016/j.oceaneng.2024.117711

Hao, Y., Li, X., Chen, B., and Zhu, Z. (2022). Marine monitoring based on triboelectric nanogenerator: Ocean energy harvesting and sensing. *Front. Mar. Sci.* 9, 1038035. doi: 10.3389/fmars.2022.1038035

He, S., Kou, L., Li, Y., and Xiang, J. (2022). Position tracking control of fully-actuated underwater vehicles with constrained attitude and velocities. *IEEE Trans. Ind. Electron.* 69, 13192–13202. doi: 10.1109/TIE.2022.3140516

Heshmati-Alamdari, S., Karras, G. C., Marantos, P., and Kyriakopoulos, K. J. (2019). A robust predictive control approach for underwater robotic vehicles. *IEEE Trans. Control. Syst. Technol.* 28, 2352–2363. doi: 10.1109/TCST.87

Hoeher, P. A., Sticklus, J., and Harlakin, A. (2021). Underwater optical wireless communications in swarm robotics: A tutorial. *IEEE Commun. Surveys. Tutorials.* 23, 2630–2659. doi: 10.1109/COMST.2021.3111984

Jahanbakht, M., Xiang, W., Hanzo, L., and Azghadi, M. R. (2021). Internet of underwater things and big marine data analytics—a comprehensive survey. *IEEE Commun. Surveys. Tutorials.* 23, 904–956. doi: 10.1109/COMST.2021.3053118

Jawhar, I., Mohamed, N., Al-Jaroodi, J., and Zhang, S. (2018). An architecture for using autonomous underwater vehicles in wireless sensor networks for underwater pipeline monitoring. *IEEE Trans. Ind. Inf.* 15, 1329–1340. doi: 10.1109/TII.2018.2848290

Jiang, S. (2018). On reliable data transfer in underwater acoustic networks: A survey from networking perspective. *IEEE Commun. Surveys. Tutorials.* 20, 1036–1055. doi: 10.1109/COMST.2018.2793964

Khalil, R. A., Saeed, N., Babar, M. I., and Jan, T. (2020). Toward the internet of underwater things: Recent developments and future challenges. *IEEE Consumer. Electron. Magazine.* 10, 32–37. doi: 10.1109/MCE.2020.2988441

Khan, S., Ullah, I., Ali, F., Shafiq, M., Ghadi, Y. Y., 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

Li, J., Zhang, G., Jiang, C., and Zhang, W. (2023). A survey of maritime unmanned search system: Theory, applications and future directions. *Ocean. Eng.* 285, 115359. doi: 10.1016/j.oceaneng.2023.115359

Li, Q., Ben, Y., Naqvi, S. M., Neasham, J. A., and Chambers, J. A. (2018). Robust student's $ t $-based cooperative navigation for autonomous underwater vehicles. *IEEE Trans. Instrumentation. Measurement.* 67, 1762–1777. doi: 10.1109/TIM.2018.2809139

Lloret, J., Sendra, S., Garcia, M., and Lloret, G. (2011). "Group-based underwater wireless sensor network for marine fish farms," in *2011 IEEE globecom Workshops (GC Wkshps)*, Houston, TX, USA, 2011, pp. 115–119.

Luo, J., Chen, Y., Wu, M., and Yang, Y. (2021). A survey of routing protocols for underwater wireless sensor networks. *IEEE Commun. Surveys. Tutorials.* 23, 137–160. doi: 10.1109/COMST.9739

Luo, H., Wu, K., Ruby, R., Liang, Y., Guo, Z., and Ni, L. M. (2018). Software-defined architectures and technologies for underwater wireless sensor networks: A survey. *IEEE Commun. Surveys. Tutorials.* 20, 2855–2888. doi: 10.1109/COMST.9739

Lyu, C., Lu, D., Xiong, C., Hu, R., Jin, Y., Wang, J., et al. (2022). Toward a gliding hybrid aerial underwater vehicle: Design, fabrication, and experiments. *J. Field Robotics.* 39, 543–556. doi: 10.1002/rob.22063

Mariani, P., Bachmayer, R., Kosta, S., Pietrosemoli, E., Ardelan, M. V., Connelly, D. P., et al. (2021). Collaborative automation and IoT technologies for coastal ocean observing systems. *Front. Mar. Sci.* 8, 647368. doi: 10.3389/fmars.2021.647368

Moradi, M., Rezazadeh, J., and Ismail, A. S. (2012). A reverse localization scheme for underwater acoustic sensor networks. *Sensors* 12, 4352–4380. doi: 10.3390/s120404352

Refulio-Coronado, S., Lacasse, K., Dalton, T., Humphries, A., Basu, S., Uchida, H., et al. (2021). Coastal and marine socio-ecological systems: A systematic review of the literature. *Front. Mar. Sci.* 8, 648006. doi: 10.3389/fmars.2021.648006

Rolland, E. S., Haji, M. N., and de Weck, O. L. (2023). Autonomous control of a prototype solar-powered offshore autonomous underwater vehicle servicing platform via a low-cost embedded architecture. *J. Field Robotics.* 40, 828–847. doi: 10.1002/rob.22155

Rymansaib, Z., Thomas, B., Treloar, A. A., Metcalfe, B., Wilson, P., and Hunter, A. (2023). A prototype autonomous robot for underwater crime scene investigation and emergency response. *J. Field Robotics.* 40, 983–1002. doi: 10.1002/rob.22164

Sendra, S., Lloret, J., Jimenez, J. M., and Parra, L. (2015). Underwater acoustic modems. *IEEE Sensors. J.* 16, 4063–4071. doi: 10.1109/JSEN.2015.2434890

Song, Y. (2020). Underwater acoustic sensor networks with cost efficiency for internet of underwater things. *IEEE Trans. Ind. Electron.* 68, 1707–1716. doi: 10.1109/TIE.41

Sozer, E. M., Stojanovic, M., and Proakis, J. G. (2000). Underwater acoustic networks. *IEEE J. Oceanic. Eng.* 25, 72–83. doi: 10.1109/48.820738

Sun, P., and Boukerche, A. (2018). Modeling and analysis of coverage degree and target detection for autonomous underwater vehicle-based system. *IEEE Trans. Vehicular. Technol.* 67, 9959–9971. doi: 10.1109/TVT.2018.2864141

Tian, W., Zhao, Y., Hou, R., Dong, M., Ota, K., Zeng, D., et al. (2023). A centralized control-based clustering scheme for energy efficiency in underwater acoustic sensor networks. *IEEE Trans. Green Commun. Networking.* 7, 668–679. doi: 10.1109/TGCN.2023.3249208

Vasilijević, A., Nađ, Đ., Mandić, F., Mišković, N., and Vukić, Z. (2017). Coordinated navigation of surface and underwater marine robotic vehicles for ocean sampling and environmental monitoring. *IEEE/ASME. Trans. Mechatronics.* 22, 1174–1184. doi: 10.1109/TMECH.2017.2684423

Wang, H., Yang, Y., Ye, X., He, Z., and Jiao, P. (2023). Combustion-enabled underwater vehicles (CUVs) in dynamic fluid environment. *J. Field Robotics.* 40, 1054–1068. doi: 10.1002/rob.22167

Wei, X., Guo, H., Wang, X., Wang, X., and Qiu, M. (2021). Reliable data collection techniques in underwater wireless sensor networks: A survey. *IEEE Commun. Surveys. Tutorials.* 24, 404–431. doi: 10.1109/COMST.2021.3134955

Yan, W., Bai, X., Peng, X., Zuo, L., and Dai, J. (2014). "The routing problem of autonomous underwater vehicles in ocean currents," in *OCEANS 2014-TAIPEI*, Taipei, Taiwan, 2014, pp. 1–6.

Yan, J., Gao, J., Yang, X., Luo, X., and Guan, X. (2019). Position tracking control of remotely operated underwater vehicles with communication delay. *IEEE Trans. Control. Syst. Technol.* 28, 2506–2514. doi: 10.1109/TCST.87

Yang, J., Huo, J., Xi, M., He, J., Li, Z., and Song, H. H. (2022). A time-saving path planning scheme for autonomous underwater vehicles with complex underwater conditions. *IEEE Internet Things. J.* 10, 1001–1013. doi: 10.1109/JIOT.2022.3205685

Yang, Y., Xiao, Y., and Li, T. (2021). A survey of autonomous underwater vehicle formation: Performance, formation control, and communication capability. *IEEE Commun. Surveys. Tutorials.* 23, 815–841. doi: 10.1109/COMST.2021.3059998

Zacchini, L., Franchi, M., and Ridolfi, A. (2022). Sensor-driven autonomous underwater inspections: A receding-horizon RRT-based view planning solution for AUVs. *J. Field Robotics.* 39, 499–527. doi: 10.1002/rob.22061

Zhou, C., Liu, M., Zhang, S., Zheng, R., Dong, S., and Liu, Z. (2023). Asynchronous localization for underwater acoustic sensor networks: A continuous control deep reinforcement learning approach. *IEEE Internet Things. J.* 11, 9505–9521. doi: 10.1109/JIOT.2023.3324392

Zhu, Z., Zhou, Y., Wang, R., and Tong, F. (2023). Internet of underwater things infrastructure: A shared underwater acoustic communication layer scheme for real-world underwater acoustic experiments. *IEEE Trans. Aerospace. Electronic. Syst.* 59, 6991–7003. doi: 10.1109/TAES.2023.3281531

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 22 October 2024
DOI 10.3389/fmars.2024.1476425

OPEN ACCESS

EDITED BY
Weimin Huang,
Memorial University of Newfoundland, Canada

REVIEWED BY
Inam Ullah,
Gachon University, Republic of Korea
Zhigang Cao,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE
Alexander Gilerson
gilerson@ccny.cuny.edu

RECEIVED 05 August 2024
ACCEPTED 23 September 2024
PUBLISHED 22 October 2024

CITATION
Gilerson A, Malinowski M, Agagliate J, Herrera-Estrella E, Tzortziou M, Tomlinson MC, Meredith A, Stumpf RP, Ondrusek M, Jiang L and Wang M (2024) Development of VIIRS-OLCI chlorophyll-a product for the coastal estuaries.
Front. Mar. Sci. 11:1476425.
doi: 10.3389/fmars.2024.1476425

Development of VIIRS-OLCI chlorophyll-a product for the coastal estuaries

Alexander Gilerson[1]*, Mateusz Malinowski[1], Jacopo Agagliate[1], Eder Herrera-Estrella[1], Maria Tzortziou[2], Michelle C. Tomlinson[3], Andrew Meredith[4], Richard P. Stumpf[3], Michael Ondrusek[5], Lide Jiang[5,6] and Menghua Wang[5]

[1]Optical Remote Sensing Laboratory, The City College of New York, CUNY, New York, NY, United States, [2]The City College of New York, CUNY, New York, NY, United States, [3]National Centers for Coastal Ocean Science, NOAA, Silver Spring, MD, United States, [4]Consolidated Safety Services, Inc., Fairfax, VA, United States, [5]NOAA NESDIS Center for Satellite Applications and Research, College Park, MD, United States, [6]CIRA at Colorado State University, Fort Collins, CO, United States

Coastal waters require monitoring of chlorophyll-a concentration (Chl-a) in a wide range of Chl-a from a few mg/m^3 to hundreds of mg/m^3, which is of interest to the fisheries industry, evaluation of climate change effects, ecological modeling and detection of Harmful Algal Blooms (HABs). Monitoring can be carried out from the Visible Infrared Imaging Radiometer Suite (VIIRS) and Ocean and Land Colour Instrument (OLCI) Ocean Color (OC) satellite sensors, which are currently on orbit and are expected to be the main operational OC sensors at least for the next decade. A Neural Network (NN) algorithm, which uses VIIRS M3-M5 reflectance bands and an I1 imaging band, was developed to estimate Chl-a in the Chesapeake Bay, for the whole range of Chl-a from clear waters in the Lower Bay to extreme bloom conditions in the Upper Bay and the Potomac River, where Chl-a can be used for bloom detection. The NN algorithm demonstrated a significant improvement in the Chl-a retrieval capabilities in comparison with other algorithms, which utilize only reflectance bands. OLCI NIR/red 709/665 nm bands red edge 2010 algorithm denoted as RE10 was also explored with several atmospheric corrections from EUMETSAT, NOAA and NASA. Good consistency between the two types of algorithms is shown for the bloom conditions and the whole range of waters in the Chesapeake Bay (with RE10 switch to OC4 for lower Chl-a) and these algorithms are recommended for the combined VIIRS-OLCI product for the estimation of Chl-a and bloom monitoring. The algorithms were expanded to the waters in Long Island Sound, demonstrating good performance.

KEYWORDS

chlorophyll-a concentration, coastal waters, neural network, VIIRS, OLCI

1 Introduction

In estuaries and adjacent coastal waters, algal blooms are both a key water quality indicator and a potential hazard (Tango and Batiuk, 2016; Karlson et al., 2021). High biomass blooms have implications for reducing water clarity and are indicators of eutrophication in coastal systems (Bricker et al., 2008; Le et al., 2013). Reduced water clarity and depleted oxygen in bottom waters can have deleterious effects on essential fish habitats such as submerged aquatic vegetation in estuaries, leading to a shift from benthic to pelagic-dominated system productivity (Bricker et al., 2008). Harmful Algal Blooms (HABs), pertaining to a class of phytoplankton that often contain toxins, occur in various coastal areas and have a strong impact on fisheries, tourism, and recreation industries, requiring improved monitoring of HABs by environmental and health programs. HABs are often difficult to locate through routine monitoring programs because of their patchiness, physical circulation of the water, and vertical migration of algal particles. As a first approximation, typically, the concentration of chlorophyll-a (Chl-a) is considered as a proxy for the strength of the algal bloom, while bloom effects can vary depending on the type of algal species (IOCCG, 2021). Satellites can support the monitoring of HABs if they provide frequent coverage and retrieve Chl-a over a wide range of concentrations. Improved temporal resolution, which could be provided by using remote sensing products from multiple satellite sensors, can improve efforts of monitoring and forecasting HABs in coastal and estuarine waters. Data should come from multiple ocean color sensors to improve coverage during periods of cloud cover or sun glint (a problem especially in spring and summer), and to provide multiple views of blooms within a day.

Ocean color algorithms are based on remote sensing reflectance, R_{rs} spectra with Chl-a dominating R_{rs} spectra in the blue in clear waters. These algorithms often fail in optically complex coastal and estuarine waters where HABs occur, due to the high absorption of colored dissolved organic matter (CDOM) and scattering from sediments. Therefore, it is important to develop Chl-a algorithms that are minimally influenced by CDOM and/or high sediment concentrations. Efforts have been made to improve Chl-a retrievals from the Visible Infrared Imaging Radiometer Suite (VIIRS) sensors operated by National Oceanic and Atmospheric Administration (NOAA) and from Sentinel-3 Ocean and Land Colour Instrument (OLCI) sensors processed by NOAA in collaboration with the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) (Wang and Son, 2016; Mikelsons and Wang, 2019; Liu and Wang, 2022; Mikelsons et al., 2022; Wynne et al., 2022). Currently, there are three VIIRS sensors (on the SNPP, NOAA-20 and NOAA-21 platforms) and two OLCI sensors on the Sentinel-3A and 3B in space with 750 m and 300 m spatial resolution (at nadir), respectively. With additional launches of VIIRS planned, these two groups of sensors are expected to provide reliable and stable multi-spectral Ocean Color (OC) data for the next decade and beyond. The NASA Phytoplankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission (Werdell et al., 2019), which was successfully launched in February 2024, has a main hyperspectral Ocean Color Instrument (OCI), but with a relatively coarse spatial resolution of 1.0 km (at nadir), which is not often sufficient for many coastal areas. The availability of consistent data in a wide range of Chl-a, with appropriate temporal resolution, will expand the number of applications and agencies, which utilize remote sensing data to complement the field data they use for decision-making regarding HABs. While the definition of HABs can be different for different water bodies, for this work we only consider high biomass blooms with Chl-a above 25-30 mg/m^3, which require the attention of coastal managers. This does not imply anything related to toxicity or deleterious effects to wildlife or public health and relies on *in situ* sampling to determine phytoplankton species.

Large uncertainties in remote sensing reflectance (R_{rs}) retrieval in blue bands remain a major problem for OC satellite sensors in coastal areas because of difficulties in atmospheric correction and low R_{rs} at this part of the spectrum (Ransibrahmanakul and Stumpf, 2006; IOCCG, 2019). In addition, due to the inability to see through clouds with OC sensors, daily imagery from current satellite sensors may be obscured. When monitoring blooms in coastal areas the combination of insufficient atmospheric correction in coastal and estuarine waters, and missing imagery due to clouds and sun glint, can often hinder the use of satellites in monitoring and forecasting efforts. Large uncertainties make an estimation of Chl-a concentration unreliable using standard OCx algorithms, which include the 443 nm band. A Neural Network (NN) Chl-a algorithm (Ioannou et al., 2014), which avoids blue bands at 412 and 443 nm for VIIRS demonstrated good performance in variable water areas (El-Habashi et al., 2016, 2017, 2019). Specifically, based on field measurements and matchups with satellite data, it has been shown that the NN Chl-a algorithm is valuable for the detection of *Karenia brevis* (KB) algal blooms near the West Florida coast (El-Habashi et al., 2016, 2017). The algorithm performs similarly to the standard OCx algorithms in the open ocean and coastal waters for Chl-a < 10 mg/m^3 (El-Habashi et al., 2019), but usually cannot detect accurately for Chl-a > 10-15 mg/m^3. A near-infrared (NIR)/red Chl-a algorithm applied to the bands available on MEdium Resolution Imaging Spectrometer (MERIS) and OLCI sensors performs well at Chl-a > 5 mg/m^3 in the field (Stumpf and Tyler, 1988; Gitelson, 1992; Moses et al., 2009; Gilerson et al., 2010; Smith et al., 2018; Neil et al., 2020). Unfortunately, applying the NIR/red algorithm to VIIRS is impossible, since it lacks a 709 nm band. A special AC has been developed by the NOAA's National Centers for Coastal Ocean Science (NCCOS) group for OLCI and has been applied to top-of-atmosphere reflectance corrected for molecular scattering (Wynne et al., 2018). Thus, an accurate estimation of high Chl-a values remains elusive from VIIRS and even from other multi-spectral sensors with a richer set of bands.

In addition to M1-M5 bands in the visible, VIIRS sensors have an imaging band I1 which integrates radiances from 600 to 680 nm, centered around 640 nm with an almost rectangular spectral transmission function. Utilization of this band on VIIRS opens additional possibilities. This 640 nm band covers R_{rs} features related to the increase of specific phytoplankton absorption from small values at 600 nm to high at 675 nm and thus can be sensitive to high Chl-a. This band as 638_ag (aggregated to 750 m spatial resolution as all M reflective bands) on SNPP VIIRS and as 642_ag on NOAA-20 was added to the images using the Multi-Sensor Level-1 to Level-2 (MSL12) data processing system (Wang and Jiang, 2018)

and distributedthrough the NOAA CoastWatch (https://coastwatch.noaa.gov/).

NN and Machine learning algorithms are based on the training of large datasets of synthetic, field, or satellite data and have recently been developed to estimate Chl-a and other water parameters on the global and regional scales (Hieronymi et al., 2017; Pahlevan et al., 2020; Liu and Wang, 2022; Werther et al., 2022; Cao et al., 2024). Their performance also depends on the applied atmospheric correction (Hieronymi et al., 2017).

The Chesapeake Bay and Long Island Sound (LIS) are large US estuaries on the US East Coast, where Chl-a needs to be monitored synoptically due to the often-occurring algal blooms and hypoxia events (Aurin et al., 2010; Wolny et al., 2020; Wynne et al., 2022). They are highly variable environments. Algal blooms are patchy and small-scale changes in Chl-a occur rapidly, making synoptic measurements essential to resolve phytoplankton biomass (Anderson and Taylor, 2001; Harding et al., 2005). While well-established monitoring programs, such as the Chesapeake Bay Program, Save the Sound, and state-lead monitoring provide routine monthly sampling at select stations, daily synoptic satellite Chl-a covering the entire estuary provide a better estimate of biomass and capture transient blooms, often missed by routine sampling.

Multiple studies characterized well water optical properties in these estuaries from field measurements and satellite observations (Stumpf and Pennock, 1989; Magnuson et al., 2004; Tzortziou et al., 2006; Aurin et al., 2010; Shi and Wang, 2013; Zheng et al., 2015; Turner et al., 2022; Menendez and Tzortziou, 2024), atmospheric correction algorithms have been assessed (Windle et al., 2022; Sherman et al., 2023; Cao and Tzortziou, 2024) and algorithms for the retrieval of Chl-a were developed (Gitelson et al., 2007; Le et al., 2013; Freitas and Dierssen, 2019; Sherman et al., 2023) for the specific sensors in these waters beyond standard OC3 and OC4 algorithms (O'Reilly et al., 1998, 2019).

The goal of this work is to extend the previously developed VIIRS NN-Chl-a algorithm for higher Chl-a by including the I1 imaging band data (600-680 nm) on VIIRS, investigate different processing schemes for the optimal use of the NIR/red (red edge) algorithm (Gilerson et al., 2010) on OLCI and develop a field validated combined OLCI-VIIRS products to improve detection and surveillance of algal blooms in complex estuarine waters such as the Chesapeake Bay and Long Island Sound. A more reliable estimation of Chl-a over the range seen along the U.S. East Coast is expected to enhance satellite coverage to improve ecological models, fisheries applications, and provide early and reliable detection of various blooms to support coastal managers in aiding aquaculture activities and protecting public health.

OLCI passes the US East Coast around 10 am EST and VIIRS around 1:30 pm EST. Data from several sensors increase coverage, however, the benefits are beyond simple statistics because bloom conditions can change in several hours with changes in tide conditions and biological processes. Multiple observations per day were the main incentive for the launch of GOCI sensors, and the development of geostationary GLIMR and Geo-XO sensors (Schaeffer et al., 2023). The product described in this paper creates a capability that would allow an approximation of the multi-scene capability offered by the geostationary satellites.

In Section 2, the bio-optical model is discussed, which is used for the generation of a large dataset for NN training and testing, different NN approaches are evaluated, and the development of the NN Chl-a algorithm for VIIRS based on M3-M5 reflectance bands and I1 imaging band is described. In Section 3, validation of the NN algorithm on field and satellite data is provided together with the comparison of VIIRS NN and OLCI RE10 algorithms for a broad range of conditions with different OLCI atmospheric correction processing schemes, and expansion of the NN-OLCI product to LIS, validation on field data. A discussion and conclusions are in Section 4.

2 Materials and methods

2.1 Field data

Field data, which were used in bio-optical modeling, comparisons of modeled and field Chl-a and other parameters, included data from several Chesapeake Bay cruises. A very comprehensive dataset was acquired by the CCNY-NOAA group in August 2013 at 43 stations, which included Chl-a, inherent optical properties (IOPs) and reflectance spectra. Attenuation and absorption of water and CDOM spectra were measured by the ac-s instrument; backscattering at 5 wavelengths was measured by the bb-9 instrument, both included in the WETLABS (Philomath, OR) package. At each station, upwelling radiance $L_u(\lambda,0^-)$ was measured using a fiber bundle placed just beneath the water surface and connected to a GER spectroradiometer (SpectraVista, NY). The downwelling radiance above the surface $L_d(\lambda,0+)$ was measured by pointing the same probe bundle onto a Spectralon plate and the downwelling irradiance was determined as $E_d(\lambda,0+) = A\cdot\pi\cdot L_d(\lambda,0+)$, where A = 0.99 is the reflectance factor of the Spectralon plate (Labsphere, NH), constant for the spectrum in the range of wavelengths from 400 to 800 nm. The underwater remote sensing reflectance R_{rs}- is then calculated as $L_u(\lambda,0^-)/E_d(\lambda,0^-)$ sr^{-1}, which was adjusted for the propagation through the water-air interface to calculate above surface R_{rs}. Chl-a from the samples that were collected during the field campaign were determined according to NASA protocol for fluorometric Chl-a determination (Ocean Optics Protocols, 2003).

Capturing the timing and location of a bloom is difficult, and often missing in routine monitoring datasets. An opportunistic sampling event occurred on May 18, 2021, during a high biomass (reaching up to 50 million cells/L) bloom of *Prorocentrum minimum*. R_{rs} and water samples for Chl-a were collected at 5 stations in the Upper Bay. Chl-a concentrations were measured in the range of 73–161 mg/m^3 and coincided in time with VIIRS and OLCI overpasses. Additional R_{rs} spectra and Chl-a were acquired in the Chesapeake Bay sporadically from 2014–2016, capturing a range of Chl-a of 9–48 mg/m^3. In all these measurements, R_{rs} were determined from below water HyperOCR depth profiles.

There was also a large dataset of NCCOS R_{rs} measurements but without corresponding Chl-a. All four R_{rs} datasets are shown below in Figure 1 in the discussion of the bio-optical model. Ranges of many parameters, necessary for the model and absorption spectra were

taken from the previous Chesapeake Bay cruises (Magnuson et al., 2004) and NASA bio-Optical Marine Algorithm Dataset (NOMAD) database (Werdell and Bailey, 2005). Finally, Chl-a data from the Chesapeake Bay program (https://www.chesapeakebay.net/) at multiple stations were used for the validation of the satellite and *in-situ* data, where most of Chl-a fell below 20 mg/m^3.

Concurrent water samples to extract Chl-a and hyperspectral R_{rs} were collected throughout LIS in 2018–2022 in collaboration with the Connecticut Department of Environmental Protection (CTDEEP) (Turner et al., 2022; Sherman et al., 2023). Additional data were collected from small boats. Hyperspectral R_{rs} were measured using a HR512-I spectroradiometer (SpectraVista, NY).

2.2 Satellite data and processing schemes

2.2.1 VIIRS data

The Level-2 science-quality data for SNPP VIIRS and near-real-time (NRT) for NOAA-20 VIIRS with the MSL12 processing were obtained from the NOAA CoastWatch site, featuring a pixel resolution of 750 meters at the nadir. This dataset included normalized water-leaving radiance spectra $nL_w(\lambda)$, which were converted to remote sensing reflectance, $R_{rs}(\lambda)$, across visible wavelengths at 410, 443, 486, 551, 638, and 671 nm on SNPP VIIRS, and 411, 445, 489, 556, 642, and 667 nm on NOAA-20 VIIRS, and Level-2 quality flags. Flag exclusion criteria were applied to pixels meeting any of the following conditions: land, cloud, sea ice, atmospheric correction failure, stray light (except for LISCO), bad navigation quality, high or moderate glint, viewing angles exceeding 60°, and solar zenith angles exceeding 70°. Selection of files required at least > 50% valid pixels in a given set, i.e., to be free of flagged conditions. Additionally, pixels with negative water-leaving radiance were excluded from averaging. In matchups of satellite to *in-situ* data from 1 pixel closest to *in-situ* measurements was considered and a 3×3-pixel grid box (2250 m × 2250 m) centered at the AERONET-OC site for the comparison with AERONET-OC data (Hlaing et al., 2013; Gilerson et al., 2022). The average $R_{rs}(\lambda)$ and standard deviation (STD) between pixels, along with their geometric and radiometric properties, were recorded. The bidirectional reflectance distribution function (BRDF) have been applied to the MSL12-derived VIIRS ocean color data as well as to OLCI data with MSL12 processing (Gordon, 2005; Wang, 2006; IOCCG, 2010).

2.2.2 OLCI data

The OLCI S3A and S3B Level-2 full-resolution data with 300-meter spatial resolution per pixel (EUMETSAT, 2021; Mikelsons et al., 2022) with the Operational Baseline Collection-3 (OBC-3) processing (Zibordi et al., 2022) were acquired from the NOAA CoastWatch website (https://coastwatch.noaa.gov/cwn/index.html), focusing on the Chesapeake Bay area and Long Island Sound. Each Level-2 file encompasses various geophysical products related to the atmosphere and ocean, including aerosol optical thickness, Angstrom exponent at 865 nm, water-leaving reflectance at 413, 443, 490, 560, 665, 681, and 709 nm, sensor zenith angle, solar zenith angle, and quality flags. The remote sensing reflectance, R_{rs} (λ), is computed by dividing the reflectance spectra by π.

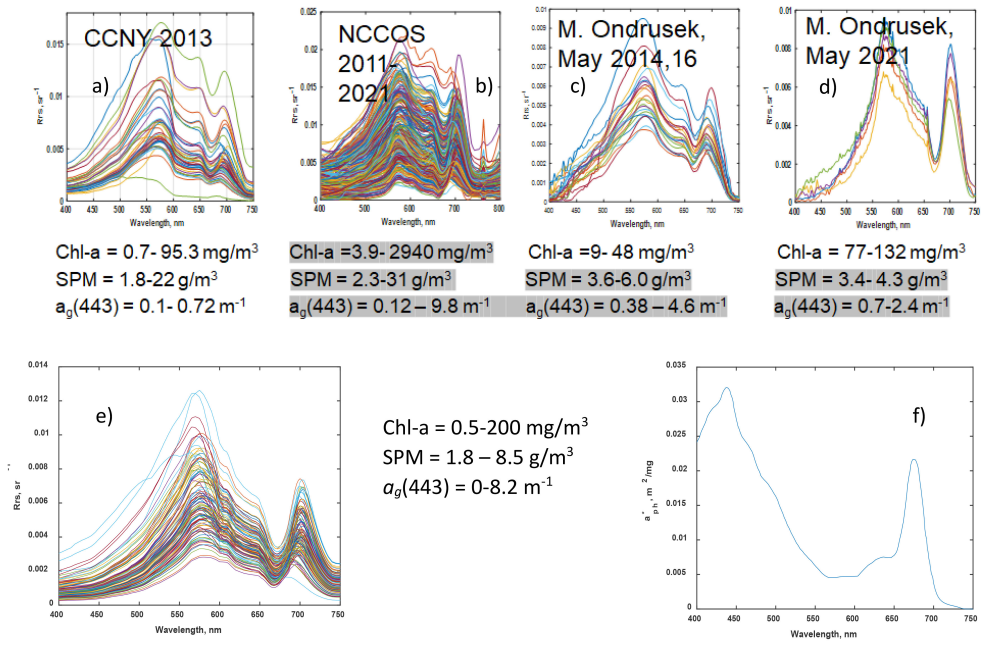


FIGURE 1
Available field R_{rs} datasets **(A–D)** and examples of simulated datasets **(E)** with main water parameters; based specific phytoplankton absorption $a^*_{ph}(\lambda)$ (Magnuson et al., 2004) used in the model **(F)**. Unshaded parameters are measured **(A, C, D)** or simulated **(E)**. Shaded parameters are estimated from different algorithms.

OLCI Level-2 operational water reflectance products do not include BRDF correction. This omission is due to historical usage patterns, with primary interest focusing on coastal and inland waters where the standard open-ocean BRDF approach is not applicable. Mikelsons et al. (2022) showed that there are some significant BRDF effects, both on the surface (Gordon, 2005; Wang, 2006) and in water BRDFs (IOCCG, 2010), over open oceans. However, because there are no established BRDF correction algorithms for a wide range of coastal waters considered in this work, BRDF correction was not applied.

Pixels flagged under any of the following conditions were excluded: invalid flag, land, cloud (including ambiguous and marginal), coastline, solar zenith angle exceeding 70°, saturated flag, moderate or high glint, whitecaps, and failed atmospheric correction. It is important to note that this flag set differs slightly from the set recommended by EUMETSAT for OLCI (EUMETSAT, 2022). A file was selected if at least 50% of valid pixels in the set were free of flags. As for VIIRS, a comparison with measured *in-situ* Chl-a was carried out for 1 closest pixel and for comparison with AERONET-OC 7×7 (2100 m × 2100 m) pixel box was considered.

R_{rs} uncertainties from OLCI in the blue part of the spectra in EUMETSAT atmospheric correction processing are higher than those from VIIRS, especially in coastal waters (Zibordi et al., 2022; Mikelsons et al., 2022; Gilerson et al., 2023). NOAA NCCOS considered a special atmospheric correction (Wynne et al., 2018) using SeaDAS and the subtraction of the Rayleigh component from the TOA radiance. Later, OLCI TOA data were processed using NOAA MSL12 and NASA l2gen algorithm. All these processing schemes were considered with a focus on $R_{rs}(\lambda)$ at the red/NIR bands necessary to apply the RE10 algorithm for the detection of algal blooms.

2.3 AERONET-OC data

Remote sensing reflectance (R_{rs}) for VIIRS and OLCI satellite sensors were assessed through comparisons with SeaPRISM instrument (CIMEL Electronique, France) data at the Chesapeake Bay and Long Island Sound (LISCO) stations, where SeaPRISM radiometers are deployed on offshore fixed platforms and are part of AERONET-OC network (Zibordi et al., 2009, 2021). Normalized water-leaving radiances, $nL_w(\lambda)$, following AERONET-OC protocols and incorporating BRDF correction based on open ocean approaches (Zibordi et al., 2009, 2021), were acquired from the AERONET-OC website for the designated sites. These radiances were transformed into remote sensing reflectance at specific wavelengths. The Long Island Sound Coastal Observatory (LISCO) site (Harmel et al., 2011) upgraded its sensor head in August 2021 to match OLCI sensors with bands at 412, 443, 490, 510, 560, 620, 667, 681, and 709 nm, for detailed comparisons with OLCI data.

The ocean color data employed in this analysis were derived from version 3 level 1.5 data, which underwent cloud screening and quality control measures to ensure data accuracy. All satellite-to-*in situ* matchups were conducted within a ±2-hour window around the satellite overpass time (Zibordi et al., 2009, 2021).

2.4 Bio-optical model

To develop the NN algorithm, datasets, which connect Chl-a, IOPs and $R_{rs}(\lambda)$, water reflectance spectra were simulated based on the bio-optical model (Gilerson and Huot, 2017) with $R_{rs}(\lambda)$ including the sum of elastic $R_{rs}^{e}(\lambda)$ component and fluorescence component $R_{rs}^{f}(\lambda)$; the latter was included because it is a part of the reflectance detected by the broad I1 600-680 nm band. $R_{rs}(\lambda)$ spectra were simulated with 1 nm resolution in the range of 400–750 nm. The maximum of the peak of the fluorescence emission was assumed at 685 nm, fluorescence quantum yield was assumed 1%; the spectral shape of fluorescence was modeled as a Gaussian spectral profile centered at 685 nm, having a full width at half maximum (FWHM) of 25 nm (Mobley, 1994; Gower et al., 2004).

Above water elastic $R_{rs}^{e}(\lambda)$ was calculated following Lee et al. (2002):

$$R_{rs}^{e}(\lambda) = 0.52\frac{R_{rs}^{-}(\lambda)}{(1 - 1.7R_{rs}^{-}(\lambda))} \quad (1)$$

where $R_{rs}^{-}(\lambda)$ is the remote sensing reflectance due to elastic scattering just below the surface, which is calculated as:

$$R_{rs}^{-}(\lambda) = g_1 u(\lambda)^2 + g_2 u(\lambda), \quad (2)$$

$$u(\lambda) = b_b(\lambda)/(a(\lambda) + b_b(\lambda)) \quad (3)$$

where $a(\lambda)$ (m^{-1}) and $b_b(\lambda)$ (m^{-1}) are the total absorption and backscattering coefficient spectra, respectively. Broadly used empirically derived parameters (Lee et al., 2009) $g_1 = 0.125$ and $g_2 = 0.089$, which work well for moderate open ocean and coastal waters were replaced with $g_1 = 0.23$ and $g_2 = 0.089$ equivalent to the relationship based on our previous studies for a broader range of water parameters (Gilerson et al., 2007, 2015).

The total spectral absorption coefficient, $a(\lambda)$, is modeled as

$$a(\lambda) = a_w(\lambda) + a_{ph}(\lambda) + a_g(\lambda) + a_{NAP}(\lambda), \quad (4)$$

where the water absorption spectrum $a_w(\lambda)$ was obtained from (Pope and Fry, 1997).

In coastal waters, $a_{ph}(443)$, $a_g(443)$ and $a_{NAP}(443)$ typically have some correlation (even often weak) with each other (IOCCG, 2006). Based on the data from the NOMAD Chesapeake Bay field campaigns (Gilerson et al., 2015) and $a_{ph}^{*}(443)$ spectra in the Upper Chesapeake Bay (Magnuson et al., 2004) the following relationships at 443 nm were used in the model:

$$a_{ph}(443) = a_{ph}^{*}(443)Chl\text{-}a = 0.031Chl\text{-}a^{-0.12}Chl\text{-}a$$
$$= 0.031Chl\text{-}a^{0.88} \text{ for } Chl\text{-}a < 60 \text{ mg/m}^3 \quad (5a)$$

$$a_{ph}(443) = a_{ph}^{*}(443)Chl\text{-}a = 0.019Chl\text{-}a \text{ for } Chl\text{-}a > 60 \text{ mg/m}^3 \quad (5b)$$

$$a_g(443) = 1.1a_{ph}(443) \quad (6)$$

$$a_{NAP}(443) = 1.32 \times 0.04Chl\text{-}a^{0.65} \quad (7)$$

According to Equation 5, $a^*_{ph}(443)$ gradually decreases with Chl-a and remains constant after 60 mg/m^3. $a_g(443)$ mostly followed $a_{ph}(443)$ and $a_{NAP}(443)$ increases with Chl-a, but less fast than Chl-a itself.

Chl-a were randomly distributed between 0.5 and 200 mg/m^3. The spectral phytoplankton absorption coefficient was obtained by multiplying the Chl-a by a specific absorption coefficient ($a^*_{ph}(\lambda)$, m^2 mg^{-1}),

$$a_{ph}(\lambda) = Chl\text{-}a \times a^*_{ph}(\lambda). \quad (8)$$

The choice of $a^*_{ph}(\lambda)$ strongly influences the corresponding remote sensing reflectance and the emission of fluorescence and was modeled as the specific phytoplankton absorption coefficient in the Upper Chesapeake Bay (Magnuson et al., 2004), shown in Figure 1F with a gradual decrease with increasing Chl-a consistent with Equation 5.

To simulate natural variability, $a^*_{ph}(443)$ were multiplied by a random number drawn from a normal distribution ($N(\mu, \sigma^2)$) with a mean μ=1 and a variance σ^2=0.04: $X_1 \sim N(1, 0.04)$. In a similar manner, $a_g(443)$ and $a_{NAP}(443)$ in Equations 6 and 7 were multiplied by $X_2 \sim N(1, 0.09)$. The ranges of variability here and below were based primarily on the published values from IOCCG (2006), NOMAD and the authors' data for the Chesapeake Bay (Gilerson et al., 2015).

The spectral absorption coefficients of both CDOM and NAP were modeled as having an exponentially decreasing shape with wavelength and referenced to 443 nm (Bukata et al., 1995; Stramski et al., 2001):

$$a_g(\lambda) = a_g(443)e^{-S_g(\lambda-443)}, \quad (9)$$

$$a_{NAP}(\lambda) = a_{NAP}(443)e^{-S_{NAP}(\lambda-443)}. \quad (10)$$

S_g was modeled as a normal distribution $0.017N(1, 0.02^2)$ and S_{NAP} as $0.010N(1, 0.01^2)$. Equation 7 was also used to determine the concentration of NAP, [NAP] (g m^{-3}):

$$[NAP] = a_{NAP}(443)/a^*_{NAP}(443), \quad (11)$$

where $a^*_{NAP}(443)$ (m^2 g^{-1}) is the mass-specific absorption coefficient of NAP at 443 nm, which was simulated as a uniformly distributed random number $0.03 \le a^*_{NAP}(443) \le 0.05$ (m^2 g^{-1}). The [NAP] was typically in the range of 0–30 g m^{-3}.

The total scattering coefficient ($b(\lambda)$, m^{-1}) was simulated as a sum of three components:

$$b(\lambda) = b_w(\lambda) + b_{ph}(\lambda) + b_{NAP}(\lambda). \quad (12)$$

Scattering by NAP was modeled using a power law function (Stramski et al., 2001; Twardowski et al., 2001) as follows:

$$b_{NAP}(\lambda) = b_{NAP}(550)\left(\frac{550}{\lambda}\right)^{\gamma_2}, \quad (13)$$

$$b_{NAP}(550) = b^*_{NAP}(550)[NAP], \quad (14)$$

where $b^*_{NAP}(550) = 0.5N(1, 0.04)$ (m^2 g^{-1}) is the mass-specific scattering of non-algal particles at 550 nm, and $\gamma_2 = 0.8N(1, 0.0049)$.

The scattering by phytoplankton was calculated as the difference between their attenuation and absorption coefficients (Voss, 1992; Roesler and Boss, 2003):

$$b_{ph}(\lambda) = c_{ph}(\lambda) - a_{ph}(\lambda). \quad (15)$$

The attenuation coefficient itself was modeled as a power law function (Voss, 1992),

$$c_{ph}(\lambda) = c_{ph}(550)\left(\frac{550}{\lambda}\right)^{\gamma_1}, \quad (16)$$

where $c_{ph}(550) = 0.3Chla^{0.57}$ and $\gamma_1 = 0.8$.

In the simulations, the backscattering coefficient ($b_b(\lambda)$, m^{-1}) was modeled as the sum of the contributing components,

$$b_b(\lambda) = b_{bw}(\lambda) + \tilde{b}_{b\,ph}b_{ph}(\lambda) + \tilde{b}_{b\,NAP}b_{NAP}(\lambda), \quad (17)$$

where $b_{bw}(\lambda)$ is obtained according to Morel, 1974 and $\tilde{b}_{b\,ph}$ and $\tilde{b}_{b\,NAP}$ are backscattering ratios for phytoplankton and non-algal particles assumed to be independent of the wavelength (Twardowski et al., 2001; Sydor and Arnone, 1997). Typical values were used as $\tilde{b}_{b\,ph}(\lambda) = 0.006$ and $\tilde{b}_{b\,NAP}(\lambda) = 0.02$.

120000 different conditions were simulated using this model with 70% used in generation and 30% in testing and validation.

As was discussed above, several field R_{rs} datasets were available for analysis together with (or without) some measurements of water parameters. Four R_{rs} sets are shown in Figure 1 with corresponding water parameters; some of these parameters (shown in grey) were not measured directly but estimated using available algorithms. Examples of simulated R_{rs} spectra are also shown in this figure. It should be noted that there was a relatively small flexibility in the selection of parameters described above, which produce spectra similar to the ones in the bloom areas with typical high CDOM and corresponding low R_{rs} in the blue, spectral features in green-red and a very strong peak around 700 nm comparable with the peak in the green.

In the model development, $R_{rs}(\lambda)$ spectra were supposed to be similar not only to the field spectra in Figure 1, but there were also supposed to be consistent with the good performance of blue-green algorithms for Chl-a retrievals. This should be true at least in the waters with low to moderate Chl-a and RE10 NIR/red bands algorithm for a broad range of waters and Chl-a concentrations, which were observed previously for the Chesapeake Bay (Gilerson et al., 2015).

The ranges of water parameters in the Chesapeake Bay are Chl-a = 0.06–165 mg/m^3, CDOM absorption at 443 nm $a_g(443)$ = 0.015–2.0 m^{-1}, absorption of non-algal particles $a_{NAP}(443)$ = 0.001–3.0 m^{-1}, scattering at 443 nm $b(443)$ = 0.3–40.3 m^{-1} with the lowest value typically in the Lower Bay and the highest in the Upper Bay (Magnuson et al., 2004). For LIS Chl-a = 1–25 mg/m^3, $a_g(440)$ = 0.012–0.5 m^{-1}, $a_{NAP}(440)$ = 0.02–0.42 m^{-1}, particulate backscattering at 650 nm b_{bp} = 0.005–0.06 m^{-1} with the lowest value in the eastern part of the Sound and the highest in the western part (Aurin et al., 2010). In the model Chl-a values were randomly distributed between 0.5 and 200 mg/m^3, $a_g(443)$ were mostly in the range of 0–3 m^{-1} with decreasing quantities till 6.5 m^{-1} and $a_{NAP}(443)$ = 0–2.5 m^{-1}.

Several metrics were used in the evaluation of Chl-a algorithms performance which includes a coefficient of determination R^2, root

mean square error (RMSE), relative error e = RMSE/mean as well as recently suggested metrics (Seegers et al., 2018) mean absolute error

$$MAE = 10 \wedge \left(\frac{\sum_{i=1}^{n} |\log_{10}(M_i) - \log_{10}(Q_i)|}{n} \right), \quad (18)$$

and bias

$$bias = 10 \wedge \left(\frac{\sum_{i=1}^{n} \log_{10}(M_i) - \log_{10}(Q_i)}{n} \right). \quad (19)$$

It should be noted that in some figures Chl-a values are shown in the logarithmic scale, while RMSE and e were calculated based on the linear scale.

2.5 NN algorithm development, analysis of the optimized structure and validation

In continuation of the approach used by (El-Habashi et al., 2016), their simple one-hidden layer multilayer perceptron (MLP) structure was first applied to a newly developed synthetic dataset, to produce a minimum benchmark against which to improve with the introduction of the VIIRS imaging I1 band to complement the 486, 551 and 671 nm band inputs as well as with modifications to the neural network itself. Variables $a_{ph}(443)$, $a_g(443)$, $a_d(443)$ and $b_b(443)$ were kept as outputs. Chl-a was determined also as an independent output parameter. Performance results are visible in Table 1. The introduction of the imaging I1 band immediately provided a large performance boost on all four output parameters. However, changes in the neural network structure with the introduction of more neurons in the single hidden layer and the introduction of Rectified Linear Units (ReLU) as the activation function produced a negligible change in the network performance. Similarly, the introduction of a second hidden layer also produced a negligible change in the network performance, indicating that the simpler neural network utilized in previous studies is already capable of capturing the relationships between inputs and outputs well.

In original tests, the bio-optical model was slightly different from the one described above (specific phytoplankton absorption consisted of the micro- and picoplankton absorption with a weighting factor from Ciotti and Bricaud, 2006). In the final version, R^2 coefficients were higher as shown in Table 1 in parentheses. Figure 2 contrasts the performance of the NNs in the 3-band and 4-band versions against the expected values for $a_{ph}(443)$, $a_g(443)$, $a_d(443)$, and $b_b(443)$ as measured during the CCNY 2013 cruise in the Chesapeake Bay. In all cases, including the VIIRS imaging I1 band noticeably improves the retrieval quality. In these tests, Chl-a were determined from $a_{ph}(443)$. When Chl-a were used directly as one of the retrieval parameters, R^2 for Chl-a became 0.984.

If large datasets of Chl-a and R_{rs} are available for relevant water conditions, the training can be carried out directly to retrieve Chl-a and other water parameters from R_{rs} spectra (Hieronymi et al., 2017; Pahlevan et al., 2020). While only 70 points of the field data were available for the Chesapeake Bay, the training gave results quite similar to the ones from the bio-optical modeling, however, some additional tuning was still required, and this option was not further explored.

3 Results

3.1 Preliminary studies

3.1.1 Performance of different Chl-a algorithms

A Satlantic HyperSAS (Halifax, Canada) system was installed from 2009 to 2014 at the LISCO site (Harmel et al., 2011) together with the SeaPRISM instrument on top of a retractable tower at approximately 12 m above the water surface. Three spectrometers observed downwelling irradiance E_d, sky radiance L_s, and total radiance L_t in the wavelength range of 305–905 nm with 180 equally spaced channels. HyperSAS data were processed by the 3C model (Groetsch et al., 2017, 2020) to minimize the impact of the sky reflectance from the windy surface and to produce reliable R_{rs} data. Several algorithms to determine Chl-a were applied to analyze water conditions in the area of LISCO during the year of 2013, which included conditions of algal blooms. Algorithms included standard 3 bands OC3V algorithm (based on 443, 486 and 551 nm), 6 bands OC6P algorithm (O'Reilly and Werdell, 2019), NN algorithm (El-Habashi et al., 2019), and NIR/red (red edge) (Gilerson et al., 2010), further referred to as RE10, based on $R_{rs}(709)/R_{rs}(665)$ ratio. The latter algorithm proved to perform well in a very broad range of Chl-a > 5 mg/m^3 and other water components (Smith et al., 2018; Pahlevan et al., 2022). All algorithms except RE10 performed similarly at Chl-a < 10 mg/m^3 and substantially underestimated Chl-a in bloom conditions in 2013, where only RE10 indicated Chl-a up to 40 mg/m^3.

3.1.2 R_{rs} uncertainties

It has been well known for a long time that main R_{rs} uncertainties over coastal waters occur at the blue bands 412 and 443 nm (IOCCG, 2019), which motivated the development of other algorithms avoiding the 443 nm band on VIIRS sensors (Ioannou et al., 2014; Gilerson et al., 2015; El-Habashi et al., 2016) and NIR/red algorithms on MERIS and OLCI sensors, which have 709 nm band (Gitelson, 1992; Moses et al., 2009; Gilerson et al., 2010). While main uncertainties in the blue were usually attributed to inaccurate aerosol models in the atmospheric correction process (IOCCG, 2019), a recent analysis based on the decomposition of R_{rs} uncertainties spectra showed that some uncertainties may be associated with Rayleigh-type components and thus might be related to small variability (about 1.5%) of the Rayleigh radiance (Gilerson et al., 2022, 2023) or Rayleigh noise (Malinowski et al., 2024). It was also shown that OLCI uncertainties in coastal waters in EUMETSAT processing are about 50% higher than uncertainties for VIIRS in the blue (Mikelsons et al., 2022; Zibordi et al., 2022; Gilerson et al., 2023) due to the different atmospheric correction schemes (Mikelsons et al., 2022) with NOAA MSL12 OLCI processing having R_{rs} uncertainties about the same as for VIIRS. Further, NASA OLCI processing also showed the same level of uncertainties as those from VIIRS and NOAA MSL12 OLCI. These effects are additionally demonstrated in Figure 3, where matchups

TABLE 1 Performance summary in R^2 of the neural networks tested on the synthetic dataset, original (final) bio-optical model.

Description	Network structure	Activation	$a_{ph}(443)$	$a_g(443)$	$a_d(443)$	$b_b(443)$
Original MLP	3 × 6 × 4	Sigmoid	0.601 (0.726)	0.588 (0.753)	0.546 (0.738)	0.555 (0.635)
I1 band	4 × 6 × 4	Sigmoid	0.722 (0.80)	0.796 (0.823)	0.640 (0.807)	0.749 (0.77)
More neurons	4 × 36 × 4	ReLU	0.719	0.794	0.635	0.743
2 hidden layers	4 × 36 × 30 × 4	ReLU	0.722	0.798	0.639	0.746

are shown for VIIRS, OLCI EUMETSAT, and OLCI MSL12 data processing at the LISCO AERONET-OC site.

High uncertainties can be clearly seen at the 443 nm band for VIIRS with a much more stable 490 nm band. Results are similar in OLCI MSL12 data processing. In EUMETSAT data processing, all bands below 560 nm show high uncertainties. Uncertainties at 665 nm and 709 nm are also high but these R_{rs} are related to low Chl-a < 10 mg/m^3 conditions in LIS, they are not of the main interest for the application of the NIR/red algorithm, which works reliably mostly for higher Chl-a. At the Chesapeake Bay AERONET-OC station with waters clearer around the AERONET-OC station than at the LISCO site, correlations were higher for OLCI (not shown).

3.1.3 Evaluation of the performance of Chl-a algorithms in algal bloom conditions

Blooms often occur near salinity fronts in the Upper Bay and Potomac River. Satellite imagery for the Chesapeake Bay with bloom conditions in the Upper Bay, processed with OC3V for VIIRS, with RE10 using EUMETSAT OLCI imagery with default and NCCOS atmospheric corrections together with the Chl-a distributions received with an additional band ratio algorithm. Chl-a in the bloom areas from different algorithms were 27–140 mg/m^3 for the Upper Bay and 30–200 mg/m^3 for the Potomac River. These data had to be reconciled between different satellite sensors and algorithms to develop a combined VIIRS-OLCI product for bloom detection.

At the beginning, Chl-a were estimated in the bloom areas in the Chesapeake Bay and in the Potomac River for May 13, 2020, using four different algorithms, including the standard three bands OC algorithm for VIIRS OC3V, the band ratio VIIRS algorithm with I1 band (chlC) (Gilerson et al., 2021) described below in Section 3.1.4, OLCI RE10 algorithm with standard OLCI AC and with NCCOS AC (Wynne et al., 2018). Two bloom areas have been identified: in the Upper Bay and in the Potomac River. While the shapes of the bloom areas on satellite images looked similar, it was found that OC3V had the lowest Chl-a values around 30 mg/m^3 and RE10 = 50 – 140 mg/m^3 in the Upper Bay and above 200 mg/m^3 in the Potomac River with chlC values were in the middle of these ranges. The focus of this work was a more detailed evaluation of these algorithms and the newly developed NN algorithm in various bloom conditions.

3.1.4 Band ratio algorithm with I1 band

The first tests (Gilerson et al., 2021) proved the utility of I1 band in detecting higher concentrations of Chl-a values. Because of the

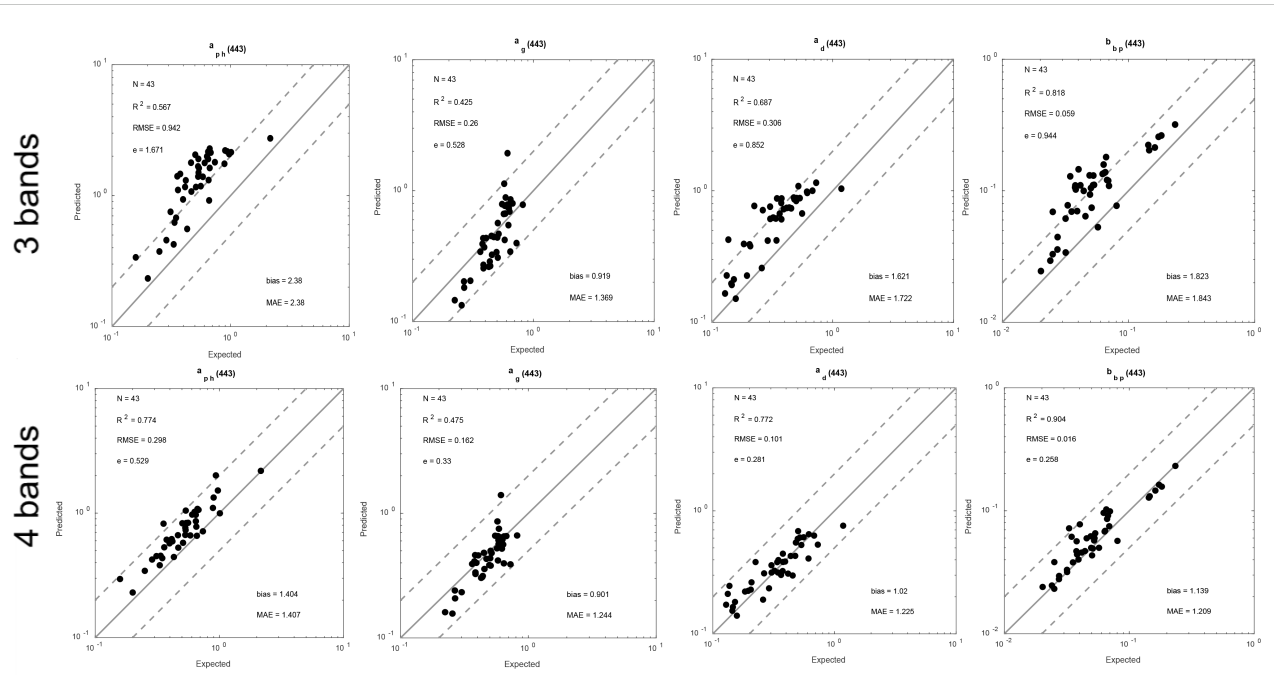


FIGURE 2
Results of a and b_b retrievals in m^{-1} using NN with 3 bands (top) and 4 bands (bottom) based on field data from CCNY 2013 cruise.

10.3389/fmars.2024.1476425

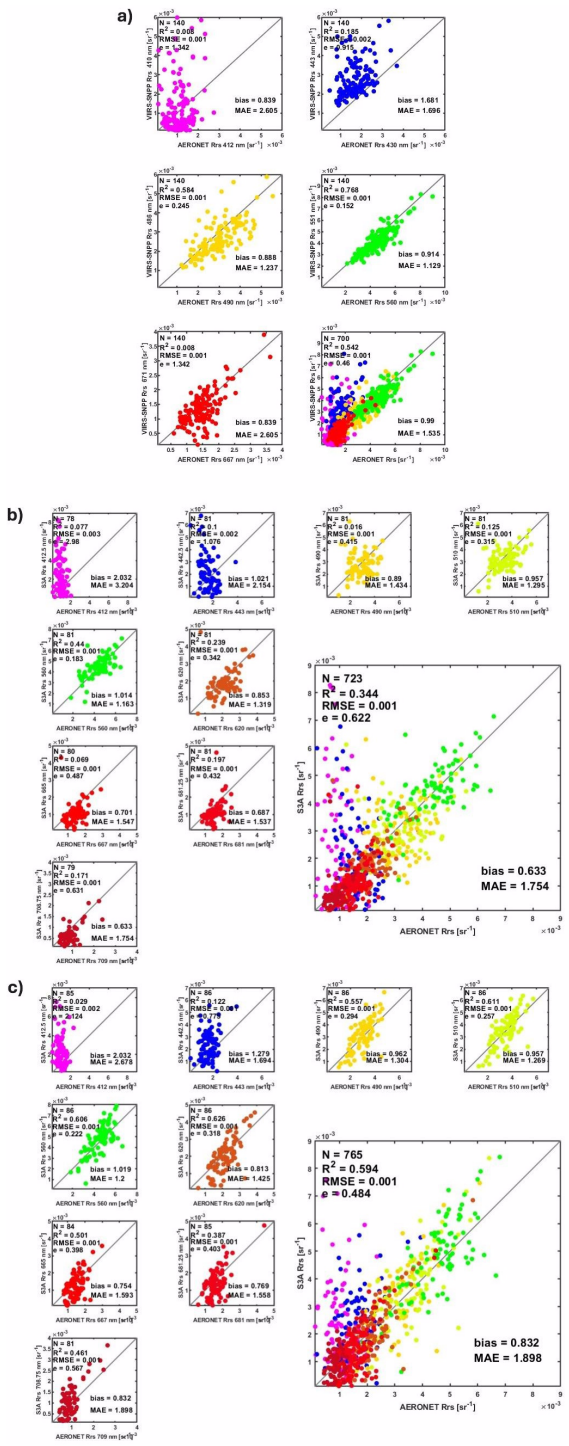


FIGURE 3
Satellite and AERONET-OC matchups at the LISCO site for the matching wavelengths available on the SeaPRISM and on the sensor: **(A)** SNPP VIIRS, **(B)** S3A OLCI with EUMETSAT (OBC-3), and **(C)** S3A OLCI with NOAA MSL12 data processing.

complexity of water IOPs spectra in the I1 range, including variability of CDOM and mineral concentrations in various areas, it was clear that the algorithm eventually needs to be implemented in a NN format. But, it appeared useful to evaluate a multi-band algorithm for the estimation of Chl-a in a wide range of water conditions. The algorithm was developed using available band ratios, which include I1 band. A proper band combination was determined by tests on the synthetic dataset discussed above.

Application of the first version of the algorithm with I1 band, which was calibrated on the field data showed a strong dependence of the estimated Chl-a on the concentration of suspended particulate matter (SPM) with sediment concentrations estimated

from (Nechad et al., 2010) based on R_{rs} at 671 nm. In the next iteration, the algorithm was corrected for the impact of SPM concentration. It was also found that the algorithm often underestimates Chl-a at Chl-a < 10–15 mg/m^3 and it was therefore combined with the standard OC3V algorithm at Chl-a ≤ 15 mg/m^3.

The algorithm was tuned using MATLAB curve fitting toolbox on 43 R_{rs}-Chl-a combinations from the CCNY 2013 cruise and then further on field data from M. Ondrusek's measurements in 2014-2021 (see Figures 4D and F) with a total of 70 points. It was implemented with the final result as chlC:

$$SPM = 384.11 \times \pi R_{rs}(671)/(1 - \pi R_{rs}(671)/0.1747) + 1.44 \quad (20)$$

$$ratio = (R_{rs}(486) + R_{rs}(551))/R_{rs}(638) \times SPM^{0.3} \quad (21)$$

$$ChlC = k \times 4604 \times ratio^{(-4.252)} \quad (22)$$

$$chlC = ChlC \; if \; ChlC > 10 \; mg/m^3 \quad (23a)$$

$$chlC = OC3V \; if \; ChlC \leq 10 \; mg/m^3 \quad (23b)$$

Coefficient k in Equation 22 is a tuning parameter, which can be further changed. In this version, coefficients are different from the original version (Gilerson et al., 2021), when the algorithm was tuned only on the data from the CCNY 2013 cruise. The performance of the algorithm with k =1.0 is demonstrated below in Figures 4D, F.

3.2 Validation of VIIRS algorithms on satellite and field data

A total of 70 matchups were included in the tests (43 from CCNY 2013, 22 from Ondrusek 2014-16, and 5 from Ondrusek 2021 measurements) for the validation of NN3, NN4 and VIIRS standard OC3V algorithms on the field data collected across the Chesapeake Bay. Results are shown in Figures 4A–C. The performance of chlC and RE10 on the same field dataset is shown in Figures 4D–F. Among the first three algorithms in Figure 4 the NN4 algorithm shows better performance, although it is worse than the performance of chlC, where all points were used in the tuning and RE10, for which 709 nm band is not available on VIIRS. In Figures 4D–F chlC is plotted against field

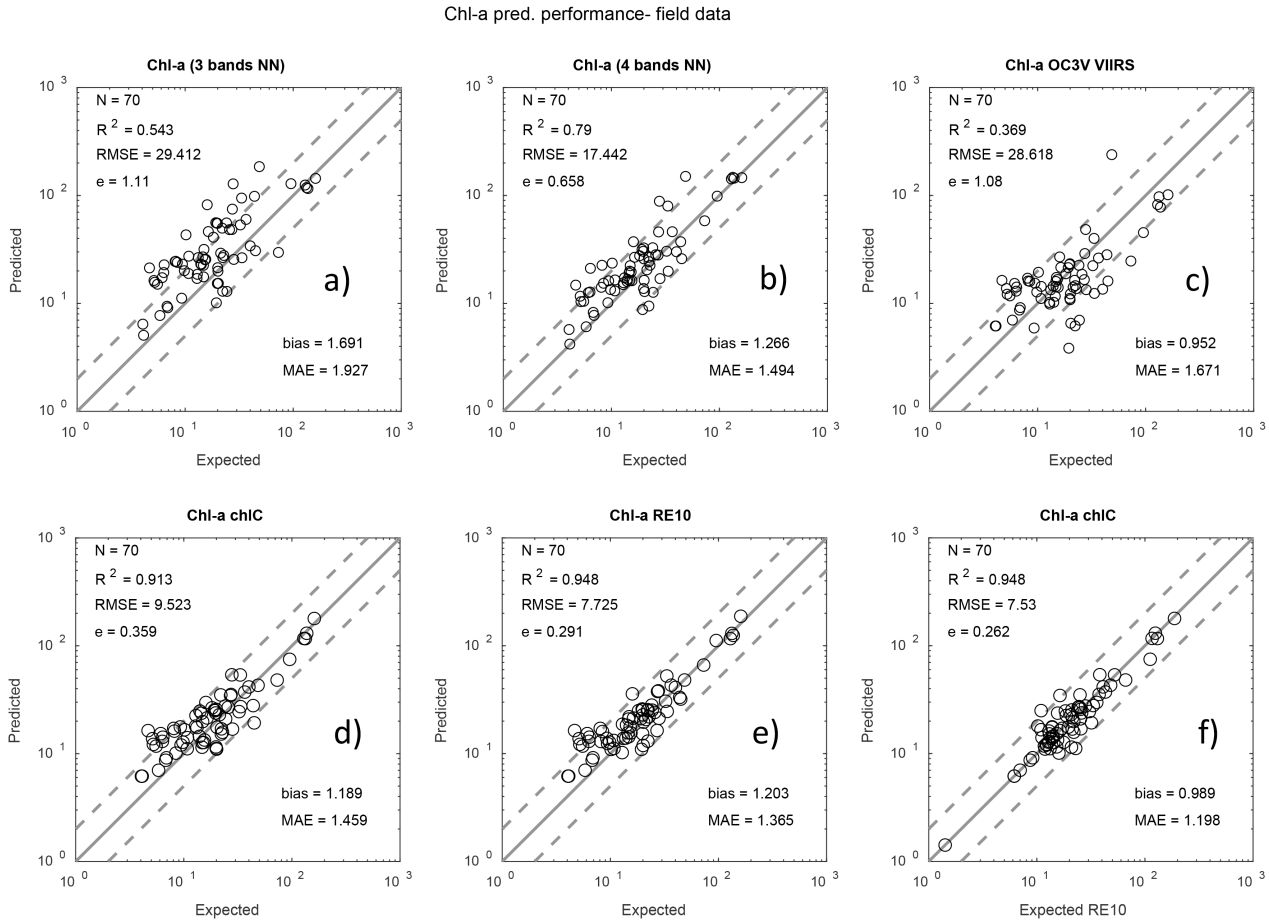


FIGURE 4
Test of algorithms on the field data: **(A)** 3 bands NN; **(B)** 4 bands NN; **(C)** OC3 VIIRS; **(D)** chlC vs *in-situ* Chl-a; **(E)** RE10 vs *in-situ* Chl-a, **(F)** chlC vs RE10.

Chl-a and against RE10; RE10 against Chl-a is also shown for the comparison. High correlations exist for all comparisons in a broad range of conditions in the Chesapeake Bay, but these relationships are not always valid for other types of waters. RE10 was also considered as OC3 VIIRS if RE10 <10 mg/m^3.

RE10 was used with the expression (24), which matches the original version in Gilerson et al., 2010, but does not produce complex numbers at low Chl-a

$$RE10 = 46.0676(R_{rs}(709)/R_{rs}(665))^{1.2260} - 22.6012 \quad (24)$$

Further tests were performed on SNPP VIIRS data 2012–2022 (NOAA MSL12 data processing) compared with *in-situ* data from the Chesapeake Bay program (https://www.chesapeakebay.net) and there were 2021 measurements at 5 locations. Results are shown in Figure 5. The stray light flag was on, HIGLINT and MODGLINT flags were suspended since they did not change the algorithm performance significantly. Most of the points are in the Chl-a range below 20 mg/m^3. However, all algorithms, including the OC3 algorithm, retrieve high Chl-a values reasonably well; good performance of OC3 is most likely due to the specific combination of the water parameters in bloom areas, which is not typical for coastal waters with high Chl-a. The time window between satellite and *in-situ* measurements was ±4 hours. Based on our studies in the Chesapeake Bay, stricter time limits would reduce the number of points but would not improve statistics.

Here and in the figures below the solid grey line marks the 1:1 relationship, while the upper and lower dashed lines mark the limit of Y = X*2 and Y = X/2, respectively, where Y are predicted values and X are expected values.

3.3 Comparison of Chl-a retrievals by VIIRS and OLCI algorithms

Performance of the RE10 algorithm for OLCI sensors was evaluated with NCCOS, EUMETSAT, MSL12 and NASA atmospheric correction by the comparison with VIIRS Chl-a in bloom areas with a very broad range Chl-a from 2 mg/m^3 to over 100 mg/m^3. Because the RE10 algorithm does not provide accurate retrievals for low Chl-a and the OC4 algorithm for OLCI was found not to be always reliable in the waters of the Chesapeake Bay, comparisons were carried out using the RE10M algorithm, where RE10 was replaced with OC3V Chl-a for Chl-a < 6 mg/m^3. It was found that the most consistent matchups between VIIRS and OLCI retrievals come from EUMETSAT and MSL12 processing. Examples of such matchups for the Upper Bay and Potomac River bloom areas are shown in Figure 6. NN4 versus RE10M shows better results than other algorithms. For low Chl-a < 6 mg/m^3, OC3V and chlC matchups with RE10M are along 1:1 line because OC3V retrievals are used in all these cases. Since VIIRS algorithms matchups vs RE10M in EUMETSAT and MSL12 matchups produce similar results, both processing approaches from EUMETSAT and MSL12 were recommended for the combined OLCI product. It should be noted that, according to Mikelsons et al. (2022), EUMETSAT processing is more sensitive to the sun glint, which was shown in our comparisons.

3.4 Combined products, and satellite imagery

Based on the whole study, NN4 VIIRS and OLCI RE10 algorithms were recommended for the combined VIIRS-OLCI product. RE10 was used in combination with OC4 (with OC4 if RE10< 10 mg/m^3 and OC4 <10 mg/m^3 or OC4 < 10 mg/m^3 and clear water conditions based on the diffuse attenuation coefficient threshold $K_d(490)$ < 0.25 m^{-1}). Examples of the imagery from both algorithms are shown in Figure 7 for May 18, 2021, when field measurements were also available at 5 locations with the coordinates shown in Table 2, together with measured Chl-a at these points and retrieved from OC3, chlC, NN3, and NN4 algorithms from VIIRS and RE10 from OLCI. Note that part of the area on the OLCI image is masked because of clouds. As before, a slight overestimation of Chl-a is seen in both images in very turbid waters in the Upper Bay, Delaware Bay, and some tributaries. Adjustment coefficients for chlC, NN3, and NN4 are also given in Table 2. Relative spectral response (RSR) functions were not taken into account in the NN algorithms development to simplify tuning of the algorithms based on the bio-optical model only and comparison with field measurements; for the same reason RSR for the I1 band was considered as RSR = 1.0 for the whole range of wavelengths 600–680 nm. The actual RSR for this

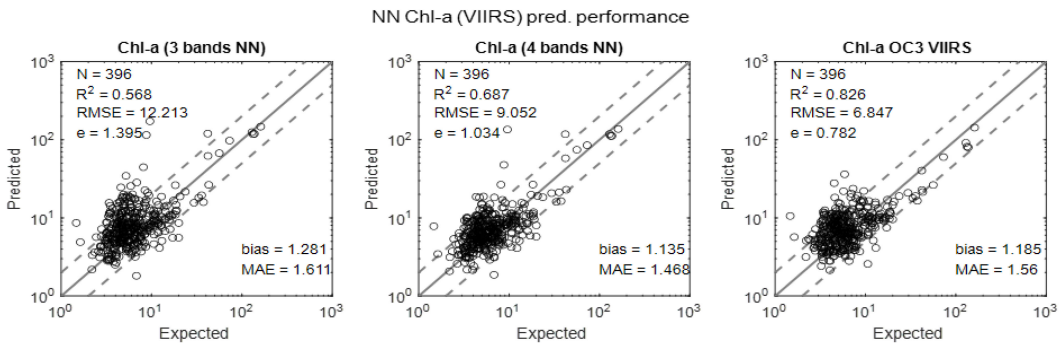


FIGURE 5
Comparison of satellite and *in-situ* data for the Chesapeake Bay. Expected and predicted Chl-a as determined by the one-hidden layer MLP in both its 3-band and 4-band versions and OC3 VIIRS algorithms.

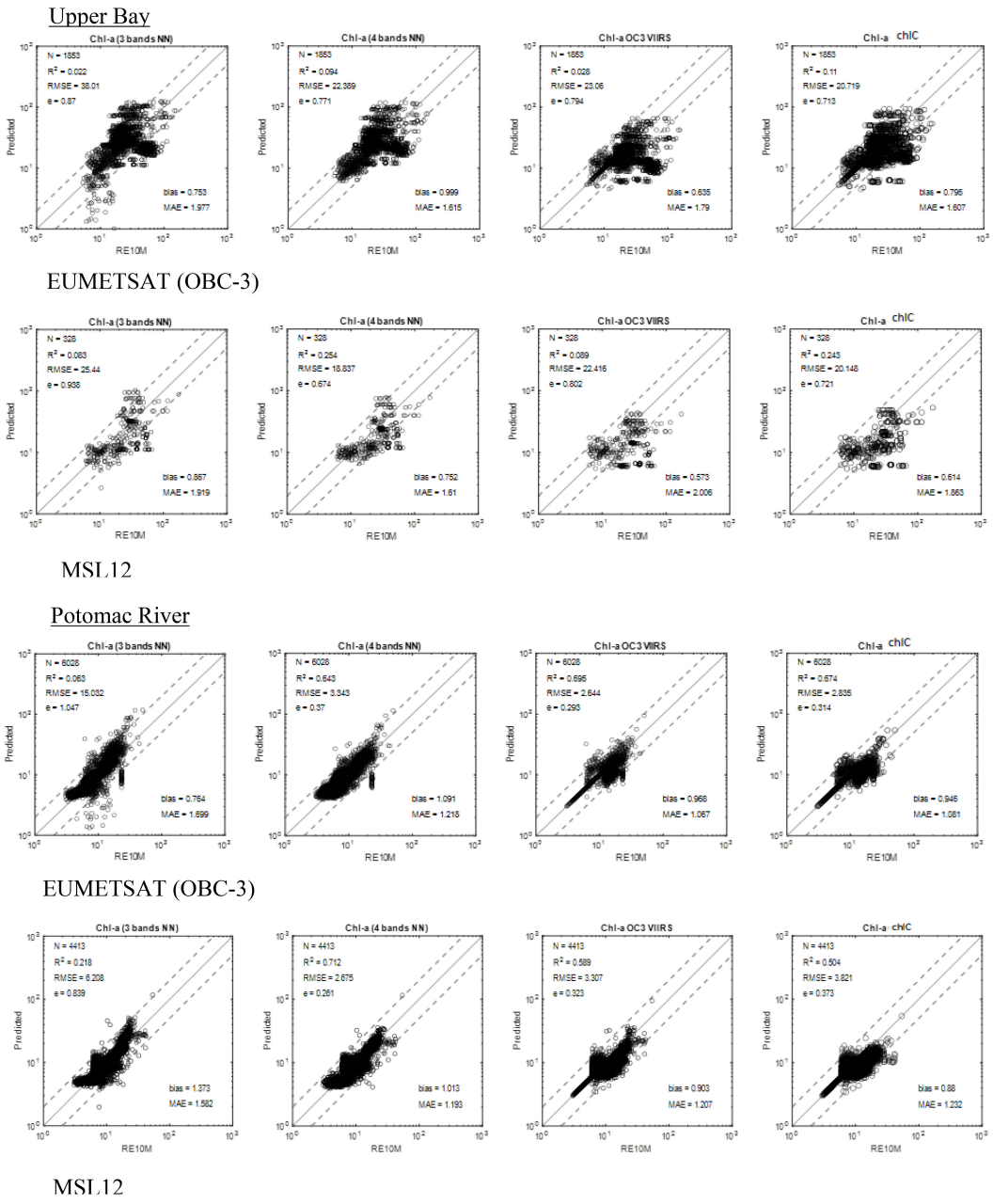


FIGURE 6
Matchups between VIIRS and OLCI Chl-a retrievals in bloom areas.

band is close to RSR ≈ 0.9, which matches the adjustment coefficient for NN4. The NN3 and chlC algorithms provided similar images but with some adjustments of coefficients, which were less stable than those from the NN4 algorithm. Other examples of images from VIIRS and OLCI for bloom conditions on May 21, 2021, and non-bloom conditions on April 4, 2024, are shown in Figure 8.

The distribution of absorption and backscattering coefficients at 443 nm retrieved from NN4 together with the SPM concentration based on Equation 20 for May 18, 2021, are shown in Figure 9, providing additional information about water parameters in the Chesapeake Bay and specifically in the bloom areas, which helps to understand bloom conditions in more details. As can be expected, $a_d(443)$, $b_b(443)$ and SPM have similar patterns since they are mostly proportional to the concentrations of non-algal particles, $a_{ph}(443)$ and $a_g(443)$ are high in the bloom areas.

The NN4 algorithm was developed based on SNPP VIIRS bands, VIIRS on NOAA-20 has several slightly different bands as was shown above, specifically for the NN algorithm there are M3–M5 bands centered at 489, 556, and 667 nm, and I1 band centered at 642 nm and NN4 algorithm required additional tuning. While the effects of

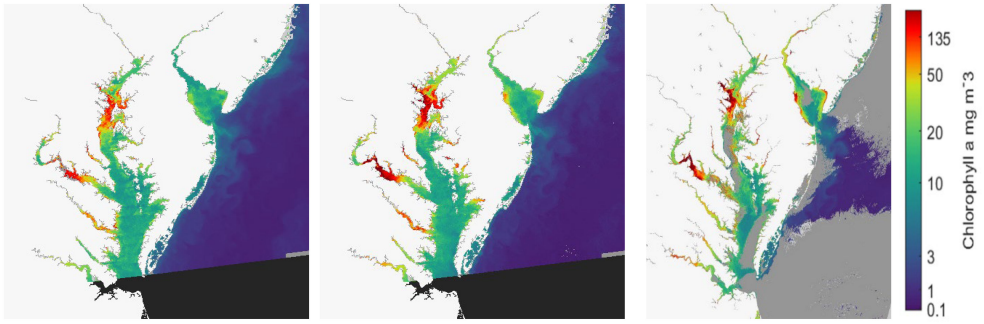


FIGURE 7
OC3 VIIRS (left panel), NN4 VIIRS (middle panel) and OLCI (right panel) Chl-a retrievals in bloom areas on May 18, 2021.

TABLE 2 Chl-a measurements and retrieval comparison for May 18, 2021.

Lat/Lon (West)	Ondrusek	SNPP VIIRS				OLCI	N20 VIIRS	
		OC3V	chlC (1.6)	NN3 (0.85)	NN4 (0.9)	RE10	OC3V	NN4 (0.7)
39.046 76.392	133	82	128	125	126	128	253	110
39.053 76.405	129	91	135	133	133	114	NaN	126
39.055 76.423	161	143	143	156	154	286	NaN	118
39.073 76.403	137	80	121	126	126	164	NaN	163
38.964 76.452	73	60	77	104	96	85	313	104

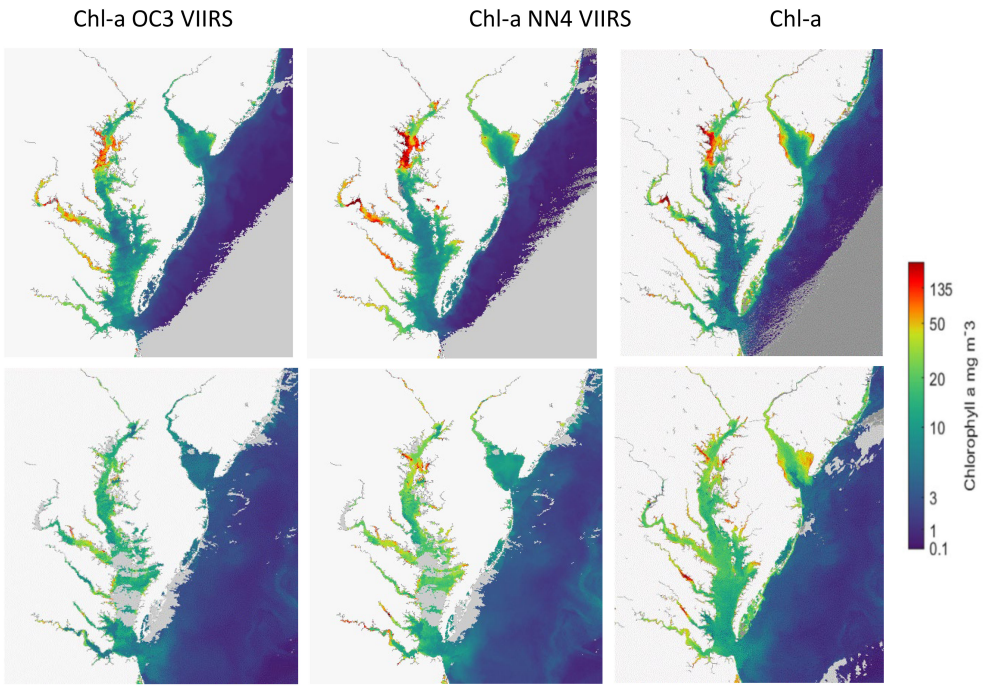


FIGURE 8
OC3 VIIRS, NN4 VIIRS and OLCI Chl-a retrievals with bloom conditions on May 21, 2021 (top row) and non-bloom conditions on April 15, 2024 (bottom row).

10.3389/fmars.2024.1476425

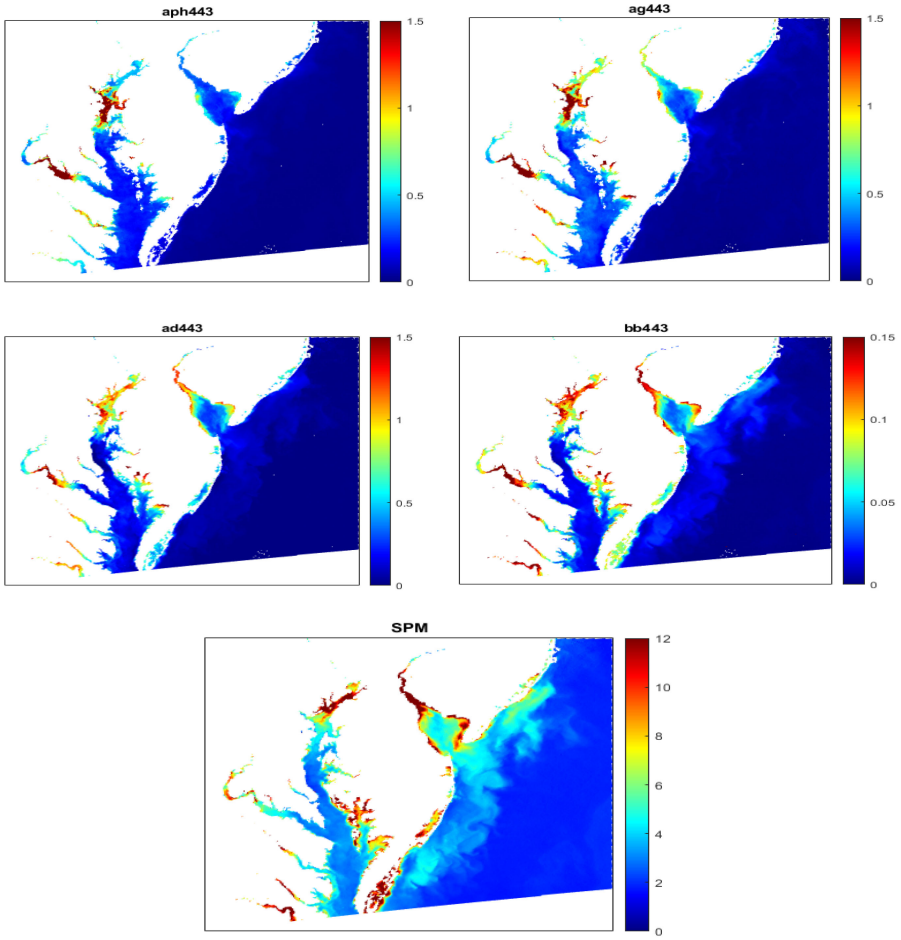


FIGURE 9
Distributions of absorption, backscattering coefficients (m^{-1}), in the Chesapeake Bay from NN4 VIIRS algorithm, SPM (g/m^3) from (NeChad et al., 2010).

spectral differences between VIIRS-SNPP and VIIRS-NOAA-20 at the blue bands are negligible (e.g., within ~0.1% at M2 band), there are large differences at M4 (green) and M5 (red) bands (e.g., ~16% at M4 band for open oceans) (Wang et al., 2020). Over coastal regions, there are important effects of M4 band difference between VIIRS-SNPP and VIIRS-NOAA-20, because R_{rs} from NOAA-20 (at 556 nm) is usually much closer to the R_{rs} peak than that from SNPP (at 551 nm). The same NN4 algorithm was used for VIIRS NOAA-20

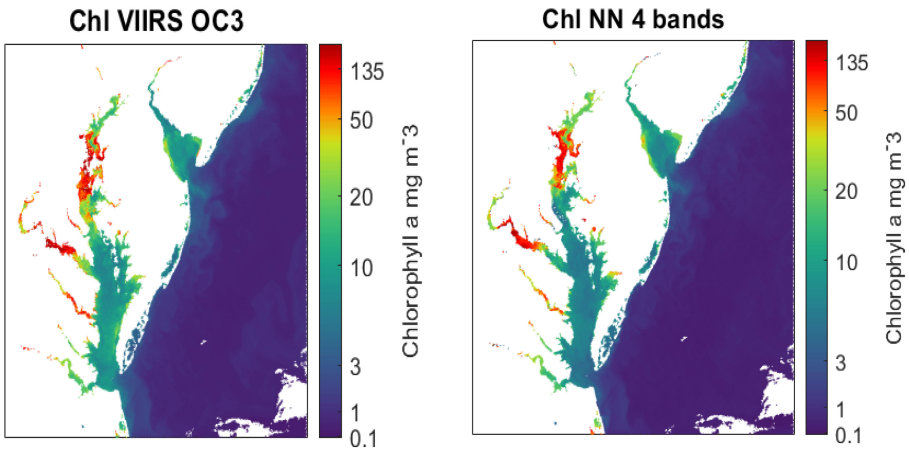


FIGURE 10
Comparison of NOAA-20 VIIRS images retrieved with OC3V and NN4 algorithms.

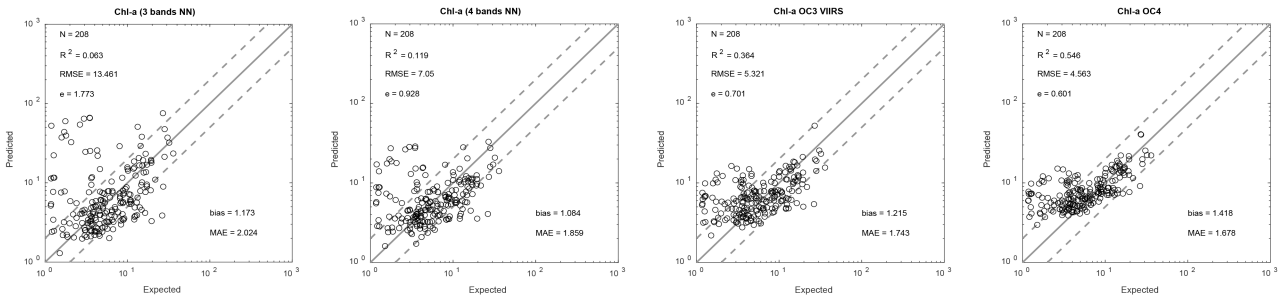


FIGURE 11
Performance of algorithms on R_{rs} and Chl-a data in LIS.

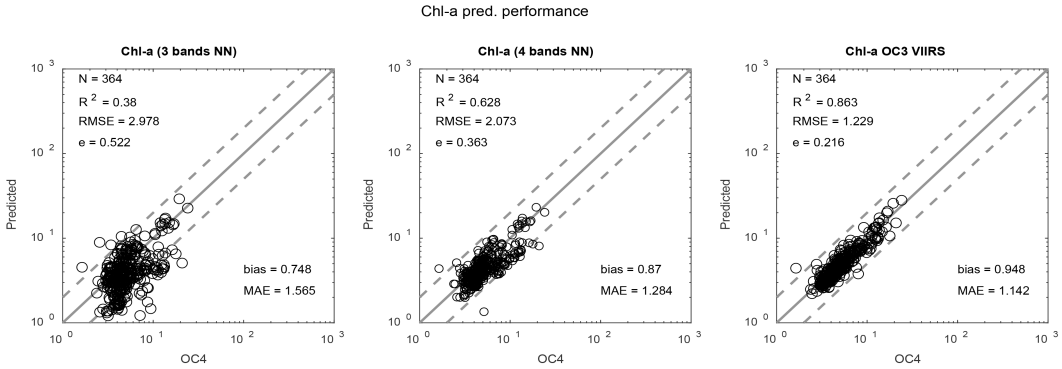


FIGURE 12
Performance of several VIIRS algorithms in comparison with OC4 based on the LISCO radiometric data.

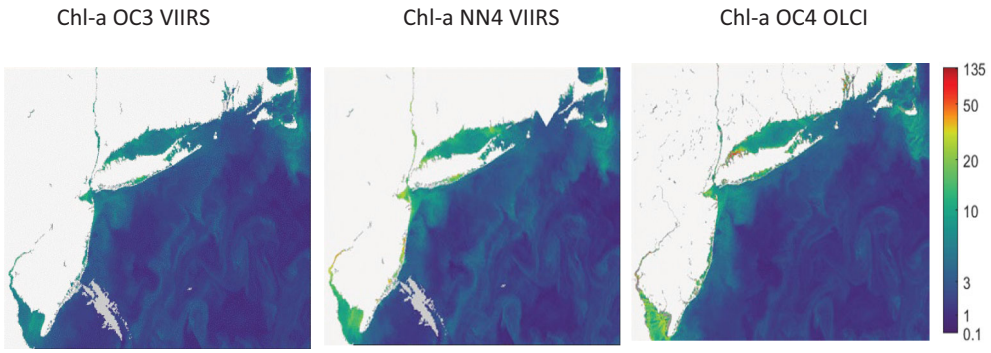


FIGURE 13
OC3, NN4 Chl-a from VIIRS and OC4 from OLCI in LIS on April 16, 2024.

bands but with the coefficient 0.65. Images for OC3V and NN4 for NOAA-20 VIIRS are shown in Figure 10 and Chl-a are added to Table 2. Chl-a from OC3V and NN4 in the same scale looks similar to SNPP Chl-a distributions. For OC3V Chl-a values at *in-situ* measured points match less accurately with a strong overestimation at two points and were not processed at three other points. NN4 for NOAA-20 is less accurate than for SNPP but can be also recommended for the joint product.

3.5 Applications of the developed algorithms to the waters in Long Island Sound

The performance of the NN4 algorithm was validated on the field data in Long Island Sound. Field data were acquired during cruises in 2018–2023 and included radiometric measurements and Chl-a (Sherman et al., 2023). Results for different algorithms are

shown in Figure 11. Most of Chl-a values are below 25 mg/m^3, the range that was not the main focus of the NN4 algorithm. The best performing algorithm is OC4 followed by OC3 and NN4. However, all these algorithms perform quite well for Chl-a > 2 mg/m^3 and much worse below this value. The NN4 algorithm was used with a coefficient of 0.6, while it was 0.9 in the Chesapeake Bay for SNPP VIIRS. The difference in coefficients might be explained by differences in $a^*_{ph}(\lambda)$ with shifts in phytoplankton species (including size and Chl-a packaging), between the time periods in LIS and in the Chesapeake Bay. Optical differences in the water may also influence the bio-optical model. More details about this difference should be further studied.

There were few matchups with VIIRS for field data used in Figure 11. Sherman et al. (2023) had OLCI retrievals corrected with the Polymer atmospheric correction algorithm and a bio-optical model for moderately turbid waters (Steinmetz et al., 2011), which resulted in good agreement with field observation across the Sound. The performance of algorithms was evaluated at the LISCO site for the period of August 2021–May 2022. The SeaPRISM instrument has bands similar to OLCI bands, and there were no direct field Chl-a measurements. Chl-a were estimated by the OC4 algorithm and compared with those from the NN4 and OC3 algorithms with VIIRS bands, with R_{rs} determined from the SeaPRISM bands using an adjustment based on the relationship between bands from the synthetic dataset. The NN4 and OC3 algorithms perform very consistently in the whole range of Chl-a from 2–25 mg/m^3 as shown in Figure 12. However, there were no *in-situ* Chl-a data to confirm these retrievals. Images of Chl-a in LIS based on OC3 and NN4 retrievals for VIIRS and OC4 for OLCI are shown in Figure 13 and are very consistent with each other generally confirming the good performance of algorithms in Figure 12. Both NN4 and OC3 algorithms for VIIRS can be recommended for the joint product with OLCI OC4.

4 Discussion and conclusions

Satellite data and imagery from SNPP and NOAA-20 VIIRS sensors and Sentinel-3A and 3B OLCI sensors were analyzed together with field data to develop the combined product for the estimation of Chl-a in two large US estuaries: the Chesapeake Bay and Long Island Sound to improve detection of algal blooms. The bio-optical model was developed to satisfy a broad range of conditions in waters from low Chl-a and corresponding absorption and backscattering coefficients in fresher reaches of the estuaries, with a switch for higher values in areas with high Chl-a and phytoplankton bloom conditions. The neural network (NN4) algorithm was developed for the retrieval of Chl-a and other water parameters from VIIRS in the Chesapeake Bay, which reasonably matches *in-situ* data. All VIIRS imagery used was from NOAA processing using MSL12 atmospheric correction. Based on the long-time knowledge about the vulnerability of the R_{rs} at 412 and 443 nm bands over coastal turbid waters, these bands were excluded from potential algorithms. The NN4 algorithm utilizes SNPP VIIRS four bands centered at 486, 551, 638, and 671 nm, which includes data from the imaging I1 600-680 nm band centered at 638 nm. It is demonstrated that the inclusion of this band data significantly improved retrieval of Chl-a and other water parameters in comparison with the previous versions of similar algorithms, which utilized only three 486, 551, and 671 nm bands. Analysis of several atmospheric correction and processing approaches from EUMETSAT (OBC-3), NOAA (MSL12), and NASA (L2gen) for OLCI for the application of the NIR/red RE10 Chl-a algorithm that requires accurate R_{rs} values at 665 and 709 nm bands showed that both MSL12 and OBC-3 data can be recommended for the combined product.

The NN4 and RE10 algorithms were analyzed in various water types demonstrating consistency during algal bloom conditions. These algorithms were selected for the multi-sensor product to support algal bloom detection in the Chesapeake Bay. The OC4 algorithm replaces RE10 for Chl-a < 10 mg/m^3, so VIIRS and OLCI Chl-a retrievals are consistent for the broad range of conditions in the Chesapeake Bay. The R_{rs} from the bio-optical model were re-trained to develop a NN4 algorithm for NOAA-20 VIIRS, which showed mostly Chl-a similar to those from the NN4 for SNPP VIIRS. In LIS during the whole period of study, there were no *in-situ* Chl-a above 30 mg/m^3. The NN4, OC3 and OC4 algorithms showed approximately the same performance and can be recommended for the estimation of Chl-a in LIS with the switch to RE10 for OLCI in case of higher Chl-a.

Further examination is recommended to determine if the combined NN4, OLCI with a switch to OC4 under low Chl-a conditions is accurate and provides the best estimate of Chl-a when switching water classes from coastal to offshore. This ability to provide consistent Chl-a from coastal to offshore, with improved cloud clearing capability through a multi-sensor approach, would support improved fisheries modeling capability, improved bloom monitoring, and the development of an improved long-time-series data of Chl-a to determine changes in primary productivity under changing climate conditions and in response to managing nutrient loading into coastal systems.

Data availability statement

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

Author contributions

AG: Writing – original draft, Writing – review & editing, Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision, Validation. MM: Investigation, Software, Validation, Writing – review & editing. JA: Conceptualization, Investigation, Methodology, Software, Writing – review & editing. EH-E: Investigation, Software, Validation, Writing – review & editing. MTz: Conceptualization, Data curation, Investigation, Methodology, Validation, Writing – review & editing. MTo: Conceptualization, Data curation, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing. AM:

Data curation, Investigation, Software, Validation, Writing – review & editing. RS: Writing – original draft, Writing – review & editing, Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation. MO: Data curation, Investigation, Methodology, Validation, Writing – review & editing. LJ: Investigation, Software, Validation, Writing – review & editing. MW: Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by the Joint Polar Satellite System (JPSS) funding including NOAA grant NA19NES4320002 (CISESS). We thank EUMETSAT for providing OLCI Level-1B data and ocean color products. AG, MM and EH-E were also supported by NASA award 80NSSC21K0562 and NOAA CESSRST center.

Acknowledgments

The authors are grateful to the reviewers, whose comments led to the significant improvement of the manuscript.

Conflict of interest

Author AM was employed by the company Consolidated Safety Services, Inc.

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.

Author disclaimer

The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the author(s) and do not necessarily reflect those of NOAA or the Department of Commerce.

References

Anderson, T. H., and Taylor, G. T. (2001). Nutrient pulses, plankton blooms, and seasonal hypoxia in western Long Island Sound. *Estuaries* 24, 228–243. doi: 10.2307/1352947

Aurin, D. A., Dierssen, H. M., Twardowski, M. S., and Roesler, C. S. (2010). Optical complexity in Long Island Sound and implications for coastal ocean color remote sensing. *JGR Oceans* 115, C7. doi: 10.1029/2009JC005837

Bricker, S. B., Longstaff, B., Dennison, W., Jones, A., Boicourt, K., Wicks, C., et al. (2008). Effects of nutrient enrichment in the nation's estuaries: A decade of change. *Harmful Algae* 8, 21–32. doi: 10.1016/j.hal.2008.08.028

Bukata, R. P., Jerome, J. H., Kondratyev, K. Y., and Pozdnyakov, D. V. (1995). *Optical properties and remote sensing of inland and coastal waters* (Boca Raton, FL: CRC Press).

Cao, F., and Tzortziou, M. (2024). Impacts of hydrology and extreme events on dissolved organic carbon dynamics in a heavily urbanized estuary and its major tributaries: a view from space. *JGR Biosci.* 129. doi: 10.1029/2023JG007767

Cao, Z., Wang, M., Ma, R., Zhang, Y., Duan, H., Jiang, L., et al. (2024). A decade-long chlorophyll-a data record in lakes across China from VIIRS observations. *Remote Sens. Environ.* 301, 113953. doi: 10.1016/j.rse.2023.113953

Ciotti, A. M., and Bricaud, A. (2006). Retrievals of a size parameter for phytoplankton and spectral light absorption by colored detrital matter from water-leaving radiances at SeaWiFS channels in a continental shelf region off Brazil. *Limnol. Oceanogr. Methods* 4, 237-253. doi: 10.4319/lom.2006.4.237

El-Habashi, A., Ahmed, S., Ondrusek, M., and Lovko, V. (2019). Analyses of satellite ocean color retrievals show advantage of neural network approaches and algorithms that avoid deep blue bands. *J. Appl. Remote Sens.* 13, 024509. doi: 10.1117/1.JRS.13.024509

El-Habashi, A., Duran, C. M., Lovko, M., Tomlinson, M. C., Stumpf, R. P., and Ahmed, S. (2017). Satellite retrievals of Karenia brevis harmful algal blooms in the West Florida Shelf using neural networks and impacts of temporal variabilities. *J. Appl. Remote Sens.* 11, 032408. doi: 10.1117/1.JRS.11.032408

El-Habashi, A., Ioannis, I., Tomlinson, M. C., Stumpf, R. P., and Ahmed, S. (2016). Satellite retrievals of Karenia brevis harmful algal blooms in the West Florida Shelf using neural networks and comparisons with other techniques. *Remote Sens* 8, 377. doi: 10.3390/rs8050377

EUMETSAT (2021). *Sentinel-3 OLCI L2 report for baseline collection ol_l2m_003.* Available online at: https://www.eumetsat.int/media/47794 (Accessed December 20, 2022).

EUMETSAT (2022). *Recommendations for sentinel-3 OLCI ocean colour product validations in comparison with in situ measurements – matchup protocols.* Available online at: https://www.eumetsat.int/media/44087 (Accessed December 20, 2022).

Freitas, F., and Dierssen, H. M. (2019). Evaluating the seasonal and decadal performance of red band difference algorithms for chlorophyll in an optically complex estuary with winter and summer blooms. *Rem. Sens. Env.* 231, 11228. doi: 10.1016/j.rse.2019.111228

Gilerson, A. A., Gitelson, A. A., Zhou, J., Gurlin, D., Moses, W., Ioannou, I., et al. (2010). Algorithms for remote estimation of chlorophyll-a in coastal and inland waters using red and near infrared bands. *Optics Express* 18, 24109-24125. doi: 10.1364/OE.18.024109

Gilerson, A., Herrera-Estrella, E., Agagliate, J., Foster, R., Gossn, J. I., Dessailly, D., et al. (2023). Determining the primary sources of uncertainty in the retrieval of marine remote sensing reflectance from satellite ocean color sensors II. Sentinel 3 OLCI sensors. *Front. Remote Sens.* doi: 10.3389/frsen.2023.1146110

Gilerson, A., Herrera-Estrella, E., Foster, R., Agagliate, J., Hu, C., Ibrahim, A., et al. (2022). Determining the primary sources of uncertainty in retrieval of marine remote sensing reflectance from satellite ocean color sensors. *Front. Remote Sens.* doi: 10.3389/frsen.2022.857530

Gilerson, A., and Huot, Y. (2017). "Sun-induced chlorophyll-a fluorescence," in *Bio-optical modelling and remote sensing of inland waters* (Elsevier). doi: 10.1016/B978-0-12-804644-9.00007-0

Gilerson, A., Malinowski, M., Herrera, E., Tomlinson, M., Stumpf, R., and Ondrusek, M. (2021). "Estimation of chlorophyll-a concentration in complex coastal waters from satellite imagery," in *Proc. of SPIE 11752, Ocean Sensing and Monitoring XIII.* doi: 10.1117/12.2588004

Gilerson, A., Ondrusek, M., Tzortziou, M., Foster, R., El-Habashi, A., Tiwari, S. P., et al. (2015). Multi-band algorithms for the estimation of chlorophyll concentration in the Chesapeake Bay. *Proc. SPIE* 9638. doi: 10.1117/12.2195725

Gilerson, A., Zhou, J., Fortich, R., Ioannou, I., Hlaing, S., Gross, B., et al. (2007). "Spectral dependence of the bidirectional reflectance function in coastal waters and its impact on retrieval algorithms," in *Proc. of IEEE 2007 International Geoscience and Remote Sensing Simposium (IGARSS 2007)*, Barcelona, Spain.

Gitelson, A. A. (1992). The peak near 700 nm on radiance spectra of algae and water: relationships of its magnitude and position with chlorophyll concentration. *Int. J. Remote Sens.* 13, 3367-3373. doi: 10.1080/01431169208904125

Gitelson, A. A., Schalles, J. F., and Hladik, C. M. (2007).Remote chlorophyll-a retrieval in turbid, productive estuaries: Chesapeake Bay case study. *Rem. Sens. Env.* 109. doi: 10.1016/j.rse.2007.01.016

Gordon, H. R. (2005). Normalized water-leaving radiance: revisiting the influence of surface roughness. *Appl. Opt.* 44, 241–248. doi: 10.1364/AO.44.000241

Gower, J. F. R., Brown, L., and Borstad, G. A. (2004). Observation of chlorophyll fluorescence in west coast waters of Canada using the MODIS satellite sensor. *Can. J. Rem. Sens.* 30, 1725. doi: 10.5589/m03-048

Groetsch, P., Foster, R., and Gilerson, A. (2020). Exploring the limits for sky and sun glint correction of hyperspectral above-surface reflectance observations. *Appl. Optics* 59, 2942–2954. doi: 10.1364/AO.385853

Groetsch, P. M. M., Gege, P., Simis, S. G. H., Eleveld, M. A., and Peters, S. W. M. (2017). Validation of a spectral correction procedure for sun and sky reflections in above-water reflectance measurements. *Opt. Express* 25, A742–A761. doi: 10.1364/OE.25.00A742

Harding, L. W., Magnuson, A., and Mallonee, M. (2005). SeaWiFS retrievals of chlorophyll in Chesapeake Bay and the mid-Atlantic bight Estuarine. *Coast. Shelf Sci.* 62, 75–94. doi: 10.1016/j.ecss.2004.08.011

Harmel, T., Gilerson, A., Hlaing, S., Tonizzo, A., Legbandt, T., Weidemann, A., et al. (2011). Long island sound coastal observatory: assessment of above-water reflectance measurement uncertainties using collocated multi and hyper-spectral radiometers. *Appl. Optics* 50, 5842–5860. doi: 10.1364/AO.50.005842

Hieronymi, M., Müller, D., and Doerffer, R. (2017). The OLCI neural network swarm (ONNS): A bio-geo-optical algorithm for open ocean and coastal waters. *Front. Mar. Sci.* 4. doi: 10.3389/fmars.2017.00140

Hlaing, S., Harmel, T., Gilerson, A., Foster, R., El-Habashi, A., Bastani, K., et al. (2013). Evaluation of the VIIRS ocean color monitoring performance in coastal regions. *Remote Sens. Environ.* 139, 398–414. doi: 10.1016/j.rse.2013.08.013

Ioannou, I., Gilerson, A., Ondrusek, M., Foster, R., El-Habashi, A., et al. (2014). Algorithms for the remote estimation of chlorophyll-a in Chesapeake Bay. *Proc. SPIE*, 9111. doi: 10.1117/12.2053753

IOCCG (2006). "Remote sensing of inherent optical properties: fundamentals, tests of algorithms, and applications," in *Reports of the international ocean-color coordinating group, no. 5*. Ed. Z.-P. Lee(Dartmouth, Canada).

IOCCG (2010). "Atmospheric correction for remotely-sensed ocean-colour products," in *Reports of the international ocean-color coordinating group, no. 10, IOCCG*. Ed. M. Wang(Dartmouth, Canada). doi: 10.25607/OBP-101

IOCCG (2019). "Uncertainties in ocean colour remote sensing," in *Reports no. 18 of the international ocean-colour coordinating group*. Ed. F. Mélin (IOCCG, Dartmouth, NS). doi: 10.25607/OBP-696

IOCCG (2021). "Observation of harmful algal blooms with ocean colour radiometry," in *IOCCG report series, no. 20*. Eds. S. Bernard, R. Kudela, L. Robertson Lain and G. C. Pitcher (International Ocean Colour Coordinating Group, Dartmouth, Canada). doi: 10.25607/OBP-1042

Karlson, B., Andersen, P., Arneborg, L., Cembella, A., Eikrem, W., John, U., et al. (2021). Harmful algal blooms and their effects in coastal seas of Northern Europe. *Harmful Algae* 102, 101989. doi: 10.1016/j.hal.2021.101989

Le, C., Hu, C., Cannizzaro, J., English, D., and Muller-Karger, F. E. (2013). Evaluation of chlorophyll-A remote sensing algorithms for an optically complex estuary. *Remote Sens. Environ.* 129, 75-89. doi: 10.1016/j.rse.2012.11.001

Lee, Z., Carder, K. L., and Arnone, R. (2002). Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep water. *Appl. Opt.* 41, 5755–5772. doi: 10.1364/AO.41.005755

Lee, Z. P., Lubac, B., Werdell, J., and Arnone, R. (2009). *An update of the quasi-analytical algorithm (QAA_v5)*. Available online at: www.ioccg.org/groups/Software_OCA/QAA_v5.pdf (Accessed October 1, 2024).

Liu, X., and Wang, M. (2022). Global daily gap-free ocean color products from multi-satellite measurements. *Int. J. Appl. Earth Oserv. Geoinf.* 108, 102714. doi: 10.1016/j.jag.2022.102714

Magnuson, A., Harding, L. W.Jr., Mallonee, M. E., and Adolf, J. E. (2004). Bio-optical model for Chesapeake Bay and the middle Atlantic bight," Estuarine. *Coast. Shelf Sci.* 61, 403–424. doi: 10.1016/j.ecss.2004.06.020

Malinowski, M., Herrera-Estrella, E., Foster, R., Agagliate, J., and Gilerson, A. (2024). Estimation of uncertainties in above-water radiometric measurements from hyperspectral and polarimetric imaging. *Ocean Sens. Monit. XVI* 13061, 1306103. doi: 10.1117/12.3014923

Menendez, A., and Tzortziou, M. (2024). Driving factors of colored dissolved organic matter dynamics across a complex urbanized estuary. *Sci. Total Environ.* 921, 171083. doi: 10.1016/j.scitotenv.2024.171083

Mikelsons, K., and Wang, M. (2019). Optimal satellite orbit configuration for global ocean color product coverage. *Opt. Express* 27, A445–A457. doi: 10.1364/OE.27.00A445

Mikelsons, K., Wang, M., Kwiatkowska, E., Jiang, L., Dessailly, D., and Gossn, J. I. (2022). Statistical evaluation of sentinel-3 OLCI ocean color data retrievals. *IEEE Trans. Geosci. Remote Sens.* 60, 4212119. doi: 10.1109/tgrs.2022.3226158

Mobley, C. D. (1994). *Light and water: radiative transfer in natural waters* (San Diego, CA: Academic Press).

Morel, A. (1974). Light and water: radiative transfer in natural waters. In: N. G. Jerlov and E. S. Nielsen (Eds.), *Optical aspects of oceanography*. New York: Academic Press, pp. 1-24.

Moses, W. J., Gitelson, A. A., Berdnikov, S., and Povazhnyy, V. (2009). Satellite estimation of chlorophyll-a concentration using the red and NIR bands of MERIS—The azov sea case study. *IEEE Geosci. Remote Sens. Lett.* 6, 845–849. doi: 10.1109/LGRS.2009.2026657

Nechad, B., Ruddick, K., and Park, Y. (2010). Calibration and validation of a generic multi-sensor algorithm for mapping of total suspended matter in turbid waters. *Remote Sens. Environ.* 114, 854–866. doi: 10.1016/j.rse.2009.11.022

Neil, C., Spyrakos, E., Hunter, P. D., and Tyler, A. N. (2020). Corrigendum to "A global approach for chlorophyll-a retrieval across optically complex inland waters based on optical water types. *Remote Sens. Environ.* 229, 159–178. doi: 10.1016/j.rse.2019.04.027

Ocean Optics Protocols for Satellite Ocean Color Sensor Validation. (2003). *Ocean optics protocols for satellite ocean color sensor validation*, NASA/TM-2003-21621. Volume 5. G. S. Fargion and J. L. Mueller (Eds.) Goddard Space Flight Space Center, Greenbelt, Maryland 20771

O'Reilly, J. E., Maritorena, S., Mitchell, B.G., Siegel, D.A., Carder, K.L., Garver, S.A., et al. (1998). Ocean color chlorophyll algorithms for SeaWiFS. *J. Geophys. Res.* 103, 24937-24953. doi: 10.1029/98JC02160

O'Reilly, J. E., and Werdell, P. J. (2019). Chlorophyll algorithms for ocean color sensors - OC4, OC5 & OC6. *Rem. Sens. Env.* 229, 32-47. doi: 10.1016/j.rse.2019.04.021

Pahlevan, N., Smith, B., Alikas, K., Anstee, J., Barboso, C., Binding, C., et al. (2022). Simultaneous retrieval of selected optical water quality indicators from Landsat-8, Sentinel-2, and Sentinel-3. *Rem. Sens. Env.* 270, 112860. doi: 10.1016/j.rse.2021.112860

Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., et al. (2020). Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach. *Rem. Sens. Env.* 240, 111604. doi: 10.1016/j.rse.2019.111604

Pope, R., and Fry, E. (1997). Absorption spectrum (380-700 nm) of pure waters: II. Integrating cavity measurements. *Appl. Opt.* 36, 87108723. doi: 10.1364/AO.36.008710

Ransibrahmanakul, V., and Stumpf, R. P. (2006). Correcting ocean colour reflectance for absorbing aerosols. *Int. J. Remote Sens.* 27, 1759–1774. doi: 10.1080/01431160500380604

Roesler, C. S., and Boss, E. (2003). Spectral beam attenuation coefficient retrieved from ocean color inversion. *Geophys. Res. Lett.* 30, 1468. doi: 10.1029/2002GL016185

Schaeffer, B. A., Whitman, P., Vandermeulen, R., et al. (2023). Assessing potential of the Geostationary Littoral Imaging and Monitoring Radiometer (GLIMR) for water quality monitoring across the coastal United States. *Mar. pollut. Bull.* 196, 115558. doi: 10.1016/j.marpolbul.2023.115558

Seegers, B. N., Stumpf, R. P., Schaeffer, B. A., Loftin, K. A., and Werdell, P. J. (2018). Performance metrics for the assessment of satellite data products: an ocean color case study. *Opt. Express* 26, 7404–7422. doi: 10.1364/OE.26.007404

Sherman, J., Tzortziou, M., Turner, K. J., Goes, J., and Grunert, B. (2023). Chlorophyll dynamics from Sentinel-3 using an optimized algorithm for ecological monitoring in complex urban estuarine waters. *Intern. J. @ Appl. Earth Observ. Geoinform.* 118, 103223. doi: 10.1016/j.jag.2023.103223

Shi, W., and Wang, M. (2013). Tidal effects on ecosystem variability in the Chesapeake Bay from MODIS-Aqua. *Remote Sens. Environ.* 138, 65–76. doi: 10.1016/j.rse.2013.07.002

Smith, M. E., Robertson Lain, L., and Bernard, S. (2018). An optimized Chlorophyll a switching algorithm for MERIS and OLCI in phytoplankton-dominated waters. *Remote Sens. Environ.* 215, 217-227. doi: 10.1016/j.rse.2018.06.002

Steinmetz, F., Deschamps, P. Y., and Ramon, D. (2011). Atmospheric correction in presence of sun glint: Application to MERIS. *Opt. Express* 19, 9780–9800. doi: 10.1364/OE.19.009783

Stramski, D., Bricaud, A., and Morel, A. (2001). Modeling the inherent optical properties of the ocean based on the detailed composition of the planktonic community. *Appl. Opt.* 40, 29292945. doi: 10.1364/AO.40.002929

Stumpf, R. P., and Pennock, J. R. (1989). Calibration of a general optical equation for remote sensing of suspended sediments in a moderately turbid estuary. *J. Geophys. Res.* 94, 14363–14371. doi: 10.1029/JC094iC10p14363

Stumpf, R. P., and Tyler, M. A. (1988). Satellite detection of bloom and pigment distributions in estuaries. *Remote Sens. Environ.* 24, 385–404. doi: 10.1016/0034-4257(88)90014-4

Sydor, M., and Arnone, R. A. (1997). Effect of suspended particulate and dissolved organic matter on remote sensing of coastal and riverine waters. *Appl. Opt.* 36, 69056912. doi: 10.1364/AO.36.006905

Tango, P. J., and Batiuk, R. A. (2016). Chesapeake Bay recovery and factors affecting trends: Long-term monitoring, indicators, and insights. *Region. Stud. Mar. Sci.* 4, 12–20. doi: 10.1016/j.rsma.2015.11.010.2016

Turner, K. J., Tzortziou, M., Grunert, B. K., Goes, J., and Sherman, J. (2022). Optical classification of an urbanized estuary using hyperspectral remote sensing reflectance. *Optics Express* 30, 41590–41612. doi: 10.1364/OE.472765

Twardowski, M. S., Boss, E., Macdonald, J. B., Pegau, W. S., Barnard, A. H., and Zaneveld, J. V. (2001). A model for estimating bulk refractive index from the optical

backscattering ratio and the implications for understanding particle composition in case I and case II waters. *J. Geoph. Res.* 106, 1412914142. doi: 10.1029/2000JC000404

Tzortziou, M., Herman, J. R., Gallegos, C. L., Neale, P. J., Subramaniam, A., Harding, L. W. Jr., et al. (2006). Bio-optics of the Chesapeake Bay from measurements and radiative transfer closure. *Estuarine Coast. Shelf Sci.* 68, 348-362. doi: 10.1016/j.ecss.2006.02.016

Voss, K. J. (1992). A spectral model of the beam attenuation coefficient in the ocean and coastal areas. *Limnol. Oceanogr.* 37, 501509. doi: 10.4319/lo.1992.37.3.0501

Wang, M. (2006). Effects of ocean surface reflectance variation with solar elevation on normalized water-leaving radiance. *Appl. Opt.* 45, 4122–4128. doi: 10.1364/AO.45.004122

Wang, M., and Jiang, L. (2018). VIIRS-derived ocean color product using the imaging bands. *Remote Sens. Environ.* 206, 275–286. doi: 10.1016/j.rse.2017.12.042

Wang, M., Jiang, L., Son, S., Liu, X., and Voss, K. J. (2020). Deriving consistent ocean biological and biogeochemical products from multiple satellite ocean color sensors. *Opt. Express* 28, 2661–2682. doi: 10.1364/OE.376238

Wang, M., and Son, S. (2016). VIIRS-derived chlorophyll-a using the ocean color index method. *Remote Sens. Environ.* 182, 141–149. doi: 10.1016/j.rse.2016.05.001

Werdell, P. J., and Bailey, S. W. (2005). An improved *in-situ* bio-optical data set for ocean color algorithm development and satellite data product validation. *Rem. Sens. Env.* 98, 122-140. doi: 10.1016/j.rse.2005.07.001

Werdell, P. J., Behrenfeld, M. J., Bontempi, P. S., Boss, E., Davis, E.T., Franz, B. A., et al. (2019). The Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission: Status, science, advances," Bull. *Am. Meteorol. Soc* 100, 1775–1794. doi: 10.1175/BAMS-D-18-0056.1

Werther, M., Odermatt, D., Simis, S. G. H., Gurlin, D., Daniel, S. F., Jorge, D. S. F., et al. (2022). Characterizing retrieval uncertainty of chlorophyll-*a* algorithms in oligotrophic and mesotrophic lakes and reservoirs. *ISPRS J. Photogramm. Remote Sens.* 190, 279–300. doi: 10.1016/j.isprsjprs.2022.06.015

Windle, A. E., Evers-King, H., Loveday, B. R., Ondrusek, M., and Silsbe, G. M. (2022). Evaluating atmospheric correction algorithms applied to OLCI Sentinel-3 data of Chesapeake Bay Waters. *Remote Sens.* 14, 1881. doi: 10.3390/rs14081881

Wolny, J. L., Tomlinson, M. C., Schollaert Uz, S., Egerton, T. A., McKay, J. R., Meredith, A., et al. (2020). Current and future remote sensing of harmful algal blooms in the Chesapeake Bay to support the shellfish industry. *Front. Mar. Sci.* 7, 337. doi: 10.3389/fmars.2020.00337

Wynne, T. T., Meredith, A., Briggs, T., and Litaker, W. (2018). Harmful algal bloom forecasting branch ocean color satellite imagery processing guidelines. *NOAA Tech. Memo. NOS NCCOS* 252, 48. doi: 10.25923/twc0-f025

Wynne, T. T., Tomlinson, M. C., Briggs, T. O., Mishra, S., Meredith, A., Vogel, R. L., et al. (2022). Evaluating the efficacy of five chlorophyll-a algorithms in chesapeake bay (USA) for operational monitoring and assessment. *J. @ Mar. Sc. Eng.* 10, 1104. doi: 10.3390/jmse10081104

Zheng, G., Stramski, D., and DiGiacomo, P. M. (2015). A model for partitioning the light absorption coefficient of natural waters into phytoplankton, nonalgal particulate, and colored dissolved organic components: A case study for the Chesapeake Bay. *J. Geophys. Res. Oceans* 120, 2601–2621. doi: 10.1002/2014JC010604

Zibordi, G., Holben, B. N., Talone, M., D'Alimonte, D., Slutsker, I., Giles, D. M., et al. (2021). Advances in the ocean color component of the aerosol robotic network (AERONET-OC). *Ocean Technol.* 38, 725–746. doi: 10.1175/JTECH-D-20-0085.1

Zibordi, G., Kwiatkowska, E., Melin, F., Talone, M., Cazzaniga, I., Dessailly, D., et al. (2022). Assessment of OLCI-A and OLCI-B radiometric data products across European seas. *Remote. Sens. Environ.* 272, 112911. doi: 10.1016/j.rse.2022.112911

Zibordi, G., Mélin, F., Berthon, J. F., Holben, B., Slutsker, I., Giles, D., et al. (2009). AERONET-OC: A network for the validation of ocean color primary products. *J. Atmos. Ocean Technol.* 26, 1634–1651. doi: 10.1175/2009JTECHO654.1

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 04 December 2024
DOI 10.3389/fmars.2024.1463531

OPEN ACCESS

EDITED BY
Chao Chen,
Suzhou University of Science and Technology, China

REVIEWED BY
Li Yineng,
Chinese Academy of Sciences (CAS), China
Ming Li,
National University of Defense Technology, China

*CORRESPONDENCE
Lei Zhang
stone333@tom.com;
333_stone@sina.com

RECEIVED 12 July 2024
ACCEPTED 15 November 2024
PUBLISHED 04 December 2024

CITATION
Ma X, Zhang L, Xu W and Li M (2024) AB-LSTM: a mesoscale eddy feature prediction method based on an improved Conv-LSTM model. *Front. Mar. Sci.* 11:1463531. doi: 10.3389/fmars.2024.1463531

AB-LSTM: a mesoscale eddy feature prediction method based on an improved Conv-LSTM model

Xiaodong Ma, Lei Zhang*, Weishuai Xu and Maolin Li

Department of Military Oceanography and Surveying, Dalian Naval Academy, Dalian, China

Mesoscale eddies are the most important mesoscale phenomena in the oceans, and determining how to predict their spatial and temporal characteristics is a very challenging task. Most previous studies focused on the accuracy of full-domain prediction and ignored the accuracy of single-eddy prediction. To solve this problem, in this paper, we first apply multi-year sea surface height data to produce a spatiotemporal sequence sample dataset with a bidirectional prediction mechanism. Then, we introduce an adversarial generative mechanism through stacked spatiotemporal prediction blocks and rely on the strong generative ability of the generative adversarial network models to construct an adversarial bidirectional long- and short-term memory model (AB-LSTM). Next, the mesoscale eddy mixing algorithm is used to extract the matching eddy pair features from the real and predicted data, and several evaluation metrics are used to conduct error analysis. The experiments yield the following results. Prediction sequence days 1–7: the root mean square error (RMSE) values are 1.97–7.70 cm, the structural similarity index (SSIM) values are >0.61, the accuracy is >54.6%, and the eddy centre distance error is 6.34 km. The result is 11.61 km, which is consistent with many spatiotemporal prediction models and passes the generalisation test in many different sea areas. Finally, we carry out single eddy prediction on the basis of the evaluation of the entire prediction of the sea surface height and also obtain a more satisfactory experimental effect. This method has a better prediction ability than the original spatiotemporal method and has a certain reference significance for mesoscale eddy spatiotemporal feature prediction technology and subsequent underwater reconstruction.

KEYWORDS

mesoscale eddies, spatiotemporal sequence prediction, generative adversarial networks, deep learning, sea surface height prediction, long short-term memory

1 Introduction

Mesoscale eddies (MEs) are a special phenomenon that widely occurs in the oceans. Their spatial scales are usually tens and hundreds of kilometres, and their lifetimes vary from tens of days to hundreds of days (Chelton et al., 2011). MEs are widely distributed in the global oceans and have become an important topic in research on ocean dynamics. According to the rotational direction of the eddy, mesoscale eddies can be divided into two categories: cold eddies (cyclonic eddies) and warm eddies (anticyclonic eddies). In the Northern Hemisphere, cyclonic eddies (CEs) rotate counterclockwise and anticyclonic eddies (AEs) rotate clockwise. In the Southern Hemisphere, these two types of eddies rotate in opposite directions (Zhang et al., 2013). These types of eddies are widely distributed in the global oceans. This rotation not only affects the fluid motion inside the eddy but also has a significant impact on the ocean's thermohaline properties. Mesoscale eddies have an all-encompassing effect on the marine environment. By adjusting local water masses, they cause a huge difference in the thermohaline properties inside and outside of their area. This difference not only affects the pattern of ocean circulation but also influences the exchange of materials and energy transfer in the ocean (Dong et al., 2014). In addition, mesoscale eddies have a significant impact on marine environment variability and are important drivers of dynamic changes in marine ecosystems. The characteristics of mesoscale eddies are particularly evident in specific oceanic regions, such as the Kuroshio Extension (KE) region. Detailed statistics presented by Itoh et al. (Itoh and Yasuda, 2010) indicate that the northern side of the KE is dominated by a large number of anticyclonic eddies, and these eddies usually have long life cycles. However, on the southern side of the KE and near the flow axis, there are more CEs, and these eddies usually have stronger intensities. Further analysis has revealed that more than 85% of the anticyclonic eddies have high-salt warm cores, whereas only 15% of the anticyclonic eddies have cold cores. These features not only reveal the unique nature of the mesoscale eddies in the KE region but also provide important clues for understanding dynamic ocean processes in this region.

With the launch of ocean observation satellites, abundant large-scale, long time-series, and high-precision ocean remote sensing observation data have been obtained and processed, among which long time-series observation data accumulated through many years of observations have been widely used in analyses and forecasts of oceanic phenomena (Oka and Qiu, 2012; Qiu and Chen, 2013). Liu et al. (Liu et al., 2012) conducted a multi-year statistical analysis of the number, life cycle, amplitude, and radius of mesoscale eddies in the North West (NW) Pacific Ocean. Wang et al. (Wang et al., 2016) found that the interannual characteristics of the KE region may be affected by the instability of the main flow axis of the KE under the effect of the topography, and the results of their experiment were also affected by the instability of the main flow axis of the Kuroshio under the effect of the topography. Qiu et al. (Qiu and Chen, 2005) used the linear vorticity dynamics method to back-project the high- and low-pressure signals and reached the conclusion that the changes in the circulation characteristics of the KE are associated with the high- and low-pressure anomalies in the eastern North Pacific Ocean. In terms of prediction of the characteristics of mesoscale eddies, roughly classified, most scholars have adopted two approaches. The first is to make predictions using ocean numerical prediction models. Shriver et al. (Shriver et al., 2007) successfully improved the resolution of the prediction system by combining the Naval Layered Ocean Model (NLOM) with the optimal interpolation method, which in turn enhances the accuracy of the ME prediction. Trott et al. (Trott et al., 2023) used the hybrid coordinate ocean model (HYCOM) to simulate future sea-level anomaly (SLA) data and then adopted an SLA-based identification technique to identify MEs and predict their future distribution. The second method is to make predictions that are purely data-driven. This type of method can be subdivided into the direct prediction of ME features (often multi-feature one-dimensional sequence prediction). For example, Ashkezari et al. (Ashkezari et al., 2016) successfully predicted ME lifetimes under stable evolutionary conditions by employing an extreme random forest regression method. Wang et al. (Wang et al., 2020) combined extreme random trees and a long short-term memory (LSTM) network based on mesoscale eddy trajectory and feature datasets to predict several key features, including the latitude and longitude coordinates. Wang et al. (Wang et al., 2021) incorporated meso-historical latitude and longitude sequence data, sea surface height data, sea surface temperature data, and other additional information using a gated recurrent unit (GRU) network combined with a temporal attention mechanism to improve the prediction accuracy of the future centre coordinates of the ME. Ge et al. (Ge et al., 2023) developed a neural network for predicting the trajectory of an ME in compliance with the physical constraints, providing a more reliable and comprehensible method for the prediction of the trajectories of MEs. Another prediction method is to reconstruct a large sea surface height field (2-D) and accordingly to use a mesoscale eddy identification algorithm to obtain mesoscale eddy features in the predicted spatiotemporal sequence. For example, Ma et al. (Ma et al., 2019) obtained an accuracy higher than that of HYCOM for predicting the 7-day sea surface height field using a more mature convolutional LSTM. Nian et al. (Nian et al., 2021) proposed a neural network equipped with a Memory In Memory (MIM) model and a spatial attention module and obtained higher experimental results than those of many spatiotemporal prediction methods. However, according to the current state of research, the limitations of numerical modelling methods in terms of prediction performance should not be ignored. These limitations mainly stem from the nonlinear nature of MEs and the sensitivity of numerical models to initial conditions. Furthermore, these models mainly focus on the prediction of the marine environment rather than directly targeting the ME, so it is difficult to achieve a direct prediction. However, the pure data-driven approach has a lower demand for the initial field, and the current sea surface height observation data have the natural advantages of being large, continuous, and accurate, making the data sufficient to support the model computation. This also lays a solid foundation for the pure data-driven deep learning network prediction model.

The spatial and temporal smoothing properties of mesoscale eddy trajectory and feature prediction enable continuous

observations with a high accuracy, which often causes the spatial and temporal properties between sequence units to have a nonlinear correlation. However, previous studies tended to focus on the predecessor sequence to the successor sequence prediction, which inevitably leads to the propagation of the errors generated by the predecessor prediction resulting in backward cumulative propagation. Although Nian et al. (Nian et al., 2021) utilized corresponding improvement measures for the non-stationary state and error accumulation problems in sea surface height anomaly (SLA) prediction, including optimising the memory and planned sampling methods, and achieved lower prediction errors, the error accumulation effect still occurred and was significant. This was due to the fact that the planned sampling method is only used to correct the weights via jump verification during the learning process, thus turning the continuous error into the accumulation of the stage error, rather than considering the entire range of errors in the prediction sequence as a whole. In addition, since most long-lived mesoscale eddies (more than 7 days) have strong continuity and physical interpretability of the sea surface height field with and without eddy features, we can make predictions from past measurements and can also make predictions from past measurements in the reverse direction. However, the related work has not been carried out so far. Currently, the models commonly used for spatiotemporal prediction are generally based on stacked recurrent neural network (RNN) models or LSTM models. Thus, the former links the correlation between the temporal and spatial attributes, while the latter is more prominent in solving the challenge of gradient explosion, leading to its wider use compared with the RNN. However, native LSTM models tend to focus more on non-Markovian attributes in the time series rather than spatial feature variations in dealing with long time-series prediction problems. For mesoscale eddy prediction tools that are time-varying and highly dependent on variations in spatial feature attributes (Yunbo Wang et al., 2017), one or the other is important. Second, the mesoscale eddy prediction process is often accompanied by eddy generation and elimination, as well as fusion, and existing prediction tools pay more attention to the description of high-value features rather than those of low-level features, which is acceptable in semantic recognition-related applications, but neither of them can be neglected in mesoscale eddy prediction. To solve the problem of the continuity of the prediction caused by unidirectional inputs and the problem of complex spatiotemporal feature description, in this paper, we propose an adversarial bidirectional LSTM (AB-LSTM) and a set of evaluation criteria for mesoscale eddy prediction, which obtained a good comparison effect compared with various spatio-temporal prediction models and numerical ocean prediction models.

2 Data and methods

2.1 Data

2.1.1 AVISO satellite altimeter data

The SLA data used in this paper were obtained from a gridded product provided by the Satellite Ocean Archive Data Centre (AVISO) of the Centre national d'études spatiales (CNES). This dataset combines altimetry data from several satellites, such as Jason-1, Topex/Poseidon, Envisat, GFO, and ERS-1&2, interpolated to a 1/4°×1/4° grid spatial resolution on the Mercator projection. The temporal resolution is interpolated from the original resolution of 7 d to 1 d, the spatial range of the selected data is 25–45°N, 150–170°E, and the time span is from January 1993 to December 2022. These data have been widely used by many scholars (Dong et al., 2014; Duo et al., 2019; Eden and Dietze, 2009), are the most important sample and training data used in this paper, and are also an important indicator for evaluating the quality of the prediction data.

2.1.2 Marine model data

The HYCOM is a data-assimilated hybrid isodensity sigma pressure (generalised) coordinate ocean model (Chassignet et al., 2009, 2007). The subset of HYCOM global sea surface height forecasts hosted in GEE (Google Earth Engine) has been plugged into a 1/12 degree latitude/longitude grid and has been widely used in several previous studies (Metzger et al., 2010; Wallcraft et al., 2007).

2.2 Research methods

2.2.1 Mesoscale eddy identification methods

Since the launch of the T/P satellite on 25 September 1992 and the output of data, the study of ocean mesoscale phenomena using ocean altimetry data has been taking place for more than 30 years. Mesoscale eddy identification algorithms have attracted the attention of several scholars, who have successively proposed physical parameters (Isern-Fontanet et al., 2004), flow field geometry (McWilliams, 2016; Nencioli et al., 2010), and machine vision algorithms (Franz et al., 2018; Xu et al., 2019). Each of the above-described algorithms has its own advantages, and in combination with the reality of this paper, in this paper, we refer to Ma et al (Ma et al., 2024).'s hybrid algorithm that combines flow field geometry and closed contours as the mesoscale eddy identification algorithm. Before carrying out the identification process of the hybrid algorithm, we need to convert the SLA data into the geostrophic flow field, which is calculated as follows:

$$u = -\frac{g}{f}\frac{\partial h}{\partial y}, \quad v = -\frac{g}{f}\frac{\partial h}{\partial x} \tag{1}$$

where u and v are the latitudinal and longitudinal components of the geostrophic anomalies, respectively, g is gravitational acceleration, f is the Koch parameter, and h is the height of the sea surface anomaly.

The flow field geometry method is based on the geometric characteristics of mesoscale eddies, which are defined as regions with rotating velocity vectors, a centre at the velocity extremum, and symmetrically rotating surrounding vectors. The SLA closure curve method focuses on the detection of sea surface altitude closure curves, which reduces the likelihood of non-closed eddies. To reduce the effect of the subjectivity of the sea surface height difference threshold and to balance the identification effect with

the subjective threshold sensitivity, a hybrid algorithm that combines the two methods is used to analyse the sea surface flow field and the SLA data. When the goal is to detect mesoscale eddy pairs with the largest overlapping boundaries, the stable identification of the same eddy using both identification methods is determined by setting generic custom thresholds (intersecting area more than 50% and eddy centre distance of less than 1/12°). The eddy centre of this eddy determined using the flow field geometry method is considered the actual centre (Figure 1).

In addition, in order to demonstrate the advantages of the recognition effect of the hybrid algorithm, 1000 days were randomly selected from the sample data set (daily sea surface height data within the time span of the data), and the flow field geometry method, closed contour method and hybrid recognition algorithm were respectively adopted for recognition. In addition, most experts and scholars in this field conducted artificial recognition and judged the recognition effect. The recognition accuracy and the proportion that should be recognized but not recognized were evaluated by horizontal comparison. The results are shown in Table 1.

2.2.2 Determination of input frame data resolution

For the identification of a mesoscale eddy, since all the current data-driven mesoscale eddy identification algorithms are based on feature identification of grid point data, the selection of the region and the determination of the data resolution play crucial roles, and too large or too small a resolution will have a great impact on the eddy identification results. Thus, in this paper, to ensure that the steps of the data extraction, model training, metric evaluation, and testing of the generalisation capability are characterised by continuity and referability, we fixed the study area as 25–45°N, 150–170°E. Since the resolution of the original altimetry data is 1/4°×1/4°, i.e., the dimension of the data in this part of the region is 80×80, to retain the details of the original data and facilitate the construction of the model, we interpolate all of the input–output data to 128×128 using the *Akima* (Akima, 1970) interpolation algorithm.

2.2.3 Evaluation metrics for predicting mesoscale eddy features

In previous mesoscale eddy predictions, most scholars have tended to use the sea surface height forecast error and the mesoscale eddy trajectory prediction as the evaluation metrics and have achieved better experimental results, but these two metrics cannot evaluate the sea surface height prediction in a complete way. Thus, in this subsection, we propose a mesoscale eddy prediction evaluation framework to evaluate the mesoscale eddy prediction metrics in a complete way. It should be noted that the evaluation metrics introduced in this subsection need to be predicated based on the basic information about the eddies obtained using the mesoscale eddy mixing identification algorithm described in Section 2.2.1, except for the root mean square error (*RSME*) and structural similarity index (*SSIM*), which is a metric for regional prediction results.

The characteristics of mesoscale vortices in the prediction can be expressed in a variety of ways, and the most important ones that can be obtained from the sea surface information field can be divided into three categories: The first type is the numerical error index of eddy prediction, which is reflected as the *RSME* index of sea surface height information, which intuitively reflects the overall error level of the predicted results and the real results. The second category is the representation of the number of vortices, because deep learning network is the best solution generated based on probability theory in two-dimensional space-time prediction process, while the application of mesoscale vortices may result in low numerical error and high distortion. For this reason, *Num* index, *Accuracy* index and *Dist* index are introduced. These three indexes can directly show whether the number and location of vortices in the prediction sequence can be accurately expressed without losing the target. The third category is the performance of the overall similarity. We use the *SSIM* index to show the structural similarity of the whole selection area. This consideration is that not only the prediction level of the eddy itself needs to be reflected, but also the complex interaction field around it needs to be well predicted and expressed.

The first metric is the sea surface height prediction error. We use the two-dimensional *RMSE* as the standard for this metric:

$$RMSE = \sqrt{\frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W}(O_a(i,j) - P_a(i,j))^2} \quad i = 1, 2, 3 \ldots H; \; j = 1, 2, 3 \ldots W, \tag{2}$$

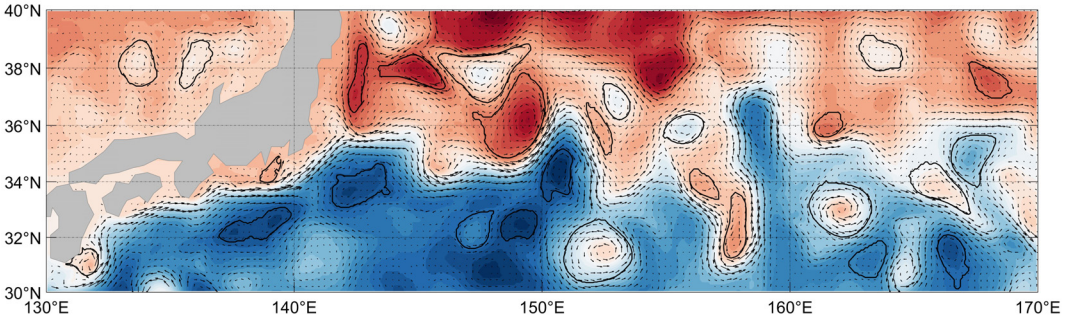


FIGURE 1
Schematic diagram of mesoscale eddy extraction in KE region utilizing the hybrid recognition algorithm.

TABLE 1 Results of horizontal comparison of recognition effects of various recognition methods.

Methods	Recognition Accuracy (%)		Failure to recognize* (%)	
	AE	CE	AE	CE
Flow field geometry	82.12	76.17	1.52	2.27
Physical parameter	73.24	70.56	2.36	3.52
Closed profile	79.38	79.01	**0.62**	**0.95**
Hybrid **(ours)**	**88.32**	**80.17**	1.97	2.34

*Represents eddies that should be detected but are not and the bolded part is the one with better value.

where H and W are the length and width of the data, respectively, and P_a and O_a are the predicted and original data, respectively.

The second metric is the mesoscale eddy prediction hit rate (*Accuracy*) and mesoscale eddy trajectory error (*Dist*). We take the distance of the same eddy centre (km) in the eddy identification results corresponding to the real dataset and the prediction dataset as the daily prediction trajectory error, in which the same eddy hit is discriminated by the fact that the area inside the two eddy profiles matches 75% or more of both the prediction results and real data in the same day. Then, we sum and average the matched eddy centre distances on that day to obtain the trajectory error indicator for that day, which is calculated as follows:

$$Dist = \frac{1}{n}\sum_{m=1}^{n}\sqrt{(x_O(m)-x_P(m))^2+(y_O(m)-y_P(m))^2},$$

$$Accuracy = \frac{N_P}{N_O}*100\,\%, \tag{3}$$

Where n is the total number of identified matching eddies on that day, x_O, y_O, x_P, and y_P are the horizontal and vertical coordinates in the real data identified eddy results, and N_P and N_O are the number of predicted eddies in the region and the number of real eddies, respectively.

The third metric is the sea surface prediction *SSIM*, which is one of the indicators used to measure the structural similarity of the data. When we have two datasets *x, y*, the structural similarity can be defined as follows:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

$$c_1 = (k_1 L)^2\ ,\ \ c_2 = (k_2 L)^2 \tag{4}$$

where μ_x is the mean of x, μ_y is the mean of y, σ_x^2 is the variance of x, $\sigma_y^2$is the variance of y, and σ_{xy} is the covariance of x and y. L is the dynamic range of the pixel value, which is set to 100 in this paper, and k_1 and k_2 are constants, which are set to 0.01 and 0.03, respectively, in this paper.

2.3 Data cleaning

Both ocean observation data and model prediction data have the advantages of wide coverage and clear grid, but they also often contain uncontrollable abnormal data. Due to the various data sources used in this paper, in order to ensure the quality of the data when forming the deep learning sample dataset, We will perform data cleaning on the data used for training, testing, verification and evaluation in this paper. Drawing on the experience of several atmospheric and oceanic researchers, we used the Mahalanobis denoising method (Eq. 5). First, the sequence data of sea surface height is obtained, and the Mahalanobis Distance (D_M) of each two-dimensional grid point in the sequence is calculated. When D_M is greater than three standard deviations of the average distance, the grid point data is considered as "abnormal"; when the number of "abnormal" grid points exceeds 1% of the total grid points, the entire sequence including the two-dimensional grid point data is discarded.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$S = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T$$

$$D_M(x_i) = \sqrt{(x_i-\mu)S^{-1}(x_i-\mu)} \tag{5}$$

3 Model

3.1 Spatiotemporal Long Short-Term Memory Model

Suppose we are monitoring a dynamic system in which each measurement is recorded at all locations in a spatial region represented by an $M \times N$ grid. From a spatial point of view, these P measurements observed at any time can be represented by the tensor $X \in \mathbb{R}^{P\times M\times N}$ (Liu et al., 2018; Wang et al., 2021). From a temporal point of view, the observations at t time steps form a tensor sequence of $X_1, X_2, X_3, X_4, \ldots, X_t$. The spatiotemporal predictive learning problem is to predict the most probable *length* – K sequence in the future given the two previous length-J sequences, including the current observation:

$$\hat{X}_{t+1},\ldots,\ \hat{X}_{t+k} = \underset{X_{t+1},\ldots,X_{t+k}}{arg\,max\,p}(X_{t+1},\ldots,X_{t+k}|X_{t-j+1},\ldots,X_t)\,. \tag{6}$$

Sequence prediction has been a popular research topic in the field of machine learning, and LSTM, as an emerging RNN model with long- and short-term memory, has led to a breakthrough in dealing with the solution of long-term-dependent problems. Shi et al. (Shi et al., 2015) creatively used the input-to-state and state-to-state methods to visually extract the inputs using stacked LSTM layers and achieved

pioneering research results in this field. However, the current problem is that this model needs to continue learning and predicting from the previous state. This means that the continuity prediction will be based on the previous prediction result, which will lead to the accumulation of error and feature bias. To solve this problem, several scholars have improved this model (Kalchbrenner et al., 2017; Patraucean et al., 2015; Villegas et al., 2017). In this paper, we utilize a spatiotemporal long- and short-term memory model (ST-LSTM) (Wang et al., 2022) as the basis of the generation of the model. Based on the stacking technique of the convolutional LSTM (Conv-LSTM), the model obtains higher experimental results than other models by proposing spatiotemporal memory flow and memory transfer across layers in several prediction results. The model's architecture is shown in Figure 2.

The formulas are as follows:

$$G_t = tanh(W_{xG} * X_t + W_{HG} * H_{t-1} + b_G)$$

$$I_t = \sigma(W_{xI} * X_t + W_{HI} * H_{t-1} + b_I),$$

$$F_t = \sigma(W_{xF} * X_t + W_{HF} * H_{t-1} + b_F),$$

$$C_t = F_t \odot C_{t-1} + I_t \odot G_t,$$

$$\boldsymbol{g_t = tanh(W_{xg} * X_t + W_{Mg} * M_{t-1} + b_g)},$$

$$\boldsymbol{i_t = \sigma(W_{xi} * X_t + W_{Mi} * M_{t-1} + b_i)},$$

$$\boldsymbol{f_t = \sigma(W_{xf} * X_t + W_{Mf} * M_{t-1} + b_f)},$$

$$\boldsymbol{M_t = f_t \odot M_t + i_t \odot g_t},$$

$$\boldsymbol{O_t = \sigma(W_{xO} * X_t + W_{HO} * H_{t-1} + W_{CO} * C_t + W_{MO} * M_t + b_O)},$$

$$\boldsymbol{H_t = O_t \odot tanh(W_{1\times1} * [C_t, M_t])}, \quad (7)$$

where σ is the activation function, W corresponds to the process weight of the corner scale, b is the bias term (distinguished by the corner scale), X is the input sequence, C is the output cell, and H is the hidden state. The most important feature of the ST-LSTM model is that the memory cell is divided into two parts, namely, the classical C_t temporal cell and the M_t spatio-temporal cell, and they are distinguished in the level of the data flow. The C_t stream is passed continuously between the same corresponding layers of different stacks according to the classical Conv-LSTM. The M_t stream is first passed layer by layer in the same stack, repeated as the input of the next stack, and finally reduced to the same dimension by a 1×1 convolutional gate and outputted as H_t. This is different from the spatiotemporal memory transfer method of the classical Conv-LSTM to a large extent.

3.2 Generative adversarial network models

The main idea of the basic model of the generative adversarial network (GAN) is to make the two neural networks continuously play the binary extremely large and extremely small game, during which the model gradually learns the real sample distribution. In general, the training is considered complete when the two networks reach a Nash equilibrium in their want confrontation (Goodfellow et al., 2014).

The basic GAN model is shown in Figure 3. The input of the generator network (denoted as G) is a random variable (denoted as z) from the hidden space (denoted as p_z) and the output of the generator samples, the training goal of which is to improve the similarity between the generator samples and real samples, so that they are indistinguishable from those of the discriminator (denoted as D) network, i.e., to make the distributions of the generator samples (denoted as p_g) and real samples (denoted as p_{data}) as identical as possible. The training objectives of the native GAN network can be summarized as follows: to minimize the distance between p_g and p_{data} and to maximize the accuracy of the samples discriminated by D, i.e., the value of $D(x)$ tends to be 1 and the value of $D(x')$ tends to be 0. This leads to the basic GAN network objective function expression:

$$\frac{min}{G}\frac{max}{D} E_{x\sim p_{data(x)}}[logD(x)] + E_{z\sim p_{z(z)}}[\log(1 - D(G(z)))]. \quad (8)$$

3.3 Adversarial bidirectional long- and short-term memory models

To solve the problems of the LSTM, namely, unidirectional prediction error and continuous accuracy of the spatiotemporal prediction, we embed a 4-layer stacked ST-LSTM model as the generative unit into the adversarial network model as the core of the generator. Then, we divide the generator inputs into forward spatiotemporal sequence inputs and inverse spatiotemporal sequence inputs and control the input streams of the two according to the discriminative results of the discriminators in a training cycle to achieve effective bi-directional training (Figure 4). To increase the learning ability of the overall trend, we train a global discriminator (Iizuka et al., 2017) to discriminate whether the output is true. The purpose of constructing the global discriminator is to strengthen the ability of the discriminator to identify the overall characteristics of the input region and to emphasise the importance of guiding the model to pay more attention to the overall trend of the sea surface data. The global discriminator consists of five consecutive convolutional layers, each of which has a step size of 2. It uses a fully connected layer and a sigmoid output layer to process the input data of size 128 × 128 into a high-dimensional vector, which is then transformed into a continuous and normalised real probability distribution by a fully connected layer and a sigmoid transfer function.

In this paper, we use a total of 10,000 days of sea surface altimetry data from 1 January 1993 to 19 May 2020 as the training (first 90%) and validation datasets (second 10%), and the sea surface altimetry data from 20 May 2020 to 20 May 2022 as the model generalisation test datasets (validation and testing sets). We process each of the three datasets into time-series blocks with a length of 10 days (structure 3-4-3: the first number is the length of the forward

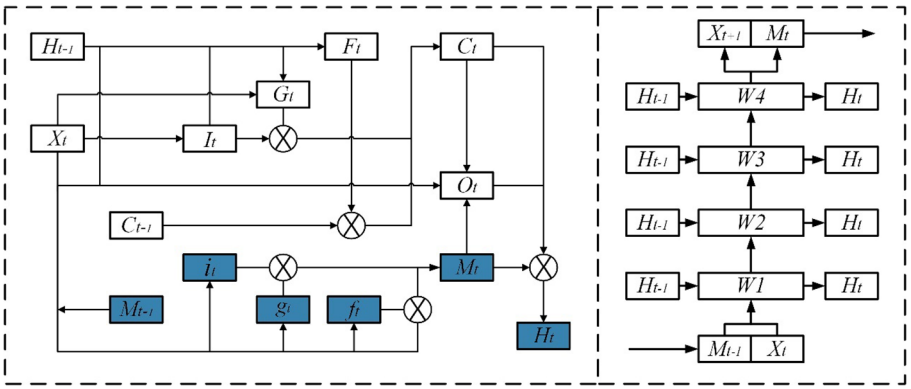


FIGURE 2
Schematic diagrams of the ST-LSTM model (left) and the stacked sequence monolayer (the dark blue marks are space-time fluid cells different from the original Conv-LSTM).

input sequence, the second number is the length of the target prediction sequence, and the third number is the length of the reverse input sequence, as shown in p_{data} in Figure 4), with 3 days of forward prediction input ($p_{forward}$) in each block, 7 days (including 3 days of reverse input) of target prediction data (x), and 3 days of reverse prediction input $p_{backward}$, corresponding to the generation results denoted as $x_{forward}$ and $x_{backward}$. The next batch of inputs in the generator is updated after the discriminator decides whether it is true or false and updates the current batch of generators (ST-LSTM cells) and the discriminator weights. The corresponding objective function is updated to

$$\frac{min}{G}\frac{max}{D}E_{x\sim p_{data(x)}}[logD(x)] + E_{x\sim p_{forward}}[\log(1 - D(G(p_{forward}))) + E_{x\sim p_{backward}}[\log(1 - D(G(p_{backward})))] \quad (9)$$

In the model proposed in this paper, we use the L1+L2 loss function and the Adam optimiser (Kingma and Ba, 2014) for the training, and in the actual training process we pre-train the GAN network and then access the ST-LSTM module. Regarding the setting of the hyperparameters, in general, the learning rate is set within 0.0001–0.1. A learning rate that is too high will make the model training effect poor, while a learning rate that is too low will make the model training convergence slow. Thus, through many adjustments, we determine the learning rate to be 0.0001, the batch is determined to be eight, and the corresponding epoch is appropriately increased to 100,000. If the dataset has a large amount of noise, we should try to minimise β_1 and β_2. Although the average coefficients converge faster, they are more susceptible to noise. In this paper, we set $\beta_1 = 0.9$ and $\beta_2 = 0.999$. All of the experiments are implemented in Pytorch = 3.10 (Paszke et al., 2019) and trained on an NVIDIA RTX4080. Additionally, it should be emphasized here that the parameter Settings of the Adam optimizer in this paper are determined by many attempts in the experiment process and previous experience of Adam optimizer parameters when applying deep learning models in the Marine field.

4 Model evaluation

In this subsection, first, we discuss the effect of different prediction lengths on prediction accuracy to confirm the optimal prediction range of the proposed model. Then, we conduct a multi-

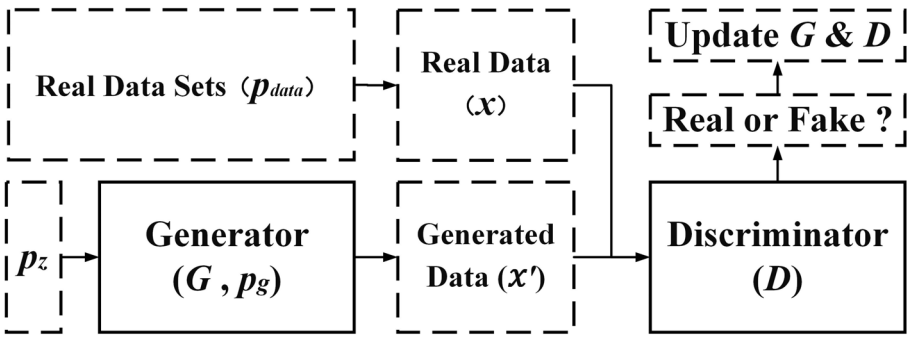


FIGURE 3
Schematic diagram of the basic GAN model.

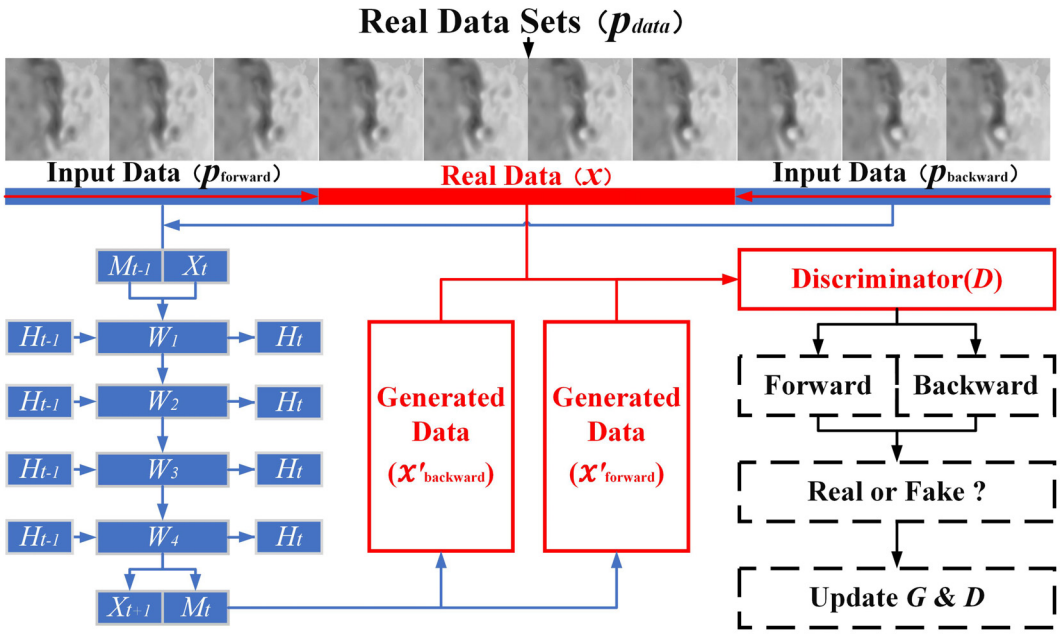


FIGURE 4
Overall schematic diagram of the AB-LSTM model.

criteria comparison with the two modal datasets and a variety of existing spatiotemporal prediction models. Finally, we test the generalisation ability of the model using the day-by-day prediction history data from HYCOM. It should be noted that all input and output data used in this process are first interpolated to 128 × 128 using the interpolation method described in Section 2.2.2. Based on the conclusion of Ma et al. (Ma et al., 2019), the polarity of the mesoscale eddies has a limited effect on the smoothness, as well as the accuracy of the prediction process, so we do not take the issue of eddy polarity into account during the training process, but we do discuss it in the evaluation process.

4.1 Prediction effect

Figure 5 shows the trend of the training loss after 100,000 iterations. The black solid line in the figure is the real value of the training loss from iteration to iteration, and the red line is the higher-order smoothing curve of the black real loss. It can be seen from Figure 5 that the training loss of the model decreases rapidly during the initial training and stabilises at 10,000 iterations. After a long period of small and slow increase, it continues to decrease slowly after 40,000 iterations and finally converges slowly after 90,000 iterations.

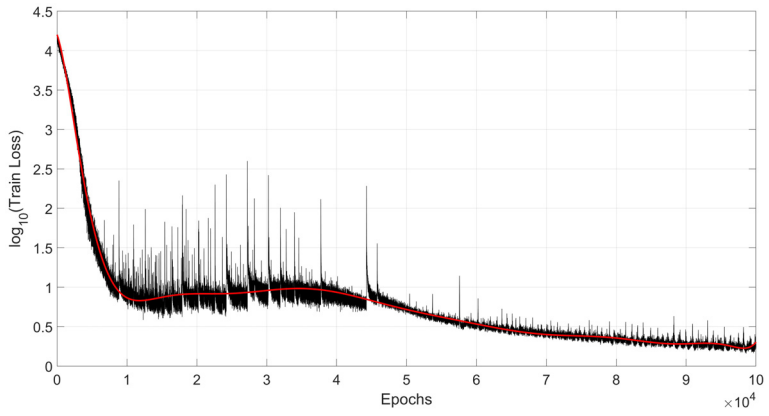


FIGURE 5
Plot of training loss versus number of iterations for the AB-LSTM and the AVISO sea surface height dataset. Due to the large span of the original training loss (Y-axis), to better show the trend of the change, we present it in logarithmic form, which results in the absence of some of the magnitude (the original magnitude is in cm), The black line is the original value of the training loss, and the red line is the error smoothing curve after 5-order Fourier fitting).

Figure 6 shows the effect of the prediction experiment for 7 days for different numbers of iterations. Intuitively, the prediction effect is good. As the number of iterations increases, the model prediction effect continues to improve. The effect tends to stabilise at 50,000 iterations, and the subsequent prediction results are not easily distinguished by the human eye. In addition, it can be seen that the prediction sequences for the different numbers of iterations exhibit good continuity of the overall trend, and the mesoscale eddy characteristics are more obvious, except for the test with 5000 iterations. This indicates that the training process is effective. Figure 7 shows the change trends of the *RMSE* and *SSIM* metrics for the AB-LSTM for the AVISO sea surface height dataset with increasing iteration numbers. It can be clearly seen that the results shown in Figure 7 are highly consistent with the prediction effect shown in Figure 6. This also shows that the selected metrics can accurately reflect the actual performance of the model in terms of the prediction process.

To discuss the effect of the forward and backward inputs on the model training in the AB-LSTM model, in this subsection we set up

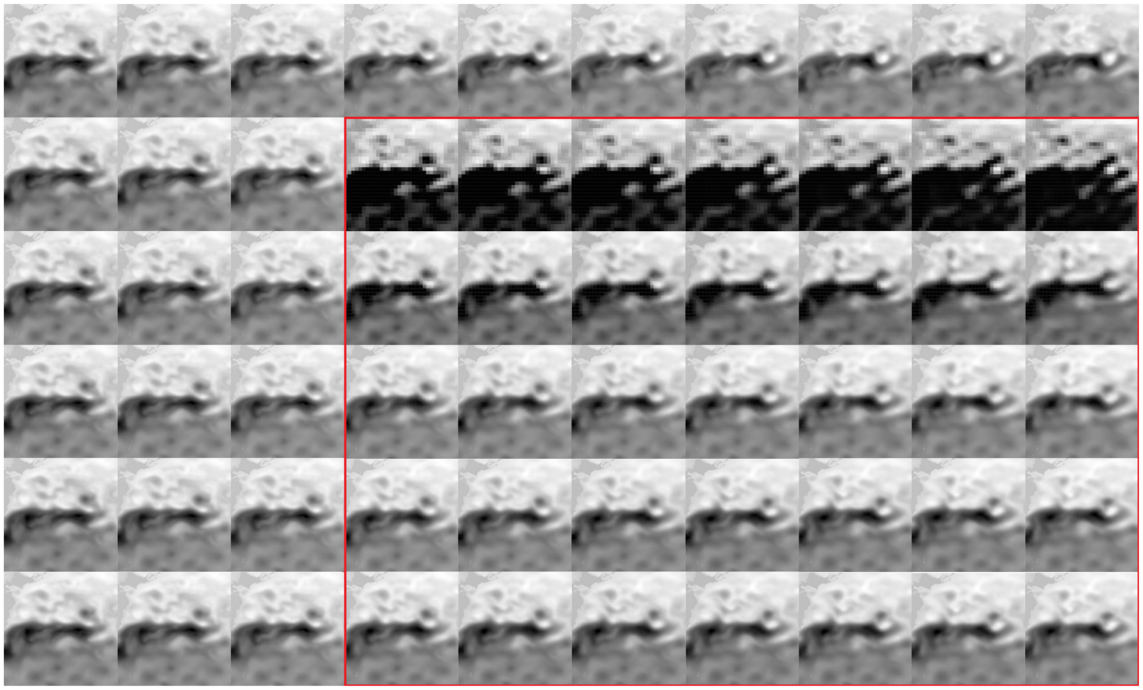


FIGURE 6
Schematic representation of the effect of the prediction experiment with different numbers of iterations. The predicted values are shown in the red boxes.

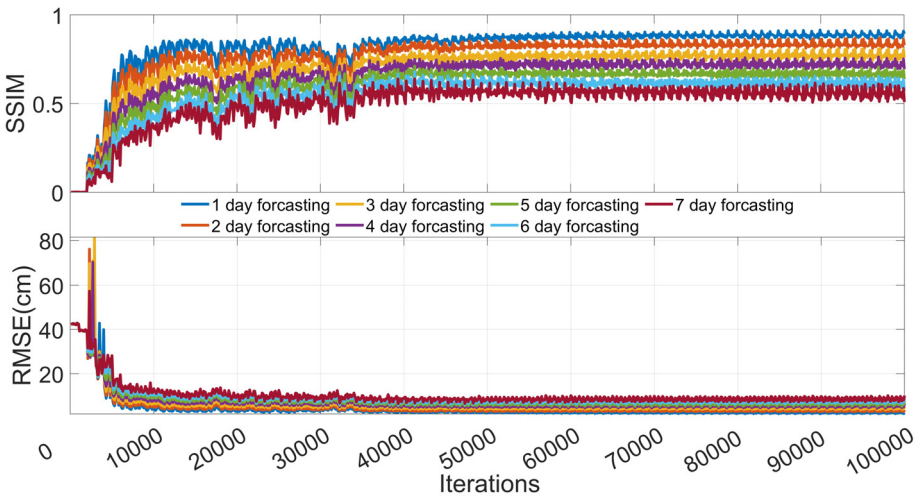


FIGURE 7
Plots of *RMSE* and *SSIM* metrics versus number of iterations for the AB-LSTM and the AVISO sea surface height dataset.

TABLE 2 Quantitative analysis of the effect of the forward and backward input conditions on the prediction.

Assessment Indicators		Forecasting Days						
		1	2	3	4	5	6	7
RMSE (*cm*)	A	1.97	2.90	3.90	4.77	5.80	6.84	7.70
	B	2.42	3.12	4.91	5.11	6.23	8.65	9.39
Accuracy (%)	A	90.0	81.3	72.7	72.7	66.7	57.1	54.6
	B	85.7	75.0	62.5	61.5	52.9	50.0	41.3
SSIM	A	0.91	0.86	0.81	0.76	0.71	0.66	0.61
	B	0.88	0.83	0.78	0.75	0.69	0.64	0.57
Dist (*km*)	A	6.34	6.52	7.40	8.18	9.27	9.80	11.61
	B	6.88	7.00	7.82	8.39	10.55	11.65	12.96

A (forward and backward) and B (forward only).

control experiment groups A (forward and backward) and B (forward only), randomly select 500 sets of experimental data from the sample dataset, which all have 3-4-3 structures, and make a 7-day prediction. The obtained results are averaged within each group according to the day-by-day prediction results (Table 2).

4.2 Model comparison validation

In this subsection, to demonstrate the feasibility and the advantages of the model, we compare the AB-LSTM with the HYCOM model forecast and the FC-LSTM (Srivastava et al., 2015), PredRNN (Wang et al., 2022), and Conv-LSTM (Shi et al., 2015) spatiotemporal prediction methods under the 3-4-3 input block conditions described in Section 4.1 and using the evaluation metrics described in Section 2.2.3. It is worth emphasizing that the horizontal comparison verification of the model should be discussed in different scenarios. For mesoscale vortices, the different properties of vortices and the setting of the research area are very important elements for scene division. Therefore, we will reflect the model verification effect under different research areas in the subsequent regional generalization verification. The model generalization test for differentiating AE and CE in a single eddy prediction scenario is also presented.

Since the prediction results of AE and CE are similar, we consider the polarity of the eddy to be less influential on the comparison experiments, so we will not distinguish between them in this subsection. To avoid the high prediction effect caused by the use of the sample dataset and the inability to effectively compare the results of the experiments, we conduct the experiments on 1000 sets of sea-surface height data that are not included in the sample data. Data structure is still 3-4-3, and the metrics are averaged within the groups. We set the prediction area to the KE region of 25–45°N and 150–170°E. The results are shown in Figure 8.

As can be seen from Figure 7, according to all the computational indexes, the AB-LSTM yields better results. The AB-LSTM's RMSE index increases from 1.97 cm on the first day to 7.70 cm on the

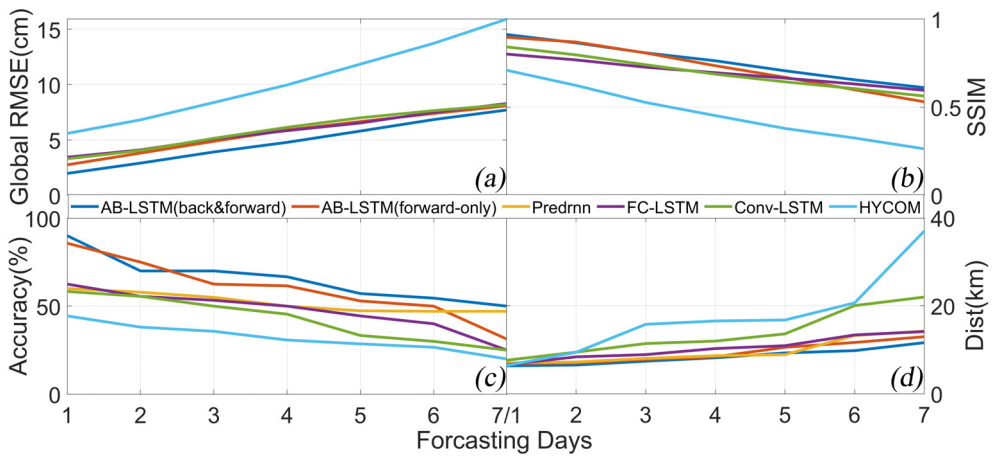


FIGURE 8
Plots of the **(A)** RMSE, **(B)** SSIM, **(C)** accuracy, and **(D)** Dist metrics for a 7-day forecast series for multiple forecasting methods and models.

seventh day, which is the same trend as the other methods, but the values are significantly lower than those of the other prediction methods. The AB-LSTM yields lower RMSE values than the unidirectional model in the control group, which distinguishes between positive and bidirectional inputs. Regarding the SSIM index, the AB-LSTM method has values similar to several of the prediction methods, except that the numerical prediction is within the range of 0.6–0.9 and only differs from the other prediction methods by about 0.05. This is since several of the deep learning-based spatial-temporal prediction models utilized as the control group in this paper are able to achieve very good results in terms of structural similarity, so it is not possible for this metric to clearly reflect the superiority of the AB-LSTM. The AB-LSTM has a much higher Accuracy, with a hit rate of more than 50% during the prediction sequence (days 1–7). This is 10–20% better than those of the other methods. On average, it achieves a 5% higher Accuracy in the control experiments and can distinguish between forward and reverse inputs. This suggests that the forward and reverse inputs are important for the model in the long mesoscale eddy time series prediction. The AB-LSTM model has a slightly better Dist value. The prediction distance error ranges from 6.34 to 11.61 km, which is generally 1–10 km lower than those of the other prediction methods for the 7-day prediction series. The AB-LSTM model with bidirectional input is more accurate in terms of the prediction results after the fourth day compared with the model with only forward input. After the fourth day, the prediction results of the AB-LSTM model with bi-directional inputs are lower, which suggests that the bi-directional inputs have a positive effect on the model in the long-term prediction of mesoscale eddies. Overall, compared with the traditional numerical prediction models' results, the spatial and temporal prediction models that use deep learning algorithms have a greater advantage in terms of the overall prediction error and mesoscale eddy-related prediction indexes. For the sample dataset introduced in this paper, according to all the metrics, the AB-LSTM has the best performance, which directly proves the superiority of the AB-LSTM. In addition, by analysing the experimental results of the control experiment group, it was found that the two-way input training of the confrontation has more advantages and positive significance compared with the one-way input.

4.3 Model generalisation test

Model generalisability refers to the model's ability to adapt to new data, i.e., whether the model can make accurate predictions for data that does not appear in the training set. A model with a strong generalisability can perform well on different datasets, not just on the training set. In summary, generalisability concerns the model's ability to adapt to unknown situations (Liu and Aitkin, 2008). To explore the generalisability of our model, we experimentally validate the AB-LSTM using data from the same region as the data in the training sample set but that are not included in the training and testing sets. We also test the model on data for other sea areas. In this paper, we take the Oyashino Extension (OE, 35–45°N, 140–150°E) region and the North Pacific Subtropical Countercurrent (STCC, 15–25°N, 130–140°E) as the validation areas. It should be noted that in this subsection, the sea surface height data for the OE and STCC regions are processed into 128×128 grids using the data processing method described in Section 2.2.2, and the data span from January 1993 to 31 December 2022. The experimental data for the KE region, which are not included in the training and testing datasets, span from 1 to 30 October 2023 and are processed in the same manner.

As can be seen from Figure 8, the prediction results of the tested models are slightly poorer than the prediction results presented in Section 4.2, but the overall effects are similar, and all of the models yield more stable and good prediction results. The AB-LSTM is still better than the other models in terms of several metrics. For the prediction results for the three regions, the RMSE index remains within the range of 2.25–9.41 cm, which is slightly higher than the prediction results of 1.97–7.70 cm obtained in Section 4.2. The SSIM indicator remains within the range of 0.52–0.85, which is slightly lower than the range of 0.61–0.90 obtained in Section 4.2. The Accuracy remains within the range of 48.35–84.03%, which is slightly lower than the range of 54.60–90.00% obtained in Section 4.2. The Dist remains within the range of 6.71–12.89 km, which is slightly higher than the range obtained in Section 4.2. The possible reason for this result is that the OE region and STCC region are not within the region of the training set, and there may be motion features that are not fully fitted by the model, which may lead to the result that the AB-LSTM fits the KE region data better and the data for the other two regions slightly worse in terms of the prediction effect. The mesoscale eddy recognition algorithm used in this paper has a better recognition ability, but it still has a slightly worse recognition ability. In addition, it still has the possibility of identification error, and the mesoscale eddies identified from the predicted sea surface height data may have the intermittent appearance or disappearance of error, which would lead to problems in estimating the distance deviation of the centre of the mesoscale eddy and will make the error falsely high.

Based on the prediction results presented in Figure 9, several of the models achieve better prediction results in several sea areas, but the performance of the AB-LSTM is the best, which proves that the AB-LSTM model has an acceptable generalization ability for different sea surface height datasets.

4.4 Single eddy prediction effect

Although regional sea surface prediction can reflect the overall prediction effect better, the prediction effect on single eddies is not fully reflected, so in this subsection, we predict multiple single eddies and use the strength at the centre of the eddy (denoted as the SSH in the centre of the eddy in this paper) and the eddy radius to describe them. Thus, the prediction effect of single eddies will be more clearly reflected in the form of data. Figure 10 shows a schematic representation of the evolution of a typical dipole pair over the course of its evolution.

We randomly select 1000 days of data in the sea surface height sample dataset used in this paper as experimental samples, and then we use the AB-LSTM model to make predictions for a period of 7 days according to the 3-4-3 structure. We use the mesoscale eddy mixing

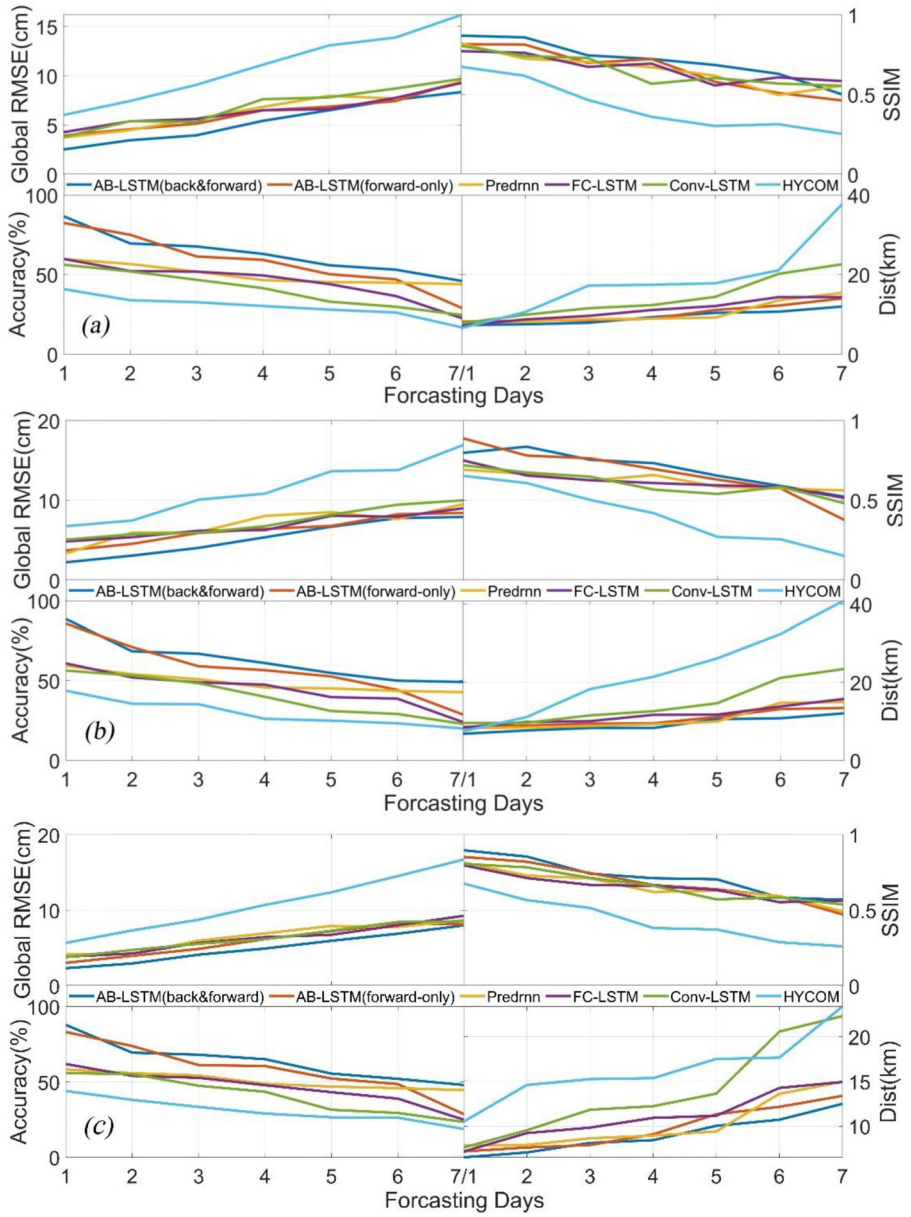


FIGURE 9
Generalisation test of the AB-LSTM model using data for the **(A)** OE region, **(B)** STCC region, and **(C)** KE region.

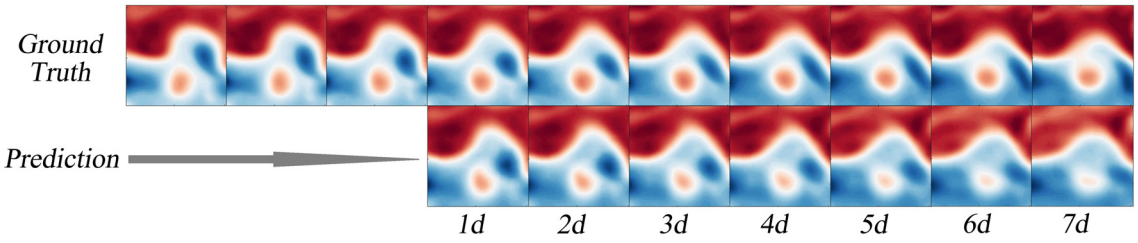


FIGURE 10
Schematic of the 7-day evolution of a pair of dipoles located in the KE region on 1 September 2017. To make the dipole evolution visually clearer, we converted the sea surface height data into a pseudo-colour map with an intercept area of 28–32°N, 146–150°E.

identification algorithm to identify the matched eddy pairs (real vs. predicted). Then, we extract their centre eddy strengths and eddy radii as the control group for the experiments. To avoid the unmatched vortices and matching errors caused by the identification algorithms (described in the previous section) and thus to more effectively reflect the prediction effect, we exclude the matching error terms. Table 3 presents the single eddy prediction errors in the form of within-group averages and the distinction between AEs and CEs.

The Radius metric is the equivalent radius of the identified eddy. As can be seen from Table 4, the AB-LSTM model also achieves relatively good prediction results in single-eddy prediction. In terms of the eddy centre height errors, the AE centre height errors are within the range of 1.12–6.59 cm, and the CE centre height errors are within the range of 1.31–7.86 cm. Overall, both increase as the prediction time increases. However, under the condition of distinguishing between the AEs and CEs, there is not a large difference in the overall errors of the two, which is similar to the conclusion of Wang et al (Wang et al., 2020). In terms of the eddy radius error, it also exhibits an overall prediction result with a trend similar to that of the eddy centre height error, which indicates that the model is more stable in the prediction process. In order to reflect the advantages of the AB-LSTM model in the single-eddy experiment, we conducted experiments according to the same sample collection method, and the results are shown in Table 4.

From Table 3, AB-LSTM is significantly better than the primary Conv-LSTM in single-eddy prediction results, and the experimental results of the AB-LSTM model are continued by the experimental results in Table 2. It is worth emphasizing that we have obtained the horizontal comparison results among multiple models above, so in the horizontal comparison experiment of single eddy prediction, we only compared the two models with similar performance as the AB-LSTM model.

4.5 Additional experiment

In the first few sections of this chapter, the AB-LSTM model proposed by us has achieved a small advantage in the image numerical evaluation index (SSIM, RMSE), and an even greater advantage in the feature prediction of mesoscale vorticity. However, its performance in the prediction error analysis results on the 7th day still makes us worried: Whether the trend of error change shown by AB-LSTM during the 7-day forecast will lead to more drastic changes over a longer forecast time horizon, making the prediction model worse than other spatio-temporal prediction models over a longer time horizon. Since the data input format we selected in the previous paper is 3-4-3 mode, which is not fully applicable for a longer prediction time, we use a longer data input mode here: 3-7-3. As a comparison, we still use RMSE, Accuracy, SSIM and Dist for error quantification (for longer forecast time series, a longer input should be selected).

After comparing the results of the two data input modes, we can find from Table 5 that there is no significant increase in prediction error on the whole. The 3-4-3 input mode has better prediction effect within 3 days, while the 3-7-3 input mode has better lasting prediction ability within a longer prediction period. This error is generally reversed on the fourth or fifth day of prediction, which also shows a relatively easy to understand trend, that is, different input data models are generated under different deep network models, and with the change of its application scenario, its prediction effect will also change.

TABLE 3 Quantification of single eddies during the 7-day forecast period.

Assessment Indicators		Forecasting Days						
		1	2	3	4	5	6	7
Amplitude (*cm*)	AE	1.12	2.54	3.12	3.27	5.67	5.21	6.59
	CE	1.31	2.83	3.76	4.35	6.33	6.34	7.86
Radius (*km*)	AE	6.82	8.32	9.37	10.17	10.62	12.64	17.24
	CE	7.77	9.36	10.24	13.69	15.24	18.56	21.54

TABLE 4 Quantification of single eddies during the 7-day forecast period (different models and no distinction between AE and CE).

Assessment Indicators		Forecasting Days						
		1	2	3	4	5	6	7
Amplitude (*cm*)	AB-LSTM	1.21	2.69	3.44	3.81	6.00	5.78	7.23
	PredRNN	1.52	2.94	3.98	4.22	5.97	6.71	8.64
	Conv-LSTM	2.99	4.12	5.32	6.64	8.24	10.58	12.21
Radius (*km*)	AB-LSTM	7.30	8.84	9.81	11.93	12.93	15.60	19.39
	PredRNN	8.15	8.99	10.10	12.52	13.17	16.02	21.10
	Conv-LSTM	12.98	14.35	16.58	19.71	21.39	24.14	26.54

TABLE 5 Quantification of eddies over a 7-day forecast period (No distinction is made between AE and CE).

Assessment Indicators		Forecasting Days									
		1	2	3	4	5	6	7	8	9	10
RMSE (*cm*)	A	1.97	2.90	3.90	4.77	5.80	6.84	7.70	8.57	9.44	11.01
	B	2.04	3.11	3.97	5.13	5.77	6.37	7.26	8.29	9.17	10.67
SSIM	A	0.91	0.86	0.81	0.76	0.71	0.66	0.61	0.55	0.50	0.43
	B	0.88	0.82	0.78	0.73	0.69	0.65	0.62	0.57	0.53	0.49
Dist(*km*)	A	6.34	6.52	7.40	8.18	9.27	9.80	11.61	12.17	14.25	15.97
	B	7.15	7.39	7.75	8.21	8.69	9.22	10.15	10.98	12.19	14.76

A (3-4-3 input format) and B (3-7-3 input format).

5 Summary and outlook

In this study, we acquired AVISO satellite altimeter data and HYCOM ocean model forecast data as the basis of our work. These data not only provide rich ocean information but also provide the necessary data support for the identification and characterization of mesoscale eddies. Then, we effectively identified and extracted the features of mesoscale eddies utilizing the hybrid mesoscale eddy identification algorithm, which has a better identification effect. We combined it with the sea surface height data and established an evaluation system for mesoscale eddy prediction, which includes four test metrics, namely, the RMSE, SSIM, Dist, and Num.

Subsequently, we combined the time-series prediction advantages of the LSTM model with those of previous studies, utilized the ST-LSTM model as the base generative model, and stacked them to form a prediction network in the same way as the Conv-LSTM. In addition to introducing the generative adversarial network model, which has a strong generative capability, the AB-LSTM model was constructed by embedding the ST-LSTM module into the generator therein. Considering that previous studies have mostly focused on unidirectional sequence prediction without using backward-assisted prediction, we incorporated backward sequence prediction into the input sequences based on the AB-LSTM model and obtained better results than when only unidirectional inputs were utilized. The RMSE was 1.97–7.70 cm, the SSIM was ≥0.61, the Accuracy was ≥54.6%, and the Dist was 6.34–11.61 km. All of the above indicators were better than those of the other models and numerical prediction products, thus achieving the goal of this study. In the training process, we used the Adam optimizer as the hyperparameter container, and through many experiments, we determined that the number of iterations should be 100,000 times and the number of batches should be 8. The experimental results show that the relevant parameters were set reasonably, and a relatively smooth trend was maintained in the training iteration loss. Then, we tested the model's generalization ability using data for a different sea area and new data for the same sea area to achieve data expansion of the non-training testing set. The experimental results show that the AB-LSTM also has a good prediction ability for data that are different from the training test sample dataset. Its prediction ability is only slightly lower than the training sample prediction results according to the indicators, and it is still able to maintain a large improvement compared with the other models. Therefore, the results of the generalization test prove that the AB-LSTM has an acceptable generalization ability. Finally, to address the problem that the full-domain prediction error cannot directly describe the single-eddy prediction effect, we conducted single-eddy prediction analyses using randomly selected pairs of identified vortices. The results show that the eddy polarity has little effect on the prediction effect and that the single-eddy prediction error tends to be smaller than the full-domain prediction error.

Although the AB-LSTM model developed in this study preforms better than other prediction models in terms of the prediction error, it still has some shortcomings. First, the mesoscale eddy identification algorithm used in this paper was found to have discrepancies in terms of matching the real eddy with the predicted eddy, and there is no matching criterion that can be adopted, which leads to the fact that we have no choice but to eliminate the eddy pairs that are incorrectly matched in our single-eddy prediction analysis. To a certain extent, this is not possible in a single eddy prediction analysis. This may make our experimental results better than the real results to a certain extent. Second, more physical parameters should be introduced into the single-eddy prediction instead of only using the eddy centre height and radius to evaluate the error. In the future, we plan to introduce vorticity, shear deformation, and tensile deformation to improve the evaluation of the single-eddy prediction effect. Third, the computational redundancy of the AB-LSTM model is greater than those of several of the prediction models cited in the paper. To achieve better results, the AB-LSTM takes longer to run, occupies more memory, and has more training iterations, which means that our model still needs to be improved in terms of performance. In the next step, we will try to introduce more mesoscale eddy physics information to improve the prediction effect while improving the model.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: AVISO,JCOPE.

Author contributions

XM: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. LZ: Conceptualization, Investigation, Software, Writing – review & editing. WX: Data curation, Methodology, Supervision, Writing – review & editing. ML: Formal analysis, Project administration, Validation, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

We are grateful for the provision of the mesoscale eddy dataset from archiving, validation, and interpretation of satellite oceanographic data (AVISO) (https://www.aviso.altimetry.fr/en/data/products/value-added-products/global-mesoscale-eddy-trajectory-product.html) and the reanalysis data from the hybrid coordinate ocean model (HYCOM) (https://www.hycom.org/), We thank Wang et al. for providing the spatial–temporal long short-term memory (ST-LSTM) model, and other scholars and organizations that helped in the research process (https://github.com/thuml/predrnn-pytorch).

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

References

Akima, H. (1970). A new method of interpolation and smooth curve fitting based on local procedures. *J. ACM (JACM)* 17, 589–602. doi: 10.1145/321607.321609

Ashkezari, M. D., Hill, C. N., Follett, C. N., Forget, G., and Follows, M. J. (2016). Oceanic eddy detection and lifetime forecast using machine learning methods. *Geophysical Res. Lett.* 43, 12,234–212,241. doi: 10.1002/2016GL071269

Chassignet, E. P., Hurlburt, H. E., Metzger, E. J., Smedstad, O. M., Cummings, J. A., Halliwell, G. R., et al. (2009). US GODAE: global ocean prediction with the HYbrid Coordinate Ocean Model (HYCOM). *Oceanography* 22, 64–75. doi: 10.5670/oceanog.2009.39

Chassignet, E. P., Hurlburt, H. E., Smedstad, O. M., Halliwell, G. R., Hogan, P. J., Wallcraft, A. J., et al. (2007). The HYCOM (hybrid coordinate ocean model) data assimilative system. *J. Mar. Syst.* 65, 60–83. doi: 10.1016/j.jmarsys.2005.09.016

Chelton, D., Schlax, M. G., and Samelson, R. M. (2011). Global observations of nonlinear mesoscale eddies. *Prog. Oceanography* 91, 167–216. doi: 10.1016/j.pocean.2011.01.002

Dong, C., McWilliams, J., Liu, Y., and Chen, D. (2014). Global heat and salt transports by eddy movement. *Nat. Commun.* 5, 3294. doi: 10.1038/ncomms4294

Duo, Z., Wang, W., and Wang, H. (2019). Oceanic Mesoscale Eddy detection method based on deep learning. *Remote. Sens.* 11, 1921. doi: 10.3390/rs11161921

Eden, C., and Dietze, H. (2009). Effects of mesoscale eddy/wind interactions on biological new production and eddy kinetic energy. *J. Geophysical Research: Oceans* 114. doi: 10.1029/2008JC005129

Franz, K., Roscher, R., Milioto, A., Wenzel, S., and Kusche, J. "Ocean Eddy Identification and Tracking Using Neural Networks," IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium. (Valencia, Spain: IEEE). (2018) pp. 6887–6890. doi: 10.1109/IGARSS.2018.8519261

Ge, L., Huang, B., Chen, X., and Chen, G. (2023). Medium-range trajectory prediction network compliant to physical constraint for oceanic eddy. *IEEE Trans. Geosci. Remote Sensing.* 61, 1–14, 2023. doi: 10.1109/TGRS.2023.3298020

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* 27.

Iizuka, S., Simo-Serra, E., and Ishikawa, H. (2017). Globally and locally consistent image completion. *ACM Trans. Graphics (ToG)* 36, 1–14. doi: 10.1145/3072959.3073659

Isern-Fontanet, J., Font, J., García-Ladona, E., Emelianov, M., Millot, C., and Taupier-Letage, I. (2004). Spatial structure of anticyclonic eddies in the Algerian basin (Mediterranean Sea) analyzed using the Okubo–Weiss parameter. *Deep Sea Res. Part II: Topical Stud. Oceanography* 51, 3009–3028. doi: 10.1016/j.dsr2.2004.09.013

Itoh, S., and Yasuda, I. (2010). Water mass structure of warm and cold anticyclonic eddies in the western boundary region of the subarctic North Pacific. *J. Phys. oceanography* 40, 2624–2642. doi: 10.1175/2010JPO4475.1

Kalchbrenner, N., Oord, A., Simonyan, K., Danihelka, I., Vinyals, O., Graves, A., et al. (2017). "*Video pixel networks*," in *Proceedings of the 34th International Conference on Machine Learning*, PMLR 70:1771–1779.

Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Liu, C. C., and Aitkin, M. (2008). Bayes factors: Prior sensitivity and model generalizability. *J. Math. Psychol.* 52, 362–375. doi: 10.1016/j.jmp.2008.03.002

Liu, J., Zhang, T., Han, G., and Gou, Y. (2018). TD-LSTM: Temporal dependence-based LSTM networks for marine temperature prediction. *Sensors* 18, 3797. doi: 10.3390/s18113797

Liu, Y., Dong, C., Guan, Y., Chen, D., McWilliams, J., and Nencioli, F. (2012). Eddy analysis in the subtropical zonal band of the North Pacific Ocean. *Deep Sea Res. Part I: Oceanographic Res. Papers* 68, 54–67. doi: 10.1016/j.dsr.2012.06.001

Ma, C., Li, S., Wang, A., Yang, J., and Chen, G. (2019). Altimeter observation-based eddy nowcasting using an improved Conv-LSTM network. *Remote Sens.* 11, 783. doi: 10.3390/rs11070783

Ma, X., Zhang, L., Xu, W., Li, M., and Zhou, X. (2024). A mesoscale eddy reconstruction method based on generative adversarial networks. *Front. Mar. Sci.* 11, 1411779. doi: 10.3389/fmars.2024.1411779

McWilliams, J. C. (2016). Submesoscale currents in the ocean. *Proc. R. Soc. A: Mathematical Phys. Eng. Sci.* 472, 20160117. doi: 10.1098/rspa.2016.0117

Metzger, E. J., Hurlburt, H., Xu, X., Shriver, J. F., Gordon, A. L., Sprintall, J., et al. (2010). Simulated and observed circulation in the Indonesian Seas: 1/12 global HYCOM and the INSTANT observations. *Dynamics Atmospheres Oceans* 50, 275–300. doi: 10.1016/j.dynatmoce.2010.04.002

Nencioli, F., Dong, C., Dickey, T., Washburn, L., and McWilliams, J. C. (2010). A vector geometry–based eddy detection algorithm and its application to a high-resolution numerical model product and high-frequency radar surface velocities in

the Southern California Bight. *J. atmospheric oceanic Technol.* 27, 564–579. doi: 10.1175/2009JTECHO725.1

Nian, R., Cai, Y., Zhang, Z., He, H., Wu, J., Yuan, Q., et al. (2021). The identification and prediction of Mesoscale Eddy variation via memory in memory with scheduled sampling for sea level anomaly. *Front. Mar. Sci.* 8, 753942. doi: 10.3389/fmars.2021.753942

Oka, E., and Qiu, B. (2012). Progress of North Pacific mode water research in the past decade. *J. Oceanography* 68, 5–20. doi: 10.1007/s10872-011-0032-5

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* 32.

Patraucean, V., Handa, A., and Cipolla, R. (2015). Spatio-temporal video autoencoder with differentiable memory. *arXiv*.

Qiu, B., and Chen, S. (2005). Eddy-induced heat transport in the subtropical north pacific from Argo, TMI, and altimetry measurements. *Gayana* 68, 499–501. doi: 10.1175/JPO2696.1

Qiu, B., and Chen, S. (2013). Concurrent decadal mesoscale eddy modulations in the western North Pacific subtropical gyre. *J. Phys. oceanography* 43, 344–358. doi: 10.1175/JPO-D-12-0133.1

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., and Woo, W.-c. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Adv. Neural Inf. Process. Syst.* 28.

Shriver, J., Hurlburt, H. E., Smedstad, O. M., Wallcraft, A. J., and Rhodes, R. C. (2007). 1/32 real-time global ocean prediction and value-added over 1/16 resolution. *J. Mar. Syst.* 65, 3–26. doi: 10.1016/j.jmarsys.2005.11.021

Srivastava, N., Mansimov, E., and Salakhudinov, R. (2015). "*Unsupervised learning of video representations using lstms*," in Proceedings of the 32nd International Conference on Machine Learning, PMLR 37:843–852.

Trott, C. B., Metzger, E. J., and Yu, Z. (2023). Luzon strait mesoscale eddy characteristics in HYCOM reanalysis, simulation, and forecasts. *J. Oceanography* 79, 423–441. doi: 10.1007/s10872-023-00686-5

Villegas, R., Yang, J., Hong, S., Lin, X., and Lee, H. (2017). Decomposing motion and content for natural video sequence prediction. *arXiv*.

Wallcraft, A., Hurlburt, H., Metzger, E., Chassignet, E., Cummings, J., and Smedstad, O. M. (2007). "Global ocean prediction using HYCOM," in *Paper presented at the 2007 DoD High Performance Computing Modernization Program Users Group Conference*. (Pittsburgh, PA, USA). 259–262. doi: 10.1109/HPCMP-UGC.2007.36

Wang, X., Wang, H., Liu, D., and Wang, W. (2020). The prediction of oceanic mesoscale eddy properties and propagation trajectories based on machine learning. *Water* 12, 2521. doi: 10.3390/w12092521

Wang, X., Wang, X., Yu, M., Li, C., Song, D., Ren, P., et al. (2021). MesoGRU: Deep learning framework for mesoscale eddy trajectory prediction. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5. doi: 10.1109/LGRS.2021.3087835

Wang, Y., Long, M., Wang, J., Gao, Z., and Yu, P. S. (2017). Predrnn: Recurrent neural networks for predictive learning using spatiotemporal lstms. *Adv. Neural Inf. Process. Syst.* 30.

Wang, Y., Wu, H., Zhang, J., Gao, Z., Wang, J., Philip, S. Y., et al. (2022). Predrnn: A recurrent neural network for spatiotemporal predictive learning. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 2208–2225. doi: 10.1109/TPAMI.2022.3165153

Wang, Y., Yang, X., and Hu, J. (2016). Position variability of the Kuroshio Extension sea surface temperature front. *Acta Oceanologica Sin.* 35, 30–35. doi: 10.1007/s13131-016-0909-7

Xu, G., Cheng, C., Yang, W., Xie, W., Kong, L., Hang, R., et al. (2019). Oceanic eddy identification using an AI scheme. *Remote Sens.* 11, 1349. doi: 10.3390/rs11111349

Zhang, Z., Zhang, Y., Wang, W., and Huang, R. X. (2013). Universal structure of mesoscale eddies in the ocean. *Geophysical Res. Lett.* 40, 3677–3681. doi: 10.1002/grl.v40.14

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 17 December 2024
DOI 10.3389/fmars.2024.1520775

OPEN ACCESS

EDITED BY
Chao Chen,
Suzhou University of Science and Technology, China

REVIEWED BY
Mengquqn Wu,
Ludong University, China
Shaojie Sun,
Sun Yat-sen University, China

*CORRESPONDENCE
Dong Liu
dliu@niglas.ac.cn
Shengqiang Wang
shengqiang.wang@nuist.edu.cn

RECEIVED 31 October 2024
ACCEPTED 25 November 2024
PUBLISHED 17 December 2024

CITATION
Yang M, Khan FA, Fang H, Maúre EdR, Ishizaka J, Liu D and Wang S (2024) Two-decade variability and trend of chlorophyll-a in the Arabian Sea and Persian Gulf based on reconstructed satellite data.
Front. Mar. Sci. 11:1520775.
doi: 10.3389/fmars.2024.1520775

Two-decade variability and trend of chlorophyll-a in the Arabian Sea and Persian Gulf based on reconstructed satellite data

Mengmeng Yang[1], Faisal Ahmed Khan[2], Hua Fang[1], Elígio de Raús Maúre[3], Joji Ishizaka[4], Dong Liu[5*] and Shengqiang Wang[6*]

[1]School of Information Science and Technology, Taishan University, Tai'an, China, [2]Institute of Environmental Studies, University of Karachi, Karachi, Pakistan, [3]Independent Researcher, Osaka, Japan, [4]Institute for Space-Earth Environmental Research (ISEE), Nagoya University, Nagoya, Japan, [5]Key Laboratory of Lake and Watershed Science for Water Security, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing, China, [6]School of Marine Sciences, Nanjing University of Information Science & Technology, Nanjing, China

The spatiotemporal variability of chlorophyll-a (Chl-a) in the Arabian Sea (AS) and Persian Gulf (PG) has been widely studied, but long-term trends and influencing factors remain less understood due to data gaps. This study investigates Chl-a variability and trends from 2001 to 2019 using reconstructed MODIS-Terra monthly Chl-a and sea surface temperature (SST) data, employing the Data Interpolating Empirical Orthogonal Functions (DINEOF) method for high-accuracy reconstruction. Results reveal pronounced seasonal variability, with Chl-a peaks exceeding 3 mg m^{-3} during southwestern monsoons and ranging between 1–3 mg m^{-3} during northeastern monsoons, with the lowest levels in transitional months. Spatially, the highest Chl-a concentrations were observed in the western and northeastern AS, influenced by summer southwestern (SW) and winter northeastern (NE) monsoons. Trend analysis using Sen's slope and the Mann-Kendall test indicates significant Chl-a declines (-0.002 to 0) along ASPG coasts, with slight increases (~0.005) in the southeastern AS and southern PG. Rising SST anomalies (SST_A) correlated with reduced Chl-a anomalies (Chl-a_A) in the western AS, while increased wind anomalies (Wind_A) enhanced Chl-a_A in the western AS but decreased it in the southern PG. These findings enhance our understanding of the complex environmental dynamics shaping the ASPG ecosystems.

KEYWORDS

chlorophyll-a, data interpolating empirical orthogonal function, Arabian Sea and Persian Gulf, MODIS, sea surface temperature, wind

1 Introduction

Chlorophyll-a (Chl-a) concentration serves as a key bioindicator of phytoplankton biomass and marine productivity, making it crucial for monitoring the health of marine ecosystems. The Arabian Sea (AS) and Persian Gulf (PG) is recogonized as one of the most productive regions in the world (Sathyendranath et al., 1996). Understanding Chl-a variability and trends over the AS and PG (ASPG) is crucial for predicting marine ecosystem health, managing fisheries sustainably, and providing early warnings for harmful algal blooms. Satellite remote sensing has proven to be an effective tool for examining the spatiotemporal dynamics of Chl-a in the ASPG (Goes et al., 2005; Prakash et al., 2012; Moradi, 2020; Sarma et al., 2012; Jayaram et al., 2018; Moradi and Moradi, 2020; Bordbar et al., 2024), thanks to its broad coverage and real-time observation capabilities.

Previous research in these areas has revealed distinct seasonal and interannual patterns in Chl-a variability, which are often associated with factors such as sea surface temperatures (SST), monsoonal winds, upwelling events, and large-scale climate phenomena like the Indian Ocean Dipole and El Niño (Jayaram et al., 2018; Nezlin et al., 2007; Sarma et al., 2012; Seelanki et al., 2022). Furthermore, some studies have reported trends in Chl-a that either increase or decrease over different time periods, typically related to changes in SST, monsoonal winds, and sea level anomalies (Prakash et al., 2012; Goes et al., 2005; Prasanna Kumar et al., 2010). However, many of these studies in the ASPG region have been limited by their relatively short time frames or their focus on specific regional areas, which may restrict the generalizability of their findings. Expanding research to cover longer time periods and broader regions could provide a more comprehensive understanding of Chl-a variability and its underlying drivers. Additionally, it has been observed that satellite-derived products in the ASPG are often affected by suboptimal conditions, such as sun-glint and cloud cover. These factors can lead to gaps in the satellite-derived geographical data, which may result in incomplete information for subsequent analyses.

Data Interpolating Empirical Orthogonal Function (DINEOF) has emerged as a powerful method for reconstructing missing geophysical data, such as SST and Chl-a. Compared to traditional methods like linear interpolation and optimal interpolation, DINEOF offers significant advantages, including faster computation, parameter-free processing, and the ability to handle multiple correlated data types without prior de-correlation scales (Miles and He, 2010). These attributes make DINEOF particularly suitable for oceanographic applications where satellite observations are often hindered by clouds, sun-glint, and aerosols, leading to data gaps. Despite alternatives like machine learning showing promise, it often requires extensive *in-situ* data for training, which limits their scalability, particularly in regions like the ASPG. DINEOF has demonstrated success in various marine environments worldwide, including the South Atlantic Bight, the coastal Gulf of Alaska, the Gulf of Maine, the Gulf of Mexico, as well as the Sargasso Sea, and is currently employed in global ocean color products by NOAA (Li and He, 2014; Shropshire et al., 2016; Liu and Wang, 2018). However, in the ASPG region, the application of DINEOF is still underexplored, with issues such as limited studies on its efficacy and its primary focus on Chl-a reconstruction (Jayaram et al., 2018), highlighting the need for further research to assess its potential in filling satellite-derived data gaps across different oceanographic parameters.

This study has two primary objectives: (1) to investigate long-term trends and associated spatiotemporal variability in Chl-a from 2001 to 2019 across the ASPG, and (2) to assess the influence of SST and wind on Chl-a variability and trends. To achieve these goals, DINEOF was employed to reconstruct missing MODIS-Terra Chl-a and SST data over the study period. Subsequently, we investigated the monthly Chl-a variability and conducted a trend analysis across the entire ASPG. Additionally, we investigated the correlation between Chl-a anomalies (Chl-a_A) and SST anomalies (SST_A), as well as between Chl-a_A and wind anomalies (wind_A), providing deeper insights into the environmental drivers of marine productivity in the ASPG.

2 Data and methods

2.1 Study area and its subregions

The AS and PG, both located in the northwestern Indian Ocean (Figure 1), exhibit distinct oceanic and atmospheric processes that are critical for regional climate regulation and marine productivity. The AS, spanning 5°N to 25°N and 55°E to 77°E, is characterized by monsoon-driven ocean dynamics, influenced by the seasonal reversal of monsoon winds and the region's unique geography. These winds generate variations in mixed layer depth, thermocline shifts, and nutrient upwelling, particularly along the coasts of Somalia and Oman during the Southwestern Monsoon, resulting in high phytoplankton biomass and biological productivity (Goes et al., 2005; Khan et al., 2023; Wiggert et al., 2005; Prasanna Kumar et al., 2010). Additional factors influencing biological activity include wind mixing, Ekman pumping, mesoscale eddies, and large-scale climate events like the Indian Ocean Dipole (IOD) and El Niño, which impact both phytoplankton blooms and surface

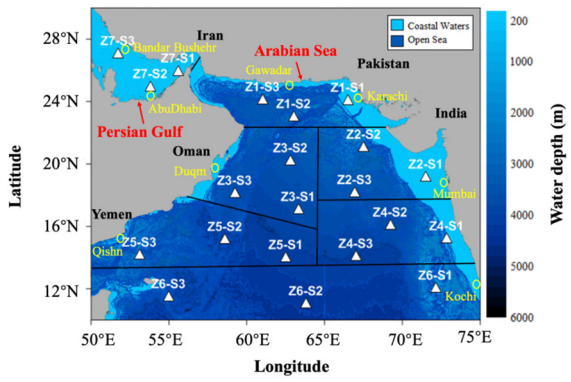


FIGURE 1
A bathymetry map of the ASPG region, showing the distribution of 21 stations across seven zones. Each station, represented by a filled triangle, is assigned to a specific zone. Z1 through Z7 represent zones 1 to 7, respectively, while S1, S2, and S3 represent stations 1, 2, and 3 within each zone. The yellow unfilled circles indicate the stations selected for calculating the upwelling index.

biomass distribution (Seelanki et al., 2022; Keerthi et al., 2013; Shafeeque et al., 2021). In contrast, the PG, situated between 24°N and 30°N and 48°E to 57°E, is a shallow, semi-enclosed sea marked by extreme salinity, temperatures, and limited water exchange. Despite these harsh conditions, it sustains a productive ecosystem, influenced by seasonal wind stress, tidal turbulence, and human activities such as coastal development and pollution (Swift and Bower, 2003; Moradi and Moradi, 2020; Khan et al., 2019). Understanding the differing productivity patterns of the ASPG is vital for assessing long-term environmental changes and the broader impact of climate change on these ecosystems.

The study area was divided into seven zones, each containing three stations strategically positioned based on geographical and oceanographic significance (Figure 1). Zone 1, along the Pakistan coastline, is crucial for its upwelling, supporting rich fisheries in the northern AS. Zone 2, along the Indian coast, is influenced by monsoons driving nutrient inflow and boosting productivity. Zone 3, near Oman, is shaped by Arabian coastal currents, while Zone 4, along southern India, is affected by monsoon-driven currents impacting nutrient dynamics. Zone 5, off Yemen, benefits from upwelling, supporting marine biodiversity. Zone 6, in the equatorial region, experiences equatorial currents and upwelling influencing Chl-a variability. Zone 7, the PG, is notable for its unique hydrological conditions and proximity to oil-producing nations. This division enabled a region-specific analysis of the factors driving Chl-a dynamics, offering insights into how geographic and climatic factors influence marine productivity across the ASPG.

2.2 Satellite data and preprocessing

The monthly composite Level-3 MODIS-Terra (hereafter referred to as MODIS) Chl-a and SST data, with a 4 km spatial resolution for 2001–2019, were obtained from NASA's Ocean Biology Processing Group (https://oceancolor.gsfc.nasa.gov/). MODIS-Terra data were selected over MODIS-Aqua due to their longer temporal coverage. A comparison of the accuracy between MODIS-Terra and MODIS-Aqua data was conducted in our subsequent research, revealing consistent seasonal variability and trends in Chl-a across the ASPG, thus confirming the reliability of MODIS-Terra for this study.

Due to factors such as cloud cover, sun glint, and other atmospheric issues, the ASPG region experiences significant gaps in the data, particularly during the summer monsoon season. For instance, a previous study reported that the missing data rate for MODIS-Aqua daily Chl-a between 2020 and 2021 fluctuated significantly in the northern AS, with an overall rate ranging from 65% to 100% (Yan et al., 2023). These data gaps can result in the loss of important local information. Therefore, it is essential to reconstruct missing Chl-a and SST data. In this study, the DINEOF method was employed to fill in the missing data over the ASPG (Section 2.4 below). Before reconstruction, all Chl-a and SST data were filtered, and images with more than 95% cloud coverage were discarded to maintain accuracy. Additionally, Chl-a data were log-transformed before reconstruction to meet DINEOF's assumption of normality, given the wide range of Chl-a values.

Once the data were reconstructed, the MODIS Chl-a and SST values at each station were obtained by averaging values from a 3 × 3 window centered on the station's location. To eliminate the seasonal cycle influence, the long-term monthly mean for each month across all years was subtracted from the corresponding monthly time series. This process generated monthly anomalies of Chl-a and SST for each station, which were then used to compute Chl-a_A and SST_A trends over the entire study period. Furthermore, correlation statistics were calculated for the time series of Chl-a_A and SST_A at each station to quantitatively analyze the relationship between these anomalies.

2.3 Reanalysis data and preprocessing

From 2001 to 2019, weekly wind data at a spatial resolution of 0.12° × 0.12° and a height of 10 meters above the surface were obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Reanalysis (ERA-Interim). To represent monthly climatological patterns, these weekly wind data were averaged for each of the 12 months, spanning from January 2001 to December 2019. ERA-Interim is a global atmospheric reanalysis product that combines model-based predictions with observations from various sources to provide a consistent, comprehensive estimate of numerous atmospheric and oceanic parameters. Furthermore, the wind data were used to compute the Ekman transport components for each month during the study period, based on the formulas provided by Kok et al. (2017) in Equations 1, 2.

$$ET_x = \frac{\rho_{air} c (u^2 + v^2)^{1/2} v}{\rho_{water} f} \quad (1)$$

$$ET_y = \frac{\rho_{air} c (u^2 + v^2)^{1/2} v}{\rho_{water} f} \quad (2)$$

where u corresponds to the wind coming from the west (with positive values indicating eastward wind) and v corresponds to the wind coming from the south (with positive values indicating northward wind). The parameter ρ_{air} represents the density of air, valued at 1.22 kg m^{-3}, while ρ_{water} represents the density of water, valued at 1025 kg m^{-3}. Additionally, c is the drag coefficient, and f is the Coriolis parameter. The calculated components of Ekman transport, ET_x and ET_y, were used to generate monthly plots of Ekman transport, providing a visual representation of its variability over time.

The analysis of Ekman transport is critical for understanding upwelling processes. This involves decomposing the movement of water masses into perpendicular components to calculate the Coastal Upwelling Index (CUI). Specifically, a "coast angle" is formed between a northward vector and the landward side of the shoreline, which is determined through geometric measurements at each coastal station. Using geometric tools, these angles are

measured and incorporated into the computation of the CUI. Specifically, θ represents the angle perpendicular to the oceanward unit vector relative to the mean shoreline location. The CUI quantifies coastal upwelling by factoring in the strength and direction of Ekman transport in relation to the coastline. In this study, the study area is divided into eight stations (as shown in Figure 1) to assess the upwelling intensity across the region. The effective angles of the coastline are calculated by averaging the angles of arbitrary coastal lines with respect to the equator at each of the eight coastal stations. The formulas used for CUI calculation is provided in Equations 3 (Kok et al., 2017).

$$UI = -\left(\sin\left(\varphi - \frac{\pi}{2}\right)\right)ET_y + \cos(\varphi - \frac{\pi}{2})ET_x) \quad (3)$$

where ϕ represents the angle between the coastline and the equator. According to the definition of CUI, a positive CUI indicates regions where upwelling conditions are favorable, while a negative CUI suggests that upwelling is unfavorable.

2.4 DINEOF reconstruction

DINEOF was employed to reconstruct missing data in the MODIS Chl-a and SST datasets over the ASPG from 2001 to 2019. We utilized the DINEOF 3.0 package (Alvera-Azcárate et al., 2005; Beckers and Rixen, 2003), available for download from the GeoHydrodynamics and Environment Research (GHER) website. The reconstruction process followed these key steps:

1. Each dataset was organized into a 3D matrix (y × x × t), where y and x represent the latitude and longitude dimensions of each image, and t is the total number of images, ensuring that y×x> t. For Chl-a data, the natural logarithm was applied to prevent negative values during reconstruction, while raw SST data were used without transformation.
2. The mean value across both spatial and temporal dimensions was subtracted from the matrix, and missing data points were initialized to zero to minimize bias in the initial guess.
3. Iterative singular value decomposition (SVD) (Toumazou and Cretaux, 2001) and cross-validation using 3% of randomly selected valid data were employed to identify the optimal empirical orthogonal function (EOF) modes.
4. The optimal EOF modes were then used to reconstruct the entire dataset. For further details on the DINEOF methodology, see Alvera-Azcárate et al. (2005) and Beckers and Rixen (2003).

To verify the accuracy of the DINEOF reconstruction, we randomly selected 1% of the valid pixels from the original Chl-a and SST datasets, treating them as "missing values" (Yang et al., 2021). The remaining valid pixels were left unchanged to ensure that only invalid pixels were involved in the reconstruction process. After performing the DINEOF method, the reconstructed values for the 1% of randomly selected pixels were compared with their original values to evaluate the accuracy of the reconstruction.

2.5 Trend calculation

The Mann-Kendall test and Sen's slope trend analysis are widely employed to assess the magnitude and significance of trends in Chl-a and SST using long-term satellite-derived datasets. The Mann-Kendall test is a non-parametric statistical method used to identify trends in time series data and is based on the variance of the data (Solidoro et al., 2009). Sen's slope (Sen, 1968), another non-parametric method, estimates the magnitude of monotonic trends over time and detects their presence at a chosen significance level. Non-parametric tests, such as these, offer higher statistical power when dealing with non-normally distributed data, which is often the case for Chl-a, and are resistant to the influence of outliers. In this study, a significance level of 95% was used to determine trend significance. Both the Mann-Kendall test and Sen's slope were calculated using MATLAB.

To further investigate relationships among Chl-a_A, SST_A, and Wind_A, Pearson's correlation coefficients (r) were calculated, and their significance was tested using Student's t-test at a 5% significance level ($p < 0.05$). Regression analyses were also conducted for each variable pair, with statistical performance evaluated through slope, coefficient of determination (R^2), bias, and root mean square error (RMSE).

3 Results

3.1 DINEOF reconstruction and validation for Chl-a and SST

MODIS monthly log-transformed Chl-a and linear SST data from 2001 to 2019 were reconstructed using the DINEOF technique. The reconstruction statistics are presented in Table 1, where the missing data rates for Chl-a and SST are 24.67% and 1.26%, respectively. This highlights the critical role of DINEOF in reconstructing Chl-a data, which has a significantly higher missing data rate. Additionally, the means of the input and output data for both Chl-a (-0.42 for input and -0.428 for output) and SST (27.27 for both input and output) are almost identical. Similarly, the standard deviations for input and output data are very close for both Chl-a (0.41 for input and 0.405 for output) and SST (1.96 for input and 1.958 for output). These similarities indicate that the distribution of the reconstructed Chl-a and SST data

TABLE 1 Statistics of the DINEOF computations.

	Log (Chl-a)	SST
Dimensions (latitude×longitude×time)	480×600×228	480×600×228
Missing data	24.67%	1.26%
Number of cross-validation points	373975	373975
Mean (input data)	-0.42	27.27
Standard deviation (input data)	0.41	1.96
Mean (output data)	-0.428	27.27
Standard deviation (output data)	0.405	1.958

closely matches that of the original data, suggesting a high accuracy of the reconstruction.

To evaluate the quality of the reconstructed data, we selected one image of the reconstructed Chl-a from August and one image of the reconstructed SST from June for comparison with the original SST and Chl-a images (see Figure 2). The original images exhibited numerous spatial gaps, particularly in the Chl-a data. In contrast, the reconstructed images were more continuous and displayed a more coherent spatial distribution.

We further conducted a cross-validation of the reconstructed Chl-a and SST data using the method described earlier. The results of the comparison between the reconstructed and original Chl-a/SST data are presented in the density plots shown in Figure 3. Both reconstructions showed strong correlations with the original data, as evidenced by favorable metrics: slope (0.86 for Chl-a, 0.95 for SST), R^2 (0.84 for Chl-a, 0.96 for SST), bias (0.002 for Chl-a, 0.04 for SST), and RMSE (0.16 for Chl-a, 1.52 for SST). Additionally, the density plots, which represent the number of data points within each 4 km × 4 km grid bin, show an increasing trend towards the 1:1 line. This suggests that the data reconstructed using the DINEOF method are both accurate and reliable.

3.2 Monthly climatology of Chl-a in the ASPG

Based on the reconstructed data, the interannual monthly climatology of Chl-a from 2001 to 2019 was generated. Hovmöller diagram (Figure 4) displays monthly Chl-a time series along latitudinal sections at 17°N, 21°N, and 25°N, as well as longitudinal sections at 61°E, 64°E, and 67°E. The interannual variability of Chl-a across both latitudinal and longitudinal gradients is further detailed in the Supplementary Materials (Supplementary Figures 1, 2). This study focuses on the monthly variability of Chl-a, with all plots in Figure 4 consistently capturing the well-established seasonal cycle in the ASPG. Chl-a concentrations peak during summer, with a secondary peak in winter, and reach their lowest levels during the transitional months. Spatially, the highest concentrations are observed near the western and northern coastlines. This seasonal cycle is driven primarily by the SW monsoon during summer and the NE monsoon in winter.

Along the latitudinal sections, chlorophyll-a (Chl-a) exhibited two annual peaks: a major peak in summer and a minor peak in winter (Figure 4). At 17°N, which is closer to the equator, Chl-a concentrations remain consistently lower throughout the year. Nevertheless, two distinct peaks are observed, one in summer (August and September) and the other in winter (February and March). As latitude increases to 21°N, Chl-a concentrations rise significantly during both seasons, with the summer peak occurring between July and September and the winter peak between February and March. Further north at 25°N, a coastal region forming the northern boundary of the AS, Chl-a levels remain high and productive throughout most of the year, with pronounced peaks during the summer (August to October) and winter (February to March). Additionally, Chl-a concentrations increase gradually with

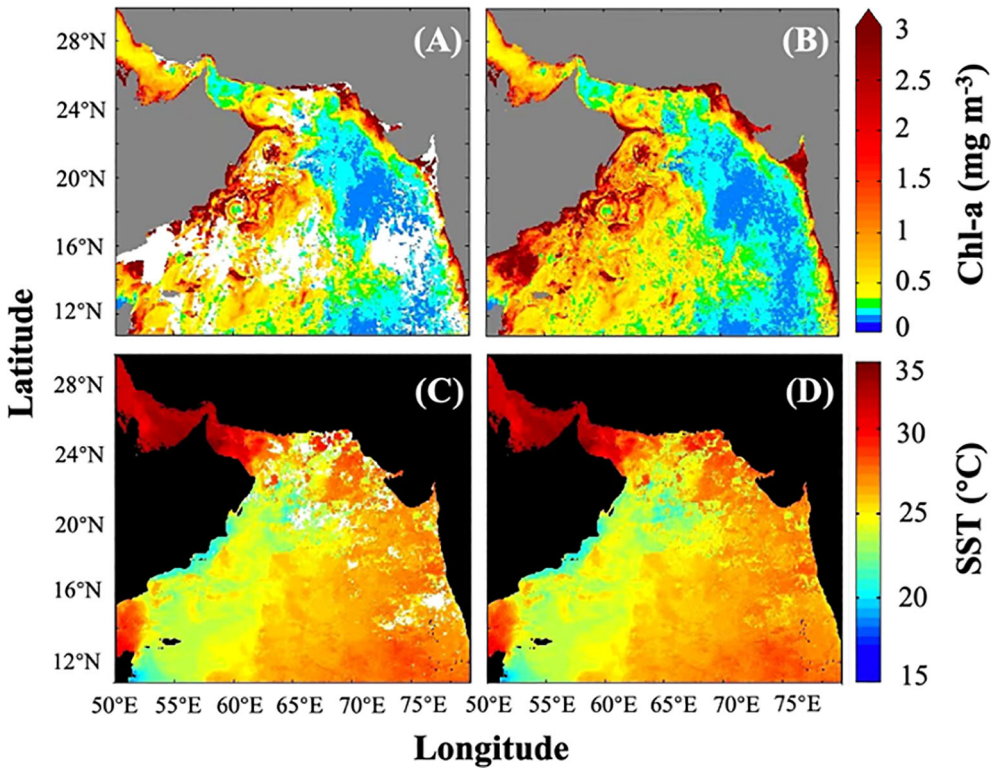


FIGURE 2
MODIS Chl-a in August and MODIS SST in June: **(A, C)** original cloudy data, and **(B, D)** data reconstructed using the DINEOF method.

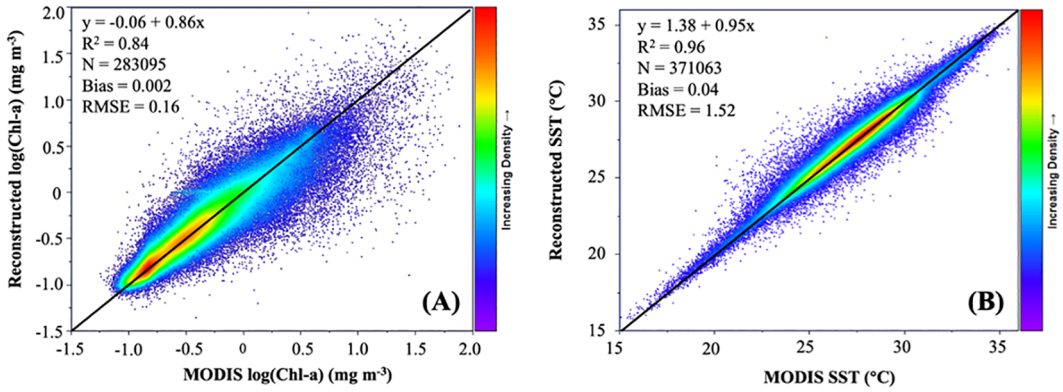


FIGURE 3
Density plots: **(A)** log(Chl-a) (reconstructed) vs Chl-a (original) and **(B)** SST (reconstructed) vs SST (original). The black solid lines are the 1:1 line.

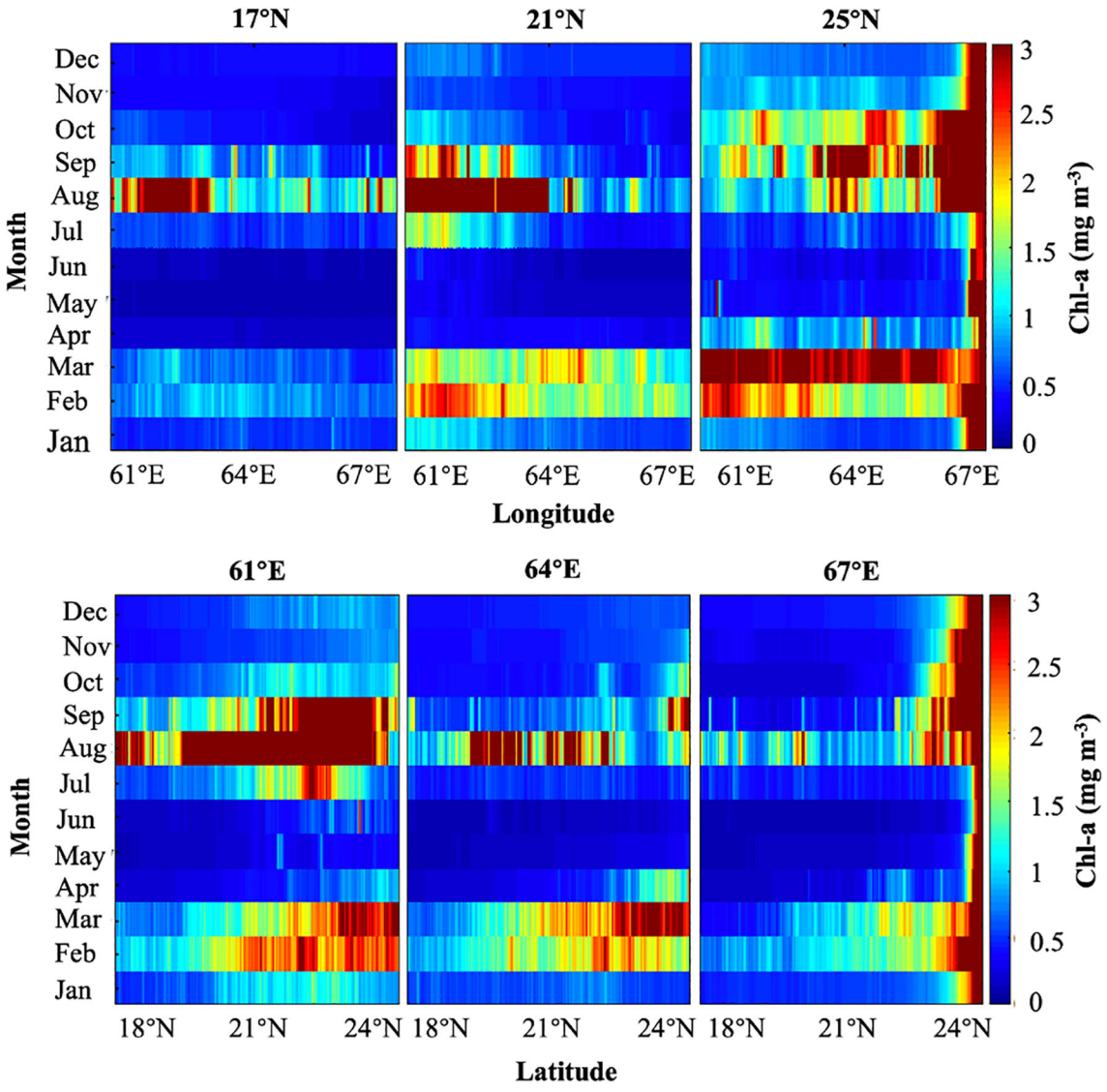


FIGURE 4
A Hovmöller diagram illustrating the monthly variability of reconstructed MODIS Chl-a from January to December at 17°N, 21°N, and 25°N, as well as at 61°E, 64°E, and 67°E.

longitude. Overall, the AS is heavily influenced by monsoonal dynamics, impacting both coastal regions and open ocean waters.

Similarly, along the longitudinal sections, Chl-a exhibited two seasonal peaks, with the exception of the northeastern region (Figure 4). At 61°E, closer to the western coast, Chl-a levels remain consistently higher throughout the year, particularly in the northern regions. Two seasonal peaks are apparent, occurring during summer (July to September) and winter (February to March). At 64°E, while the temporal and spatial patterns of Chl-a are similar, the overall concentration is slightly lower. Moving further east to 67°E, distinct spatial and temporal distribution patterns emerge. Specifically, between 21°N and 23°N, Chl-a exhibits two peaks in summer (August) and winter (February), while between 23°N and 24°N, Chl-a increases markedly and remains elevated throughout the year. These spatial variations highlight the complex interplay between monsoonal forces and the unique oceanographic characteristics of different regions within the AS.

3.3 Long-term trends of Chl-a associated with SST and wind

To analyze long-term trends in Chl-a and SST, the interannual monthly anomaly data were used to compute Sen's slope for each pixel, where positive and negative values indicate increasing and decreasing trends, respectively, and a value of zero denotes no trend. The statistical significance of Sen's slope was assessed using the Mann-Kendall (MK) test, with results coded as 1 for significant trends and 0 for non-significant trends. Non-significant Sen's slope values were masked, indicated by white areas. The spatial distributions of Sen's slope, along with the MK test results for Chl-a_A and SST_A, are illustrated in Figures 5A, B, respectively.

The Sen's slope values for Chl-a_A with statistically significant MK-test results were primarily concentrated in the coastal areas of the ASPG. Most values were negative across the entire ASPG, indicating a declining trend in Chl-a levels. In the AS, the lowest Sen's slope values were observed along the Arabian coasts, gradually increasing towards open sea waters, with some positive values in the southeastern region. In contrast, in the PG, Sen's slope values increased from the northern to the southern coasts. For SST_A, Sen's slope values with significant MK-test results were widespread across the ASPG, with all values being positive, reflecting a rising trend in SST. In the AS, larger Sen's slope values were observed along the Arabian coasts, decreasing towards open sea waters. In the PG, the highest Sen's slope values were found in the northwestern region, diminishing towards the southern part of the gulf. The detailed statistical summaries of Sen's slope values for both Chl-a_A and SST_A are presented in the Supplementary Materials (Supplementary Table 1).

For the 21 selected stations shown in Figure 1, significant Sen's slope values were identified at only four stations: Z3-S1 (open sea waters near the Oman coast), Z5-S2 (open sea waters near the Yemen coast), Z5-S3 (coastal waters near the Oman coast), and Z7-S2 (southern PG). The Sen's slope values for these stations are detailed in the Supplementary Materials (Supplementary Table 2). Additionally, Sen's slope values for SST_A and Wind anomalies (Wind_A) were calculated for these stations, as they are two key factors influencing Chl-a variability. The calculation of Wind_A followed the same methodology used for Chl-a_A and SST_A. As shown in Supplementary Table S2, all four stations exhibited a decreasing trend in Chl-a_A. The trends for SST_A were

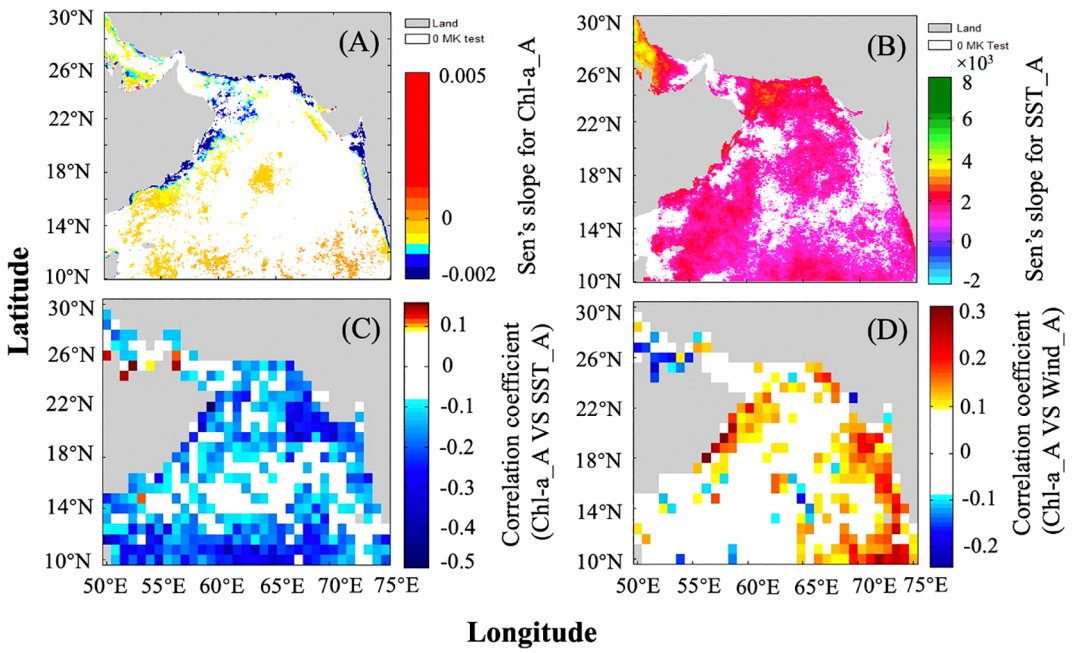


FIGURE 5
Spatial distributions of Sen's slopes and MK-test results for **(A)** Chl-a_A and **(B)** SST_A over the ASPG from 2001 to 2019, along with the spatial distributions of r values between **(C)** Chl-a_A and SST_A, as well as **(D)** Chl-a and Wind_A over the same period.

significantly positive at Z3-S1, Z5-S2, and Z5-S3, but there was no significant trend at Z7-S2. For Wind_A, significant positive trends were observed at Z5-S2 and Z7-S2, while no significant trends were found at Z3-S1 and Z5-S3. These results highlight the complex interplay between Chl-a, SST, and wind patterns across different regions of the ASPG.

To assess the impact of SST and wind on the long-term trends of Chl-a, correlation coefficients (r) between Chl-a_A and SST_A, as well as Chl-a_A and Wind_A, were calculated for each pixel (Figures 5C, D). The r values between Chl-a_A and SST_A were predominantly negative across the ASPG, indicating an inverse relationship between these variables, with a few positive values in the southern PG suggesting localized positive correlations. Additionally, the majority of these correlations were statistically significant throughout the ASPG. In contrast, the r values between Chl-a_A and Wind_A were mostly positive in the AS, signifying a positive correlation, while in the PG, the r values were generally negative. Significant correlations between Chl-a_A and Wind_A were primarily observed along the Oman coast, northeastern Arabian coast, and western Indian coast in the AS, as well as in the southern PG. The detailed statistical summaries of r values between Chl-a_A and SST_A, as well as Chl-a_A and Wind_A are presented in the Supplementary Materials (Supplementary Table 3).

Since significant trends in Chl-a_A were only detected at four stations—Z3-S1, Z5-S2, Z5-S3, and Z7-S2—the time series of Chl-a_A, SST_A, and Wind_A were extracted for these locations to further examine temporal variability and the correlations between Chl-a_A and SST_A, as well as Chl-a_A and Wind_A. A detailed statistical summary of these correlations, covering the entire study period, the southwestern monsoons, the northeastern monsoons, and the transitional months (pre- and post-southwestern monsoons), is presented in Table 2. At Z3-S1, no significant correlations between Chl-a_A and either SST_A or Wind_A were observed for any time frame. At Z5-S2, two significant correlations were found between Chl-a_A and Wind_A: one positive correlation for the entire study period and the other positive correlation during the northeastern monsoons. At Z5-S3, three significant correlations were identified between Chl-a_A and SST_A—one for the entire study period, one for the northeastern monsoons, and another during the transitional months. At Z7-S2, two significant correlations emerged between Chl-a_A and Wind_A: one for the entire study period and the other during the southwestern monsoons. These results underscore the regional differences in the relationships between Chl-a_A and SST_A, as well as Chl-a_A and Wind_A.

The time series of Chl-a_A, SST_A, and Wind_A at four stations (Z3-S1, Z5-S2, Z5-S3, and Z7-S2) are presented in Figure 6. The coefficient of variation (CV) was used to quantify variability, revealing the highest Chl-a_A variation at Z5-S3 (6.50E+16), followed by Z5-S2 (-1.77E+16), Z3-S1 (-1.47E+16), and Z7-S2 (-3.49E+15). SST_A variation was highest at Z5-S3 (-4.24E+15), followed by Z3-S1 (-2.54E+15), Z7-S2 (-1.34E+15), and Z5-S2 (1.04E+15). Wind_A exhibited the most variation at Z3-S1 (6246.99), followed by Z7-S2 (152.20), Z5-S2 (-93.41), and Z5-S3 (-93.41). Due to the low spatial resolution of wind data, Z5-S2 and Z5-S3 shared the same dataset. Large Chl-a_A outliers were observed at Z3-S1 (e.g., August 2003, February 2017), Z5-S2 (e.g., September 2001, February 2008), and Z5-S3 (e.g., August 2002, 2009, 2017). Although not fully explored, some anomalies were linked to specific oceanographic events. For example, the high Chl-a_A value in August 2003 coincided with a cold-core eddy near the Somali coast, which likely contributed to elevated Chl-a_A concentrations during this period (Prakash et al., 2012). These findings highlight the complex dynamics influencing Chl-a_A variability.

TABLE 2 Statistical summary of the significance of r values between Chl-a_A and SST_A, as well as Chl-a_A and Wind_A, at the four stations for the entire study period, southwestern monsoon seasons, northeastern monsoon seasons, and transitional months from 2001 to 2019.

Stations	Chl-a_A VS SST_A	Chl-a_A VS Wind_A	Time period
Z3-S1	0	0	All months
	0	0	Southwestern monsoons
	0	0	Northeastern monsoons
	0	0	Transitional months
Z5-S2	0	**1+**	All months
	0	0	Southwestern monsoons
	0	**1+**	Northeastern monsoons
	0	0	Transitional months
Z5-S3	**1-**	0	All months
	0	0	Southwestern monsoons
	1-	0	Northeastern monsoons
	1-	0	Transitional months
Z7-S2	0	**1-**	All months
	0	**1-**	Southwestern monsoons
	0	0	Northeastern monsoons
	0	0	Transitional months

A value of 1 indicates a significant correlation, while 0 denotes no significance. The symbols "+" and "–" represent positive and negative correlations, respectively.

4 Discussion

4.1 Advantages of using the DINEOF to fill in the data gaps

In this study, the DINEOF method was employed to reconstruct MODIS datasets of Chl-a and SST over the ASPG from 2001 to 2019. The primary source of missing data in the original datasets was adverse

weather conditions, such as cloud cover and rainfall. Specifically, 24.67% of the Chl-a data and 1.26% of the SST data were missing, as shown in Table 1. The relatively high percentage of missing Chl-a data underscores the significance of applying DINEOF for accurate reconstruction in this region. The comparison between the original and reconstructed datasets demonstrated that the mean and standard deviation values were closely aligned (Table 1), confirming the precision and reliability of the DINEOF reconstruction. Additionally, visual comparisons of the original and reconstructed Chl-a and SST data for specific dates, illustrated in Figures 2, 3, reveal smooth and plausible patterns in the reconstructed outputs. Further validation, through cross-correlation analysis (Figure 4), shows strong agreement between the reconstructed and original datasets for both Chl-a and SST, reinforcing the robustness of the reconstruction method. In future research, we aim to integrate field observations to further enhance the validation of our reconstructed data.

To the best of our knowledge, only a limited number of studies have utilized the DINEOF method to reconstruct satellite-derived Chl-a datasets in specific regions like the AS or PG. Even fewer have applied DINEOF to simultaneously reconstruct both Chl-a and SST datasets over the entire ASPG. For instance, Jayaram et al. (2018) employed DINEOF to reconstruct MODIS-Aqua Chl-a data over the AS for the period 2002–2015. This study primarily investigated the seasonal and interannual variability of Chl-a, highlighting the method's utility in regions with frequent data gaps due to cloud cover. Similarly, Huang et al. (2022) used DINEOF to reconstruct Chl-a datasets from the Ocean Colour Climate Change Initiative (OC-CCI) by the European Space Agency (ESA) over the AS from 1998 to 2017. In contrast, Khan et al. (2019) and Khan et al. (2022) extended the application of DINEOF by reconstructing both MODIS-Terra monthly Chl-a and SST datasets from 2001 to 2017. Their studies analyzed the seasonal variability and explored the correlations between Chl-a and SST over the entire study area. However, while they provided valuable insights into the seasonal dynamics of Chl-a and SST, their work did not examine the long-term trends in Chl-a.

In light of these gaps, the present study offers a more comprehensive approach by not only reconstructing both Chl-a and SST datasets using DINEOF but also performing an in-depth analysis of the spatio-temporal variability and long-term trends of Chl-a across the entire ASPG from 2001 to 2019. This extended temporal range allows us to assess the potential impacts of climate variability and oceanographic changes on Chl-a dynamics in the region. Additionally, by reconstructing both Chl-a and SST, we are able to investigate their interactions and correlations over time, providing a more holistic view of the region's marine ecosystem dynamics. Our study contributes to the broader field of oceanography by demonstrating the effectiveness of DINEOF in reconstructing multi-variable datasets and its potential application in other regions where satellite data is frequently compromised by missing observations.

4.2 Impact of SST and wind on the spatiotemporal variability of Chl-a

The seasonal variability of Chl-a, as revealed in Figure 4, aligns with findings from previous studies (Lévy et al., 2007; Sarma et al., 2012; Piontkovski et al., 2013; Jayaram et al., 2018; Khan et al., 2022), where monsoon-driven wind reversals were identified as the main drivers of phytoplankton blooms. These wind shifts significantly impact mixed-layer dynamics and promote upwelling, bringing nutrient-rich waters from the deeper ocean to the surface, which fuels phytoplankton growth during both the SW and NE monsoon seasons (Goes et al., 2005; Jayaram et al., 2018). This is further supported by the monthly climatology of wind patterns from 2001 to 2019 (Figure 7), which reveals stronger southwestern winds during the SW monsoon and weaker northeastern winds during the NE monsoon, with the weakest winds observed during the transitional periods.

Ekman transport, derived from wind data, exhibits distinct seasonal variability across the ASPG, as illustrated in the Supplementary Materials (Supplementary Figure 3). In the AS, it peaks during the summer monsoon, driving surface water offshore and promoting upwelling, with a maximum value of approximately 2 $m^3\ s^{-1}\ m^{-1}$ in July. In winter, the transport shifts southeast, resulting in downwelling. In contrast, Ekman transport in the PG remains minimal throughout the year, with the highest values observed in June, directed northeast.

To further investigate upwelling dynamics, an upwelling index was calculated using wind vectors at eight coastal stations (Figure 1). These coastal stations were strategically selected for their proximity to known upwelling regions, such as Kochi, Duqm, and Qishn, which are significantly influenced by seasonal monsoonal winds. Additional stations were chosen based on their alignment with nearshore data points within each zone to ensure comprehensive coverage. Spanning a wide latitudinal range across the ASPG, these stations provide a thorough spatial representation of upwelling zones. This selection forms a robust foundation for analyzing upwelling dynamics and their influence on regional Chl-a variability.

The time series of the monthly upwelling indices (Figure 8) reveals that upwelling was most pronounced along the western and southeastern coasts of the AS (Duqam, Qishn, Kochi), followed by the northern PG (Bandar Bushehr) and northeastern AS (Karachi) during the SW monsoon. Higher Chl-a concentrations in the western and northern AS (Figure 4) suggest that upwelling is a key driver of Chl-a variability in these regions during the SW monsoon. Notably, the monthly Chl-a data for the southeastern AS and northern PG are not depicted in Figure 4. However, a prior study by Khan et al. (2019) reported elevated Chl-a levels in the southeastern AS during the SW monsoon, whereas the northern PG did not exhibit similar increases during this period; instead, higher Chl-a concentrations were noted during the NW monsoon. This discrepancy suggests that the effects of upwelling on Chl-a variability differ between the AS and PG.

We also observed that the timing of Chl-a peaks varies across different regions of the AS (Figure 4). A previous study by Jayaram et al. (2018) reported that the northern AS was more productive during the winter monsoon, while the southern coastal regions were less productive, and vice versa. Our findings refine this observation, indicating that the southwestern AS is more productive during the summer monsoon, with reduced productivity in the northern coastal regions, except for the northeastern area. Their study also

identified intra-seasonal variability, with a primary productivity peak during the onset phase of the summer monsoon and a secondary peak during the withdrawal phase in the northern AS, in addition to a single dominant peak during the winter monsoon, based on Wavelet analysis. In contrast, our results show that the timing of Chl-a peaks varies across regions in both summer and winter in the northern AS. This regional variability aligns with the findings of Lévy et al. (2007), who similarly reported that the timing of peak productivity differs between regions within the northern AS, due to differences in local physical and oceanographic processes. These variations highlight the intricate relationship between large-scale monsoon patterns and local environmental conditions, showing that a detailed, region-specific analysis is essential for a complete understanding of Chl-a variability in this area.

Additionally, we examined the influence of SST on the spatiotemporal variability of Chl-a in the ASPG in our previous research. Khan et al. (2019) applied the DINEOF method to reconstruct monthly MODIS-Terra Chl-a and SST data from 2001 to 2017, revealing that the majority of the study area (96%) exhibited a significantly negative correlation between SST and Chl-a. Only a small portion (4%), including certain coastal areas, the PG, and parts of the southeastern AS, showed a significant positive correlation. This negative correlation is primarily driven by wind-induced upwelling, where cooler, nutrient-rich water is brought to the surface, resulting in higher Chl-a concentrations (Goes et al., 2005). Building on this, in our recent study (Khan et al., 2022), we utilized the same reconstructed MODIS-Terra Chl-a and SST datasets and found that regions with elevated Chl-a were associated with lower SST and strong Ekman transport, further validating the connection between upwelling and the negative correlation between Chl-a and SST. Our findings suggest that both SST and wind are key factors influencing the seasonal variability of Chl-a in the ASPG, with

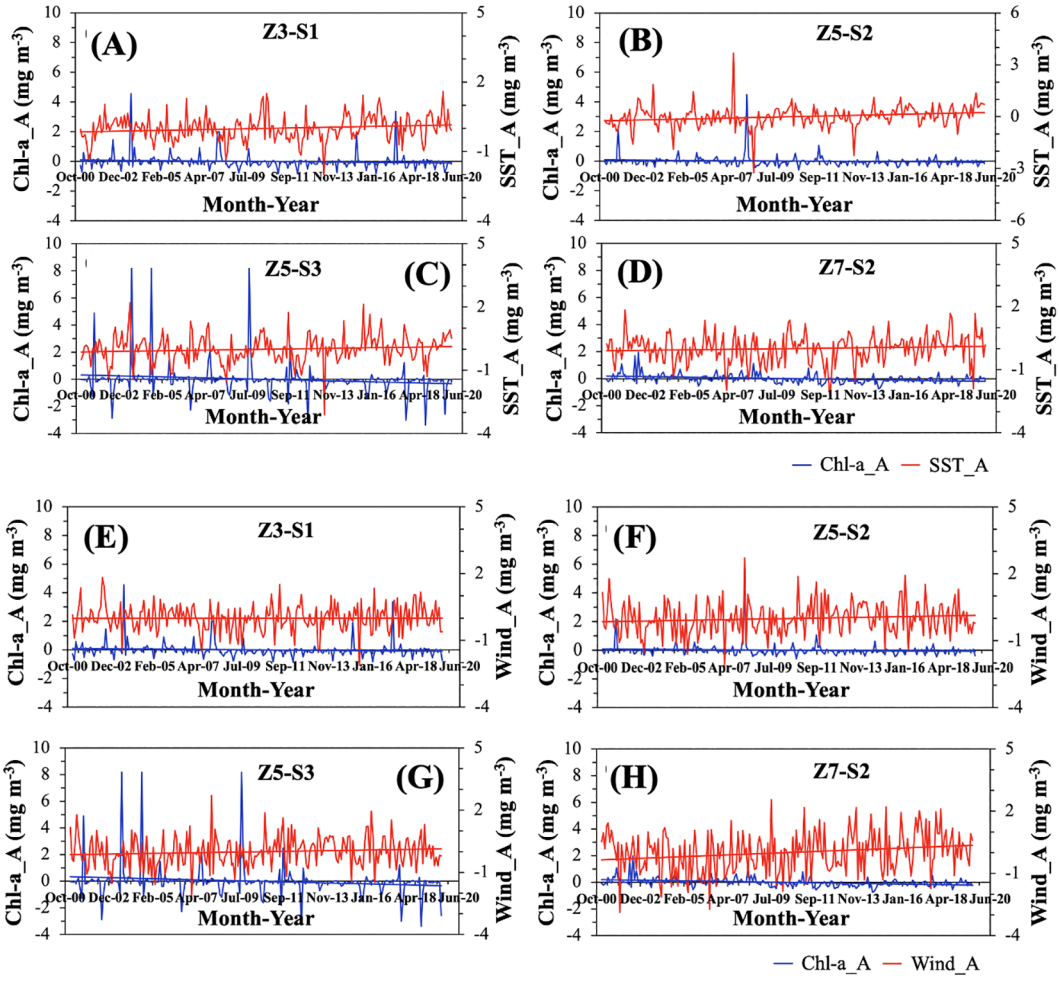


FIGURE 6
Time series of Chl-a_A and SST_A at four stations—**(A)** Z3-S1, **(B)** Z5-S2, **(C)** Z5-S3, and **(D)** Z7-S2—spanning the period from January 2001 to December 2019 are shown. The blue and red lines represent the respective trendlines for Chl-a_A and SST_A. Similarly, time series of Chl-a_A and Wind_A at the same four stations—**(E)** Z3-S1, **(F)** Z5-S2, **(G)** Z5-S3, and **(H)** Z7-S2—are presented for the same period, with blue and red lines depicting the trendlines for Chl-a_A and Wind_A, respectively.

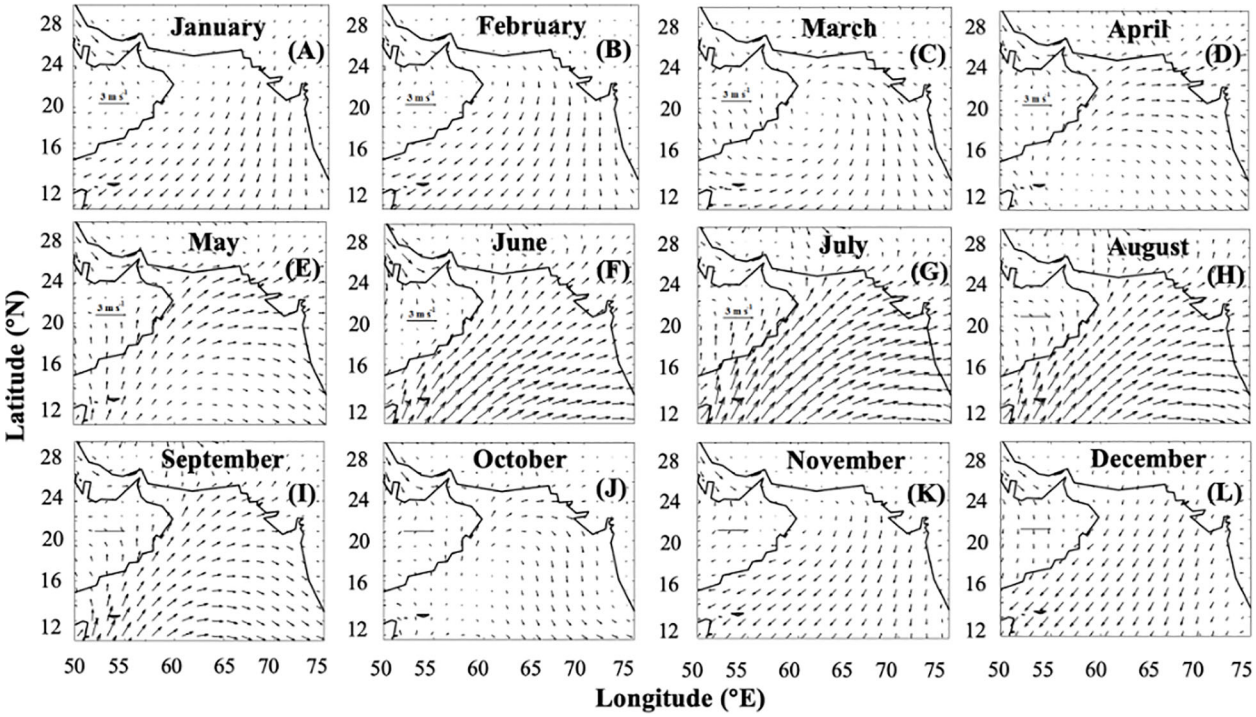


FIGURE 7
Monthly climatology of wind vectors over the ASPG from January to December **(A–L)**. Vectors in all panels have the same scaling to allow for direct comparison of wind intensity across different months. Quantitative values of wind speeds are also provided for reference.

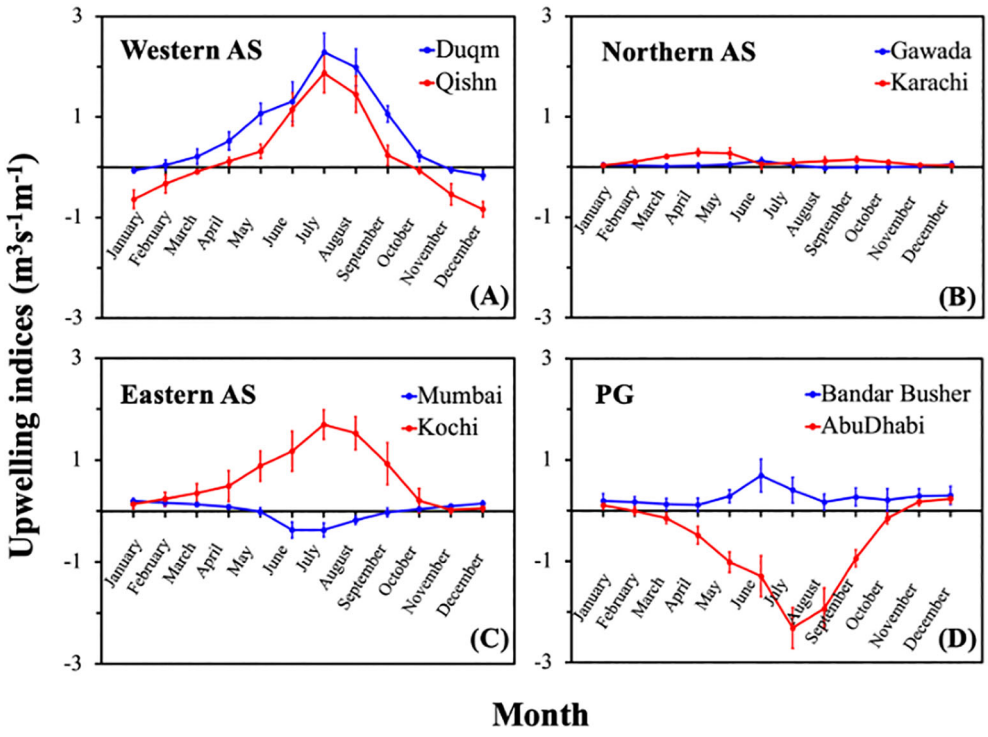


FIGURE 8
Time series of monthly averaged coastal upwelling indices from 2001 to 2019 for **(A)** Duqm and Qishn stations along the western AS coast, **(B)** Gawada and Karachi stations along the northern AS stations, **(C)** Mumbai and Kochi stations along the eastern AS stations, and **(D)** Bandar Busher and Abu Dhabi stations along the PG coast. Error bars in each plot represent one standard deviation.

upwelling playing a critical role in regulating surface productivity in response to local wind patterns.

4.3 Influence of SST_A and wind_A trends on the Chl-a_A trend

Previous studies have revealed conflicting trends in Chl-a for the AS. Goes et al. (2005) reported a more than 350% increase in Chl-a off the Somali coast during the summer, attributed to the strengthening of southwestern monsoon winds. In contrast, Prasanna Kumar et al. (2010) observed a weak basin-wide increasing trend in the monthly Chl-a during September-October and the winter monsoon, but a decreasing trend during the summer monsoon from 1997 to 2007. They linked the Chl-a increase in September-October to dust-induced iron fertilization, which enhanced productivity when sufficient nitrate accumulated in the upper ocean. During winter, intensified evaporative cooling, driven by stronger winds, promoted convective mixing and the upward transport of nutrients from deeper layers, further supported by increased dust deposition, which together explained the Chl-a increase. Prakash et al. (2012) found an increasing Chl-a trend from 1997 to 2003, similar to Goes et al. (2005), but attributed it to a cold-core eddy in 2003, which enhanced Chl-a. However, from 2004 to 2010, they observed a decline in Chl-a off the Somali coast, suggesting that SLA, rather than SST or wind, were likely the main drivers. These studies highlight the spatial and temporal variability in Chl-a trends across the AS. Given the significant seasonal-to-interannual variability in this region, identifying long-term, climate-driven trends requires an extended dataset of at least a decade or more (McClain, 2009). Therefore, we used two decades of Chl-a data in this study. We also found a decreasing trend in the western AS (Figure 5A), consistent with Prakash et al. (2012).

Our results for the Persian Gulf align with previous studies, but with some differences. Moradi (2020) reported a mostly decreasing trend in annual Chl-a from 2002 to 2018 across the Persian Gulf, except for small areas in the southern and central regions, while SST showed an increasing trend throughout the Gulf, with the exception of the Strait of Hormuz. In contrast, we found non-significant trends in Chl-a in the central Persian Gulf and similarly non-significant trends in SST in the Strait of Hormuz (Figure 5B). Bordbar et al. (2024) observed an increasing SST trend in the entire Persian Gulf from 2003 to 2021, which differs slightly from our findings. This discrepancy could be attributed to differences in datasets or trend calculation methods. Regarding the correlation between Chl-a and SST, Bordbar et al. (2024) found an inverse relationship between SST and Chl-a throughout the Gulf, except in the southern region, which is consistent with our results (Figure 5C). Concerning surface winds, the northwesterly Shamal wind, prevalent year-round in the Persian Gulf (Perrone, 1979; Pous et al., 2013; Yu et al., 2016), has shown a positive trend over the past decades (Aboobacker and Shanas, 2018), consistent with the increasing wind trend observed at station Z7-S2 (Supplementary Table 2). Moreover, as Chl-a_A at Z7-S2 exhibited a decreasing trend (Supplementary Table 2), this led to a negative correlation between Chl-a_A and Wind_A at this station (Figure 5D).

5 Conclusion

In this study, we conducted a comprehensive analysis of the spatiotemporal variability and long-term trends of Chl-a across the ASPG using reconstructed MODIS monthly Chl-a and SST data from 2001 to 2019. The validation of the reconstructed dataset confirmed its high accuracy and reliability, ensuring the robustness of our findings. Our analysis revealed significant seasonal variability in Chl-a, with distinct regional differences. Generally, a pronounced Chl-a peak occurred in summer, followed by a secondary peak in winter, with the lowest levels observed during the transitional months. Chl-a concentrations were highest in the western and northeastern Arabian Sea. This seasonal pattern is primarily driven by the SW monsoon in summer and the NE monsoon in winter. Additionally, we observed regional variability in the timing of Chl-a peaks in both summer and winter, likely due to differences in local physical and oceanographic processes, such as wind patterns, vertical mixing, and nutrient availability.

Over the two decades from 2001 to 2019, Chl-a_A exhibited a significant decreasing trend along the coasts of the ASPG, with only small areas showing increasing trend in the southeastern AS and southern PG. At the regional level, an analysis of 21 stations identified significant Chl-a trends at four locations: Z3-S1, Z5-S2, and Z5-S3 in the western AS, and Z7-S2 in the southern PG. Correlation analysis revealed predominantly negative correlations between Chl-a_A and SST_A in the western AS, while correlations between Chl-a_A and Wind_A were positive in the western AS and negative in the southern PG. Significant correlations were found in specific cases: For Z5-S2, we observed a significant positive correlation between Chl-a_A and Wind_A throughout the study period and during the northeastern monsoon. For Z5-S3, significant negative correlations between Chl-a_A and SST_A were found over the entire study period, during the northeastern monsoon, and the transitional monsoons. Similarly, Z7-S2 exhibited significant negative correlations between Chl-a_A and SST_A over the entire period and during the southwestern monsoon. These three stations also displayed significant positive trends in both SST_A and Wind_A.

This research advances our understanding of the complex dynamics of marine ecosystems in the ASPG, shaped by both local physical processes and broader climate variability. Future studies should investigate additional factors, such as sea level anomalies (SLA), wind stress curl ($curl_\tau$), and the horizontal (u) and vertical (v) components of wind vectors, and their influence on Chl-a trends, as well as explore the underlying mechanisms driving these changes. Such research will deepen our knowledge of marine productivity trends in the ASPG and their broader ecological implications.

Data availability statement

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

Author contributions

MY: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. FK: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – review & editing. HF: Writing – review & editing. EM: Formal analysis, Writing – review & editing. JI: Writing – review & editing. DL: Funding acquisition, Project administration, Writing – review & editing. SW: Project administration, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Natural Science Foundation of Shandong Province, China, (Grant #ZR202211290173), the Scientific Startup Foundation for Doctors of Taishan University (Grant #Y-03-2022016), and the Youth Innovation Promotion Association of the Chinese Academy of Sciences (Grant #2021313).

Acknowledgments

The authors express our sincere gratitude to the data service provided by the NASA's Ocean Biology Processing Group (OBPG), and the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Reanalysis (ERA-Interim).

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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.2024.1520775/full#supplementary-material

References

Aboobacker, V. M., and Shanas, P. R. (2018). The climatology of shamals in the Arabian Sea-Part 1: Surface winds. *Int. J. Clim.* 38, 4405–4416.

Alvera-Azcárate, A., Barth, A., Rixen, M., and Beckers, J. M. (2005). Reconstruction of incomplete oceanographic datasets using Empirical Orthogonal Functions: Applications to the Adriatic Sea. *Ocean Model.* 9, 325–346. doi: 10.1016/j.ocemod.2004.08.001

Beckers, I. M., and Rixen, M. (2003). EOF calculations and data filling from incomplete oceanographic datasets. *J. Atmos. Ocean Technol.* 20, 1839–1856. doi: 10.1175/1520-0426(2003)020<1839:ECADFF>2.0.CO;2

Bordbar, M. H., Nasrolahi, A., Lorenz, M., Moghaddam, S., and Burchard, H. (2024). The Persian Gulf and Oman Sea: Climate variability and trends inferred from satellite observations. *Estuarine Coast. Shelf Sci.* 296, 108588. doi: 10.1016/j.ecss.2023.108588

Goes, J. I., Thoppil, P. G., do R. Gomes, H., and Fasullo, J. T. (2005). Warming of the eurasian landmass is making the Arabian sea more productive. *Science* 308, 545–548. doi: 10.1126/science.1106610

Huang, K., Xue, H., Chai, F., Wang, D., Xiu, P., Xie, Q., et al. (2022). Inter-annual variability of biogeography-based phytoplankton seasonality in the Arabian Sea during 1998-2017. *Deep Sea Res. Part II: Top. Stud. Oceanogr.* 200, 105096.

Jayaram, C., Priyadarshi, N., Kumar, J. P., Udaya Bhaskar, T. V. S., Raju, D., and Kochuparampil, A. J. (2018). Analysis of gap-free chlorophyll-a data from MODIS in Arabian Sea, reconstructed using DINEOF. *Int. J. Remote Sens.* 39, 7506–7522. doi: 10.1080/01431161.2018.1471540

Keerthi, M. G., Lengaigne, M., Vialard, J., de Boyer Montégut, C., and Muraleedharan, P. M. (2013). Interannual variability of the Tropical Indian Ocean mixed layer depth. *Clim. Dyn.* 40, 743–759. doi: 10.1007/s00382-012-1411-2

Khan, F. A., Khan, T. M. A., and Udin, R. M. G. (2019). Satellite based Monitoring of Interactions between Chl-a and SST in the Arabian Sea and Persian Gulf area: a useful tool to identify ocean productive zones. *J. Space Technol.* 9, 9–14. doi: 10.13189/jst2019.091001

Khan, F. A., Yang, M., Khan, T. M. A., and Khan, M. A. (2022). Detection of productive oceanic areas in the Arabian Sea and Persian Gulf based on reconstructed satellite-derived sea surface temperature and chlorophyll-a. *Pak. J. Engg. Appl. Sci.* 31, 1–13.

Khan, H., Govil, P., Panchang, R., Agrawal, S., Kumar, P., Kumar, B., et al. (2023). Surface and thermocline ocean circulation intensity changes in the western Arabian Sea during ~172 kyr. *Quaternary Sci. Rev.* 311, 108133. doi: 10.1016/j.palaeo.2023.110033

Kok, P. H., Akhir, M. F. M., Tangang, F., and Husain, M. L. (2017). Spatiotemporal trends in the southwest monsoon wind-driven upwelling in the southwestern part of the South China Sea. *PloS One* 12, 1–22. doi: 10.1371/journal.pone.0171979

Lévy, M., Shankar, D., André, J.-M., Shenoi, S. S. C., Durand, F., and de Boyer Montégut, C. (2007). Basin-wide seasonal evolution of the Indian Ocean's phytoplankton blooms. *J. Geophys Res-Oceans* 112, C12014. doi: 10.1029/2007JC004090

Li, Y., and He, R. (2014). Spatial and temporal variability of SST and ocean color in the gulf of Maine based on cloud-free SST and chlorophyll reconstructions in 2003–2012. *Remote Sens. Environ.* 144, 98–108. doi: 10.1016/j.rse.2014.01.019

Liu, X., and Wang, M. (2018). Gap filling of missing data for VIIRS global ocean color products using the DINEOF method. *IEEE Trans. Geosci. Remote Sens.* 56, 4464–4476. doi: 10.1109/TGRS.2018.2820423

McClain, C. R. (2009). A decade of satellite ocean color observations. *Annu. Rev. Mar. Sci.* 1, 19–42. doi: 10.1146/annurev.marine.010908.163650

Miles, T. N., and He, R. (2010). Temporal and spatial variability of chl-A and SST on the south atlantic bight revisiting with cloud-free reconstructions of MODIS satellite imagery. *Cont Shelf Res.* 30, 1951–1962. doi: 10.1016/j.csr.2010.08.016

Moradi, M. (2020). Trend analysis and variations of sea surface temperature and chlorophyll-a in the Persian Gulf. *Mar. Pollut. Bull.* 156, 111267. doi: 10.1016/j.marpolbul.2020.111267

Moradi, M., and Moradi, N. (2020). Correlation between concentrations of chlorophyll-a and satellite derived climatic factors in the Persian Gulf. *Mar. Pollut. Bull.* 161, 111728. doi: 10.1016/j.marpolbul.2020.111728

Nezlin, N. P., Polikarpov, I. G., and Al-Yamani, F. (2007). Satellite-measured chlorophyll distribution in the Arabian Gulf: Spatial, seasonal and inter-annual variability. *Int. J. Oceans Oceanogr.* 2, 139–156. doi: 10.5376/ijms.2012.02.0001

Perrone, T. J. (1979). Winter shamal in the persian gulf. *Naval Environ. Prediction Res. Facility Monterey CA*, 79–86.

Piontkovski, S. A., Claereboudt, M. R., and Al-Jufaili, S. (2013). Seasonal and interannual changes in epipelagic ecosystem of the western Arabian Sea. *Int. J. Oceans Oceanogr.* 7, 117–130.

Pous, S., Carton, X. J., and Lazure, P. (2013). A process study of the wind-induced circulation in the Persian Gulf. *Open J. Mar. Sci.* 3, 27160. doi: 10.4236/ojms.2013.31001

Prakash, P., Prakash, S., Rahaman, H., Ravichandran, M., and Nayak, S. (2012). Is the trend in chlorophyll-a in the Arabian Sea decreasing? *Geophys. Res. Lett.* 39, L23605. doi: 10.1029/2012GL054187

Prasanna Kumar, S., Roshin, R. P., Narvekar, J., Dinesh Kumar, P. K., and Vivekanandan, E. (2010). What drives the increased phytoplankton biomass in the Arabian Sea? *Curr. Sci.* 99, 101–106. doi: 10.1007/s00227-010-0776-9

Sarma, Y. V. B., Al-Azri, A., and Smith, S. L. (2012). Inter-annual variability of chlorophyll-a in the Arabian sea and its gulfs. *Int. J. Mar. Sci.* 2, 1–11. doi: 10.5376/ijms.2012.02.0001

Sathyendranath, S., Platt, T., Stuart, V., Lrwin, B. D., Veldhuis, M. J. W., JXraay, G. W., et al. (1996). Some bio-optical characteristics of phytoplankton in the NW Indian Ocean. *Mar. Ecol. Prog. Ser.* 132, 299–311. doi: 10.3354/meps132299

Seelanki, V., Nigam, T., and Pant, V. (2022). Unravelling the roles of Indian Ocean Dipole and El-Niño on winter primary productivity over the Arabian Sea. *Deep-Sea Res. PT I* 190. doi: 10.1016/j.csr.2022.104383

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's Tau. *Am. Stat. Assoc.* 63, 1379–1389. doi: 10.1080/01621459.1968.10480934

Shafeeque, M., Balchand, A. N., Shah, P., George, G., Smitha, S. B. R., Varghese, E., et al. (2021). Spatio-temporal variability of chlorophyll-a in response to coastal upwelling and mesoscale eddies in the South Eastern Arabian Sea. *Int. J. Remote Sens* 42, 4840–4867. doi: 10.1080/01431161.2021.1899329

Shropshire, T., Li, Y., and He, R. (2016). Storm impact on sea surface temperature and chlorophyll a in the Gulf of Mexico and Sargasso Sea based on daily cloud-free satellite data reconstructions. *Geophys. Res. Lett.* 43, 12,199–12,207. doi: 10.1002/2016GL071178

Solidoro, C., Bastianini, M., Bandelj, V., Codermatz, R., Cossarini, G., Melaku Canu, D., et al. (2009). Current state, scales of variability and decadal trends of biogeochemical properties in the Northern Adriatic Sea. *J. Geophys Res.* 114, C07S91. doi: 10.1029/2008JC004838

Swift, S. A., and Bower, A. S. (2003). Formation and circulation of dense water in the Persian/Arabian Gulf. *J. Geophys. Res.* 108, 3004. doi: 10.1029/2002JC001360

Toumazou, V., and Cretaux, J. F. (2001). Using a Lanczos eigensolver in the computation of empirical orthogonal functions. *Mon. Weather Rev.* 129, 1243–1250. doi: 10.1175/1520-0493(2001)129<1243:UALEIT>2.0.CO;2

Wiggert, J. D., Hood, R. R., Banse, K., and Kindle, J. C. (2005). Monsoon-driven biogeochemical processes in the Arabian Sea. *Prog. Oceanogr.* 67, 78–121. doi: 10.1016/j.pocean.2005.08.003

Yan, X., Gao, Z., Jiang, Y., He, J., Yin, J., and Wu, J. (2023). Application of synthetic DINCAE–BME spatiotemporal interpolation framework to reconstruct chlorophyll–a from satellite observations in the Arabian sea. *J. Mar. Sci. Eng.* 11, 743. doi: 10.3390/jmse11040743

Yang, M., Khan, F. A., Tian, H., and Liu, Q. (2021). Analysis of the monthly and spring-neap tidal variability of satellite chlorophyll-a and total suspended matter in a turbid coastal ocean using the DINEOF method. *Remote Sens.* 13, 632. doi: 10.3390/rs13040632

Yu, Y., Notaro, M., Kalashnikova, O. V., and Garay, M. J. (2016). Climatology of summer Shamal wind in the Middle East. *J. Geophys. Res. Atmos.* 121, 289–305. doi: 10.1002/2015JD024063

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 08 January 2025
DOI 10.3389/fmars.2024.1480796

OPEN ACCESS

EDITED BY
Weimin Huang,
Memorial University of Newfoundland, Canada

REVIEWED BY
Mukesh Gupta,
Université du Québec à Rimouski, Canada
Jingsong Yang,
Ministry of Natural Resources, China

*CORRESPONDENCE
Ge Chen
gechen@ouc.edu.cn

RECEIVED 14 August 2024
ACCEPTED 10 December 2024
PUBLISHED 08 January 2025

CITATION
Wu J, Zheng Y, Wang T, Ma C and Chen G (2025) Oriented ice eddy detection network based on the Sentinel-1 dual-polarization data. *Front. Mar. Sci.* 11:1480796. doi: 10.3389/fmars.2024.1480796

Oriented ice eddy detection network based on the Sentinel-1 dual-polarization data

Jinqun Wu[1,2], Yiqin Zheng[2], Tingting Wang[3], Chunyong Ma[1,4] and Ge Chen[1,4]*

[1]Frontiers Science Center for Deep Ocean Multispheres and Earth System, School of Marine Technology, Ocean University of China, Qingdao, China, [2]Marine Remote Sensing Department, Piesat Information Technology Co., Ltd., Beijing, China, [3]Information Technology Department, Qingdao Earthquake Prevention and Disaster Reducion Center, Qingdao, China, [4]Laboratory for Regional Oceanography and Numerical Modeling, Laoshan Laboratory, Qingdao, China

The complex convergence of cold and warm ocean currents in the Nordic Seas provides suitable conditions for the formation and development of eddies. In the Marginal Ice Zones (MIZs), ice eddies contribute to the accelerated melting of surface sea ice by facilitating vertical heat transfer, which influences the evolution of the marginal ice zone and plays an indirect role in regulating global climate. In this paper, we employed high-resolution synthetic aperture radar (SAR) satellite imagery and proposed an oriented ice eddy detection network (OIEDNet) framework to conduct automated detection and spatiotemporal analysis of ice eddies in the Nordic Seas. Firstly, a high-quality RGB false-color imaging method was developed based on Sentinel-1 dual-polarization (HH+HV) Extra-Wide Swath (EW) mode products, effectively integrating denoising algorithms and image processing techniques. Secondly, an automatic ice eddy detection method based on oriented bounding boxes (OBB) was constructed to identify the ice eddy and output features such as horizontal scales, eddy centers and rotation angles. Finally, the characteristics of the detected ice eddies in the Nordic Seas during 2022-2023 were systematically analyzed. The results demonstrate that the proposed OIEDNet exhibits significant performance in ice eddy detection.

KEYWORDS

synthetic aperture radar, dual-polarization, ice eddy, oriented object detection, deep learning

1 Introduction

Ocean eddies are a pervasive oceanic phenomenon that plays a significant role in the transport and distribution of material, energy, heat, and freshwater in the global ocean (Chelton et al., 2011; Zhang et al., 2020). The observational advantages of SAR satellites, which operate in all weather conditions and at all times of the day, and offer high spatial

resolution, make them an important data source for the refined study of oceanic eddies. SAR satellites are essential for the study of submesoscale eddies that remain unobservable by altimeter satellites. The remote sensing imaging mechanism of SAR ocean eddies is mainly influenced by two mechanisms (ZHENG et al., 2018; Fu and Holt, 1983; Karimova et al., 2012): the wave-current interaction mechanism and the sea surface floating tracer mechanism, such as bio-oil films and ice floes. In the MIZs, surface sea ice is driven by ocean eddies, exhibiting spiral motion and eddy characteristics (Manucharyan and Thompson, 2017). This paper refers to the ice-water mixing pattern formed by surface sea ice and ocean eddies as an ice eddy (Johannessen et al., 1987; Dumont et al., 2011). The melting of surface ice is facilitated by ice eddies through the vertical transfer of heat, which affects the development of MIZs and indirectly influences global climate regulation.

Data acquisition for ice eddies relies on both *in situ* instruments and satellite sensors. In general, *in situ* observations are characterized by their high quality and reliability and include moorings (Cassianides et al., 2021; von Appen et al., 2018), ice-tethered profilers (Toole et al., 2011), and under-ice gliders. However, due to the high cost of observations and poor weather conditions, the amount and coverage of *in situ* observational data may not adequately support experimental demands. Satellite sensors theoretically possess the capability to acquire vast amounts of data, supporting ice eddy detection and characterization tasks with high spatial resolution and wide-area global observation. In the Arctic Ocean, the detection of submesoscale and small-scale eddies using satellite altimetry data is challenging due to the limited spatial and temporal coverage of both altimetry and *in situ* data. The Rossby radius of deformation in the Arctic Ocean is significantly smaller than in mid- and low-latitude seas (Bashmachnikov et al., 2020; Nurser and Bacon, 2013). Due to the presence of sea ice, the complexity of using altimeter data in the Arctic Ocean renders it nearly unsuitable for detecting ice eddies. Observational costs and adverse weather limit the quantity and coverage of *in situ* data, which may be insufficient to meet experimental demands. In contrast, SAR satellites with high spatial resolution, full-time, and all-weather capability are better suited for detecting mesoscale and submesoscale oceanic phenomena in the Arctic Ocean (Kozlov et al., 2019). SAR satellites have become essential in in-depth studies of oceanic eddies, particularly submesoscale eddies challenging to detect with altimeter satellites. The unique advantages of SAR satellites are illustrated in Figure 1.

The detection of eddies using SAR imagery has been the focus of numerous studies (Cassianides et al., 2021; Kozlov and Atadzhanova, 2021; Manucharyan and Thompson, 2017). However, most studies rely on manual visual interpretation methods for the detection of eddies from SAR images (Toole et al., 2011; Gupta and Thompson, 2022). The accumulation of massive SAR images has rendered it time-consuming and laborious to recognize ocean eddies solely through manual visual interpretation, highlighting the growing importance of automated ocean eddy detection. In recent years, several researchers have applied deep learning methods to ocean eddy detection on synthetic aperture radar (SAR) images (Zhang et al., 2023; Xia et al., 2022; Huang et al., 2017; Du et al., 2019b; Zhang et al., 2020). Du et al. (2019a) attempted to fuse a variety of features to automatically identify ocean eddies and proposed an eddy identification method based on adaptive weighted multi-feature

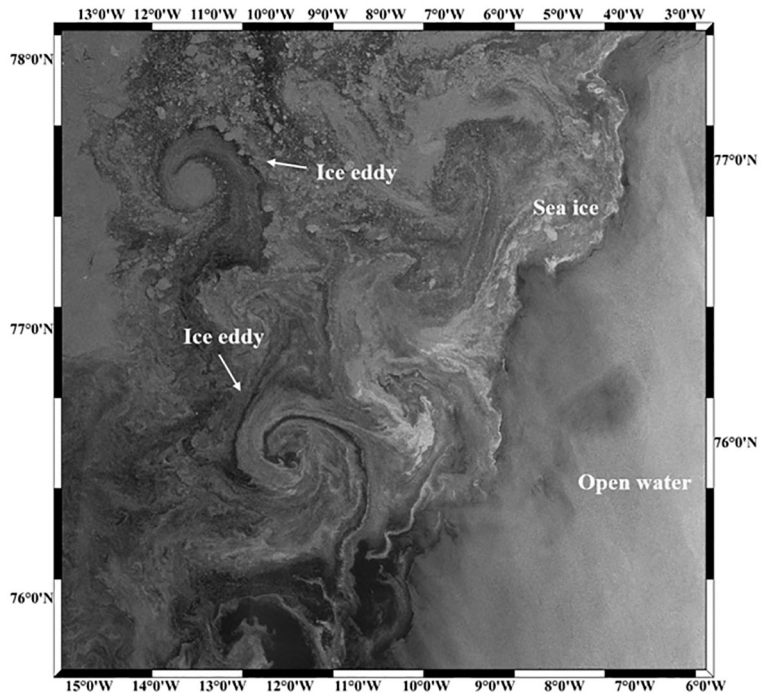


FIGURE 1
S1 EW HH-polarized SAR image, 6 August 2022,07:47 UTC.

fusion for SAR images. Considering the different importance of different features for eddy recognition, an adaptive weighted feature fusion method based on multiple kernel learning (MKL) is also proposed. Although MKL demonstrates excellent performance in addressing heterogeneous data, the method exhibits low detection efficiency. Du et al. (2019b) proposed DeepEddy, a deep learning-based ocean eddy detection method consisting of a hierarchical feature learning model and a simple Support Vector Machine (SVM) classifier. Eddy features are learned using two principal component analysis convolutional layers. Additionally, DeepEddy employs Spatial Pyramid Pooling (SPP), which addresses the complex structure and morphology of ocean eddies by fusing multi-scale features. However, this method fails to localize eddies on SAR images. Zhang et al. (2023) proposed EddyDet, a deep framework based on the Mask RCNN framework utilizing Convolutional Neural Networks for eddy detection on SAR images. Khachatrian et al. (2023) applied the YOLOv5 network to SAR ocean eddy detection and realized the automatic detection of ice eddies in the MIZs. Zi et al. (2024) proposed an EOLO network to enhance the feature fusion method by introducing a channel attention mechanism and employing an upsampling operator with a larger receptive field. Xia et al. (2022) constructed a context and edge association network (CEA-Net) based on the YOLOv3 backbone network for identifying ocean eddies in S1 interferometric wide (IW) swath mode data. While the automatic detection of eddies in SAR images using deep learning has shown promising results, current research emphasizes the detection of eddies in ice-free areas within mid- and low-latitude waters through the use of co-polarization SAR images. HH-polarized images make small-scale features more visible, while HV-polarization provided more stable large-scale features related to sea-ice morphology (Korosov and Rampal, 2017). The HV-polarized images were less sensitive to surface scattering from open water but were very sensitive to body scattering from sea ice. As a result, the contrast between sea ice and open water is higher in HV-polarized images, making ice eddy features more visible (Qiu and Li, 2022). The advantages of HH-polarized images in detecting ice eddies are due to its high sensitivity to surface scattering, its strong contrast with open water, and its high signal-to-noise ratio, particularly under low wind speed or rough surface conditions, where HH polarization can offer precise and reliable ice eddy detection results. Combining HH-polarization and HV-polarization features for ice eddy detection, compared to using a single polarization, is beneficial for reducing detection errors and improving accuracy.

Although the aforementioned methods have achieved superior results in eddy detection in SAR images, they all utilize the conventional horizontal bounding box (HBB) and still exhibit notable limitations. HBBs are not optimal for representing oceanic ice eddies with arbitrary orientations and large aspect ratios, as they provide only a rough location without accurate directional and scale information. Additionally, the HBB representation often includes excessive background or nearby object interference, which can lead to misidentification of ice eddies. Unlike HBBs, OBBs are capable of flexibly adjusting the orientation of detection boxes, allowing for the accurate enclosure of inclined or rotated ice eddies. This capability addresses issues related to redundant and overlapping detection boxes, thereby significantly reducing detection errors. The field of target detection has made remarkable progress over the past decade. Directional target detection, as an extended branch of target detection, has attracted significant attention due to its wide range of applications (Li et al., 2020; Liu et al., 2020; Han et al., 2021; Xia et al., 2018; Ma et al., 2018; Ding et al., 2019; Yang et al., 2019). Ice eddies have distinct rotational characteristics and directionality, and directional target detection can not only detect the position of eddies but also accurately estimate their rotational direction, which is highly significant for ocean dynamics research, marine environment monitoring, and marine resource development.

To address the above challenge, in this paper, we proposed OIEDNet, which is a oriented ice eddy detection network based on the Sentinel-1 dual-polarization data. The remainder of this paper is structured as follows. Section 2 provides an overview of the dataset. Section 3 describes the methodology employed in this study. Section 4 presents the experimental results and discussion. Finally, conclusions are outlined in Section 5.

2 Materials

We utilize Sentinel-1A Level-1 EW mode Ground Range Detected (GRD) product. The swath width for the EW Mode is approximately 400 km, with an incidence angle ranging from 18.9° to 47° and a pixel spacing of 40 m × 40 m. We selected 702 Sentinel-1 SAR images containing ice eddies in the marginal ice area of the Nordic Seas during January 2022-December 2023 as shown in Table 1 and Figure 2.

The bathymetric product is the 200m resolution version 4.0 of the International Bathymetric Chart of the Arctic Ocean (IBCAOv4.0) (Jakobsson et al., 2020). The relationship between the intensity of ice eddy production and the background wind velocity was analyzed using 10m u and v hourly means from the ERA-Interim reanalysis. The validation was conducted using the Level 3 (L3) products from the Surface Water and Ocean Topography (SWOT) mission, the Mesoscale Eddy Trajectory Atlas Product (META3.1exp DT), the situ data collected from OpenMetBuoys-v2021 (OMBs) deployed in the marginal ice zone (Rabault et al., 2024) and drifters 15m drogue.

TABLE 1 SAR data statistics for Nordic ice eddy detection (S1).

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
2022	18	13	12	28	39	41	48	35	33	46	45	28
2023	23	9	6	20	29	28	45	30	23	46	30	27

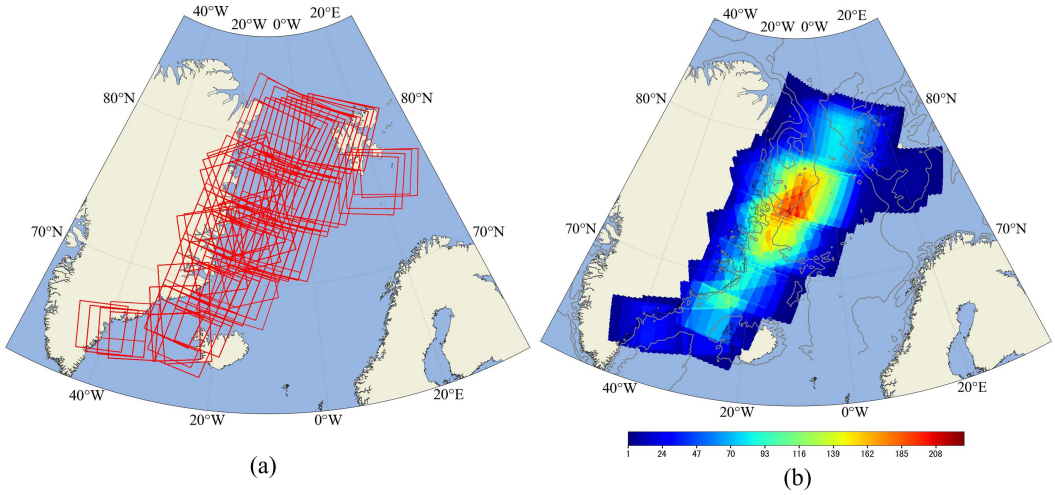


FIGURE 2
The distribution of experimental SAR images collected in the marginal ice zone of the Nordic Seas from January 2022 to December 2023. **(A)** Spatial coverage of SAR images. **(B)** The number of SAR images is represented by color intensity. The gray lines indicate the 200 m and 2000 m isobaths, derived from IBCAOv4.0.

3 Methods

The pipeline of the proposed OIEDNet framework is depicted in Figure 3. From this figure, it is evident that the proposed framework consists of three components: the Polarization Combination Enhancement Module, the Neural Network Module, and the Feature Statistical Analysis Module.

Firstly, the Sentinel-1 satellite's HH and HV dual-polarized ice eddy SAR images undergo data preprocessing, HH-polarized incidence angle correction (IAC), HV-polarized thermal noise removal (TNR), and dual-polarized false-color image synthesis to generate dual-polarized SAR false-color ice eddy images. Secondly, the ice eddy sample library is created using the data expansion method. Finally, based on the dual-polarized SAR false-color ice eddy images, a rotating frame ice eddy auto-detection model is developed and trained to achieve the automatic detection of ice eddies in the Nordic Seas MIZs.

3.1 Polarization combination enhancement method

The polarization combination enhancement method includes (1) data preprocessing; (2) HH-polarized IAC; (3) HV-polarized TNR; (4) polarized data enhancement, and (5) RGB false-color composite. Figure 4 illustrates the flowchart of the polarization

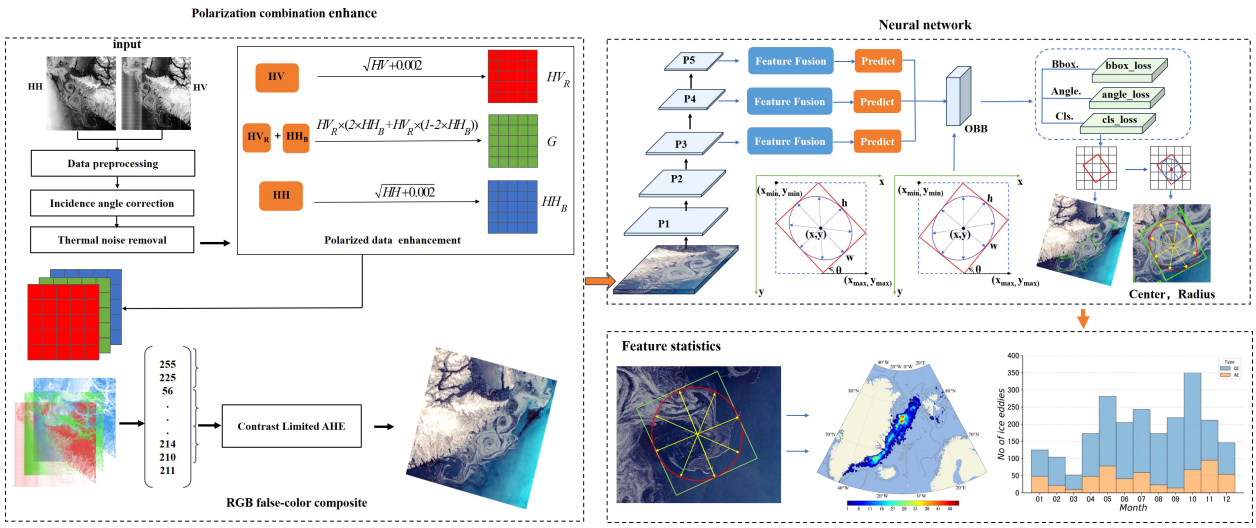


FIGURE 3
The structure of the proposed OIEDNet framework.

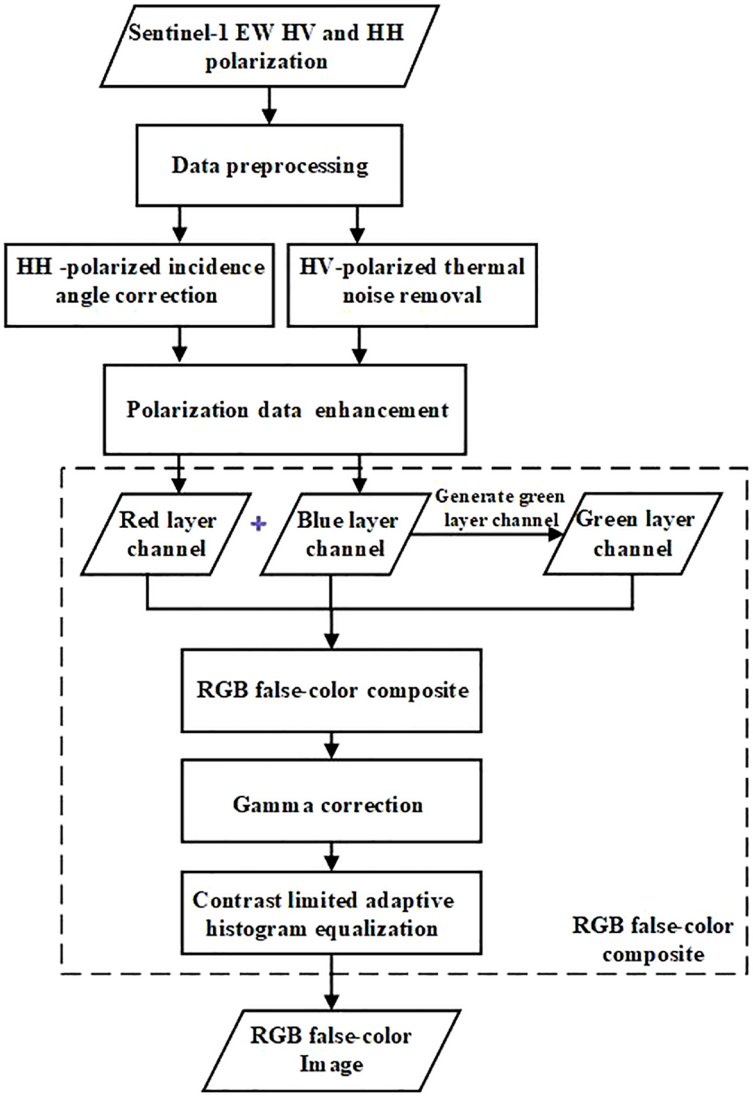


FIGURE 4
Flowchart of the polarization combination enhancement method.

combination enhancement process. Data preprocessing mainly involves orbit correction, radiometric calibration, filtering, conversion to dB values, and geocoding processing.

In the process of ice eddy detection, the variation in the backward scattering coefficient caused by changes in the incidence angle may introduce significant errors, necessitating the correction of the incidence angle for HH-polarized data. In this study, the IAC algorithm (Qiu and Li, 2022; Li et al., 2020) is utilized, and the calculation formula is presented in Equation 1.

$$\sigma = \sigma^0 + 0.200 \times (\theta - \theta_0), \quad (1)$$

where σ is the corrected backward scattering coefficient (in dB), σ^0 is the backward scattering coefficient before correction, θ is the incidence angle of the pixel, and θ_0 is the corrected standard incidence angle, which is taken as 34.5°. Figure 5D illustrates the effects following the correction of the HH polarization incidence angle.

In Sentinel-1 EW-mode SAR images that are strongly affected by scallop stripe noise in the azimuth direction and by noise gradients in the distance direction, especially in HV-polarized images, thermal noise is particularly prominent as displayed in Figures 5A, B. Although ESA provides a standard method of noise vector correction, the effect of residual noise cannot be ignored due to the narrow distribution of HV polarization backscatter.

The denoising algorithm (Park et al., 2017; Sun and Li, 2020) was improved for the removal of thermal noise. The average noise power was added to the denoised results. This adjustment enabled the conversion of noise power from a linear scale to a logarithmic scale (dB) sigma zero conversion, ensuring that these pixels did not become invalid values. By appropriately scaling and balancing the noise vectors given by ESA, the algorithm can approximate the actual noise values as much as possible by using the azimuthal antenna element pattern in the azimuthal direction, so that the effects of the scallop stripe and the noise gradient in the distance

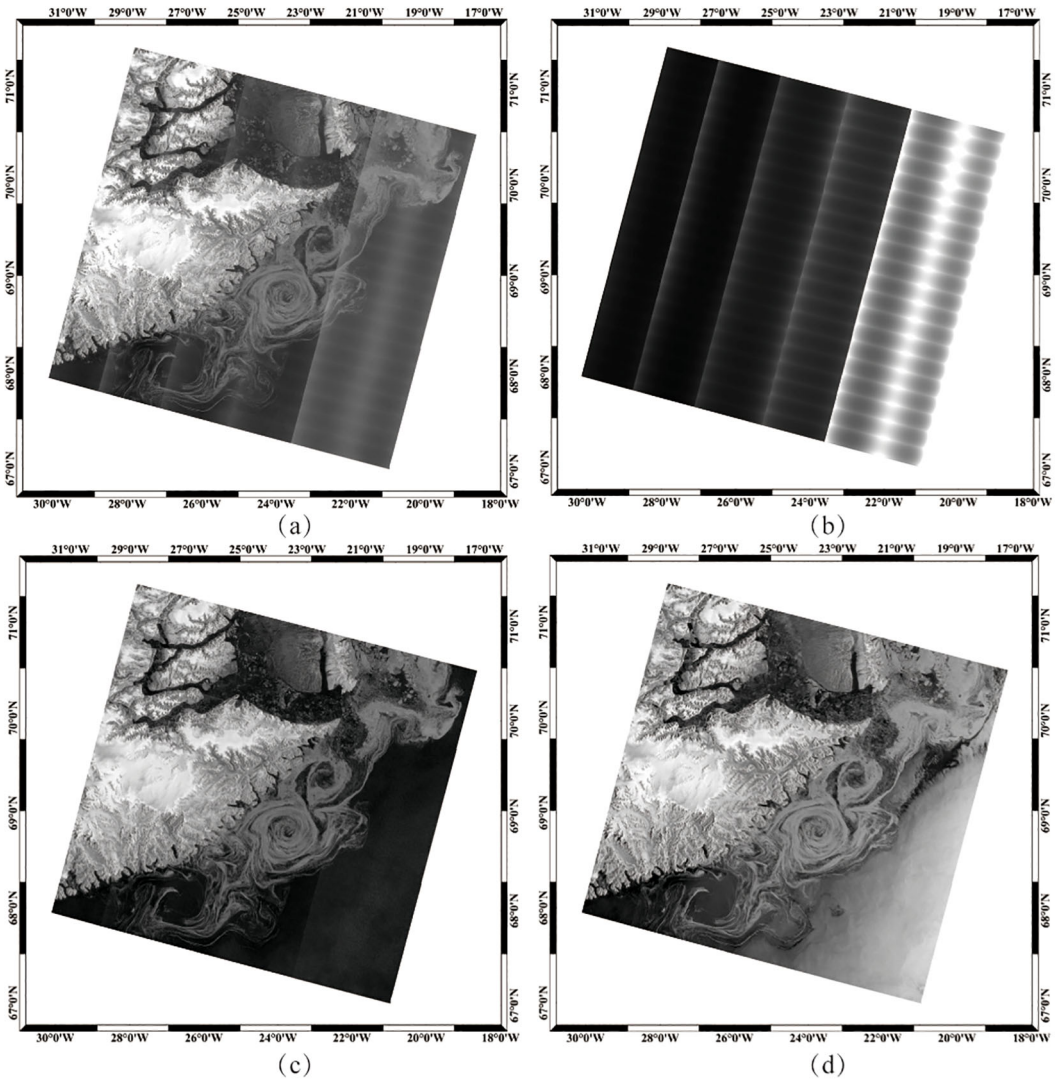


FIGURE 5
Effect of HV TNR and HH IAC. **(A)** Original HV polarized image. **(B)**The thermal noise in HV polarized images. **(C)** HV polarized image after TNR, **(D)** HH polarized image after IAC (S1 EW image taken on 1 November 2023 at 08:21:44 UTC).

direction can be effectively eliminated. The specific processing steps are as follows.

To eliminate the noise step phenomenon between sub-bands, it can be assumed that the denoising process model satisfies a linear relationship. It is calculated using Equation 2.

$$s(k) = \sigma^0_{SN} - (K_{ns,n} \cdot G \bullet \sigma^0_N + K^0_{pb,n}), \tag{2}$$

where $s(k)$ is the denoised σ^0 value. σ^0_{SN} is the uncorrected original σ^0 value. σ^0_N is the σ^0 calculated by bilinear interpolation using the thermal noise vector provided by ESA. $K_{ns,n}$ is the optimal noise scaling factor. $K^0_{pb,n}$ is the interstrip noise power balance factor. n is the number of sub-bands, n = 1, 2, 3, 4, 5. $K_{ns,n}$ can be obtained by least squares solution using a large amount of HV polarized data. $K^0_{pb,n}$ can be calculated using Equation 3.

$$K^0_{pb,n} = (\alpha_{n-1} i + \beta_{n-1}) - (\alpha_n i + \beta_n), \tag{3}$$

where α_n and β_n are the slopes and intercepts, respectively, of the linear models for the different sub-strips. i is the number of image elements in the range direction at the boundary between the strips n = 2, 3, 4, 5. Since there are only four interstrip boundaries. $K^0_{pb,1}$ is set to 0.

When the original image is subtracted from the thermal noise acquired using the described method, some image element points become negative. To eliminate the effect of negative noise power, noise compensation is required. By appropriately scaling and balancing the noise vectors provided by ESA, the algorithm can closely approximate the actual noise values using the azimuthal antenna element pattern, effectively eliminating the effects of scallop stripes and noise gradients in the range direction.

First, the Signal-Noise Ratio (SNR) is defined as the ratio of the σ^0 value (s_{0g}) after Gaussian filtering to the noise equivalent sigma zero (NESZ). The SNR is calculated using Equation 4.

$$\mathrm{SNR} = \frac{s_{0g}}{\mathrm{NESZ}}. \tag{4}$$

Subsequently, further calculations were conducted to obtain the power compensated using Equation 5.

$$s_{0o} = \frac{\mathrm{weight} \times s_{0g} + \mathrm{SNR} \times s_0}{\mathrm{weight} + \mathrm{SNR}} + s_{0\mathrm{offset}}, \tag{5}$$

where s_{0o} is the residual noise power compensated σ^0. $s_{0\mathrm{offset}}$ is the noise field compensation value, which can be taken as the average value of the reconstructed noise field.

Finally, the HV polarization grayscale image with thermal noise removed can be obtained. Figure 5C illustrates the effects after the removal of thermal noise from the HV polarization.

In this paper, a high-quality dual-polarization SAR RGB false-color ice eddy image production method is proposed, compositing HH and HV polarizations into a single false-color image. Since the ice eddy information in the Sentinel-1 EW model is primarily contained in the HH-polarized data, the HH-polarized image is used for the blue channel and the HV-polarized image for the red channel. To optimize the visual quality, the square root is applied to the HH and HV channels, with a slight offset added to mitigate the effect of grain noise on the data. The calculation formula is presented in Equation 6 and Equation 7.

$$HH_B = \sqrt{HH + 0.002}. \tag{6}$$

$$HV_R = \sqrt{HV + 0.002}. \tag{7}$$

The green channel image is produced by combining the offset-processed HH and HV polarization data, as shown Equation 8.

$$G = HV_R \times (2 \times HH_B + HV_R \times (1 - 2 \times HH_B)). \tag{8}$$

Finally, the SAR image was enhanced using SAR image stretching and contrast-limited adaptive histogram equalization (CLAHE). Figure 6 illustrates a comparison the RGB false color images before and after denoising.

The data expansion of 702 dual-polarized false-color ice eddy images was achieved through noise perturbation transformations, rotations (90°, 180°, 270°), and up-down flip transformations, resulting in dual-polarized ice eddy samples. Eddies that rotate clockwise in the northern hemisphere are referred to as anticyclonic eddies, while those that rotate counterclockwise are referred to as cyclonic eddies. Figure 7 illustrates examples of anticyclonic and cyclonic ice eddies.

3.2 Neural network

In this paper, we propose the neural network component of OIEDNet, a multiscale rotating frame model designed for the automatic detection of ice eddies. The model structure is illustrated in Figure 8. Traditional target detection algorithms typically utilize HBB, assuming that object positions in the image are calculated relative to the image center. However, this assumption is not always accurate, particularly for objects with distinct directional features, as the HBB often fails to accurately locate the true position of such objects. OIEDNet addresses this limitation by introducing OBB, which allow bounding boxes to be positioned at any arbitrary angle, making it more adaptable for detecting target objects with various orientations.

3.2.1 Feature spatial pyramid module

The backbone of OIEDNet consists of the CSPDarknet53 feature extractor, which is followed by a C2f module. The C2f module is succeeded by two segmentation heads designed to predict the semantic segmentation masks of the input images. Submesoscale ice eddies (approximately 0.1 to 10 km) and mesoscale ice eddies (approximately 10 to 100 km) can be detected by SAR satellites. To address the wide range of ice eddy target scales in SAR images, a feature fusion module is integrated into CSPDarknet53 to fuse feature maps of varying scales, enhancing the detection of ice eddies of different sizes. OIEDNet incorporates the Spatial Pyramid Pooling Faster (SPPF) module in the feature-enhanced Neck layer, which is optimized from the original SPP module structure. To obtain high-level semantic information from multiscale features and further improve detection accuracy and speed, The SPPF module is inserted between the convolutional and fully connected layers. The SPPF module integrates multiscale local feature information, providing the network with a global perspective and facilitating the extraction of rich multiscale feature representations, as illustrated in Figure 9. The original SPP module generates a final feature map by connecting three feature maps processed in parallel with 5 × 5, 9 × 9, and 13 × 13 max pooling kernels. However, this approach is time-intensive. To improve operational efficiency and detection speed, the SPPF module optimizes this process by merging the feature map processed by a mixed layer (convolutional layer + BatchNorm layer + SiLU layer) with three feature maps derived from a single 5 × 5 max pooling operation. This concatenation enables efficient extraction of the final feature map.

Traditional Feature Pyramid Networks (FPNs) enhance the representation of low-level features by transferring high-level features downwards through a top-down pathway (Lin et al., 2017). Nonetheless, traditional FPNs face challenges in effectively managing scale variations. To compensate for this deficiency, OIEDNet introduces the Progressive Asymmetric Feature Pyramid Network (PAFPN) structure (Liu et al., 2018), which enhances the performance of the target detection task by fusing features from neighboring levels and incorporating higher-level features into the fusion process in an incremental manner, enabling direct interaction between non-neighboring levels. PAFPN is applied between a feature extraction network (backbone) and a neck network (neck module). Specifically, different levels of feature maps are first extracted by the backbone, and then feature fusion is performed using PAFPN. The fused feature maps are fed into OIEDNet's head network (head module) for object classification and bounding box regression. PAFPN incorporates the Path Aggregation Network (PAN) into the Feature Pyramid Network (FPN) by employing a bottom-to-top fusion approach. OIEDNet replaces the Context Enhancement

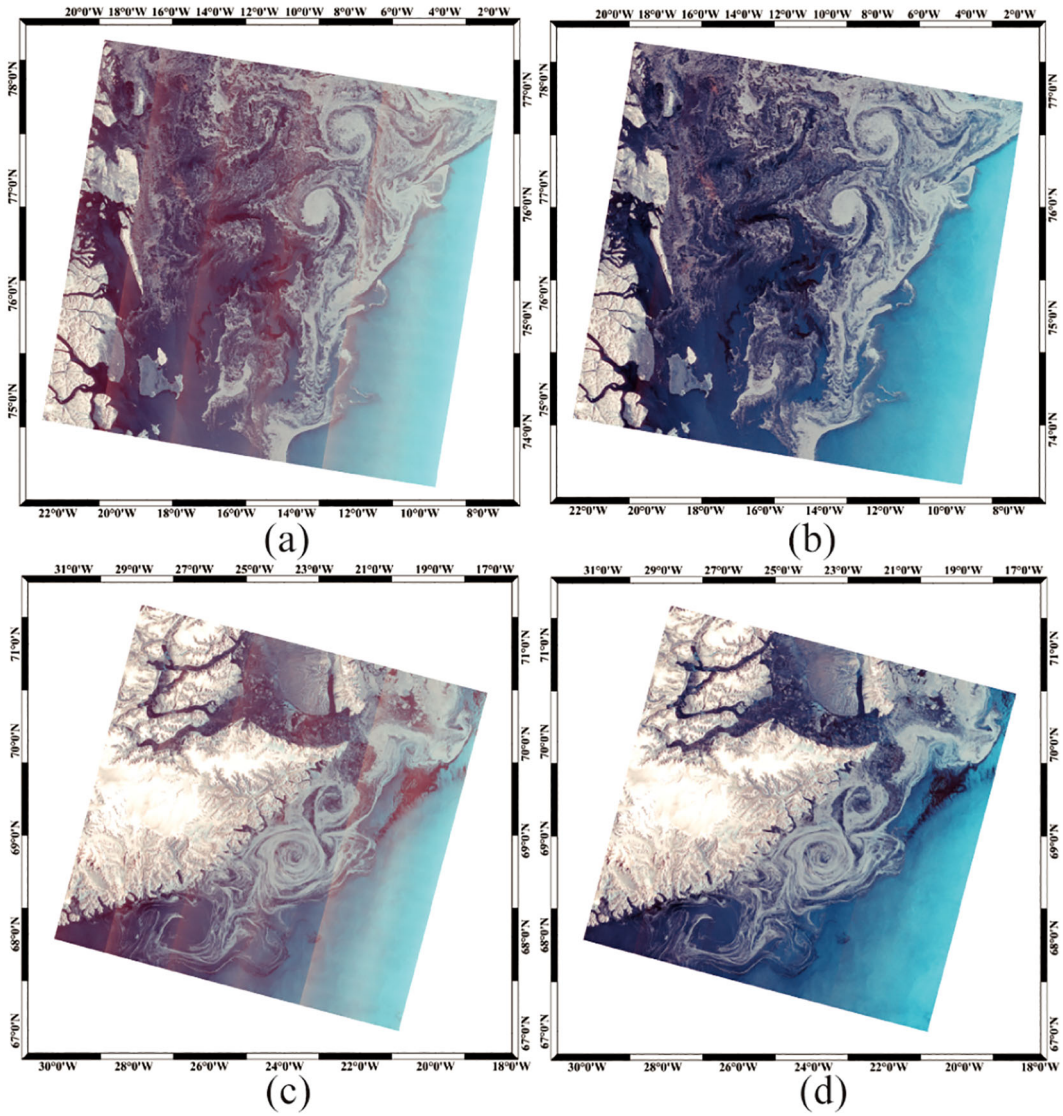


FIGURE 6
Comparison of the RGB false-color images before and after denoising. A comparison of the RGB false-color images before and after denoising is presented. **(A, C)** represent the RGB images prior to denoising, while **(B, D)** represent the RGB images following denoising. **(A, B)** correspond to the S1 EW image acquired on 28th September 2023 at 08:03:22 UTC, while Figures **(C, D)** correspond to the data acquired on 1st November 2023 at 08:21:44 UTC.

Module (C3) in PAN with a Context Enhancement Module with feature fusion (C2f) and removes the 1×1 convolution prior to upsampling. OIEDNet directly inputs the feature output from various stages of the backbone into the upsampling operation. The PAFPN network structure enables the construction of multi-scale feature maps from a single image, ensuring that each layer of the pyramid produces feature maps with robust semantic information. This approach provides richer spatial detail and high-level semantic features for detecting marine ice eddies, which exhibit complex structures, varying scales, and rapid, continuous changes.

3.2.2 Rotation bounding box

Five variables (cx, cy, w, h, θ) are used to define the bounding box with an arbitrary orientation. As shown in Figure 10, cx and cy represent the coordinates of the center point, and the rotation angle θ indicates the angle between the horizontal axis and the first edge of the rectangle after counterclockwise rotation. Here, the first edge defines the width of the bounding box, while the other edge defines its height, with the angle ranging from -90° to 0°.

3.3 Feature statistical analysis module

Based on the obtained location information, the center and diameter of the ice eddy in the predicted box can be determined, laying the foundation for subsequent ice eddy studies. The center of the tangent ellipse inside the rotating frame was used as the eddy center, and the average distance from the center of the ice eddy to all points on the fitted ellipse is used as the radius of the ice eddy.

10.3389/fmars.2024.1480796

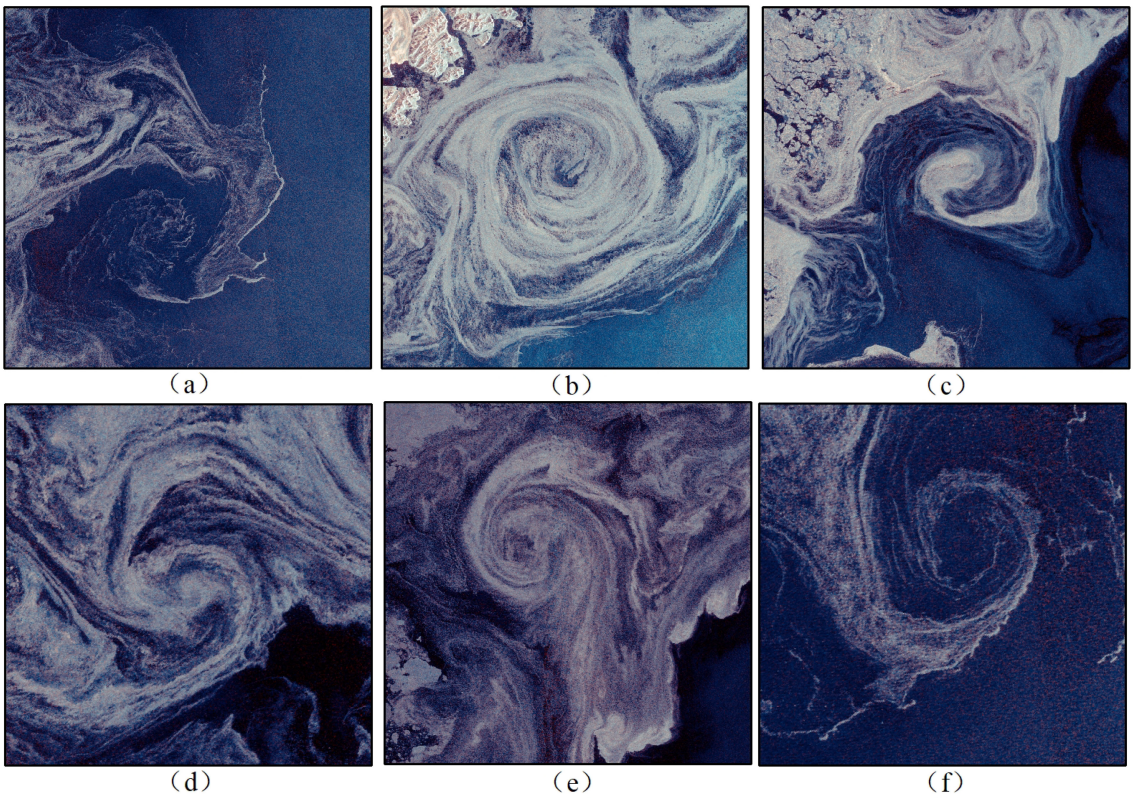


FIGURE 7
Examples of ice eddies photographed by S1. **(A–C)** are anticyclonic eddies and **(D–F)** are cyclonic eddies.

According to the ice eddy automatic detection model, the rotating frame parameters (cx, cy, w, h, θ) of the ice eddy are obtained. Using the eddy center (cx, cy) as the starting point, the coordinate positions of the four vertices A,B,C,D can be calculated.

During the data preprocessing stage, SAR images are geocoded using the WGS1984 standard, transforming pixel coordinates (rows and columns) into geographic coordinates (longitude and latitude). Consequently, it becomes possible to calculate the location of the ice eddy center and the eddy diameter. The Feature Statistical Analysis Module primarily facilitates the extraction of ice eddy center and diameter information, and performs statistical analyses to produce thematic maps of ice eddies for any time period and any region. These maps depict the spatial distribution of ice eddies and related scale histograms (see Chapter 4.4). These analyses support

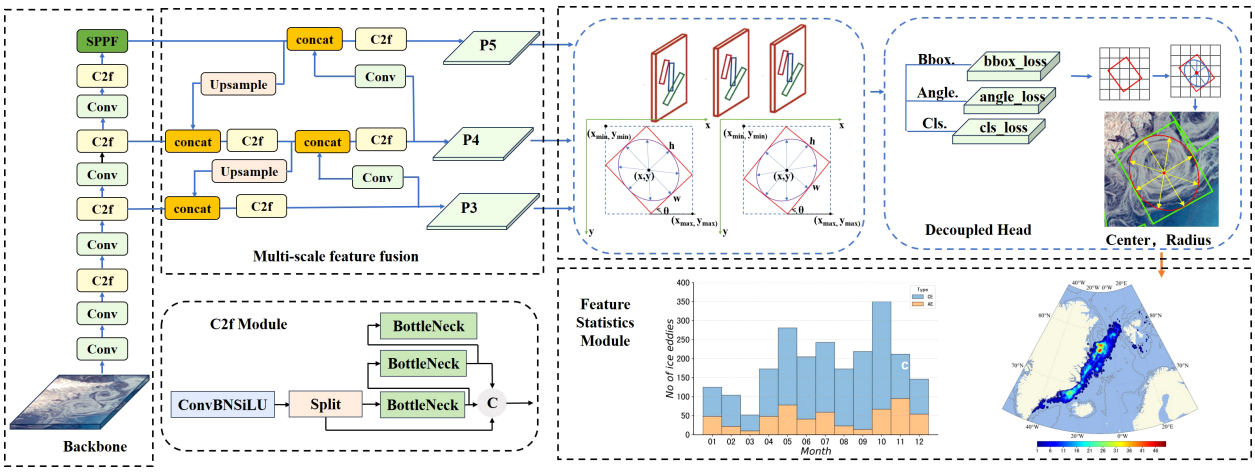


FIGURE 8
The detailed structure of the neural network part of OIEDNet.

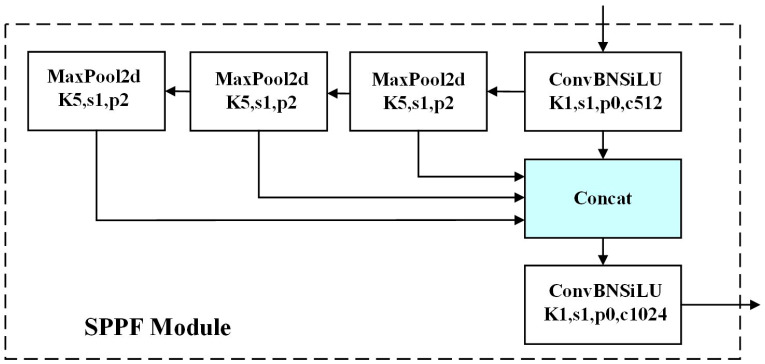


FIGURE 9
The structure of the Spatial Pyramid Pooling Faster module.

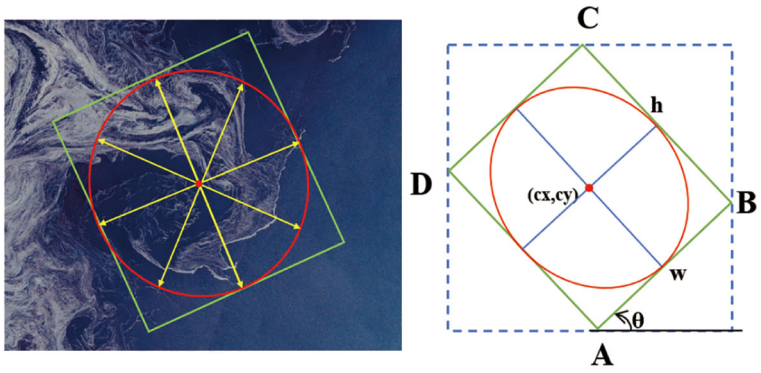


FIGURE 10
Oriented bounding boxes(green solid lines).The red ellipse represents the eddy edge, while yellow arrows show the distance from the eddy center to any point on the ellipse. A red dot marks the center of the ice eddy.

researchers in examining the generative mechanisms and evolutionary processes of ice eddies.

4 Experimental results and discussion

4.1 Experimental environment

The experimental setup configuration is provided in Table 2. Computation was performed on GPUs with 16 multithreads, and the training data share was configured to 0.75. The RGB ice eddy training set and S1 annotations were employed for training, and the model's parameters were fine-tuned based on experience and experimental results to attain optimal performance.

YOLOX is an open-source high-performance detector that builds upon YOLOv3 by introducing decoupled heads, data augmentation, anchor-free detection, and the SimOTA sample matching method, thus constructing an end-to-end anchor-free object detection framework (Zheng et al., 2021). YOLOv8 is a real-time object detection model that utilizes advanced techniques such as anchor-free detection and multi-scale feature fusion within a HBB framework (Varghese and Sambath, 2024). We conduct comparative experiments on OIEDNet, YOLOX and YOLOv8. In addition, we conduct multi-model comparison experiments to evaluate performance before and after denoising, and between single polarization and dual polarization.

The precision evaluation of the model is based on the validation set, and the evaluation metrics include the precision rate (P), the recall rate (R) and the F1-Score (F1), as shown in Equations 9–11.

$$P = \frac{TP}{TP + FP}. \tag{9}$$

$$R = \frac{TP}{TP + FN}. \tag{10}$$

TABLE 2 Experimental setup configuration.

Server	Configuration	Operating System
Bare Metal (GPU)	GPU Card: NVIDIA TESLA-A-1002 Single Card Memory Size: 32GB per card Memory: 1024GB Single Memory Module: 128GB	Kunpeng kC1

$$F1 = 2 \times \frac{P \times R}{P + R}. \quad (11)$$

TP denotes the number of correctly detected ice eddies, *FP* denotes the number of false positive detections of ice eddies, and *FN* denotes the number of missed ice eddies.

4.2 Comparison

The results of OIEDNet,YOLOX and YOLOv8 were compared using the same test set, as presented in Table 3. Both YOLOv8 and YOLOX are traditional HBB detection models. The experimental results indicate that the OIEDNet model exhibits a precision of 94.40% and a recall of 93.65%, while YOLOv8 and YOLOX exhibit precisions of 93.50% and 87.90%, as well as recalls of 92.00% and 86.51%. In comparison to YOLOv8 and YOLOX, the OIEDNet model demonstrates superior performance in detecting dense eddy regions. The rotational detection of OIEDNet more accurately detects eddies with irregular shapes and changing directions, and the inspection frame fits the eddies more closely, significantly reducing the redundancy of the horizontal inspection frame, as shown in Figure 11. For eddies with large differences in scales and similar locations, there is obvious overlapping of inspection frames in horizontal detection, while rotational detection effectively avoids overlapping of inspection frames (Figure 11B). The interaction between ocean circulation and ocean currents is accompanied by the splitting and fusion of ocean eddies, leading to the multinucleated structure of ice eddies, which the rotational detection method can detect more accurately (Figure 11D). The IEDNet model reduces the leakage and false alarms of ice eddies to a certain extent. The OIEDNet model has obvious advantages in the precision and recall of ice eddy detection, and it can effectively detect submesoscale and mesoscale ice eddies.

This study evaluates the enhancement effects of IAC, TNR, and dual-polarization RGB false color synthesis in the OIEDNet model. Comparison of four sets of ice eddy detection results for the same OIEDNet model (Figure 12). Before and after the denoising of dual-polarized false-color images, the detection accuracy increases from 88.71% to 94.40%, reflecting an improvement of 5.69%. In contrast, the detection accuracy of ice eddies in HV-polarized images without TNR is 85.04%, while the detection accuracy in HH-polarized images without IAC is 89.06%. This indicates that thermal noise significantly reduces the detection accuracy of ice eddies, whereas the incidence angle has a relatively minor effect on detection accuracy. The detection performance of the proposed model shows significant improvement with the adoption of the Polarization Combination Enhancement, resulting in an approximate 8.65% increase in the F1 score. This enhancement effectively boosts detection accuracy in noise-heavy environments.

4.3 Validation

Altimeters and SWOT satellites rely on radar echo signals for measuring sea surface height. However, sea ice leads to attenuation and scattering of radar signals, rendering the echo signals unstable, which makes it difficult to obtain accurate sea surface height data and, therefore, makes it unable to accurately detect ice eddies, as shown in Figures 13, 14. SAR, on the other hand, can clearly detect ice eddies in this environment due to its high-resolution imaging and penetration capabilities. Mesoscale eddies can be identified from sea level height data using altimetry, but the daily mesoscale eddy dataset is identified by measuring different time trajectories, which results in low spatial and temporal resolution. Figures 13B, 14B shows a comparison of eddies identified by OIEDNet and altimeters. It is clear that SAR is able to detect more submesoscale ice eddies and that SAR is even more advantageous in detecting high-latitude ice eddies.

The ice eddies detected by OIEDNet were compared and validated against *in situ* data collected from OpenMetBuoys-v2021 (OMBs) deployed in the marginal ice zone. Figure 15 illustrates the movement trajectories of two ice buoys in the marginal ice zone around Svalbard from August 18, 2022, to August 26, 2022. Red triangles are used to denote the starting positions of the buoys, while pentagrams indicate the ending positions of their trajectories. The ice buoy trajectories exhibit a counterclockwise rotation consistent with the direction of the ice eddy, indicating a cyclonic ice eddy.

4.4 Spatial and temporal distribution of ice eddies

Using the OIEDNet ice eddy detection framework, ice eddy identification and scale information extraction were performed on 702 SAR images containing ice eddies in the Nordic Seas from

TABLE 3 Accuracy evaluation of different models.

Model	HH IAC	HV TNR	HH	HV	RGB	P	R	F1
OIEDNet	×	×	✓	×	×	0.8906	0.9048	0.8976
	×	×	×	✓	×	0.8504	0.8571	0.8537
	×	×	✓	✓	✓	0.8871	0.9167	0.9017
	✓	✓	×	×	✓	**0.9440**	**0.9365**	**0.9402**
YOLOv8	✓	✓	×	×	✓	0.9350	0.9200	0.9274
YOLOX	✓	✓	×	×	✓	0.8790	0.8651	0.8720

"×" indicates that the corresponding data is not utilized by the model, whereas "✓" indicates that the data is utilized.

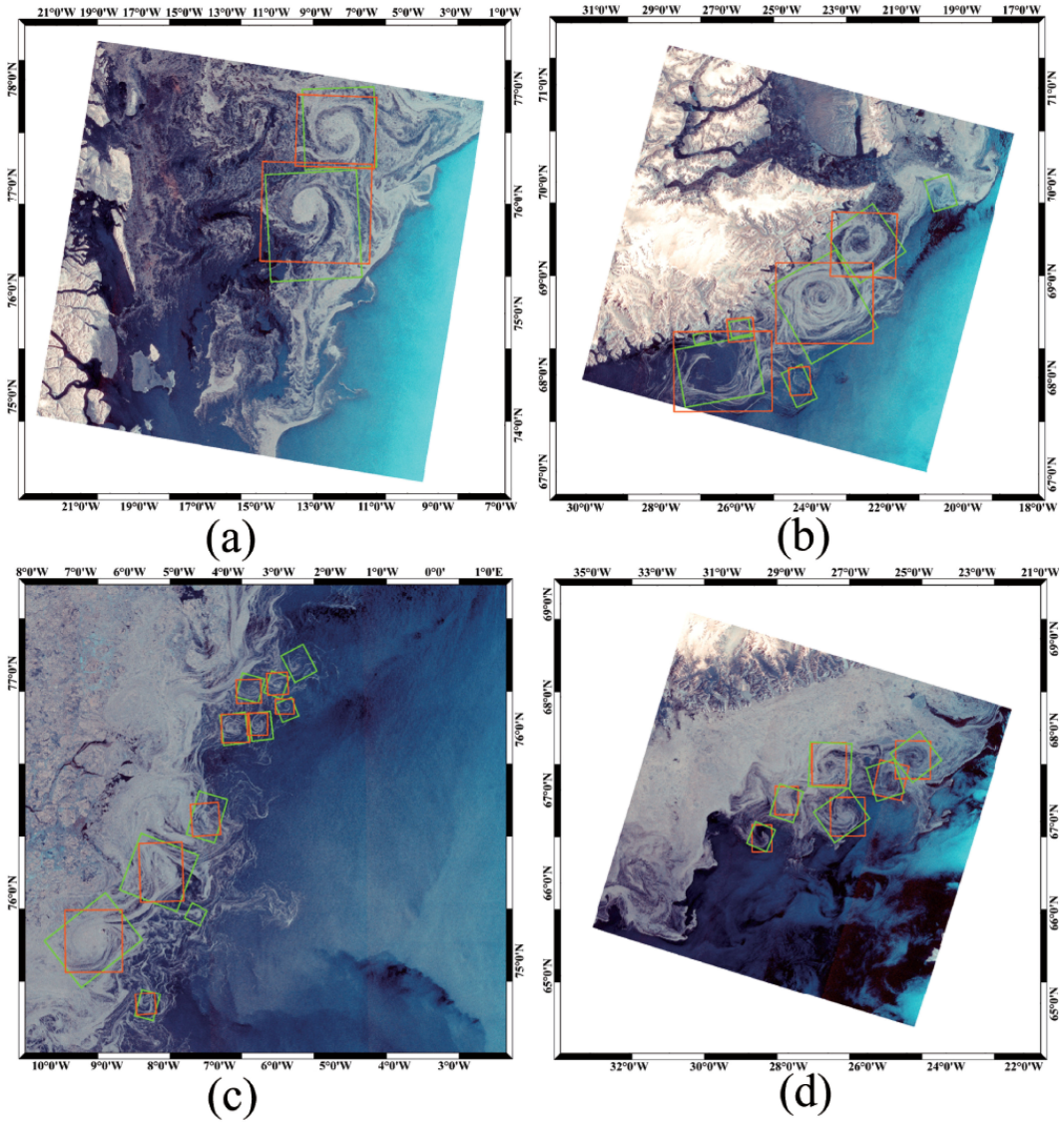


FIGURE 11
The comparison of ice eddy detection results between OIEDNet and YOLOv8 is shown in **(A–D)** where the red HBB is used to represent YOLOv8 detection results, and the green OBB is used to represent OIEDNet detection results.

January 2022 to December 2023. To ensure the accuracy of the statistical feature information of the ice eddies, the detected ice eddy types were annotated using a manual visual inspection method. A total of 2283 ice eddies were identified, including 1724 cyclonic eddies (CEs) and 559 anticyclonic eddies (AEs). The number of cyclonic ice eddies is 3.08 times that of anticyclonic eddies, which may be related to the mechanism of anticyclonic eddy generation and the interaction between the two (McWilliams, 2016).

The spatial density distribution of ice eddies was calculated using a 0.1° × 0.1° grid, as shown in Figure 16A, revealing that the densest distribution of ice eddies is located in the north-central Greenland Sea, which exhibits a high number of both cyclonic and anticyclonic ice eddies. The monthly variation is shown in Figure 16B, indicating that Nordic Seas ice eddies are present throughout the year, with two peaks in the total number of ice eddies in May and October, and a low in March. Overall, May to November is the period when ice eddies are most frequent. The formation of ice eddies in the Nordic Seas results from a combination of dynamical and thermal forces (Perovich and Jones, 2014). Spatially, areas of high ice eddy occurrence are often closely linked to the Arctic Current, with the East Greenland Cold Current flowing along the east coast of Greenland. Temporally, with the onset of the Arctic summer polar day, sea surface temperatures (SSTs) rise, and glacier melting causes the expansion of marginal ice areas, leading to high ice eddy occurrence. In contrast, the Nordic Seas ice cover decreases rapidly to reach a minimum at the beginning of October, after which the ice area starts to expand rapidly. Thus, the thermodynamic factors in the Nordic Seas are more complex in October, which is conducive to the formation of ice eddies.

Figure 17A shows that the sizes of ice eddies in the Nordic Seas are primarily concentrated in the mesoscale and submesoscale

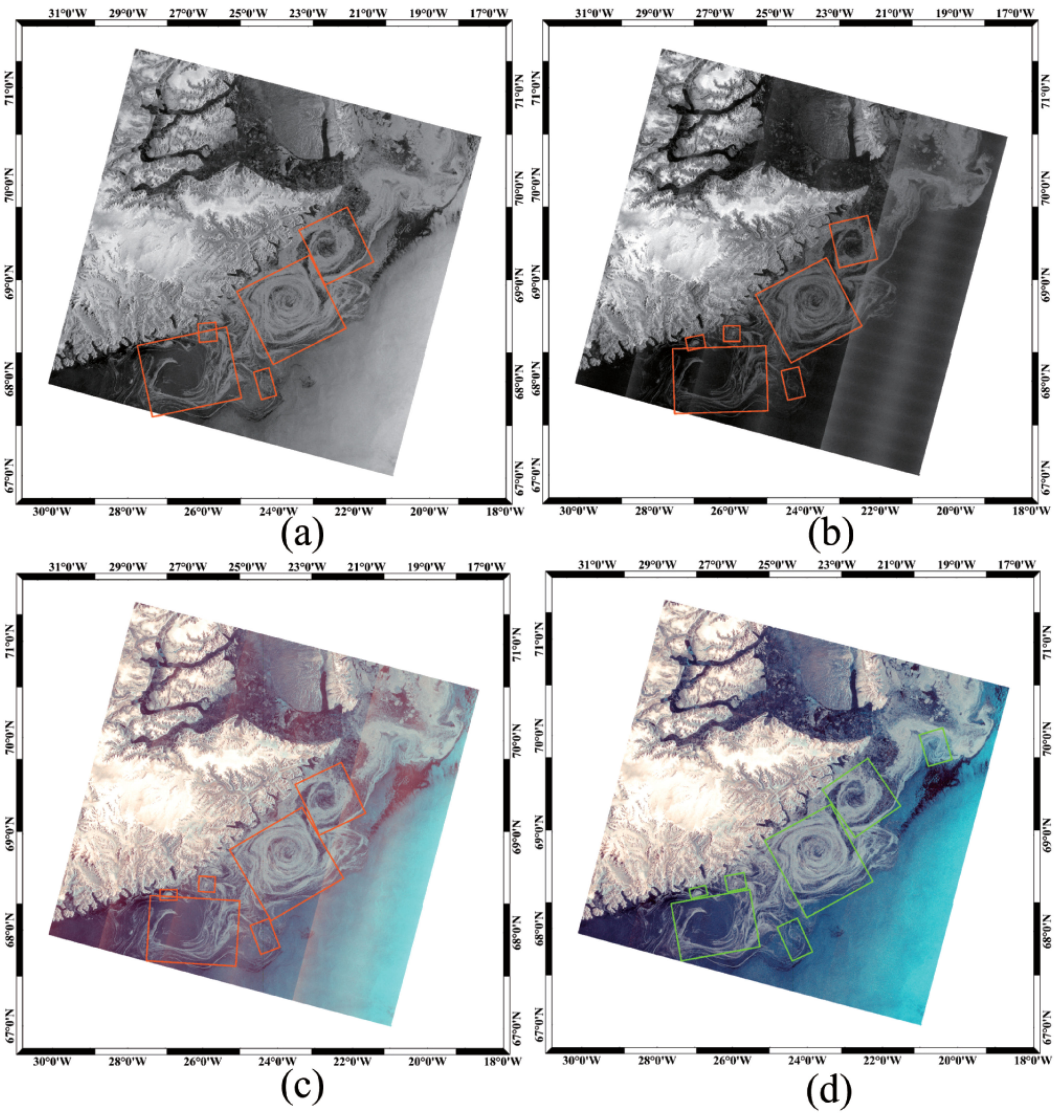


FIGURE 12
Comparison of four sets of ice eddy detection results using the OIEDNet model: **(A)** detection results without HH-polarized IAC, **(B)** detection results without HV-polarized thermal noise reduction, **(C)** detection results of dual-polarized RGB images before denoising, and **(D)** detection results of dual-polarized RGB images after denoising.

intervals. The diameters of these eddies are mostly in the range of 10-100 km. The diameter of cyclonic ice eddies is mainly between 10-60 km, while the diameter of anticyclonic ice eddies is mostly between 30-70 km, indicating that anticyclonic ice eddies tend to be larger than cyclonic ice eddies. Large ice eddies are primarily located in the north-central Greenland Sea.

From Figure 17B, we observe that the proposed model maintains detection performance despite increasing wind velocity. Although the number of ice eddy detections decreases with higher wind speeds, this does not indicate a decline in model performance; instead, it reflects the inherent difficulty of eddy formation in areas with strong winds, resulting in a reduced number of eddies. The lack of a sharp downward trend in detections further illustrates the robustness of the proposed model across varying wind speeds. Regarding wind velocity, 79.7% of the detected ice eddies formed under low wind conditions of 1-4 m/s, while about 20.3% occurred under medium wind conditions. Similarly, from Figure 18, as ice concentration increases, the number of ice eddies decreases. The rate of detected ice eddies shows a gradual decline, which further demonstrates the robustness of the proposed model under varying ice concentrations.

5 Conclusions

To accurately detect MIZs ice eddies, denoising algorithms and image processing techniques are combined to propose a high-quality RGB false-color image production method and to create a dual-polarization synthetic aperture radar false-color ice eddy dataset. Simultaneously, the OIEDNet ice eddy detection model was developed and trained, achieving a precision rate of 94.4% and a recall rate of 93.65%, highlighting significant advantages in ice eddy

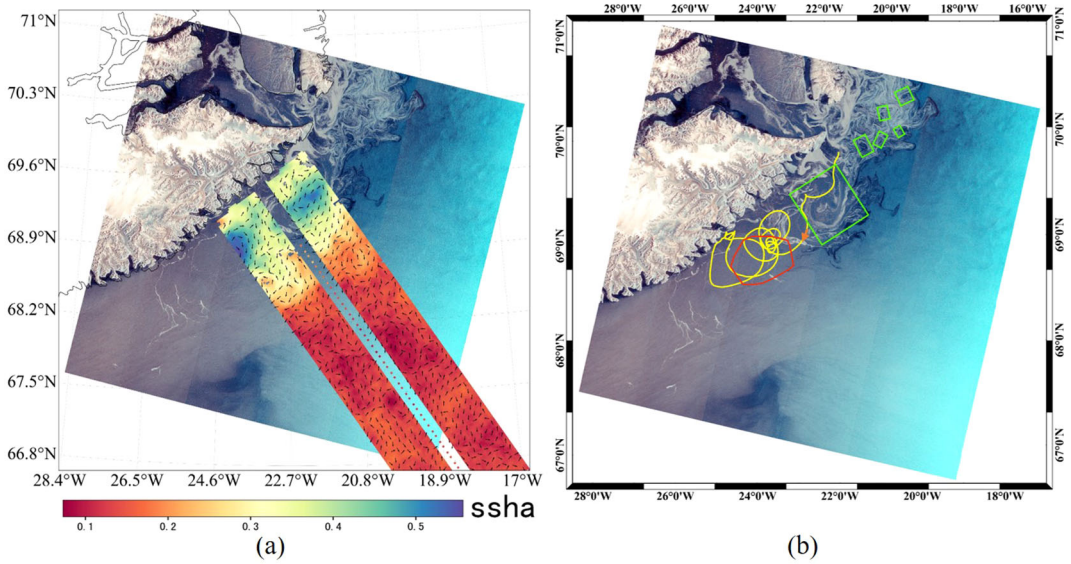


FIGURE 13
Comparison of the ice eddy identification results from OIEDNet (green) with those obtained from SWOT, drifting buoys(yellow), and the mesoscale eddy track atlas product META3.1exp DT (red). **(A, B)** data time are S1: 2023-10-15 08:13:37 UTC. SWOT: 2023-10-15 03:43:56 UTC, META3.1exp DT: 2023-10-15 UTC. drifting buoys(5801987): from 2023-10-1 to 2023-10-15 UTC. The orange arrow in **(B)** indicates the trajectory of the drifting buoy.

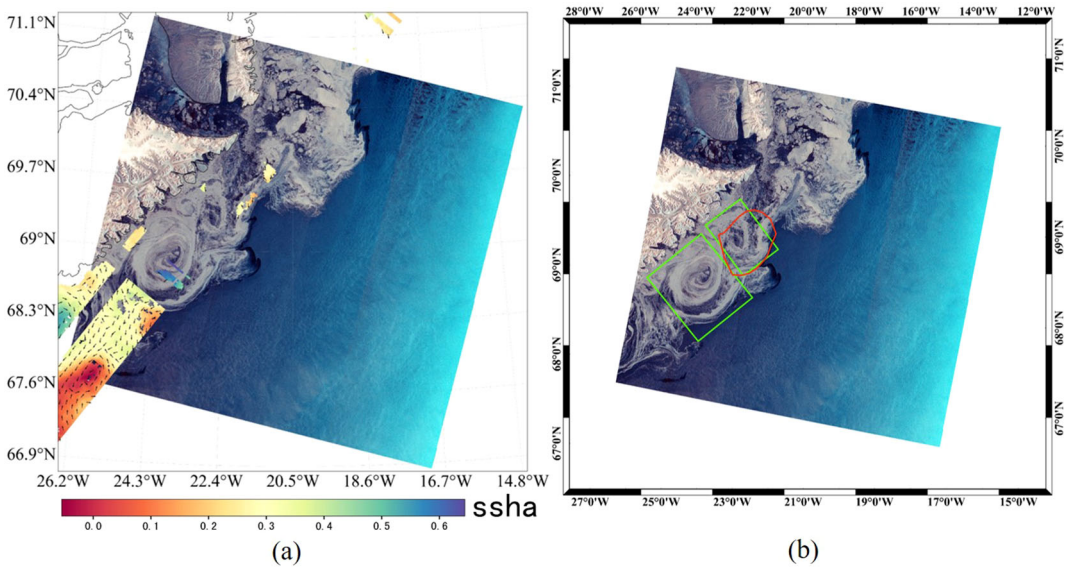


FIGURE 14
Comparison of the ice eddy identification results from OIEDNet (green) with those obtained from SWOT and the mesoscale eddy track atlas product META3.1exp DT (red). **(A, B)** data time are S1: 2023-11-03 08:05:21 UTC. SWOT: 2023-11-03 16:46:19 UTC, META3.1exp DT: 2023-11-03 UTC.

detection. The OIEDNet effectively detects dual-polarized SAR ice eddies with a small sample size, identifying Submesoscale and mesoscale ice eddies in SAR images quickly and accurately. The experimental results demonstrate that the ice eddies detected in SAR images are not as large as previously indicated. The experimental results show that the OIEDNet model excels at detecting dense eddy regions in ice eddy detection. The rotating detection frame of OIEDNet better fits the eddy, effectively avoiding overlap. The interaction between ocean circulation and currents involves the splitting and fusion of ocean eddies, leading to the multinuclear structure of ice eddies, which can be more accurately detected by the rotational detection method. The OIEDNet also significantly reduces the leakage of ice eddies and false detections, especially in dense eddy regions. The OIEDNet not only

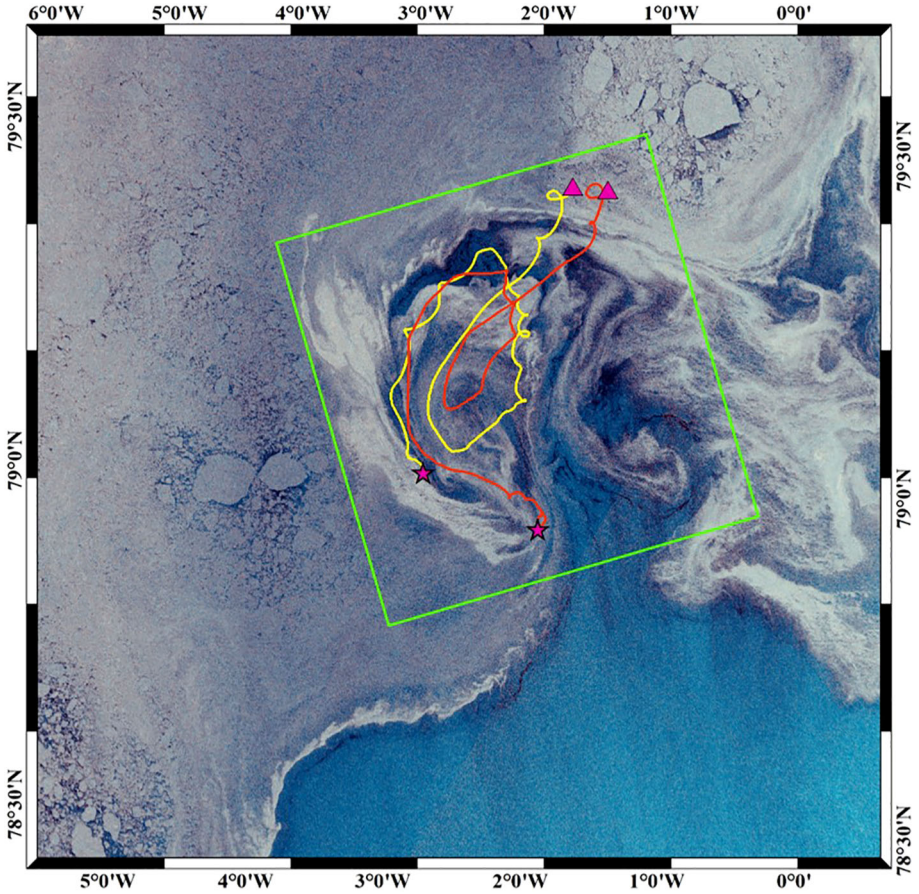


FIGURE 15
Comparison of the ice eddy identification results from OIEDNet (green) with those obtained from OMBs (red and yellow). The acquisition time for Sentinel-1 occurred on August 23, 2022, at 07:53:58 UTC. OMBs data collection spanned from August 18, 2022, to August 26, 2022, between 07:53:58 and 07:55:02 UTC. Red triangles are used to denote the starting positions of the buoys, while pentagrams indicate the ending positions of their trajectories.

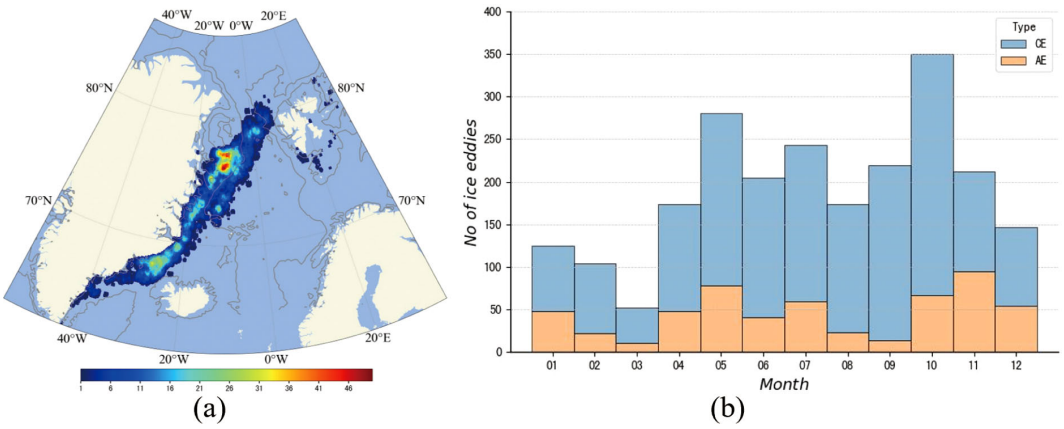


FIGURE 16
Spatial distribution of ice eddies with histograms of months. **(A)** The spatial distribution of ice eddy frequency (gray lines represent the 200 m and 2000 m bathymetry lines of IBCAOv4.0). **(B)** Monthly variation in the number of ice eddies detected during 2022-2023.

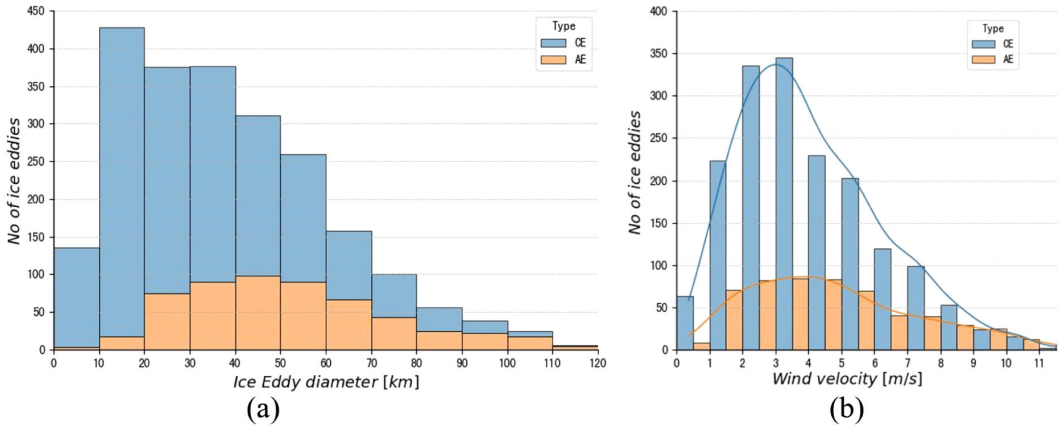


FIGURE 17
Histogram distributions of ice eddy number, ice eddy diameter, and wind velocity. **(A)** Distributions of the number and diameter of ice eddies, with cyclones shown in blue and anticyclones in orange. **(B)** Distributions of the number of ice eddies and wind velocity (m/s), based on data from the ERA5 Interim Reanalysis.

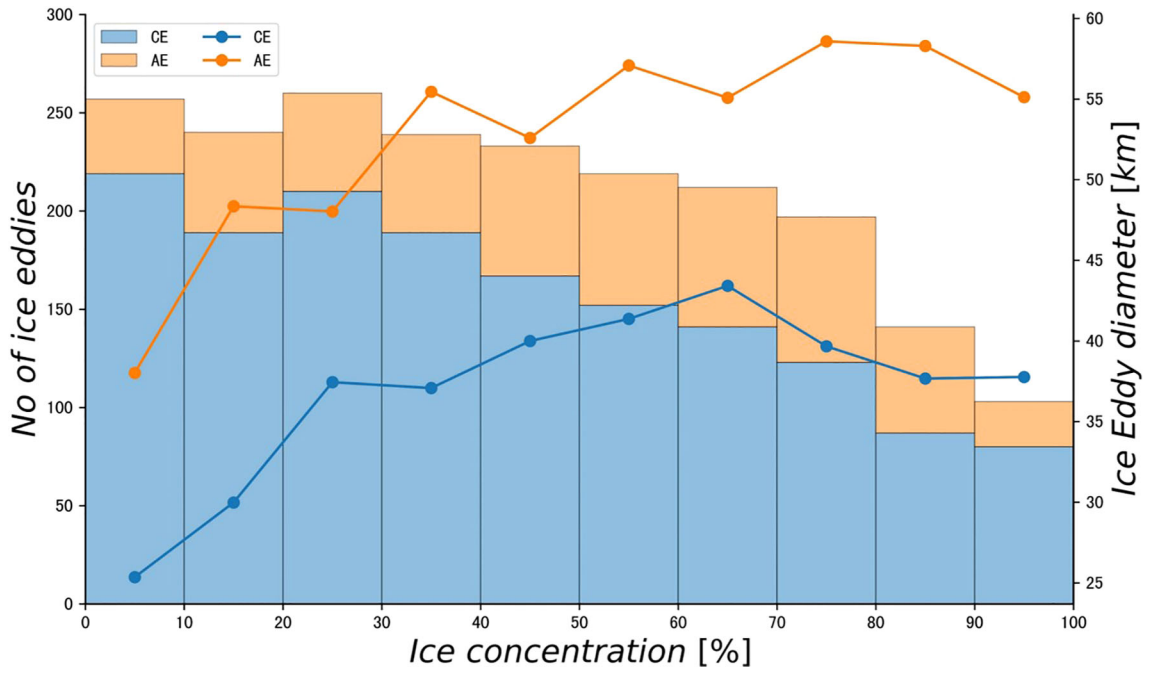


FIGURE 18
Distribution of the number of ice eddies and their average diameter (km) across ice concentration (%) for cyclonic (blue) and anticyclonic (orange) ice eddies.

accurately detects ice eddies in the Arctic MIZs but also addresses the traditional HBB's limitations in identifying ice eddies of different scales and forms. This work lays a solid foundation for future research on the automatic detection and quantification of ice eddies.

Despite the OIEDNet model's high performance in ice eddy detection, certain limitations persist. For instance, minor changes in the rotation angle can lead to significant alterations in the detection frame, increasing instability and difficulty in the detection and regression process. Exploring new angle representations could reduce ambiguity. Furthermore, we will incorporate multi-polarization and multi-frequency SAR images for model training to enhance the accuracy of ice eddy detection. The identification of ice eddy drift using Synthetic Aperture Radar (SAR) images holds significant potential for enhancing the understanding of sea ice eddy dynamics.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Author contributions

JW: Data curation, Investigation, Methodology, Software, Validation, Writing – original draft, Conceptualization, Project administration, Visualization, Writing – review & editing. YZ: Software, Visualization, Writing – review & editing, Formal analysis. TW: Methodology, Writing – review & editing, Writing – original draft, Investigation. CM: Funding acquisition, Resources, Writing – review & editing, Investigation, Supervision, Project administration. GC: Funding acquisition, Methodology, Resources, Writing – review & editing, Conceptualization, Project administration, Supervision.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported in part by Laoshan Laboratory Science and Technology Innovation Projects under Grant LSKJ202204301 and LSKJ202201302, the National Natural Science Foundation of China under Grant 42030406 and 42276179, and in part by the Taishan Scholars Program.

Conflict of interest

Authors JW and YZ were employed by Piesat Information Technology Co., Ltd.

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

References

Bashmachnikov, I. L., Kozlov, I. E., Petrenko, L. A., Glok, N. I., and Wekerle, C. (2020). Eddies in the north Greenland sea and fram strait from satellite altimetry, SAR and high-resolution model data. *J. Geophysical Research: Oceans* 125, e2019JC015832. doi: 10.1029/2019JC015832

Cassianides, A., Lique, C., and Korosov, A. (2021). Ocean eddy signature on sar-derived sea ice drift and vorticity. *Geophysical Res. Lett.* 48, e2020GL092066. doi: 10.1029/2020GL092066

Chelton, D. B., Schlax, M. G., and Samelson, R. M. (2011). Global observations of nonlinear mesoscale eddies. *Prog. oceanography* 91, 167–216. doi: 10.1016/j.pocean.2011.01.002

Ding, J., Xue, N., Long, Y., Xia, G.-S., and Lu, Q. (2019). "Learning RoI transformer for oriented object detection in aerial images," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2849–2858.

Du, Y., Liu, J., Song, W., He, Q., and Huang, D. (2019a). Ocean eddy recognition in sar images with adaptive weighted feature fusion. *IEEE Access* 7, 152023–152033. doi: 10.1109/Access.6287639

Du, Y., Song, W., He, Q., Huang, D., Liotta, A., and Su, C. (2019b). Deep learning with multi-scale feature fusion in remote sensing for automatic oceanic eddy detection. *Inf. Fusion* 49, 89–99. doi: 10.1016/j.inffus.2018.09.006

Dumont, D., Kohout, A., and Bertino, L. (2011). A wave-based model for the marginal ice zone including a floe breaking parameterization. *J. Geophysical Research: Oceans* 116. doi: 10.1029/2010JC006682

Fu, L.-L., and Holt, B. (1983). Some examples of detection of oceanic mesoscale eddies by the SEASAT synthetic-aperture radar. *J. Geophysical Research: Oceans* 88, 1844–1852. doi: 10.1029/JC088iC03p01844

Gupta, M., and Thompson, A. F. (2022). Regimes of sea-ice floe melt: Ice-ocean coupling at the submesoscales. *J. Geophysical Research: Oceans* 127, e2022JC018894. doi: 10.1029/2022jc018894

Han, J., Ding, J., Li, J., and Xia, G.-S. (2021). Align deep features for oriented object detection. *IEEE Trans. Geosci. Remote Sens.* 60, 1–11. doi: 10.1109/TGRS.2021.3062048

Huang, D., Du, Y., He, Q., Song, W., and Liotta, A. (2017). "DeepEddy: A simple deep architecture for mesoscale oceanic eddy detection in SAR images," in *2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC)*. 673–678.

Jakobsson, M., Mayer, L. A., Bringensparr, C., Castro, C. F., Mohammad, R., Johnson, P., et al. (2020). The international bathymetric chart of the arctic ocean version 4.0. *Sci. Data* 7, 176. doi: 10.1038/s41597-020-0520-9

Johannessen, J., Johannessen, O., Svendsen, E., Shuchman, R., Manley, T., Campbell, W., et al. (1987). Mesoscale eddies in the fram strait marginal ice zone during the 1983 and 1984 marginal ice zone experiments. *J. Geophysical Research: Oceans* 92, 6754–6772. doi: 10.1029/JC092iC07p06754

Karimova, S., Lavrova, O. Y., and Solov'Ev, D. (2012). Observation of eddy structures in the baltic sea with the use of radiolocation and radiometric satellite data. *Izvestiya Atmospheric oceanic Phys.* 48, 1006–1013. doi: 10.1134/S0001433812090071

Khachatrian, E., Sandalyuk, N., and Lozou, P. (2023). Eddy detection in the marginal ice zone with sentinel-1 data using YOLOv5. *Remote Sens.* 15, 2244. doi: 10.3390/rs15092244

Korosov, A. A., and Rampal, P. (2017). A combination of feature tracking and pattern matching with optimal parametrization for sea ice drift retrieval from SAR data. *Remote Sens.* 9, 258. doi: 10.3390/rs9030258

Kozlov, I. E., Artamonova, A. V., Manucharyan, G. E., and Kubryakov, A. A. (2019). Eddies in the western arctic ocean from spaceborne SAR observations over open ocean and marginal ice zones. *J. Geophysical Research: Oceans* 124, 6601–6616. doi: 10.1029/2019JC015113

Kozlov, I. E., and Atadzhanova, O. A. (2021). Eddies in the marginal ice zone of fram strait and svalbard from spaceborne SAR observations in winter. *Remote Sens.* 14, 134. doi: 10.3390/rs14010134

Li, K., Wan, G., Cheng, G., Meng, L., and Han, J. (2020). Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS J. photogrammetry Remote Sens.* 159, 296–307. doi: 10.1016/j.isprsjprs.2019.11.023

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*. 2117–2125.

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, 261–318. doi: 10.1007/s11263-019-01247-4

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.

Ma, J., Shao, W., Ye, H., Wang, L., Wang, H., Zheng, Y., et al. (2018). Arbitrary-oriented scene text detection via rotation proposals. *IEEE Trans. multimedia* 20, 3111–3122. doi: 10.1109/TMM.2018.2818020

Manucharyan, G. E., and Thompson, A. F. (2017). Submesoscale sea ice-ocean interactions in marginal ice zones. *J. Geophysical Research: Oceans* 122, 9455–9475. doi: 10.1002/2017JC012895

McWilliams, J. C. (2016). Submesoscale currents in the ocean. *Proc. R. Soc. A: Mathematical Phys. Eng. Sci.* 472, 20160117. doi: 10.1098/rspa.2016.0117

Nurser, A., and Bacon, S. (2013). Eddy length scales and the rossby radius in the arctic ocean. *Ocean Sci. Discussions* 10.5, 1807-1831. doi: 10.5194/osd-10-1807-2013

Park, J.-W., Korosov, A. A., Babiker, M., Sandven, S., and Won, J.-S. (2017). Efficient thermal noise removal for sentinel-1 TOPSAR cross-polarization channel. *IEEE Trans. Geosci. Remote Sens.* 56, 1555–1565. doi: 10.1109/TGRS.2017.2765248

Perovich, D. K., and Jones, K. F. (2014). The seasonal evolution of sea ice floe size distribution. *J. Geophysical Research: Oceans* 119, 8767–8777. doi: 10.1002/2014JC010136

Qiu, Y., and Li, X. (2022). Retrieval of sea ice drift from the central arctic to the fram strait based on sequential sentinel-1 SAR data. *IEEE Trans. Geosci. Remote Sens.* 60, 1–14. doi: 10.1109/TGRS.2022.3226223

Rabault, J., Taelman, C., Idžanović, M., Hope, G., Nose, T., Kristoffersen, Y., et al. (2024). An openmetbuoy dataset of marginal ice zone dynamics collected around svalbard in 2022 and 2023. *arXiv*. doi: 10.48550/arXiv.2409.04151

Sun, Y., and Li, X. (2020). Denoising sentinel-1 extra-wide mode cross-polarization images over sea ice. *IEEE T. Geosci. Remote.* 3, 2116-2131. doi: 10.20944/preprints202006.0092.v1

Toole, J. M., Krishfield, R. A., Timmermans, M.-L., and Proshutinsky, A. (2011). The ice-tethered profiler: Argo of the arctic. *Oceanography* 24, 126–135. doi: 10.5670/oceanog.2011.64

Varghese, R., and Sambath, M. (2024). "Yolov8: A novel object detection algorithm with enhanced performance and robustness," in *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*. 1–6.

von Appen, W.-J., Wekerle, C., Hehemann, L., Schourup-Kristensen, V., Konrad, C., and Iversen, M. H. (2018). Observations of a submesoscale cyclonic filament in the marginal ice zone. *Geophysical Res. Lett.* 45, 6141–6149. doi: 10.1029/2018GL077897

Xia, G.-S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., et al. (2018). "DOTA: A large-scale dataset for object detection in aerial images," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3974–3983.

Xia, L., Chen, G., Chen, X., Ge, L., and Huang, B. (2022). Submesoscale oceanic eddy detection in SAR images using context and edge association network. *Front. Mar. Sci.* 9, 1023624. doi: 10.3389/fmars.2022.1023624

Yang, X., Yang, J., Yan, J., Zhang, Y., Zhang, T., Guo, Z., et al. (2019). "Scrdet: Towards more robust detection for small, cluttered and rotated objects," in *Proceedings of the IEEE/CVF international conference on computer vision*. 8232–8241.

Zhang, D., Gade, M., Wang, W., and Zhou, H. (2023). EddyDet: A deep framework for oceanic eddy detection in synthetic aperture radar images. *Remote Sens.* 15, 4752. doi: 10.3390/rs15194752

Zhang, D., Gade, M., and Zhang, J. (2020). "Sar eddy detection using mask-rcnn and edge enhancement," in *IGARSS 2020-2020 IEEE international geoscience and remote sensing symposium*. 1604–1607.

Zheng, P., Chong, J., and Wang, Y. (2018). A method of automatic shape depiction and information extraction for oceanic eddies in sar images. *J. Measurement Sci. Instrumentation* 9, 241. doi: 10.3969/j.issn.1674-8042.2018.03.006

Zheng, G., Songtao, L., Feng, W., Zeming, L., and Jian, S. (2021). Yolox: Exceeding yolo series in 2021. *arXiv*. doi: 10.48550/arXiv.2107.08430

Zi, N., Li, X.-M., Gade, M., Fu, H., and Min, S. (2024). Ocean eddy detection based on yolo deep learning algorithm by synthetic aperture radar data. *Remote Sens. Environ.* 307, 114139. doi: 10.1016/j.rse.2024.114139

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 21 January 2025
DOI 10.3389/fmars.2024.1536363

OPEN ACCESS

EDITED BY
Chao Chen,
Suzhou University of Science and Technology, China

REVIEWED BY
Inam Ullah,
Gachon University, Republic of Korea
Inayat Khan,
University of Engineering and Technology, Pakistan

*CORRESPONDENCE
Ling Ke
keling@cma.gov.cn

RECEIVED 28 November 2024
ACCEPTED 30 December 2024
PUBLISHED 21 January 2025

CITATION
Liu D, Ke L, Zeng Z, Zhang S and Liu S (2025) Machine learning-based analysis of sea fog's spatial and temporal impact on near-miss ship collisions using remote sensing and AIS data. *Front. Mar. Sci.* 11:1536363. doi: 10.3389/fmars.2024.1536363

Machine learning-based analysis of sea fog's spatial and temporal impact on near-miss ship collisions using remote sensing and AIS data

Dan Liu[1], Ling Ke[2*], Zhe Zeng[1,3], Shuo Zhang[1] and Shanwei Liu[1,3]

[1]College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao, Shandong, China, [2]National Satellite Meteorological Center, China Meteorological Administration, Beijing, China, [3]Technology Innovation Center for Maritime Silk Road Marine Resources and Environment Networked Observation, Ministry of Natural Resources, Qingdao, Shandong, China

Sea fog is a severe marine environmental disaster that significantly threatens the safety of maritime transportation. It is a major environmental factor contributing to ship collisions. The Himawari-8 satellite's remote sensing capabilities effectively bridge the spatial and temporal gaps in data from traditional meteorological stations for sea fog detection. Therefore, the study of the influence of sea fog on ship collisions becomes feasible and is highly significant. To investigate the spatial and temporal effects of sea fog on vessel near-miss collisions, this paper proposes a general-purpose framework for analyzing the spatial and temporal correlations between satellite-derived large-scale sea fog using a machine learning model and the near-miss collisions detected by the automatic identification system through the Vessel Conflict Ranking Operator. First, sea fog-sensitive bands from the Himawari-8 satellite, combined with the Normalized Difference Snow Index (NDSI), are chosen as features, and an SVM model is employed for sea fog detection. Second, the geographically weighted regression model investigates spatial variations in the correlation between sea fog and near-miss collisions. Third, we perform the analysis for monthly time series data to investigate the within-year seasonal dynamics and fluctuations. The proposed framework is implemented in a case study using the Bohai Sea as an example. It shows that in large harbor areas with high ship density (such as Tangshan Port and Tianjin Port), sea fog contributes significantly to near-miss collisions, with local regression coefficients greater than 0.4. While its impact is less severe in the central Bohai Sea due to the open waters. Temporally, the contribution of sea fog to near-miss collisions is more pronounced in fall and winter, while it is lowest in summer. This study sheds light on how the spatial and temporal patterns of sea fog, derived from satellite remote sensing data, contribute to the risk of near-miss collisions, which may help in navigational decisions to reduce the risk of ship collisions.

KEYWORDS

Himawari-8 satellite data, sea fog, near miss, ship collision, spatio-temporal pattern

1 Introduction

Sea fog is a frequent and dangerous meteorological phenomenon, significantly threatening marine activity safety. This phenomenon drastically reduces the horizontal visibility of the sea surface to less than one kilometer (Gultepe et al., 2007). Unlike land-based scenarios, reduced visibility at sea poses a heightened risk due to the intricate nature of maritime navigation (Sim and Im, 2023), substantially increasing the likelihood of ship collisions and thus endangering lives, property, and the environment. Ship collisions, as one of the primary maritime accidents, can inflict substantial eco-nomic losses and adverse social impacts. Using non-accident information to understand maritime transportation safety is an effective strategy. This often involves identifying near-miss collision events from Automatic Identification System (AIS) data. Since near-miss collisions occur more frequently than actual accidents, near-miss collisions can provide richer insights for maritime traffic risk analysis than actual accident data (Zhou et al., 2021). Due to sea fog on 22 May 1922, the Peninsular & Oriental Steam Navigation Company's Egypt collided with the French cargo ship Seine en route from London to Bombay, India. The ship sank, killing 86 passengers and crew members. Because sea fog occurs geographically heterogeneously and temporally seasonally, it is crucial to analyze how it affects near-miss collisions over time and space.

In 2000, the International Maritime Organization (IMO) adopted a new requirement for all ships to carry an Automatic Identification System (AIS) that automatically communicates information among ships and coastal authorities. The AIS system transmits the ship's static, dynamic, and voyage information to the surrounding ships and AIS base stations via a specific Very High Frequency (VHF). Because of the rich positional and temporal information provided by AIS, it has become a valuable tool in maritime studies, including maritime traffic (Harun-Al-Rashid et al., 2022; Yang et al., 2024; Zhang et al., 2019), marine observing (Almunia et al., 2021; Wright et al., 2019), and ship collisions (Cai et al., 2021; Liu et al., 2023; Zhang et al., 2016), etc. The AIS is popular because of its ability to conduct in-depth studies of ship near-miss collisions.

Nowadays, water traffic safety studies are focusing on incidents narrowly susceptible to collisions, often termed "near-miss collisions". In the maritime sector, a near-miss collision refers to a scenario where two vessels pass each other in close proximity (Du et al., 2020). A prevalent method for detecting near-miss collisions involves using navigation information from AIS data (Zhang et al., 2015, 2016). The few maritime accidents so far limit the possibility of conducting large-scale collision studies. However, near-miss collisions studies can help overcome this limitation (Prastyasari and Shinoda, 2020). To prevent ship collisions more effectively, numerous studies have been conducted on the spatial geographic distribution of near-miss collisions to identify high-risk areas (Du et al., 2021; Zhixiang et al., 2019; Zhou et al., 2021). However, previous studies have primarily focused on visualizing the spatial distribution of near-miss collisions without delving deeply into the relevant influencing factors. From the maritime traffic safety perspective, the factors contributing to collisions can be categorized into human, vessel, and environment domains. Among these, environmental factors are the primary causes of accidents (Zhang and Hu, 2009). Variations in environmental conditions can significantly increase collision risks. Given that marine environmental factors exhibit stability, regularity, and spatial heterogeneity, it is crucial to optimally use the rich geographic information associated with near-miss collisions. Integrating these marine environmental factors into the research framework for near-miss collisions would enable more comprehensive and insightful studies.

Among various marine environmental factors, sea fog is a frequently occurring catastrophic weather. Studies have shown that poor visibility, often associated with fog, exerts the most significant impact on maritime traffic safety, predisposing vessels

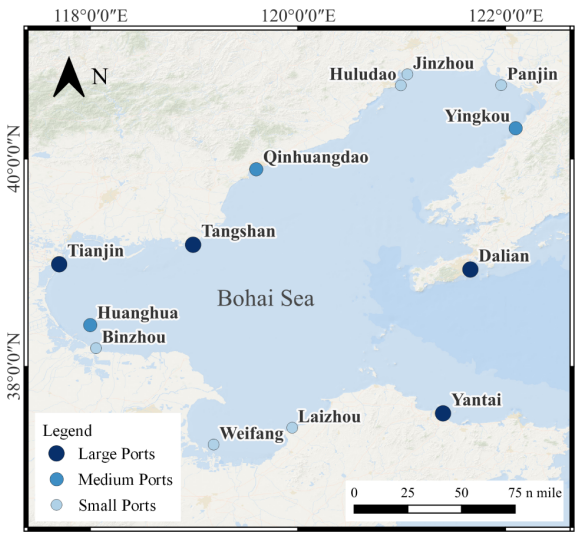


FIGURE 1
Overview of the study area.

to collision accidents (Bye and Aalberg, 2018; Gultepe et al., 2006). Approximately 70% of ship collisions are attributed to foggy conditions (Wu et al., 2015). Moreover, the consequences of ship collisions are most severe during the foggy season (Zhang and Hu, 2009). Investigating the influence of sea fog on near-miss collision risk is essential for enhancing the supervision and management of critical maritime areas and periods to ensure secured marine navigation.

Traditional sea fog detection methods rely on meteorological stations and buoys, which are sparse in spatial and temporal distributions (Kim et al., 2020). In recent years, remote sensing technology has been widely applied in ocean environment monitoring (Ullah et al., 2024; Khan et al., 2023). And, the advent of satellite remote sensing technology enables long-term and large-scale sea fog detection results. Using remote sensing for sea fog detection started in the 1970s when Hunt (Hunt, 1973) discovered significant differences in brightness temperatures between the mid-infrared (MIR) channel of 3.7 μm and the thermal infrared (TIR) channel of 11 μm for low clouds or fog with small particle size. Based on this theory, several studies have explored sea fog detection techniques, leveraging the difference between mid-infrared and thermal infrared channels (Cermak, 2012; Eyre et al., 1984; Wu and Li, 2014; Yibo et al., 2016; Zhang and Yi, 2013). Also, the sea fog detection accuracy can be enhanced with spectral indices, such as Normalized Snow Deposition Index, NDSI (Ryu and Hong, 2020), Normalized Difference Water Index, NDWI (Wu and Li, 2014), and Normalized Difference Flow Index, NDFI (Shi et al., 2023) and environmental factors such as air–sea temperature difference (Han et al., 2022). Due to the challenges of determining optimal thresholds with traditional methods, various machine-learning techniques are also widely employed in sea fog detection. With its unique vertically resolved measurement capability that provides accurate sea surface cloud information, the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (Calipso) has been widely used for sea fog detection (Badarinath et al., 2009; Cermak, 2012; Wu et al., 2015; Xiao et al., 2023; Xiaofei et al., 2021). Sea fog based on remote sensing satellites can conduct spatial analyses of ship near-miss collisions.

Many studies have examined ship near-miss collisions to achieve a safe and reliable maritime transportation system (Chai et al., 2017; Rawson and Brito, 2021; Szlapczynski and Szlapczynska, 2016). Most recent studies infer that sea fog positively affects collisions (Heo et al., 2014; Rømer et al., 1995). However, sea fog occurrences are spatially heterogeneous and temporally seasonal. Therefore, it is necessary to explore the impact of sea fog on near miss-collision risk in time and space. Conventional global regression analysis methods, such as least squares regression, assume independence and identical distribution of observations, rendering them unsuitable for analyzing spatially unevenly distributed data. Geographically weighted regression (GWR), a local linear regression method based on spatial variation relationships, is widely applied in various fields, such as meteorology (Li et al., 2024; Wahiduzzaman et al., 2022), ecology (Wang et al., 2021; Xiao et al., 2023), and economics (Cellmer et al., 2020; Shang and Niu, 2023). The model generates a regression equation at each local location, enabling spatial analysis of sea fog's impact on near-miss collisions (Yongtian et al., 2023).

However, few studies have focused on the spatial and temporal variations in the relationship between ship near-miss collisions and

TABLE 1 AHI observation bands details on Himawari-8 satellite.

Channel	Spatial resolution (μm)	Central wavelength (μm)	Main detection category
1	1	0.47	Vegetation, aerosol
2	1	0.51	Vegetation, aerosol
3	0.5	0.64	Low cloud (fog)
4	1	0.86	Vegetation, aerosol
5	2	1.6	Cloud phase recognition
6	2	2.3	Cloud droplet effective radius
7	2	3.9	Low cloud (fog), natural disaster
8	2	6.2	Water vapor density from troposphere to mesosphere
9	2	6.9	Water vapor density in the mesosphere
10	2	7.3	Water vapor density in the mesosphere
11	2	8.6	Cloud phase discrimination, sulfur dioxide
12	2	9.6	Ozone content
13	2	10.1	Cloud image, cloud top
14	2	11.2	Cloud image, sea surface temperature
15	2	12.4	Cloud image, sea surface temperature
16	2	13.3	Cloud height

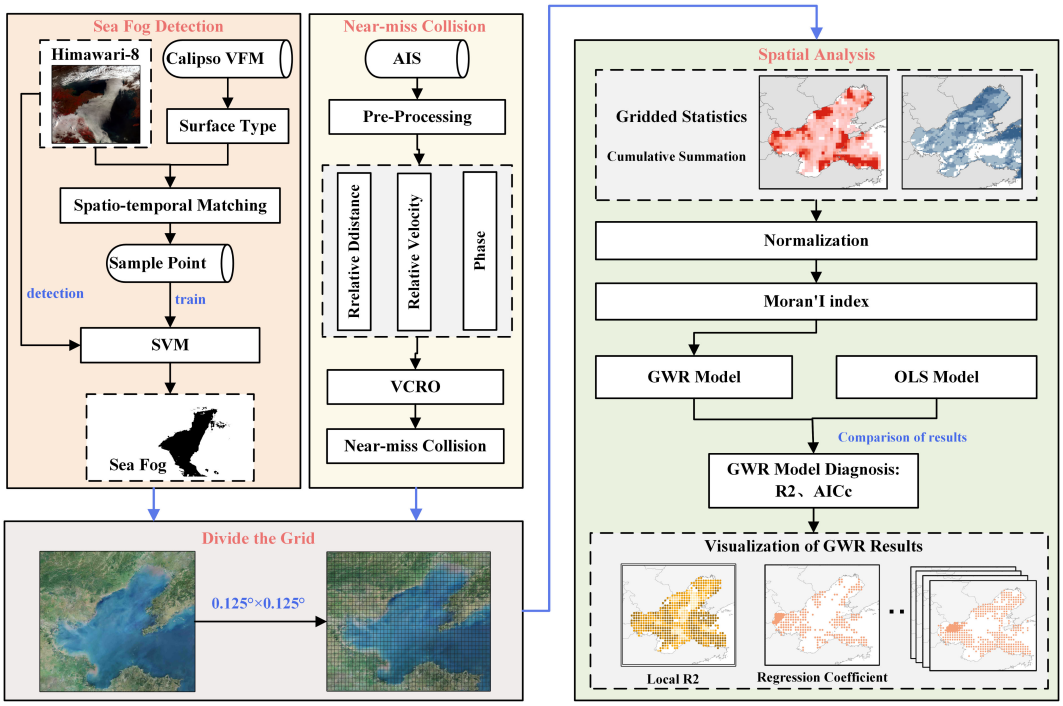


FIGURE 2
Workflow of the analytical procedure.

sea fog because traditional ocean observation data are usually in point form, which limits studying the relationship between sea fog and ship near-miss collisions in terms of spatial and temporal variations. To address the issue, this paper presents a novel approach of exploring the spatial and temporal variations in the relationship between ship near-miss collisions and sea fog. The primary contribution of the paper lies in proposing a framework for measuring spatial and temporal variation in the correlations between large-scale sea fog, which is detected using satellite remote sensing data instead of traditional point-based data from meteorological stations, and near-miss collisions which are derived from AIS data by the VCRO model. The GWR model measures the spatial variation of near-miss collisions influenced by sea fog while an average coefficient analysis of monthly data is used to describe the temporal variation of those collisions. The Bohai Sea is chosen as a case study to illustrate the approach. This study provides insights into the spatial heterogeneity and intra-annual seasonal variations of near-miss collisions influenced by sea fog. The

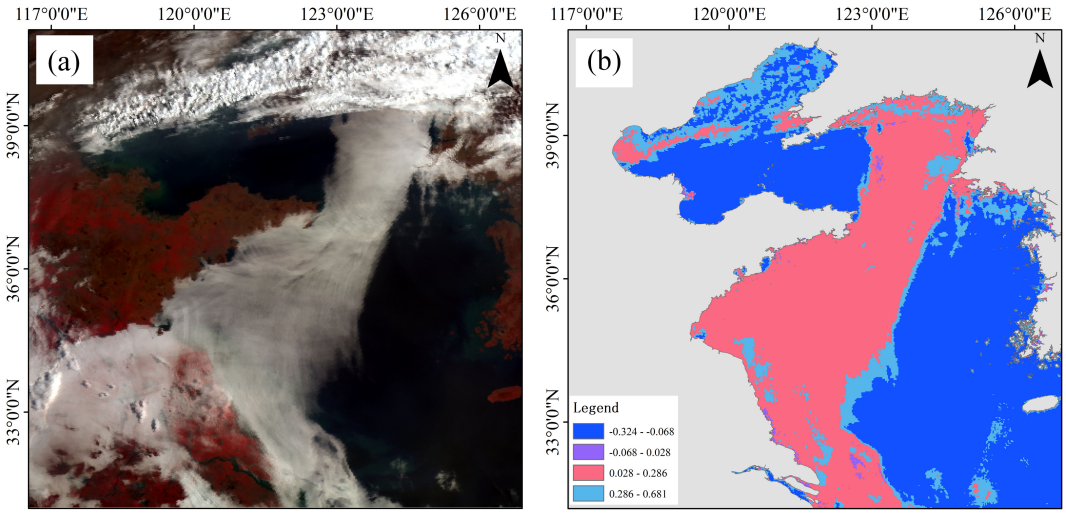


FIGURE 3
Himawari-8-image NDSI index distribution chart **(A)** Original image of Himawari-8 **(B)** NDSI calculation results shown in graded classes.

approach can support decision-making for navigation and enhance maritime safety.

2 Study area and datasets

2.1 Study area

This study selected the Bohai Sea area (37°07′~41°00′N117°35′~121°10′E) as the study area (Figure 1). This region represents the northernmost offshore area of China, surrounded by land on three sides, characterized as an almost enclosed inland sea. The Bohai Sea is particularly susceptible to sea fog. Sea fog in the Bohai Sea primarily occurs during spring and less frequently in summer. Renowned for its abundance of fisheries and mineral resources and its dense concentration of ports and harbors, the Bohai Sea emerges as one of the busiest maritime regions for shipping activities.

In 2018, the major ports in the Bohai Sea (including Tangshan, Tianjin, Dalian, Yantai, Yingkou, and Huanghua) ranked among the world's top 20 ports in terms of cargo throughput. The total port throughput size reflects a port's transport capacity. According to the 2018 port data from the China Port Yearbook, the annual throughput (in million tons) of Tianjin, Tangshan, Huanghua, Qinhuangdao, Dalian, Yantai, Yingkou, Jinzhou, Huludao, Panjin, Binzhou, Dongying, Weifang, and Laizhou Ports was 507, 637, 288, 231, 468, 443, 370, 110, 31.9, 40.91, 12, 58.25, 46.57, and 22.7, respectively. The total throughput of each port is categorized into large, medium, and small sizes based on mean and standard deviation breakpoints. Large ports include Tangshan, Tianjin, Dalian, and Yantai Ports; medium ports include Yingkou, Huanghua, and Qinhuangdao Ports; and small ports include Jinzhou, Huludao, Panjin, Dongying, Binzhou, Weifang, and Laizhou Ports.

2.2 Data

2.2.1 Himawari-8

This study's remote sensing satellite data were obtained from the Himawari-8 satellite, a third-generation geostationary meteorological satellite operated by the Japanese Meteorological Office and equipped with Advanced Himawari Imager (AHI). It covered sixteen spectral bands, including three visible light channels, three near-infrared channels, and ten infrared channels (Table 1). Its quality of cloud imagery, number of spectral bands, and clarity were substantially improved over those of previous generations. Additionally, its full-disk observation frequency of every 10 min provided excellent time resolution, thereby facilitating the study of time-series sea fog events.

2.2.2 The AIS data

The Automatic Identification System (AIS) is a shipboard monitoring system that provides vital information about a ship's position, speed, heading, and other relevant data. Being less impacted by meteorological conditions, sea surface states, and other environmental factors, AIS has gradually become a mainstream data source for ship trajectory research. The primary data used in this study is the ship's position, timestamp, direction toward the earth, and sailing speed.

This paper used the 42.6 GB of 2018 Bohai Sea area AIS data, containing a substantial data volume. To ensure the usability of the data, we initially performed preliminary cleaning to remove records with abnormal critical information, such as speed, heading, longitude, and latitude. Since analyzing the encounter process is unpractical when the shipping speed is low or in a moored state, we filtered out low-speed data and data indicating a moored sailing state. The remaining trajectory data were then divided into several sub-trajectories for detailed analysis.

3 Methodologies

Figure 2 provides the study workflow. We explored the effect of sea fog on collision risk and the key factors influencing the collision risk as explanatory variables, such as ship density. As shown in Figure 2, the main steps include identifying sea fog, calculating collision risk, dividing the sea area to be studied into grids, counting the monthly frequency of sea fog and the total collision risk, and performing spatial analyses. The main steps are further described in detail.

3.1 Sea fog detection

The advantages of remote sensing satellite data include wide coverage and continuous observation, enabling constant monitoring of sea fog over a wide range and an extended period. In this study, we used Himawari-8 satellite data, which is equipped with the Advanced Himawari Imager (AHI), a next-generation sensor with 16 spectral bands ranging from visible to infrared wavelengths. The spectral characterization of Himawari-8 data identified the bands B1, B2, B3, and B14 as the most suitable for the task. To enhance the differentiation between sea fog and other features, the Normalized Snow Deposition Index (NDSI) was constructed as follows:

$$NDSI = \frac{B_3 - B_5}{B_3 + B_5} \quad (1)$$

where B3 is the third-band reflectance and B5 is the fifth-band reflectance. Figure 3 shows the spatial distribution of NDSI index, and it can be found that most of the sea fog pixels in the Bohai Sea and Yellow Sea can be distinguished according to the NDSI index. The selected feature bands are normalized to address the varied data magnitudes in each band, which could induce low accuracy and slow computation.

In this study, only sea fog is dichotomized, i.e., into fog and non-fog categories. In sea fog remote sensing detection, visual interpretation is the conventional approach to sample selection. It involves analyzing the texture or spectral characteristics of features on satellite remote sensing images to identify those that meet the pre-defined interpretation criteria. Among the visual interpretation criteria for sea fog, the following features are essential: uniform, smooth, and delicate texture, milky white color, darker and less variable brightness, and more apparent and precise boundaries.

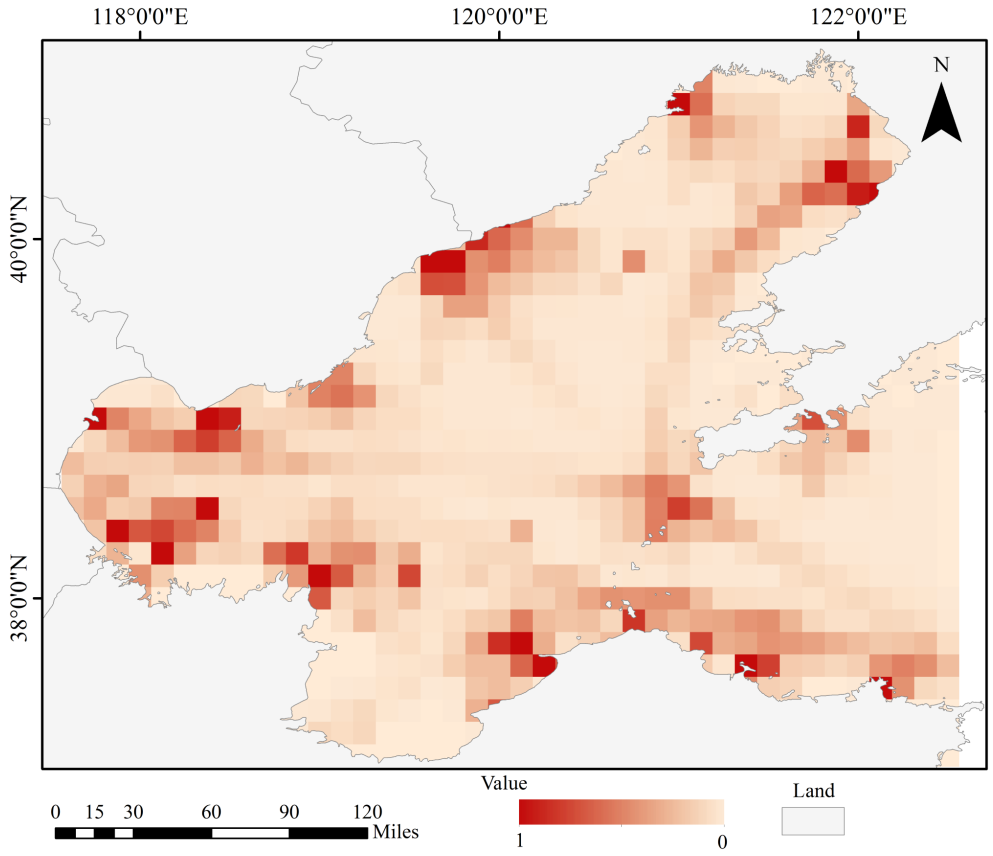


FIGURE 4
Spatial distribution of near-miss collisions in 2018.

Nevertheless, low-altitude stratocumulus clouds and sea fog are essentially clouds, with no significant difference in their physical properties. Therefore, selecting sea fog samples solely based on visual interpretation of satellite remote sensing images is subjective.

Vertical Feature Mask (VFM) data, a secondary product of CALIOP data, can differentiate among several feature types, including cloud, sea surface, subsurface, stratosphere, aerosol, and no-signal data, within the range of satellite subsurface points. The data is widely used in cloud and fog detection research. Based on the CALIOP VFM data, those connected to the sea surface were considered sea fog. The synchronized transit of CALIOP VFM data and Himawari-8 satellite images are taken. Here, synchronization is a transit time difference between the two data sets of no more than 10 minutes. Samples of sea fog and non-fog conditions have been identified through visual interpretation and are further corroborated with CALIOP Vertical Feature Mask (VFM) data. Four types of feature samples, sea fog, medium-high clouds, low clouds, and sea surface, were selected through visual interpretation and in combination with CALIOP VFM data. The samples were selected by the following cases: 1) Sea fog samples are clouds in contact with the sea surface or anomalous sea surface above sea level in the VFM data. 2) Low cloud samples are clouds with cloud base heights lower than 2 km in the VFM. 3) Medium-high cloud samples are clouds with cloud base heights greater than 2 km in the VFM. The sample selection process resulted in the following types and corresponding pixel counts: 6725 pixels for sea fog, 7267 pixels for sea surface, 6961 pixels for low-level clouds, and 9367 pixels for mid-high level clouds.

The classification model in the study is the Support Vector Machine (SVM), a novel pattern recognition method initially proposed by Vapnik and Cortes in 1995 (Vapnik, 1995). The SVM is widely used in numerous domains, including feature extraction, pattern recognition, and regression analysis. Additionally, the SVM exhibits several advantageous characteristics, such as its suitability for small-sample training, robustness, stability, and automation. It has been extensively adopted, demonstrating high efficacy in remote sensing image classifications. The system randomly generates a hyperplane in the binary classification of linearly divisible data. It moves it until the points belonging to different categories in the training set are precisely on both sides of the hyperplane, thus achieving the optimal classification with the minimum difference between similar categories and vice versa. In the case of nonlinear problems, it is necessary to map the input samples to a high-dimensional feature space and construct the optimal classification surface in this feature space. As the dimensionality of the feature space increases exponentially, computing the optimal classification plane directly in this high-dimensional space becomes challenging. The SVM addresses this issue by defining a kernel function, which

translates the problem to the input space. SVM can effectively divide sea fog and non-sea fog regions in high-dimensional feature space, especially suitable for complex data features in sea fog detection. SVM can accurately capture the distribution features of different regions by constructing the decision hyperplane to improve the classification accuracy. Unlike deep learning methods that usually rely on a large amount of labeled data, SVM can still provide good classification performance with limited sample size. In view of the difficulty and high cost of acquiring sea spray labeled data, CALIPSO data is used for labeling in this study, and SVM is able to give full play to its classification advantages with limited labeled samples. SVM has strong robustness to noise and outliers, which effectively improves the stability of the detection of sea spray, and reduces the classification error of the traditional methods in complex environments. Therefore, the SVM method can realize efficient processing while ensuring accuracy, and is an ideal choice for the sea fog detection task in this study.

This study selected the radial basis function (RBF) as the kernel function, with 70% of the samples used as training data and 30% as test data.

$$k(x, x') = \phi(x)^T \phi(x') = \sum_{i=1}^{M} \phi_i(x)\phi_i(x') \quad (2)$$

3.2 Near miss collisions

There are two main approaches for calculating collision risk based on historical AIS data. The first method utilizes Distance at Closest Point of Approach (DCPA) and Time to Closest Point of Approach (TCPA). The technique identifies near-miss collisions by establishing criteria for DCPA and TCPA within a defined vessel domain (Fukuto and Imazu, 2013; Langard et al., 2015; Yoo, 2018). Nevertheless, collision risk assessment, solely based on DCPA/TCPA, ignores the heading information between ship pairs and thus cannot detect the collision risk during head-on encounters. The second method involves constructing a model to calculate the near-miss collisions based on factors that directly influence ship collisions.

The Vessel Conflict Ranking Operator (VCRO) model assessed the collision risks between ships, with the input variables including distance, relative speed, and phase difference between the two ships (Zhang et al., 2015). The equation is as follows:

$$\mathbf{VCRO}(x, y, z) = ((kx^{-1}y)(m \cdot \sin(z) + n \cdot \sin(2z)) \quad (3)$$

where x is the distance between the two ships, y is the relative speed, z is the phase, k, m, n are the model parameters. The parameter values used in this study are based on Zhang, with k;=3.87, m=1, and n=0.386.

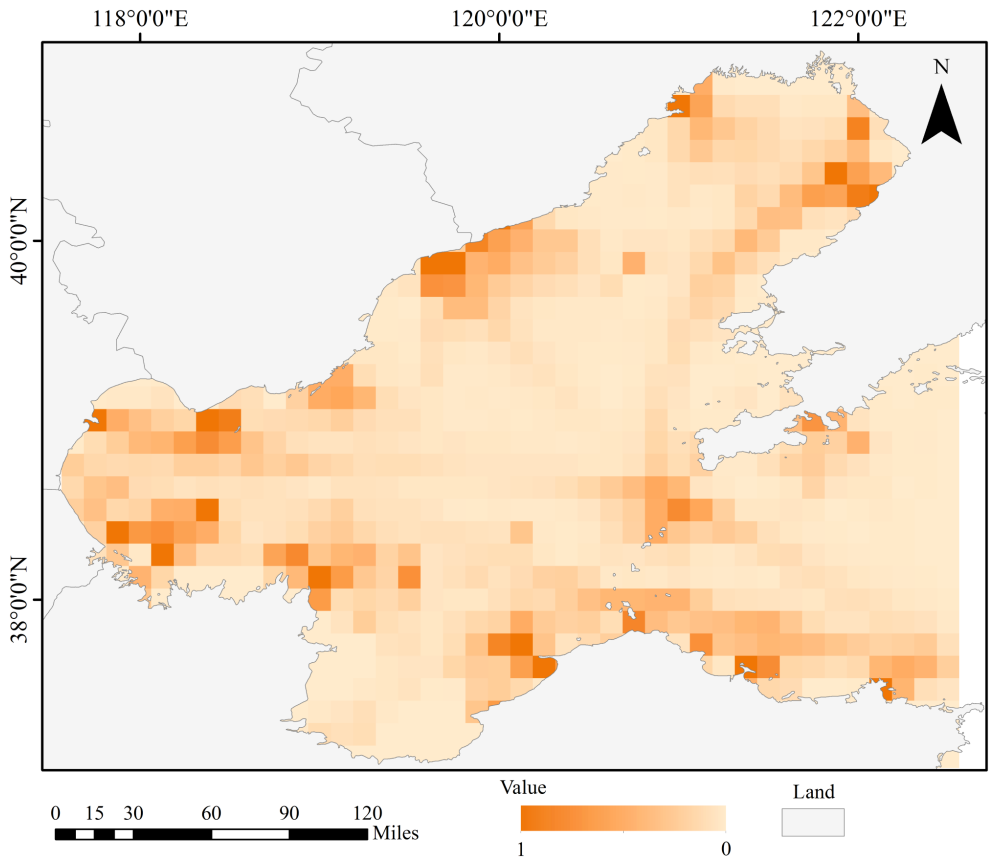


FIGURE 5
Spatial distribution of vessel density in 2018.

The relative distance between ships is calculated using Equation 4, where (x_1, y_1) represents the coordinates of ship A, (x_2, y_2) is the coordinates of ship B, and d is the distance between the centers of the two ships.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{4}$$

The relative velocity between ships is calculated using Equation 5, where V_a and V_b represent the speed of ship A and ship B, respectively, HDG_a and HDG_b represent the heading of ship A and ship B, and α represents the heading angle of the ship.

$$y_{(a,b)} = \sqrt{V_a^2 + V_b^2 - 2V_a V_b \cos\alpha} \tag{5}$$

The phase describes the relative position of the ships, denoted by angle and direction. The phase range is [-π, π], where a negative value indicates a concluded encounter and the two ships move away from each other, posing no collision risk. Conversely, a positive value indicates that the ships are approaching each other, heightening their collision risks.

To analyze the law governing ship collision risk on spatial and temporal scales, the study area must be gridded. Considering its size, the Bohai Sea is divided into grid cells of 0.125°, and the sum of near-miss collisions of each grid cell is counted as the value of this grid near-miss collisions:

$$Risk_{sum} = \sum VCRO_n \tag{6}$$

3.3 Global Moran's I

Global Moran's I is the most frequently employed statistic in global correlation analysis. It is a comprehensive measure of spatial autocorrelation across the study area (Moran, 1948). It is expressed as Equation 7, where w_{ij} represents the weight between observations i and j, and S_0 denotes the total sum of w_{ij}, given as Equation 8

$$I = \frac{n}{S_0} \times \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{7}$$

$$S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \tag{8}$$

A Moran's I > 0 indicates a positive spatial correlation, described as a "high-high, low-low" aggregation trend between neighboring elements. The larger the value, the more pronounced the spatial correlation. Conversely, Moran's I< 0 signifies a negative spatial correlation, characterized as a "high-low, low-high" distribution trend among neighboring elements. However, there is a random distribution when Moran's I = 0, indicating spatial randomness. After calculating Moran's I index, it is impossible to judge the spatial correlation directly based on its positive or negative value. The significance of the index must be assessed in combination with the p-value and Z-score.

3.4 Geographically weighted regression

According to the first law of geography, anything is spatially correlated. Geographically weighted regression is a local linear regression method that involves modeling spatially varying relationships to solve spatial heterogeneity of the variables by assigning weights to different locations (Brunsdon et al., 1996). Its Equation 9 is as follows:

$$y_i = \beta_0(\mu_i, \nu_i) + \sum_k \beta_k(\mu_i, \nu_i)x_{ik} + \varepsilon_i \tag{9}$$

where (μ_i, ν_i) denotes the position of grid cell i, $\beta_0(\mu_i, \nu_i)$ is the intercept term, $\beta_k(\mu_i, \nu_i)$ is the regression coefficient of the parameter k on the grid cell, and ε_i is the model random error. The parameter vector at location i is estimated using the weighted least square approach as follows Equation 10:

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i)X)^{-1}X^T W(u_i, v_i)y \tag{10}$$

The GWR model is adjusted using a distance decay weighted function modified by a bandwidth. The three most commonly used weighting functions are Gaussian-based, bi-square, and tri-cube kernels. Bandwidth includes fixed and adaptive types. We used a geographically weighted regression model with the dependent variable as near-miss collisions, while the explanatory variables were the frequency of sea fog, ship density. We employed a Gaussian kernel spatial weight matrix, where the weight between observation points i and j is calculated as Equation 11, where d_{ij} represents the geographical distance between the two points and b is the bandwidth parameter. We used the adaptive bandwidth specified by the Akaike information criterion (AICc) due to the uneven distribution of the near-miss collision data.

$$w_{ij} = exp\left(-\frac{d_{ij}^2}{2b^2}\right) \tag{11}$$

Further, the AICc and R^2 values evaluated the performance of the developed models. Higher R^2 indicates a better fit, while lower AICc indicates a poorer fit. The GWR model has significant advantages over the OLS model in its ability to optimize the global model on a local scale and to visualize the spatial distribution of the local regression coefficients. It enables the analyses of each factor's local contribution and non-stationarity characteristics through local coefficient variations, which are unavailable in the OLS model.

4 Result and discussion

4.1 Spatial and temporal differences in near-miss collision

Figure 4 displays the grid statistics for near-miss collisions in 2018. The value for each grid represents the total values for all near-

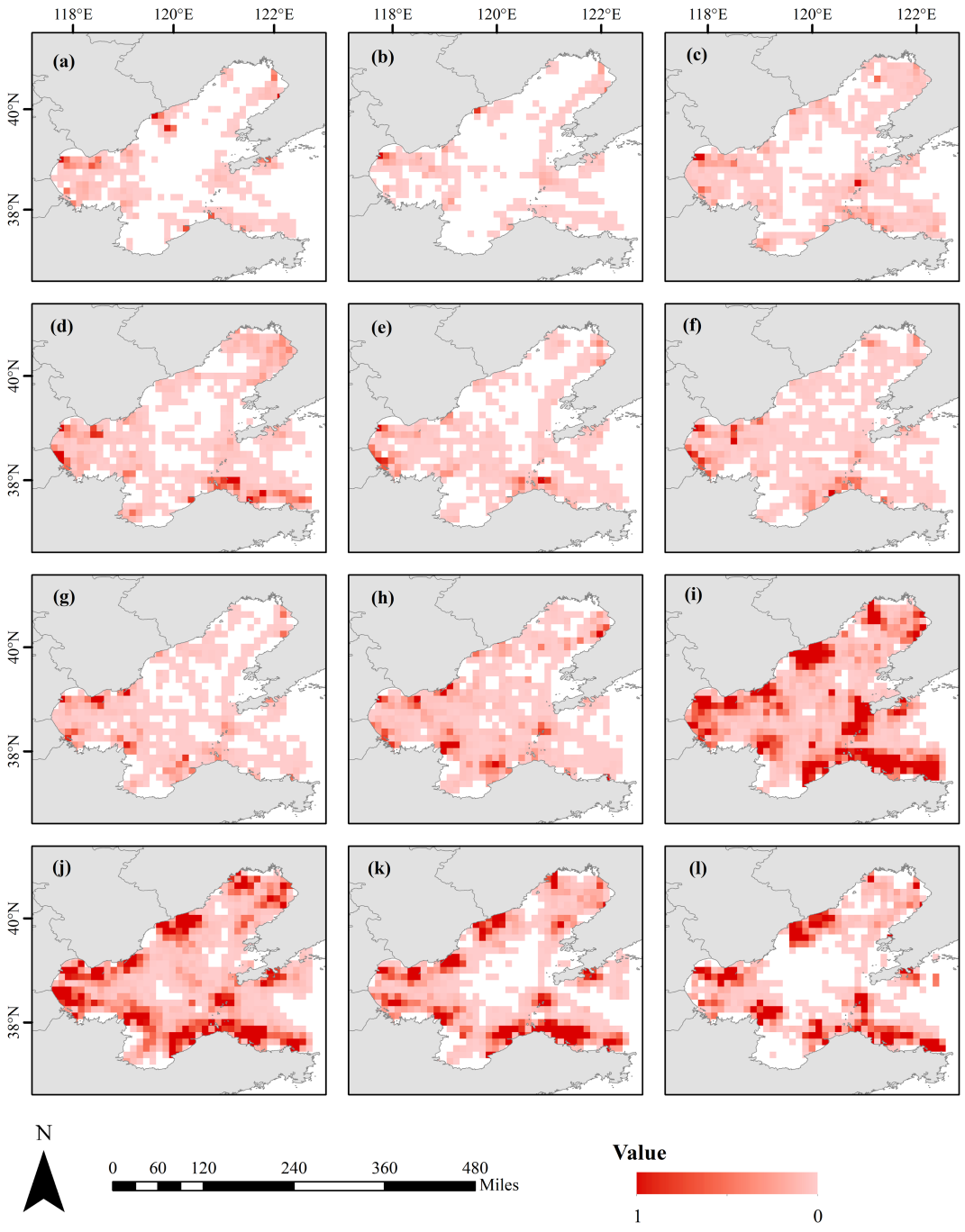


FIGURE 6
Spatial distribution of near-miss collisions, **(A–I)** represent the spatial distribution of near-miss collisions for each month from January to December 2018 respectively.

miss collisions occurring within that grid as calculated using Eq (6). Areas of high near-miss collision are concentrated around ports because of the confined navigable space and the high density of ships in these areas (Figure 5), while fewer near miss collisions were observed in the central waters of Bohai Sea. Notably, the Laotieshan Channel, located at the northernmost end of the Bohai Strait, is a major maritime transport hub in the Bohai Sea. It experiences substantial maritime traffic, resulting in a heightened risk of near-miss collision risks in the area.

In addition, Figure 6 shows the spatial distribution of near-miss collisions from January to December 2018. We observed that the fishing moratorium in the Bohai Sea, lasting from May to August, results in fewer near-miss collisions during this period. The number of near collisions starts to increase in September. By January, vessel

activity decreases as the temperature drops and the icing period begins, leading to a corresponding decline in near collisions.

4.2 Spatial and temporal differences in sea fog

The spatial distribution of sea fog in the Bohai Sea is significantly heterogeneous, with most occurrences concentrated in the southwestern and northern regions (Figure 7).

Figure 8 illustrates the monthly distribution of sea fog frequency in the Bohai Sea in 2018. The data indicate that sea fog is significantly higher in winter and spring. Despite this seasonal peak, the overall frequency of sea fog remained relatively low, with almost no occurrences in summer.

In summary, the sea fog in the Bohai Sea in 2018 has obvious spatial and temporal distribution differences, showing the characteristics of "high in spring, low in summer, high along the coast, and low in the distant sea". Spring is the high incidence of sea fog, with a wide spatial distribution; while in summer, sea fog is significantly reduced and concentrated in local coastal areas. Understanding the spatial and temporal variability in the distribution of sea fog is critical to maritime safety and the development of effective navigation strategies.

4.3 Spatial autocorrelation

Before performing the GWR model, a spatial autocorrelation analysis of sea fog occurrence was conducted using the Moran's I index, along with z-scores (indicating the distance from the mean in standard deviations) and p-values (assessing the statistical significance of the index). Table 2 presents these results for each month of 2018, as well as for the entire year. All the Moran' I index values (bounded by 1.0 and 1.0) are positive and high (> 0.25), indicating a high degree of spatial positive autocorrelation. Also, the p-values are all less than 0.01 (reaching 99% confidence level), and the z-scores are significantly higher than 2.58, indicating that the spatial autocorrelation results are statistically significant. Consequently, the linear regression model is inadequate for analyzing the impact of sea fog on collision risk. In contrast, the GWR model is well-suited to address these spatial dependencies. Using the GWR model enables an in-depth analysis, better capturing the spatial impact of sea fog on near-miss collision risks across the region.

4.4 GWR model diagnosis

The GWR models were constructed for 2018 and each month therein, with near-miss collisions as the dependent variable, while sea

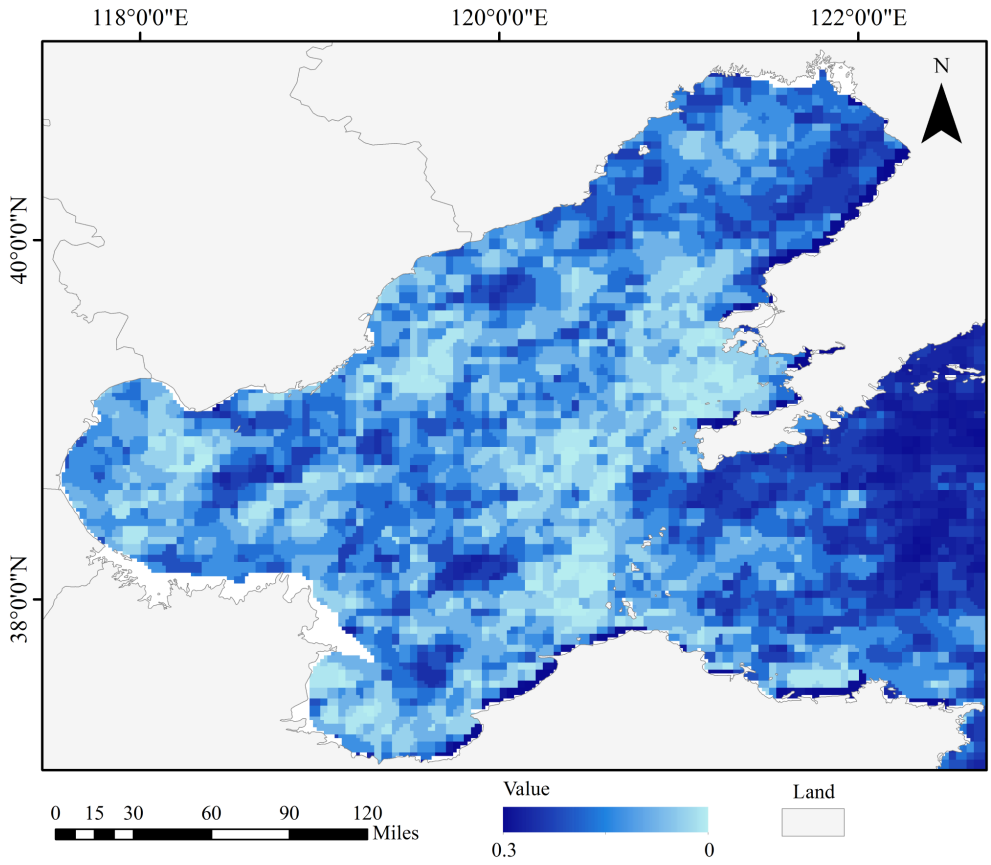


FIGURE 7
Spatial distribution of sea fog occurrence days in 2018.

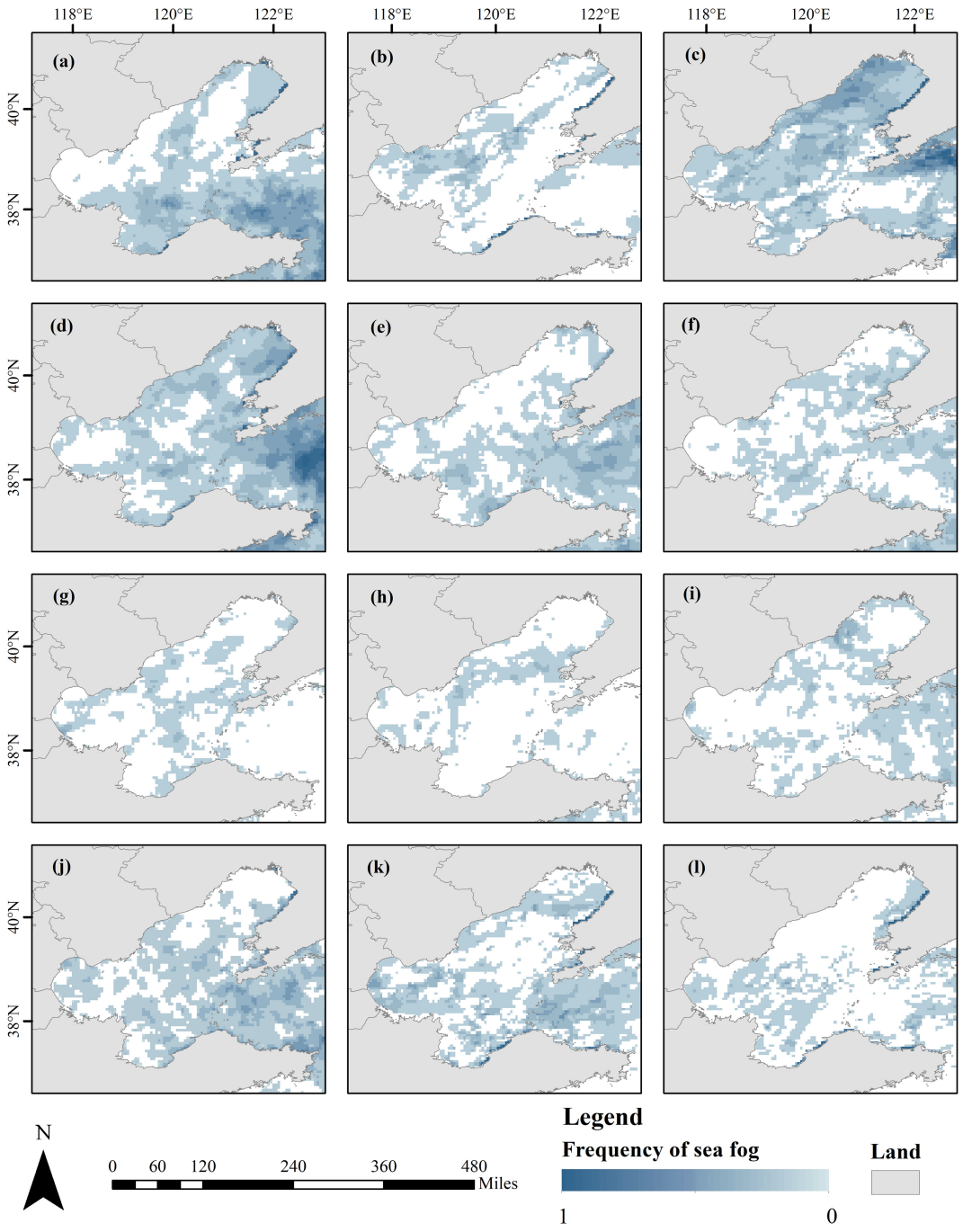


FIGURE 8
Spatial distribution of the frequency of sea fog occurrence, **(A–I)** represent the spatial distribution of the frequency of sea fog occurrence in each month from January to December 2018, respectively.

fog frequency and ship density were explanatory variables. Prior to constructing the GWR models, all values were normalized to ensure consistent scale and improve model accuracy. To assess the effectiveness of the GWR model, an OLS model was also established for comparison. The model results (Table 3) showed that the R^2 values of the OLS model are generally lower than 0.6, indicating that it explains less than 60% of the variance in near-miss collision incidents. For instance, in January, February, and March, the OLS R2 values are low at 0.10, 0.21, and 0.19, respectively, suggesting limited explanatory power. In contrast, the GWR model significantly outperforms the OLS model with R2 values above 0.7 for most months, indicating that its effectiveness in dealing with spatially heterogeneous data. Similarly, the Akaike Information Criterion corrected (AICc) values further validate the GWR model's superiority. AICc is a measure of model quality where lower values indicate better fit. The AICc values of the GWR model are lower than those of the OLS model. These results indicate that the GWR model, which accounts for spatial heterogeneity, fits the data more effectively and provides more accurate regression analyses.

TABLE 2 The spatial autocorrelation test results obtained using moran' I index combined with the z-score and p-value of sea fog (as GWR-independent variables).

Month	Moran' I	Z	P
1	0.307	11.4	0.00
2	0.314	11.67	0.00
3	0.444	16.46	0.00
4	0.524	19.28	0.00
5	0.419	15.44	0.00
6	0.271	10.03	0.00
7	0.284	10.45	0.00
8	0.264	9.74	0.00
9	0.258	9.52	0.00
10	0.358	13.25	0.00
11	0.314	11.62	0.00
12	0.303	10.98	0.00
Year	Moran' I	Z	P
2018	0.406	14.98	0

TABLE 3 Performance evaluation of the GWR and OLS model.

Month	OLS		GWR	
	R^2	AICC	R^2	AICC
1	0.10	-760.77	0.81	-1141.87
2	0.21	-651.09	0.87	-1035.79
3	0.19	-1415.51	0.93	-2378.54
4	0.15	-1493.52	0.84	-2198.11
5	0.76	-1730.34	0.95	-2278.35
6	0.12	-1515.95	0.82	-2197.67
7	0.68	-1847.73	0.94	-2590.70
8	0.47	-1642.87	0.74	-1918.78
9	0.60	-4550.58	0.78	-4689.05
10	0.30	-1399.03	0.71	-1702.76
11	0.59	-1903.69	0.84	-2302.02
12	0.57	-1596.99	0.86	-1970.90
Year	OLS		GWR	
	R^2	AICC	R^2	AICC
2018	0.265	-2084.45	0.82	-2729.98

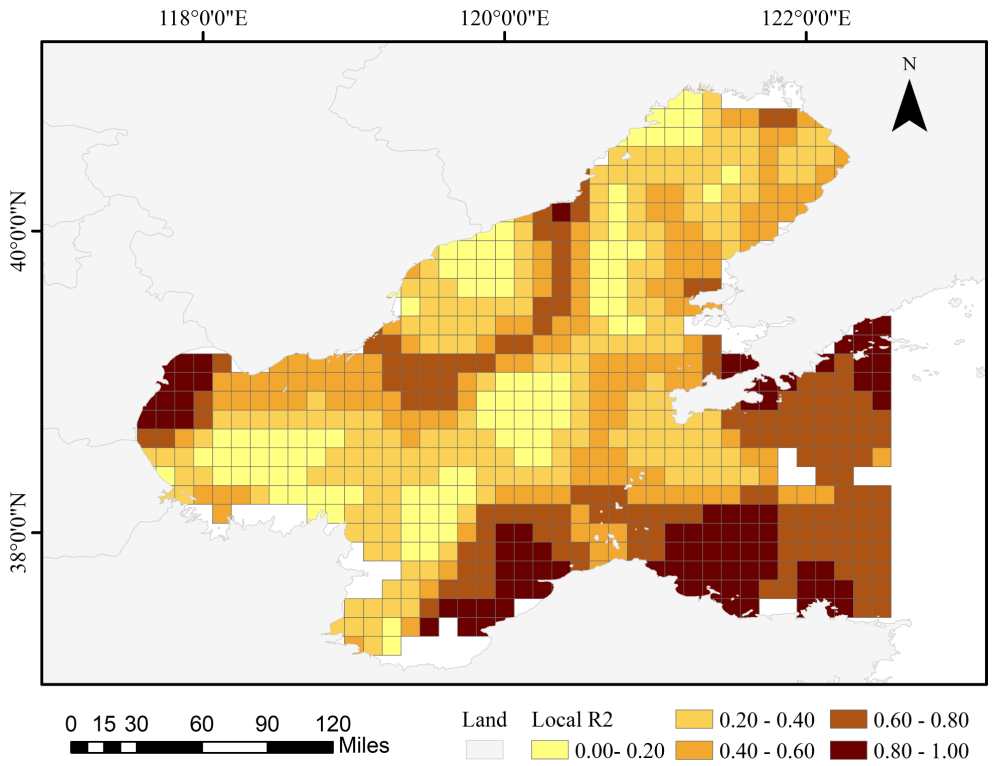


FIGURE 9
Spatial distribution of the local R2 values by GWR model in 2018.

The seasonal patterns also suggest that the GWR model performs especially well in winter and spring, when sea fog occurrences are more frequent. For example, in February through May, when sea fog events are prevalent, the GWR model R2 values range from 0.87 to 0.95. This result reinforces that sea fog, as an environmental factor, has a significant spatially variable impact on near-miss collisions during these months.

Figure 9 shows the spatial distribution of local R^2 values of GWR for the 2018 annual data. The values generally exceed 0.4, indicating that the sea fog and ship density can fit the GWR model well. Notably, the areas with higher R^2 (> 0.8) are concentrated in large port areas, such as Tianjin Port, Tangshan Port, Yantai Port, and Dalian Port. In contrast, the rest of the medium ports, such as Qinhuangdao and Yingkou Port, also have R2 between 0.6 and 0.8.

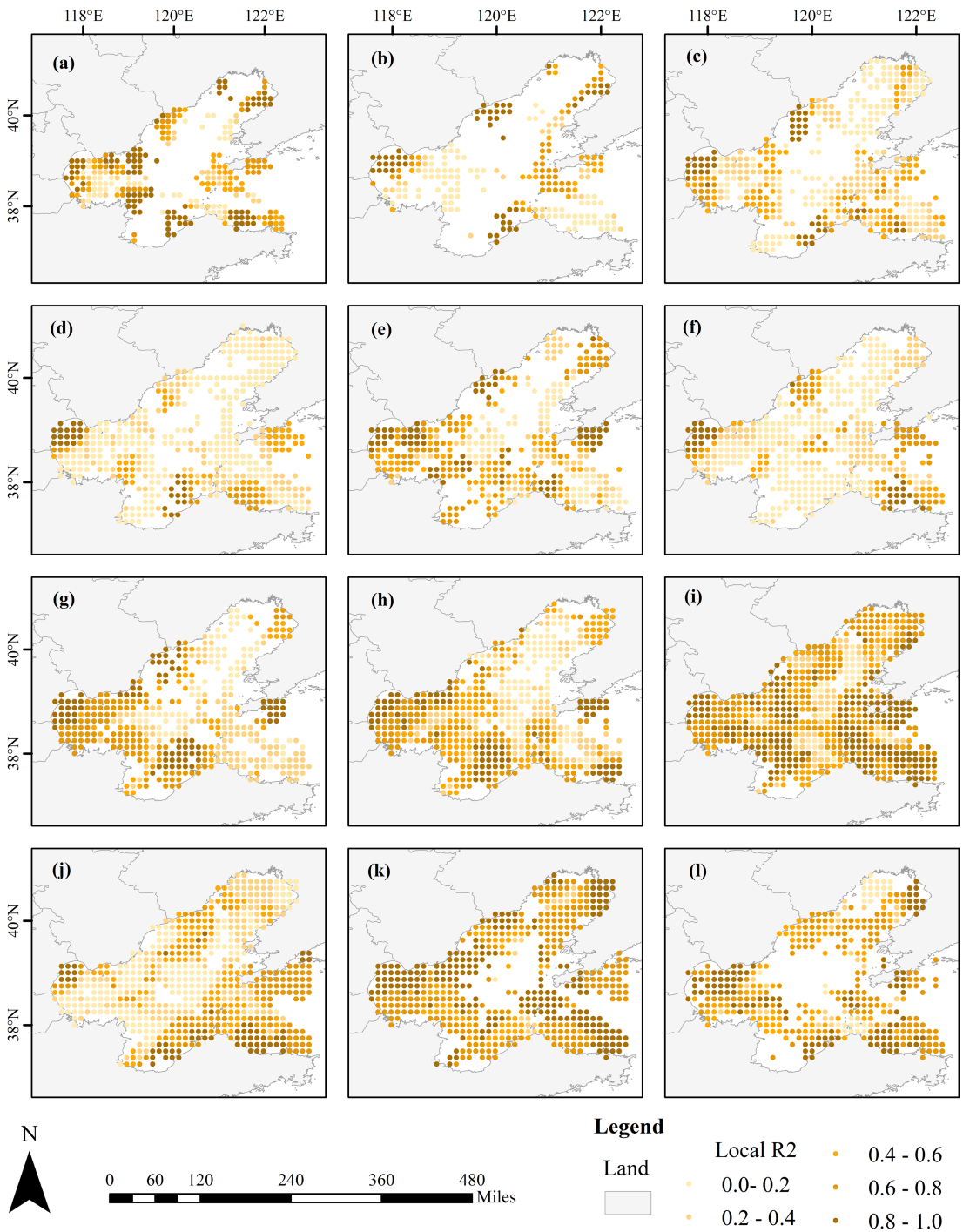


FIGURE 10
Spatial distribution of the local R2 values by GWR model, **(A–I)** represent the spatial distribution of the local R2 values in each month from January to December 2018, respectively.

Suggests that sea fog and ship density are more strongly correlated with ship near-miss collisions in ports areas.

Figure 10 displays the local R2 values for different locations in the GWR model over the 12 months of 2018, highlighting temporal variation in the model's performance across different locations. The GWR model performs well in essentially all months, with local R2 values generally exceeding 0.6, although it varies monthly for different locations. This temporal variability suggests that the influence of sea fog and ship density on collision risks may shift over time, potentially due to seasonal changes in weather conditions, maritime traffic, or operational patterns in these port areas.

Overall, the R^2 values are consistently high for most regions of the Bohai Sea. This deduction indicates that the driving factors used in the model effectively explain the spatial heterogeneity in near-miss collision risk.

4.5 Spatial relationship between sea fog and collision

The local regression coefficients of the GWR model (Figure 11) highlight the spatial variation in the effect of sea fog on near-miss collisions. The regression coefficients are generally greater than 0, indicating that sea fog positively affects near-miss collisions, thus the occurrence of sea fog contributing to collision risk. Generally, the impact of sea fog on near-miss collisions shows significant spatial inhomogeneity. The areas with the highest impact by sea fog are predominantly near the ports in the western part of the Bohai Sea, mainly concentrated around Tianjin Port and Tangshan Port. The high density of ships and heavy traffic in these harbors increase the likelihood of collision accidents when encountering sea fog due to reduced visibility and increased difficulty in ship handling. Further from these large ports, the coefficients decrease, indicating a relatively lower but still positive effect of sea fog on near-miss incidents. The areas with moderate coefficients (0.3 to 0.5) include regions around medium ports, where the collision risk remains elevated during fog but to a lesser extent than in the large ports. Therefore, near-miss collisions at key shipping nodes, such as ports, significantly increase during sea fog scenarios. Consequently, port authorities in large ports, such as Tianjin and Tangshan, should enhance navigation monitoring and optimize ship scheduling during foggy conditions to mitigate the increased risk of collisions. Implementing real-time navigation assistance and optimizing traffic flow in these key nodes can further reduce the risk of incidents under low-visibility conditions.

Figure 12 illustrates the monthly spatial distribution of local regression coefficients from the GWR model. Throughout the year, sea fog consistently shows a positive effect on near-miss collision risk, but the intensity and spatial distribution of this impact fluctuate

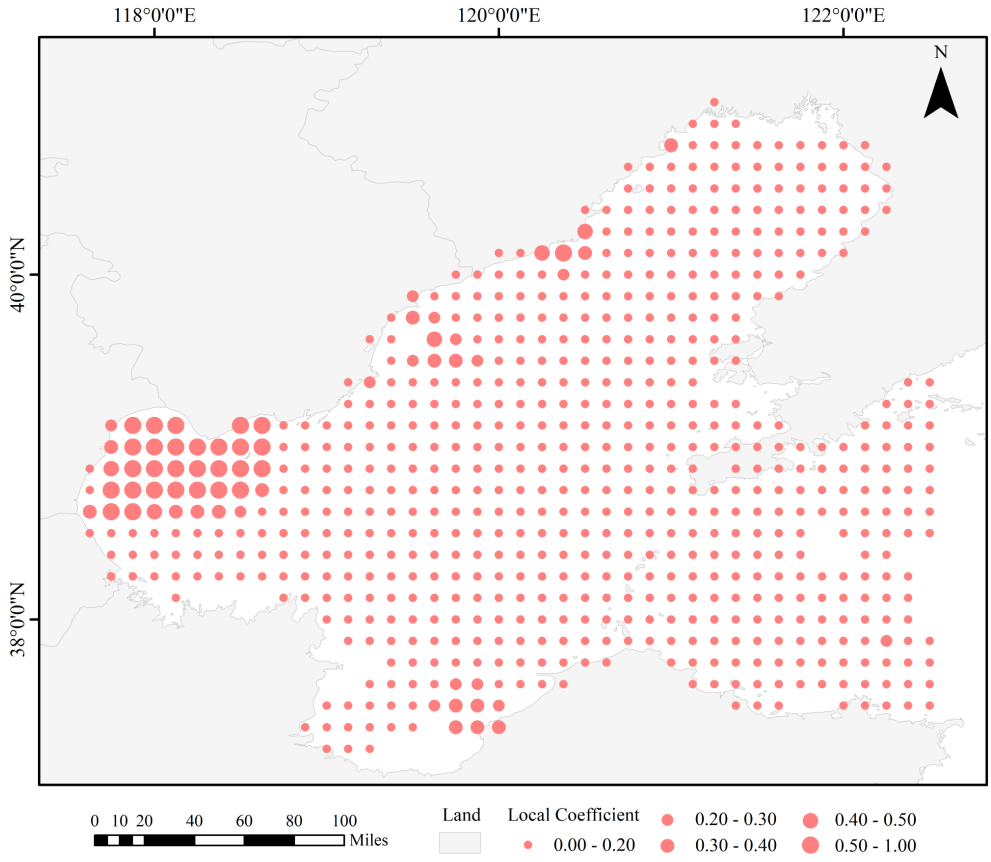


FIGURE 11
Spatial distribution of regression coefficient values of sea fog in 2018 using GWR models.

FIGURE 12
Spatial distribution of regression coefficient values of sea fog by GWR models, **(A–I)** represent the spatial distribution of regression coefficient values of sea fog in each month from January to December 2018, respectively.

significantly. Specifically, the contributions of sea fog were more significant in January, February, March, April, and June, with high-impact areas concentrated near the large, medium-sized ports in the western Bohai Sea, such as Tianjin and Yingkou Port. In contrast, May, July, and September display a more even distribution of lower local coefficients, with values generally below 0.1. This pattern suggests that during these months, the effect of sea fog on near-miss collisions is less severe across the region. In August, some changes occurred in the geographical distribution of the contribution of sea fog, with Dongying and Huludao harbors being more affected in localized areas. The effect of sea fog in the Bohai Sea intensified again from October to December, with several high-impact zones. Particularly in October, the effect was more significant, affecting the ports of Tianjin, Qinhuangdao, Laizhou, and Dongying. In November, Tianjin and Qinhuangdao ports were more affected, while in December, the port of Tianjin experienced the most significant impact.

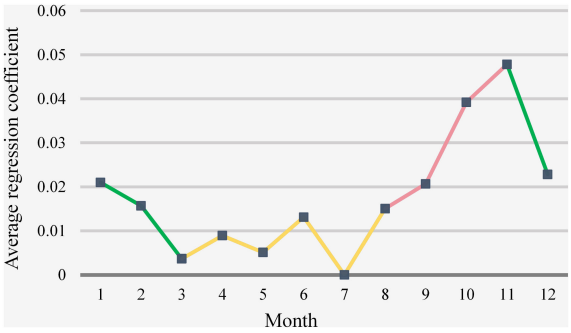


FIGURE 13
Average regression coefficients for January-December 2018.

4.6 Temporal relationship between sea fog and collision

Here, we present line plots of the average regression coefficients for each month in Figure 13, providing a visual comparative time-series analysis of how much the Bohai Sea area is affected by sea fog in different months. The results demonstrate that sea fog in autumn and winter most significantly impacts ships' near-miss collisions, while spring has the second-highest impact. In contrast, the effect of sea fog on near misses is minimal in summer. The seasonal difference can be explained in two ways. First, sea fog is less frequent in summer, which directly reduces the adverse effects of sea fog on navigational conditions. Secondly, the fishing moratorium in the Bohai Sea area coincides with summer, and the reduced activity of fishing vessels leads to a relative decrease in the number of vessels, thus reducing the risk of collision due to sea fog. Nevertheless, it is crucial to note that, although May and June also fall within the fishing moratorium period, commercial vessel activity is higher at this time than from January to April. This increased activity can still contribute to collision risks, even with the reduced traffic of fishing vessels.

Specifically, May to August is the closed season for fishing in the Bohai Sea, so the mean regression coefficient increases from September (Figure 13 red line), indicating that sea fog has started to affect ship collisions significantly. However, as winter approaches (December-March), the number of active ships decreases due to the lowering of temperatures and the freezing period, and the mean regression coefficient starts to decrease, indicating less impact by sea fog (Figure 13 green line). The regression coefficients remain smoother but slowly increase in spring and summer (March-August) (Figure 13 yellow line). In July, sea fog had almost no effect on collision risk because it hardly occurred, and the number of vessels was low during the fishing moratorium in the Bohai Sea. Collisions are more significantly affected by sea fog when vessel traffic is high. This observation suggests that navigation safety strategies should focus on periods with high vessel traffic and frequent sea fog to mitigate collision risks effectively.

5 Conclusions

This paper presents a new framework for analyzing the spatial and temporal effects of sea fog on ship near-miss collisions. Data from the Himawari-8 satellite is used to detect sea fog, with a Support Vector Machine (SVM) model applied for identification. Near-miss collisions between vessels are analyzed using the Vessel Conflict Ranking Operator (VCRO) model, which is based on Automatic Identification System (AIS) data. Spatial autocorrelation analysis by Moran 's I index reveals significant spatial heterogeneity in the distribution of sea fog. To account for this variability, a geographically weighted regression model (GWR) is employed, which enables measuring the spatial variation of sea fog's effect on ship near-miss collisions through local regression coefficients. Additionally, further conduct regression analysis on the monthly time series data to investigate the intra-annual seasonal dynamics and variations by calculating the mean regression coefficients. This temporal analysis can help us understand how the sea fog factor influences ship near-miss collisions over time. The proposed framework is implemented in a case study focused on the Bohai Sea, and the results are as follows.

According to the performance metrics (AICc and R^2), the GWR model performs much better than the OLS model. The R^2 of the GWR model ranges from 0.70 to 0.95, suggesting that GWR is more suitable for data where spatial non-stationarity exists. Regression coefficients generally greater than 0 indicate a positive influence of sea fog on ship near-miss collisions. Visualizing the local regression coefficients can intuitively reveal the spatial differences in the contribution of sea fog to ship near-miss collisions. Overall, sea areas near large and medium ports along the coast of the Bohai Sea with high ship densities, such as Tangshan Port and Tianjin Port, are more susceptible to sea fog. However, the impact on the central Bohai Sea is minimal due to the vast expanse of the water area. We estimate the mean regression coefficients for each month to explore temporal differences. It reveals that the contribution of sea fog intensifies in the autumn after the end of the fishing moratorium. In winter, the contribution of sea fog decreases due to the low number of vessel activities. However, the contribution rises steadily by spring, while it is lowest in summer due to its low occurrence frequency. Future studies should explore the spatial and temporal correlation between sea fog and ship near-miss collisions in more detail in response to multi-year data analysis. This research demonstrates that sea fog data derived from remote sensing satellite observations allows for a more comprehensive understanding of relationships and patterns in space and time.

Data availability statement

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

Author contributions

DL: Data curation, Software, Validation, Writing – original draft. LK: Formal Analysis, Methodology, Writing – review & editing. ZZ: Investigation, Methodology, Writing – original draft, Writing – review & editing. SZ: Methodology, Software, Validation, Writing – original draft. SL: Software, Validation, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

Almunia, J., Delponti, P., and Rosa, F. (2021). Using automatic identification system (ais) data to estimate whale watching effort. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.635568

Badarinath, K. V. S., Kharol, S. K., Sharma, A. R., and Roy, P. S. (2009). Fog over indo-gangetic plains-a study using multisatellite data and ground observations. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2, 185–195. doi: 10.1109/JSTARS.2009.2019830

Brunsdon, C., Fotheringham, A. S., and Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geogr. Anal.* 28, 281–298. doi: 10.1111/j.1538-4632.1996.tb00936.x

Bye, R. J., and Aalberg, A. L. (2018). Maritime navigation accidents and risk indicators: an exploratory statistical analysis using ais data and accident reports. *Reliab. Eng. Syst. Saf.* 176, 174–186. doi: 10.1016/j.ress.2018.03.033

Cai, M., Zhang, J., Zhang, D., Yuan, X., and Soares, C. G. (2021). Collision risk analysis on ferry ships in Jiangsu section of the Yangtze River based on ais data. *Reliab. Eng. Syst. Saf.* 215, 10790. doi: 10.1016/j.ress.2021.107901

Cellmer, R., Cichulska, A., and Belej, M. (2020). Spatial analysis of housing prices and market activity with the geographically weighted regression. *Isprs Int. J. Geoinf.* 9, 380. doi: 10.3390/ijgi9060380

Cermak, J. (2012). Low clouds and fog along the south-western African coast - satellite-based retrieval and spatial patterns. *Atmos Res.* 116, 15–21. doi: 10.1016/j.atmosres.2011.02.012

Chai, T., Weng, J., and De-qi, X. (2017). Development of a quantitative risk assessment model for ship collisions in fairways. *Saf. Sci.* 91, 71–83. doi: 10.1016/j.ssci.2016.07.018

Du, L., Banda, O. A. V., Goerlandt, F., Kujala, P., and Zhang, W. (2021). Improving near miss detection in maritime traffic in the northern Baltic Sea from ais data. *J. Mar. Sci. Eng.* 9, 180. doi: 10.3390/jmse9020180

Du, L., Goerlandt, F., and Kujala, P. (2020). Review and analysis of methods for assessing maritime waterway risk based on non-accident critical events detected from ais data. *Reliab. Eng. Syst. Saf.* 200, 106933. doi: 10.1016/j.ress.2020.106933

Eyre, J. R., Brownscombe, J. L., and Allam, R. J. (1984). Detection of fog at night using advanced very high resolution radiometer (avhrr) imagery. *Meteorological Magazine.* 113, 266–271.

Fukuto, J., and Imazu, H. (2013). New collision alarm algorithm using obstacle zone by target (ozt). *IFAC Proc. Volumes* 46, 91–96. doi: 10.3182/20130918-4-JP-3022.00044

Gultepe, I., Mueller, M. D., and Boybeyi, Z. (2006). A new visibility parameterization for warm-fog applications in numerical weather prediction models. *J. Appl. Meteorol. Climatol.* 45, 1469–1480. doi: 10.1175/JAM2423.1

Gultepe, I., Tardif, R., Michaelides, S. C., Cermak, J., Bott, A., Bendix, J., et al. (2007). Fog research:: a review of past achievements and future perspectives. *Pure Appl. Geophys.* 164, 1121–1159. doi: 10.1007/s00024-007-0211-x

Han, L., Zhang, S., Xu, F., Lu, J., Lu, Z., Ye, G., et al. (2022). Simulations of sea fog case impacted by air-sea interaction over South China Sea. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.1000051

Harun-Al-Rashid, A., Yang, C., and Shin, D. (2022). Detection of maritime traffic anomalies using satellite-ais and multisensory satellite imageries: application to the 2021 Suez Canal obstruction. *J. Navig.* 75, 1082–1099. doi: 10.1017/S0373463322000364

Heo, K., Park, S., Ha, K., and Shim, J. (2014). Algorithm for sea fog monitoring with the use of information technologies. *Meteorol. Appl.* 21, 350–359. doi: 10.1002/met.1344

Hunt, G. E. (1973). Radiative properties of terrestrial clouds at visible and infrared thermal window wavelengths 99, 346–369. doi: 10.1002/qj.49709942013

Khan, S., Ullah, I., Ali, F., Shafiq, M., Ghadi, Y. Y., 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. doi: 10.3389/fmars.2022.1094915

Kim, D., Park, M., Park, Y., and Kim, W. (2020). Geostationary ocean color imager (goci) marine fog detection in combination with himawari-8 based on the decision tree. *Remote Sens. (Basel)* 12, 149. doi: 10.3390/rs12010149

Langard, B., Morel, G., and Chauvin, C. (2015). Collision risk management in passenger transportation: a study of the conditions for success in a safe shipping company. *Psychol. Fr* 60, 111–127. doi: 10.1016/j.psfr.2014.11.001

Li, F., Shi, X., Wang, S., Wang, Z., de Leeuw, G., Li, Z., et al. (2024). An improved meteorological variables-based aerosol optical depth estimation method by combining a physical mechanism model with a two-stage model. *Chemosphere* 363, 142820. doi: 10.1016/j.chemosphere.2024.142820

Liu, Z., Zhang, B., Zhang, M., Wang, H., and Fu, X. (2023). A quantitative method for the analysis of ship collision risk using ais data. *Ocean Eng.* 272, 113906. doi: 10.1016/j.oceaneng.2023.113906

Moran, P. A. P. (1948). The interpretation of statistical maps. *J. R. Stat. Society: Ser. B (Methodological)* 10, 243–251. doi: 10.1111/j.2517-6161.1948.tb00012.x

Prastyasari, F. I., and Shinoda, T. (2020). Near miss detection for encountering ships in sunda strait. *IOP Conf. Series: Earth Environ. Sci.* 557(1),12039. doi: 10.1088/1755-1315/557/1/012039

Rawson, A., and Brito, M. (2021). A critique of the use of domain analysis for spatial collision risk assessment. *Ocean Eng.* 219, 108259. doi: 10.1016/j.oceaneng.2020.108259

Rømer, H., Petersen, H. J. S., and Haastrup, P. (1995). Marine accident frequencies – review and recent empirical results. *J. Navig.* 48, 410–424. doi: 10.1017/S037346330001290X

Ryu, H., and Hong, S. (2020). Sea fog detection based on normalized difference snow index using advanced himawari imager observations. *Remote Sens. (Basel)* 12, 1521. doi: 10.3390/rs12091521

Shang, X., and Niu, H. (2023). Analysis of the spatiotemporal evolution and driving factors of China's digital economy development based on esda and gm-gwr model. *Sustainability* 15, 11970. doi: 10.3390/su151511970

Shi, X., Liu, X., Zhao, S., and Gu, Y. (2023). Applicability comparison of three classical remote sensing retrieval methods for nighttime sea fog in Shandong offshore. doi: 10.1117/12.2668131

Sim, S., and Im, J. (2023). Improved ocean–fog monitoring using himawari-8 geostationary satellite data based on machine learning with shap-based model interpretation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 16, 7819–7837. doi: 10.1109/JSTARS.2023.3308041

Szlapczynski, R., and Szlapczynska, J. (2016). An analysis of domain-based ship collision risk parameters. *Ocean Eng.* 126, 47–56. doi: 10.1016/j.oceaneng.2016.08.030

Ullah, I., Ali, F., Sharafian, A., Ali, A., Naeem, H. M. Y., and Bai, X. (2024). Optimizing underwater connectivity through multi-attribute decision-making for underwater iot deployments using remote sensing technologies. *Front. Mar. Sci.* 11. doi: 10.3389/fmars.2024.1468481

Vapnik, V. (1995). *The nature of statistical learning theory* (Springer, Berlin, Germany).

Wahiduzzaman, M., Cheung, K. K., Luo, J., and Bhaskaran, P. K. (2022). A spatial model for predicting north Indian Ocean tropical cyclone intensity: role of sea surface temperature and tropical cyclone heat potential. *Weather Clim Extrem* 36, 100431. doi: 10.1016/j.wace.2022.100431

Wang, D., Wan, R., Li, Z., Zhang, J., Long, X., Song, P., et al. (2021). The non-stationary environmental effects on spawning habitat of fish in estuaries: a case study of coilia mystus in the Yangtze estuary. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.766616

Wright, D., Janzen, C., Bochenek, R., Austin, J., and Page, E. (2019). Marine observing applications using ais: automatic identification system. *Front. Mar. Sci.* 6. doi: 10.3389/fmars.2019.00537

Wu, C., Dong, H., and Ai, W. (2015). "Security countermeasures on ships sailing in the fog," in *International Conference on Electrical Engineering and Mechanical Automation (ICEEMA 2015)*. USA: DEStech Publications, Inc. 341–345.

Wu, X., and Li, S. (2014). Automatic sea fog detection over Chinese adjacent oceans using terra/modis data. *Int. J. Remote Sens.* 35, 7430–7457. doi: 10.1080/01431161.2014.968685

Wu, D., Lu, B., Zhang, T., and Yan, F. (2015). A method of detecting sea fogs using caliop data and its application to improve modis-based sea fog detection. *J. Quant Spectrosc Radiat. Transf* 153, 88–94. doi: 10.1016/j.jqsrt.2014.09.021

Xiao, Y., Liu, R., Ma, Y., and Cui, T. (2023). Merra-2 reanalysis-aided sea fog detection based on caliop observation over north pacific. *Remote Sens. Environ.* 292, 113583. doi: 10.1016/j.rse.2023.113583

Xiao, G., Wang, T., Luo, Y., and Yang, D. (2023). Analysis of port pollutant emission characteristics in United States based on multiscale geographically weighted regression. *Front. Mar. Sci.* 10. doi: 10.3389/fmars.2023.1131948

Xiaofei, G., Jianhua, W., Shanwei, L., Mingming, X., Hui, S., and Muhammad, Y. (2021). A scse-linknet deep learning model for daytime sea fog detection. *Remote Sens. (Basel)* 13, 5163. doi: 10.3390/rs13245163

Yang, J., Bian, X., Qi, Y., Wang, X., Yang, Z., and Liu, J. (2024). A spatial-temporal data mining method for the extraction of vessel traffic patterns using ais data. *Ocean Eng.* 293, 116454. doi: 10.1016/j.oceaneng.2023.116454

Yibo, Y., Zhongfeng, Q., Deyong, S., Shengqiang, W., and Xiaoyuan, Y. (2016). Daytime sea fog retrieval based on goci data: a case study over the yellow sea. *Opt Express* 24, 781-801. doi: 10.1364/OE.24.000787

Yongtian, S., Zhe, Z., Dan, L., and Pei, D. (2023). Exploring spatial non-stationarity of near-miss ship collisions from ais data under the influence of sea fog using geographically weighted regression: a case study in the Bohai Sea, China. *Hai Yang Xue Bao* 42, 77–89. doi: 10.1007/s13131-022-2137-7

Yoo, S. (2018). Near-miss density map for safe navigation of ships. *Ocean Eng.* 163, 15–21. doi: 10.1016/j.oceaneng.2018.05.065

Zhang, W. B., Goerlandt, F., Kujala, P., and Wang, Y. H. (2016). An advanced method for detecting possible near miss ship collisions from ais data. *Ocean Eng.* 124, 141–156. doi: 10.1016/j.oceaneng.2016.07.059

Zhang, W., Goerlandt, F., Montewka, J., and Kujala, P. (2015). A method for detecting possible near miss ship collisions from ais data. *Ocean Eng.* 107, 60–69. doi: 10.1016/j.oceaneng.2015.07.046

Zhang, J. P., and Hu, S. P. (2009). Application of formal safety assessment methodology on traffic risks in coastal waters & harbors. doi: 10.1109/IEEM.2009.5373103

Zhang, L., Meng, Q., and Fang Fwa, T. (2019). Big ais data based spatial-temporal analyses of ship traffic in Singapore port waters. *Transportation Res. Part E: Logistics Transportation Rev.* 129, 287–304. doi: 10.1016/j.tre.2017.07.011

Zhang, S., and Yi, L. (2013). A comprehensive dynamic threshold algorithm for daytime sea fog retrieval over the Chinese adjacent seas. *Pure Appl. Geophys.* 170, 1931–1944. doi: 10.1007/s00024-013-0641-6

Zhixiang, F., Hongchu, Y., Ranxuan, K., Shih-Lung, S., and Guojun, P. (2019). Automatic identification system-based approach for assessing the near-miss collision risk dynamics of ships in ports. *IEEE Trans. Intell. Transp Syst.* 20, 534-543. doi: 10.1109/TITS.2018.2816122

Zhou, Y., Yang, J., Bian, X., Ma, L., and Kang, Z. (2021). Macroscopic collision risk model based on near miss. *J. Navig.* 74, 1104–1126. doi: 10.1017/S0373463321000321

frontiers | Frontiers in Marine Science

TYPE Methods
PUBLISHED 03 March 2025
DOI 10.3389/fmars.2025.1528921

OPEN ACCESS

EDITED BY
Muhammad Yasir,
China University of Petroleum, China

REVIEWED BY
Aiqin Han,
Ministry of Natural Resources, China
Liangming Wang,
Chinese Academy of Fishery Sciences (CAFS), China

*CORRESPONDENCE
Chan Shu
shuchan16@mails.ucas.ac.cn
Peng Xiu
pxiu@xmu.edu.cn

†These authors have contributed equally to this work

RECEIVED 15 November 2024
ACCEPTED 12 February 2025
PUBLISHED 03 March 2025

CITATION
Fang W, Li A, Jiang H, Shu C and Xiu P (2025) Leveraging ResUnet, oceanic and atmospheric data for accurate chlorophyll-a estimations in the South China Sea. *Front. Mar. Sci.* 12:1528921. doi: 10.3389/fmars.2025.1528921

Leveraging ResUnet, oceanic and atmospheric data for accurate chlorophyll-a estimations in the South China Sea

Weiwei Fang[1†], Ao Li[2,3†], Haoyu Jiang[3,4], Chan Shu[5*] and Peng Xiu[1*]

[1]State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Xiamen University, Xiamen, China, [2]College of Marine Science and Technology, China University of Geosciences, Wuhan, China, [3]Shenzhen Research Institute, China University of Geosciences, Shenzhen, China, [4]College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, China, [5]College of Mathematics and Statistics, Huanggang Normal University, Huanggang, China

Chlorophyll-a (Chl-a) plays a vital role in assessing environmental health and understanding the response of marine ecosystems to physical factors and climate change. *In situ* sampling, remote sensing, and moored buoys or floats are commonly employed methods for obtaining Chl-a in marine science research. Although *in situ* sampling, buoys, and floats could provide accurate data, they are limited by the spatial and temporal resolution. Remote sensing offers continuous and broad spatial coverage, while it is often hindered by cloud cover in the South China Sea (SCS). This study discussed the feasibility of a predictive model by linking the physical factors [e.g., wind field, surface currents, sea surface height (SSH), and sea surface temperature (SST)] with surface Chl-a in the SCS based on the ResUnet. The ResUnet architecture performs well in capturing non-linear relationships between variables, with the model achieving a prediction accuracy exceeding 90%. The results indicate that (1) the combination of oceanic dynamical and meteorological data could effectively estimate the Chl-a based on deep learning methods; (2) the combination of meteorological and SST effectively reproduces Chl-a in the northern SCS, while adding surface currents and SSH improves model performance in the southern SCS; (3) With the addition of surface currents and SSH, the model effectively captures the high Chl-a patches induced by eddies. This research presents a viable method for estimating surface Chl-a concentrations in regions where they are highly correlated with dynamic factors, using deep learning and comprehensive oceanic and atmospheric data.

KEYWORDS

ResUnet, chlorophyll-a, deep learning, South China Sea, physical factors

1 Introduction

Phytoplankton chlorophyll-a (Chl-a) is a key indicator of marine phytoplankton biomass and primary productivity (Fernández-González et al., 2022). The SCS is characterized by diverse biogeochemical regimes, which is related to the dynamical process over the SCS. Nutrients from rivers such as the Pearl River and the Mekong River typically dominate the shelf regions (Dai et al., 2022). While the central SCS exhibits oligotrophic conditions with low productivity and depths exceeding 5000 m (Chen, 2005). The East Asian monsoon largely drives the circulation in the South China Sea (SCS), forming the South China Sea Western Boundary Current influencing the distribution of nutrients (Fang et al., 2012). Under northeasterly monsoon and stronger Kuroshio intrusion, a cyclonic circulation prevails in the upper layer during winter (Qu, 2000; Gan et al., 2006). However, some studies indicate an anticyclonic circulation pattern (Chu et al., 1999; Xue et al., 2004; Fang et al., 2009), while others describe a cyclonic circulation in the northern SCS (NSCS) and an anticyclonic circulation in the southern SCS (SSCS) (Figure 1; we recreated it based on the Shu et al., 2018; Liu et al., 2008).

In the NSCS, Chl-a concentrations display a marked seasonal cycle, with high levels in winter and low levels in summer (Ning et al., 2004; Xian et al., 2012). The SCS connects to the Pacific Ocean through the Luzon Strait, allowing the Kuroshio to intrude into the SCS and contribute to its circulation (Xue et al., 2004; Qian et al., 2018; Cai et al., 2020). Winter phytoplankton blooms in Luzon Strait are often attributed to the interaction between monsoon-driven or current-induced upwelling, vertical mixing, meso-scale eddies, and fronts (Peñaflor et al., 2007; Shen et al., 2008; Wang et al., 2010, 2023; Shang et al., 2012; Lu et al., 2015; Xiu et al., 2016; Guo et al., 2017; Chang et al., 2022; Lao et al., 2023). The Luzon Cold Eddy, generally prevailed in winter and spring near the northwestern coast of Luzon Island, would alter the distribution of the Chl-a near the Luzon Island (Lu et al., 2015; Huang et al., 2019; Sun et al., 2023). During the summer, when the southwest monsoon prevails, upwelling and a northeastward jet are induced along the coast of Vietnam (Kuo, 2000; Fang et al., 2002; Xie et al., 2003; Lin et al., 2009; Ma et al., 2012). The upwelling elevates nutrients into shallow layers, supporting phytoplankton growth, resulting in the surface high Chl-a (Yang et al., 2012; Chen et al., 2021). With the transport of this jet in nutrients and biomass, the Chl-a off the east of the Vietnam significantly was enhanced. The interaction between cyclonic and anticyclonic eddies with the jet stream formed a high Chl-a belt (Liang et al., 2018).

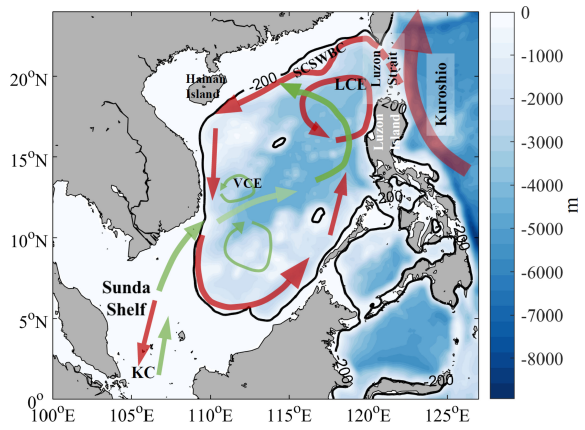


FIGURE 1
Diagram of the surface current patterns (based on data from Shu et al. 2018; Liu et al., 2008). Red (Green) means current pattern in winter (summer). LCE, Luzon Cold Eddy; SCSWBC, South China Sea Western Boundary Current; VCE, Vietnam Cold Eddy; KC, Karimata Current.

There are several methods to measure Chl-a concentrations in the ocean, each with its own limitations. Traditionally, *in situ* ship-based, autonomous profiling float, and remote sensing satellites are the primary means of acquiring Chl-a data in the ocean (Kishino et al., 1997; Wright, 1997; Dierssen, 2010; Rykaczewski and Dunne, 2011; Boyce et al., 2012; Wernand et al., 2013). *In situ* ship-based and floats generally have low spatial or temporal resolution. Remote sensing satellite, offering high spatial and temporal resolution data, is easily affected by cloud cover (Shropshire et al., 2016). Considering the difficulties in acquiring the Chl-a, simulating the Chl-a or phytoplankton with marine ecological numerical model was an excellent method. However, the accuracy of numerical model results depends on the parameterization scheme of ecological (or biogeochemical) processes and the optimization of parameters. Developing a robust ocean ecological model requires substantial time for construction, calibration, and computation.

Recently, machine learning techniques, particularly deep learning, have advanced rapidly. The application of machine learning in ocean science has provided new insights into predicting key environmental or hydrodynamic indicators (Jouini et al., 2013; Aleshin et al., 2024; Krestenitis et al., 2024). Due to its strong capabilities in nonlinear regression, deep learning has been extensively utilized in oceanography, for tasks such as predicting sea surface temperature (SST), eddies, waves, and Chl-a (Liu et al., 2021; Liu and Li, 2023; Roussillon et al., 2023; Zhao et al., 2024). Ding and Li (2024) compared the performance of CNN, LSTM, and hybrid CNN-LSTM models for Chl-a prediction, concluding that the hybrid CNN-LSTM model outperformed standalone models with an R-squared, R^2 = 0.72. Similarly, Zhou et al. (2024) contributed further insights into the application of machine learning for ecological predictions. However, in some cases, machine-learning was not performed well than empirical algorithms. Bygate and Ahmed (2024) combined observational data and Landsat 8 surface reflectance to evaluate empirical and machine learning models for retrieving water quality indicators in Matagorda Bay, highlighting the limitations of traditional machine learning models in water quality inversion. Yang et al. (2024) developed a self-attention mechanism-based deep learning model to estimate nine phytoplankton pigment concentrations within the upper 300 m of the ocean, achieving $R^2 > 0.8$ and revealing a positive correlation between the maximum phytoplankton layer location and the Niño 3.4 index in the Equatorial Pacific Niño 3.4 region. Roussillon et al. (2023) introduced a multi-mode CNN to globally reconstruct phytoplankton biomass by learning region-specific responses to physical forcing. Their model achieved an $R^2 > 0.87$, highlighting the capacity of multi-mode approaches to uncover spatially consistent responses to ocean dynamic.

On the one hand, previous studies have revealed various complex dynamical processes related with the surface Chl-a in the SCS (Dai et al., 2022; Xian et al., 2012; Wang et al., 2023; Guo et al., 2017; Ma et al., 2012; Yu et al., 2019). On the other hand, machine learning has the advantage of finding complex nonlinear relationship among variables in an environmental setting (Song and Jiang, 2023). Hence, machine learning can provide a powerful support in elucidating the complex quantitative relationship between the physical factors (such as wind, SST) and the surface Chl-a. A few studies have used machine learning or deep learning to build a model link the physical factors and surface Chl-a with monthly data (Li et al., 2023; Roussillon et al., 2023). However, the possibility and performance by using the atmospheric and oceanic physical data to predict surface Chl-a with daily data remains unclear. This study discussed the feasibility of a predictive model based on the ResUnet architecture (Diakogiannis et al., 2020) to predict daily Chl-a concentrations in the SCS (100°E-124°E, 0°N-25°N) by atmospheric and oceanic dynamic factors. The ResUnet model enables the capture of the effects of multiple ocean dynamical processes on Chl-a evolution from the data. This approach yields accurate results while significantly reducing computational costs compared to traditional ocean ecological modeling methods.

2 Data and methods

2.1 Data

The dataset used in this study was derived from the atmosphere and ocean reanalysis datasets, European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA-5; https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5) and Hybrid Coordinate Ocean Model (HYCOM; https://www.hycom.org/). The 10 m wind fields were derived from the ERA-5, with spatial resolution as 0.25°×0.25° and the temporal resolution as 1-hourly. We calculated the mean value per 24 hours for acquiring the daily air forcing data to keep the same temporal resolution in our study. The SST, surface currents (eastward and northward velocity) and sea surface height (SSH) were derived from the HYCOM. The original spatial resolution is 0.08° and temporal resolution is 3-hourly. We interpolated the original data to the ERA-5 resolution and calculated the daily data every 8 times layer. These physical factors, such as wind, current, SSH, and SST, have been shown to be closely related to the variation in surface Chl-a in previous studies (Yu et al., 2019; Xiu et al., 2016; Geng et al., 2019).

This study focuses on discussing feasibility of a predictive model capable of forecasting future Chl-a concentrations by establishing a link between oceanic and atmospheric dynamic variables (e.g., wind fields, sea surface temperature, and current fields) and surface Chl-a. The predictive model requires complete and valid Chl-a as the label to ensure the effectiveness of the model. However, there is a number of missing values in the SCS from the remote sensing satellite data. Therefore, the Chl-a data used as the target variable (True) was derived from the Ye et al. (2024). The data covers the period from January 1, 2013, to December 31, 2017, with a temporal resolution of daily averages. This dataset was reconstructed using a combination of satellite and observational data, employing optimal interpolation and the SwinUnet method. Ye et al. (2024) successfully reconstructed a high-quality surface Chl-a dataset; however, the approach relies heavily on satellite remote sensing data, which limited the application in short-term prediction. In contrast, numerical models, such as HYCOM and ERA5, could provide oceanic and atmospheric dynamic factors, which can be leveraged to predict short-term variations in surface Chl-a. For this purpose, we considered the datasets from Ye et al. (2024) as the true Chl-a to train a model with physical factors. More information is listed in Table 1.

2.2 Methods

2.2.1 Data pre-processing

In order to achieve spatial resolution consistency across all predictor variables, we employed linear interpolation to adjust predictor variables from HYCOM to a resolution of 0.25° × 0.25°. Each predictor variable contained 97 × 101 data grid points, covering the period from 2013 to 2017. To maintain consistency among the variables, data standardization was applied. The daily

TABLE 1 Introduction of the datasets used in this study.

DataSets	Unit	Min	Max	Spatial Resolution	Time Period	Data Sources
Chl-a	$mg\ m^{-3}$	0.0012	4.9×10^{33}	0.0105°	2013.01 – 2017.12	Ye et al. (2024)
Wind speed	$m\ s^{-1}$	1.4	15.4	0.25°		ERA5 (Wind stress curl is calculated based on the Equations 1, 2)
Wind stress curl	$N\ m^{-3}$	-2×10^{-7}	2.5×10^{-7}			
10m v wind	$m\ s^{-1}$	-32.6	32.9			
10m u wind	$m\ s^{-1}$	-31.3	32.2			
Sea surface temperature	°C	12.35	34.05	0.08°		HYCOM
u-velocity	$m\ s^{-1}$	-1.7	1.8			
v-velocity	$m\ s^{-1}$	-2.0	1.8			
Sea surface height	m	-0.1	1.6			

predictor variables, represented as two-dimensional arrays of 97 × 101, were then concatenated to form a three-dimensional array with dimensions N × 97 × 101, with each variable occupying a separate channel within the data structure. In our experimental design, the predictors include data points for all available variables on a given day, which are subsequently used to forecast the Chl-a concentration (predictand) for that same day. To align the predictand data with the model output, Chl-a data was resampled to 0.25° × 0.25° before model training and was standardized thereafter. Following training, the model outputs were denormalized to retrieve the predicted Chl-a values. The experimental results demonstrated that this methodology effectively enhances the model's fitting performance. The wind stress and wind stress curl in Table 1 are calculated as follows:

$$\vec{\tau} = \rho C \vec{u} \cdot |\vec{u}| \tag{1}$$

$$\nabla \times \vec{\tau} = \frac{\partial \tau_y}{\partial x} - \frac{\partial \tau_x}{\partial y} \tag{2}$$

The $\vec{u}$ is the wind vector, and $\vec{\tau}$ is the wind stress. τ_x and τ_y represent the eastward and northward component of the wind stress. The ρ and C are the air density and drag coefficient, respectively. The C is estimated based on Large and Pond (1981).

2.2.2 Residual U-Net model

The UNet is a deep learning architecture for image segmentation that utilizes a symmetric encoder-decoder structure with skip connections to effectively capture and preserve detailed spatial information (Ronneberger et al., 2015). In this study, we employed a modified UNet architecture to enhance effectiveness, as shown in Figure 2. The model features a U-shaped structure with four encoder-decoder modules. To enhance the model's ability to handle non-linear relationships, the traditional ReLU activation function was replaced with the Sigmoid Linear Unit (SiLU) activation function due to its advantage in smooth activation (Elfwing et al., 2017). To address overfitting and mitigate issues of exploding or vanishing gradients, Batch Normalization (BN) was applied after the convolutional layers. Furthermore, the AdamW optimizer was employed to improve training stability and performance by effectively managing weight decay (Loshchilov and Hutter, 2019). Consistent with most regression tasks, Mean Squared Error Loss (MSELoss) was utilized as the loss function. These modifications were implemented to collectively improve the model's performance, accuracy, and computational efficiency.

The basic module of the UNet network is a residual module, each of which consists of two 3 × 3 two-dimensional convolutional layers, two BatchNorm2d layers, and two SiLU activation functions. The encoder part (left half of Figure 2) consists of a residual module and a max pooling layer. This configuration gradually reduces the feature mapping dimensions in length and width, thereby enhancing higher-order features. Following the encoder, the same number of decoders (right half of Figure 2) decode the features, including up-sampling to double the size of the feature map and skip connections. This process produces a feature map of size [64, 97, 101]. The final layer of the model is a 1 × 1 convolutional layer that reduces the number of channels to 1, producing the final 97 × 101 Chl-a outputs of the model. Definitions of deep learning terms, including Residual Block, SiLU, and max pooling, are provided in the Appendix.

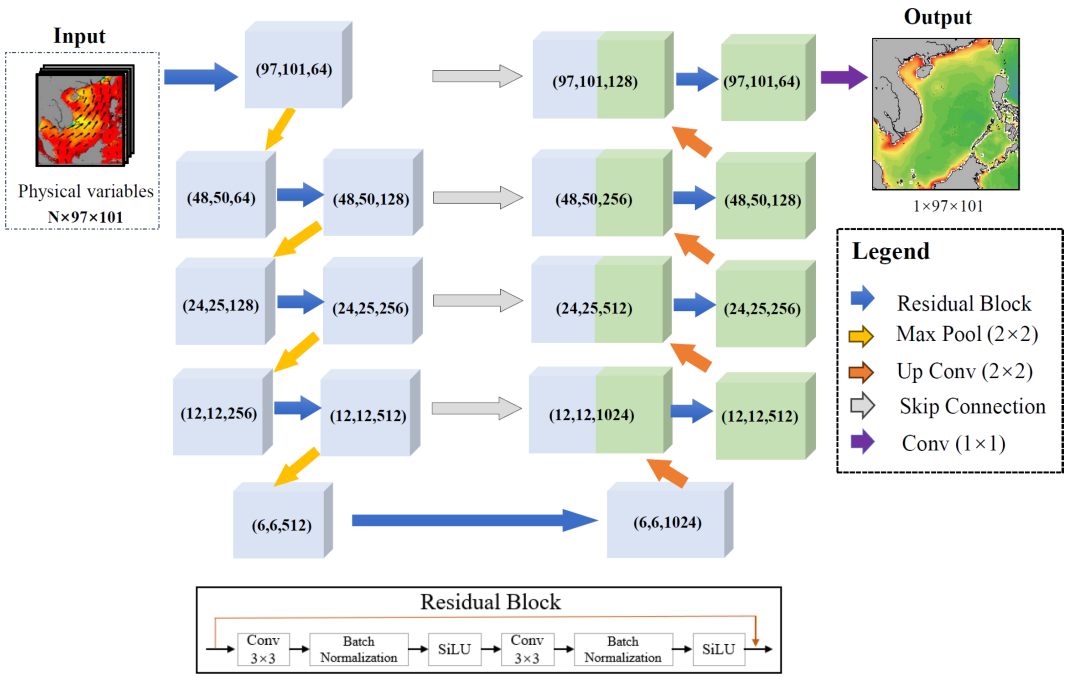


FIGURE 2
An illustration of the ResUNet architecture. Each colored cube symbolizes a feature map, with the numbers within the parentheses indicating the (width × height × channels).

2.2.3 Data split and model accuracy metrics

In this study, Chl-a data from 2013 to 2016 were allocated for model training, testing, and validation at proportions of 70%, 20%, and 10%, respectively. Data from 2017 was subsequently utilized to evaluate the model's effectiveness in applications. There are some extremely large anomalies (> 10^{10}) in Chl-a data from Ye et al. (2024). Therefore, during data preprocessing, we conducted thorough data cleaning and identified anomalies in the Chl-a data for a total of 26 days, which were removed to maintain the accuracy and consistency of the dataset. To comprehensively evaluate model performance, we employed three key metrics: the correlation coefficient (r), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). These metrics offer a quantitative assessment of the correlation and discrepancies between predicted and True data, thus providing valuable insights into the model's performance and reliability.

1. Correlation Coefficient (r): It measures the strength and direction of the linear relationship between predicted and True values, calculated as:

$$r = \frac{\sum(y_i - \overline{y})(\hat{y}_i - \overline{\hat{y}})}{\sqrt{\sum(y_i - \overline{y})^2}\sqrt{\sum(\hat{y}_i - \overline{\hat{y}})^2}}$$

2. Root Mean Square Error (RMSE): RMSE quantifies the average deviation of predictions from actual values, given by:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

3. Mean Absolute Error (MAE): MAE provides a straightforward interpretation of the average prediction error:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

The symbols used in the equations are defined as follows: y_i represents the True value, $\hat{y}_i$ denotes the predicted value, $\overline{y}$ is the mean of the True values, $\overline{\hat{y}}$ is the mean of the predicted values, and n refers to the number of observations. It is already known that there is a certain correlation between atmospheric and oceanic dynamic data and surface Chl-a in the SCS (Yu et al., 2019). The temporal and spatial variation of Chl-a are influenced by factors such as wind fields, ocean currents, and SST. To test the accuracy of model in different predictors, we conducted two sets of experiments: (1) using 10 m wind field, wind speed, wind stress curl, and SST (Exp1), and (2) using 10 m wind field, wind speed, wind stress curl, SST, surface current, and SSH (Exp2). In the SCS, wind fields and SST are strongly correlated with surface Chl-a (Yu et al., 2019). Therefore, the goal of the Exp1 was to explore the feasibility of building a robust model. On the other hand, surface current and SSH are related to the horizontal advection process and vertical structure of density to some extent (e.g., mesoscale eddies), which, to some extent, influence the distribution of nutrients and phytoplankton (Xiu et al., 2016). The goal of the Exp2 was to explore the performance of the model when considering the currents and SSH.

3 Results and discussion

3.1 Model evaluation using statistical indicators

The comparisons between predicted and true Chl-a concentrations of two experiments based on the Chl-a from 2013 to 2016, separated into three parts (training, testing, and validation sets), are shown in Figure 3. In general, the data points are primarily distributed along the 1:1 line, with correlation coefficients between predicted and true Chl-a exceeding 0.9 across all datasets (Figure 3). It indicated that both Exp1 and Exp2 could well predict the surface Chl-a in the SCS. However, there were some discrepancies in performances between these two experiments. The Exp2 showing higher correlation coefficient (Figures 3a–f) among training (0.929 in Exp1 versus 0.935 in Exp2), testing (0.911 in Exp1 *vs* 0.918 in Exp2) and validation datasets (0.913 in Exp1 versus 0.925 in Exp2). And the RMSE of Exp2 were 0.1, 0.112, and 0.107 for the training, testing, and validation datasets, respectively (Figures 3d–f). It also indicated that the deviation between the predicted values and the true values of the model is smaller. The comparison of MAE between Exp1 and Exp2 also denoted the Exp2 might be better. Li et al. (2023) employed four machine learning methods to predict the Chl-a using physical factors with Random Forests demonstrating the best performance (R^2 ~ 0.8). Aleshin et al. (2024) applied LightGBM and ResNet-18 to predict the Chl-a with an R^2 ~ 0.7. Roussillon et al. (2023) used a multi-mode convolutional neural network to reconstruct satellite-derived Chl-a with monthly physical drivers, such as SST, with R^2 ~ 0.85. In comparison, our model exhibited superior performance in predicting the Chl-a in the SCS.

Further, the residuals between predicted and true Chl-a, separated into training, testing, and validation sets, from 2013 to 2016 were calculated and shown in Figure 4. The results showed that frequency of the residuals shown normal distribution (Figure 4). The average of the residuals is -0.00039, -0.00095, -0.00045 for training, testing, and validation datasets in Exp1, respectively (Figures 4a–c). While the averages of the residuals are -0.00015, -0.00163, and -0.00061 for training, testing, and validation datasets in Exp2, respectively (Figures 4d–f). Although the mean residuals in Exp2 was less than Exp1, both Exp1 and Exp2 had small mean residuals (< 1%), which indicated a good performance of the model without significant systematic bias. This reflected the robustness and reliability of the model in capturing the surface Chl-a. In addition, the σ were about 0.14, 0.22, and 0.21 for training, testing, and validation datasets in Exp1, respectively (Figures 4a–c). They were slightly higher than the corresponding parts in Exp2 (Figures 4d–f). It denotes the results of Exp2 are more stable compared to the result of Exp1.

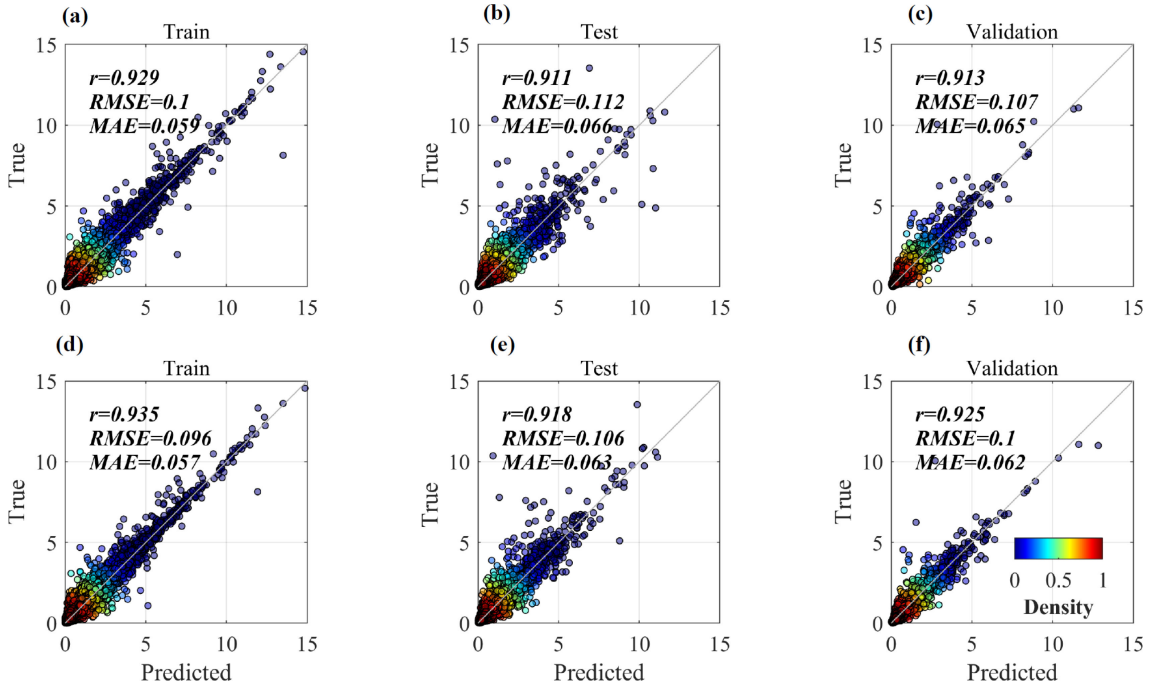


FIGURE 3
Scatter plots between predicted and the counterpart locations of the True Chl-a in training **(a, d)**, testing **(b, e)**, and validation **(c, f)**. The first (second) row represents Exp1 (Exp2).

3.2 Model evaluation in terms of Chl-a temporal and spatial distributions

The model performance was evaluated using correlation coefficients, RMSE, and MAE, all of which indicated good performance for this deep learning model. Experimental results suggested that surface currents (eastward and northward velocities) and SSH slightly enhance the model's performance. The model's ability to predict the spatial distribution and seasonal variation of surface Chl-a requires further evaluation.

To represent seasonal variations (Spring, Summer, Autumn, and Winter), surface Chl-a values from the validation dataset on the dates 2013/03/05, 2013/06/15, 2013/09/28, and 2013/12/11 were selected. Figure 5 illustrates the spatial distributions of Chl-a for these selected dates across the true, Exp1, and Exp2. Generally, surface Chl-a exhibits high concentrations on the shelf, particularly along the coast, and low concentrations in the basin of the SCS (Liu et al., 2002, 2012; Shen et al., 2008; Fang et al., 2014). The high Chl-a on the shelf is typically attributed to riverine inputs, such as nutrients, biomass, terrestrial transport, and upwelling (Li et al., 2018; Lu and Gan, 2015). Both the Exp1 and Exp2 effectively captured the prominent feature of the higher Chl-a along the coast and lower Chl-a in the basin (Figures 5e–l).

Meanwhile, seasonal Chl-a variation were exhibited significantly (Figures 5a–d). Along the coast, the area with high Chl-a (e.g., > 0.4) were more prominent in the Spring and Winter (Figures 5a, d), while they were lower in the Summer and Autumn (Figures 5b, c). And the Chl-a in the basin were lowest during the Summer (Figures 5b). This feature was also captured by the model in both Exp1 and Exp2 (Figures 5f, j). Additionally, the Luzon Strait, as a major pathway between the Pacific and the SCS, shows significant blooms in winter and spring when northeasterly winds prevail (Peñaflor et al., 2007; Shen et al., 2008). The true Chl-a data includes a notable phytoplankton bloom on the western side of the Luzon Strait (see arrow in Figures 5a, d). Both Exp1 and Exp2 predicted similar phytoplankton blooms, although the area might be slightly larger.

In terms of the overall Chl-a distribution, both Exp1 and Exp2 successfully captured the high Chl-a on the shelf and low Chl-a in the basin, and the seasonal variation of the surface Chl-a. They also reproduced the relatively high Chl-a concentration on the northwest side of Luzon Island (Figures 5e, h, i, l) in Spring and Winter. Based on the evaluation of the Chl-a spatial pattern and seasonal variation, the two experiments demonstrated good performance.

3.3 Spatial distribution of temporal correlation coefficients

The model well captured the spatial pattern and seasonal variation of the Chl-a in both Exp1 and Exp2. However, the temporal correlation between true Chl-a and model predicted Chl-a was unclear. To evaluate the model's performance in capturing Chl-a temporal variation, the Pearson correlation coefficients between true Chl-a and model predicted Chl-a for

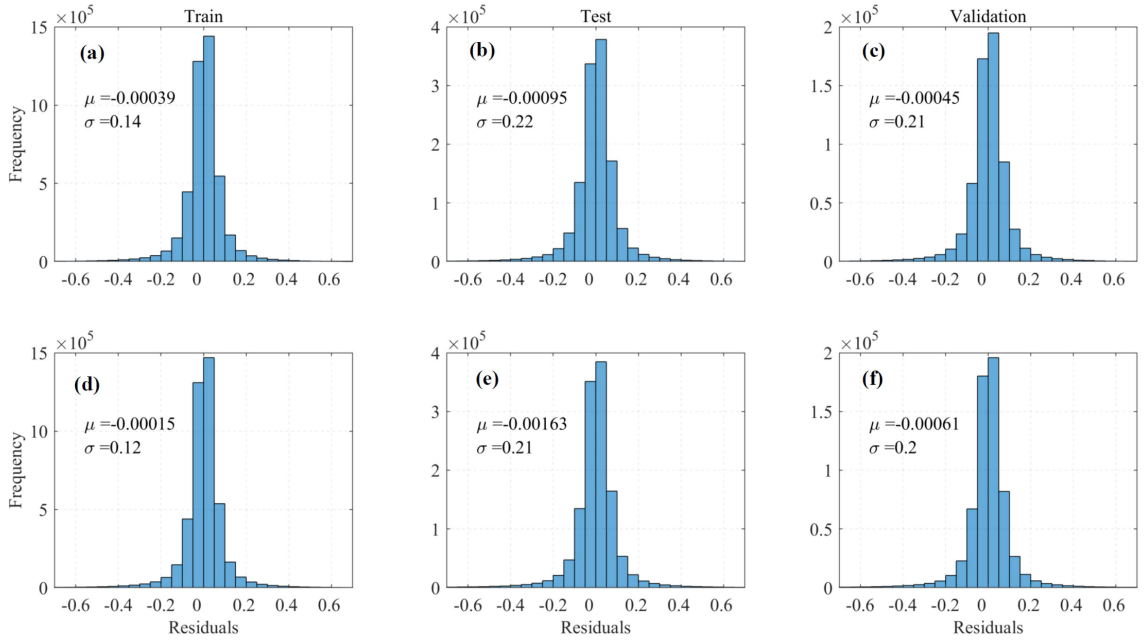


FIGURE 4
Frequency plots with x-axis as residuals (model results – True value) in training **(a, d)**, testing **(b, e)**, and validation **(c, f)**. The first (second) row represents Exp1 (Exp2).

each grid were calculated (Figure 6). Figure 6 illustrates the spatial distribution of correlation coefficients in training (Figures 6a, d), testing (Figures 6b, e), and validation (Figures 6c, f).

In general, the correlation coefficients in the training dataset (Figures 6a, d) were the highest, which is reasonable given that the training dataset was used to train the model. Regarding the spatial pattern of the correlation coefficients, whether in the Exp1 or Exp2, the values to the north of 16°N were notably higher than those to the south of 16°N in training, testing, and validation (Figures 6a–f). Specifically, the correlation coefficients in the NSCS were generally above 0.8, while in the SSCS, they typically ranged from 0.6 ~ 0.8, with the highest values observed in the training dataset (Figures 6a, d). This discrepancy might be caused by the strength of the relationship between physical factors and surface Chl-a in the NSCS and SSCS. Significant seasonal and inter-seasonal variability of Chl-a is observed in the NSCS (Shen et al., 2008; Palacz et al., 2011; Tang et al., 2014), which is generally associated with the seasonal dynamics of factors such as the monsoon and Kuroshio intrusion (Xue et al., 2004; Xian et al., 2012; Chang et al., 2022; Sun et al., 2023). Previous studies have shown a high correlation between SST and Chl-a (Shen et al., 2008; Tang et al., 2014; Yu et al., 2019). In summer, the mixed layer depth (MLD) is shallow, and the presence of strong stratification due to high SST and weaker winds inhibits the supply of nutrient-rich subsurface water. However, in winter, the MLD usually deepens due to intensified northeasterly monsoons and buoyancy flux, accompanied by a reduction in SST (Tang et al., 2003). As the MLD deepens, nutrient-rich water from the subsurface is transported to the surface layer. With sufficient nutrient support, phytoplankton flourishes during winter. Consequently, Exp1 performs well in capturing the temporal variability of surface Chl-a in NSCS (Figures 6a–c). However, in the SSCS, Geng et al. (2019) revealed that wind- and buoyancy-induced mixing are less intense in the central SCS than in the NSCS, limiting vertical nutrient transport to above the subsurface Chl-a maximum layer. This may explain the lower correlation coefficients in the SSCS (Figures 6a–f).

In respect of the comparison between Exp1 and Exp2, the correlation coefficients in the Exp2 were generally slightly higher than that in the Exp1 in the SCS (Figures 6g–i). However, in the Exp1, the correlation coefficients in the NSCS were comparable with those of Exp2, especially in the training dataset, with increasing correlation coefficients less than 0.03 (Figures 6g–i). It indicated that atmospheric data and SST are crucial factors for simulating the Chl-a in the NSCS. However, between 12°N and 16°N, Exp2 performed well in capturing the temporal variation of Chl-a, with Δr ($r_{Exp2} - r_{Exp1}$) exceeding 0.04 (Figures 6h, i). Generally, Exp2 performed better than Exp1, although there were small areas with decreased correlation coefficients to the south of 16°N. In the basin of SSCS, the correlation coefficients were higher than that on the shelf. For Exp1, the correlation coefficient in the Sunda Shelf were not as strong as in Exp2, with $r < 0.7$ (Figure 6c). However, the Exp2 showed slightly improvement in the Sunda Shelf with slightly higher r (Figure 6i). Comparisons between Exp1 and Exp2 demonstrated that the model achieved the best performance when SSH and currents were included as an input variable, especially in the SSCS.

3.4 Model performance in capturing local important features

We evaluated the model based on spatial distribution of Chl-a and the temporal correlation by Pearson correlation coefficients

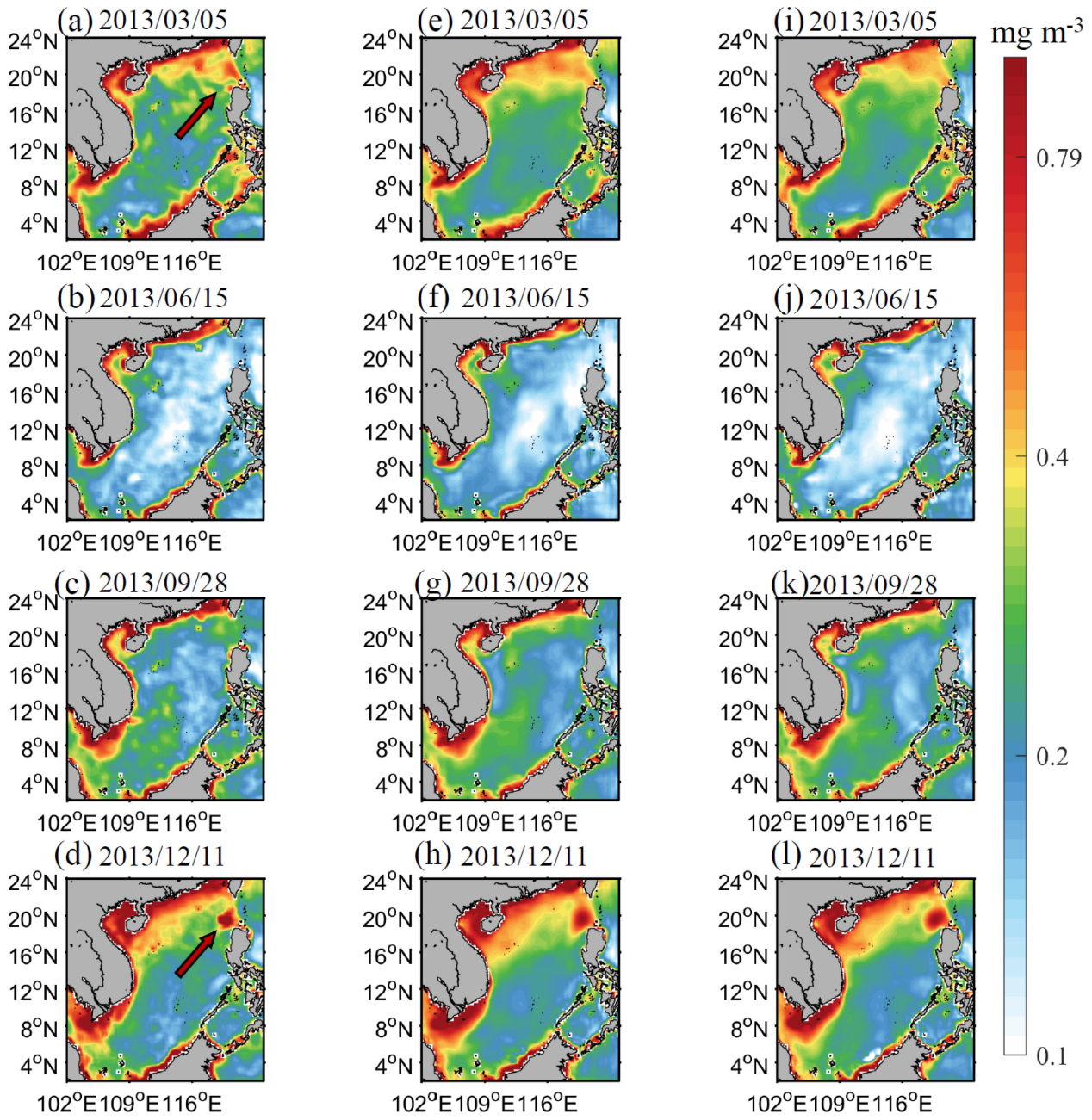


FIGURE 5
Spatial distributions of Chl-a in 2013/03/05 **(a, e, i)**, 2013/06/15 **(b, f, j)**; 2013/09/28 **(c, g, k)**; 2013/12/11 **(d, h, l)**. The first column is the true Chl-a, while the second and third column represent Chl-a in Exp1 and Exp2, respectively.

between true Chl-a and model predicted Chl-a. It denoted the performance of the model was excellent, especially for the NSCS. However, the model's ability to reproduce local spatial characteristics of Chl-a required further assessment. We selected typical high surface Chl-a patches near the Luzon Strait, Hainan Island, and Vietnam (see red arrows in Figures 7a, d, g) to validate the model's ability in capturing details from validation datasets (2014/01/30, 2013/10/20, 2014/7/23). Figures 7a, d showed a high surface Chl-a patch surrounded by low surface Chl-a. Previous studies have demonstrated that cold eddies contribute to this phenomenon (Wang et al., 2010; Lu et al., 2015; Sun et al., 2023). In fact, these high Chl-a patches were generally closed to the cold eddies, as indicated by SSH (0.4 contours in Figures 7a, d). Off the coast of Vietnam, high Chl-a concentrations usually followed the jet during the summer (Liang et al., 2018), as shown in Figure 7g (see red arrow). The high Chl-a patch off the Vietnam closely matched the location of the strengthened current velocity.

Both Exp1 and Exp2 captured the main features of these high Chl-a patches. To the northwest of Luzon Island, while Exp1 predicted high Chl-a patch (Figure 7b), the Chl-a concentration

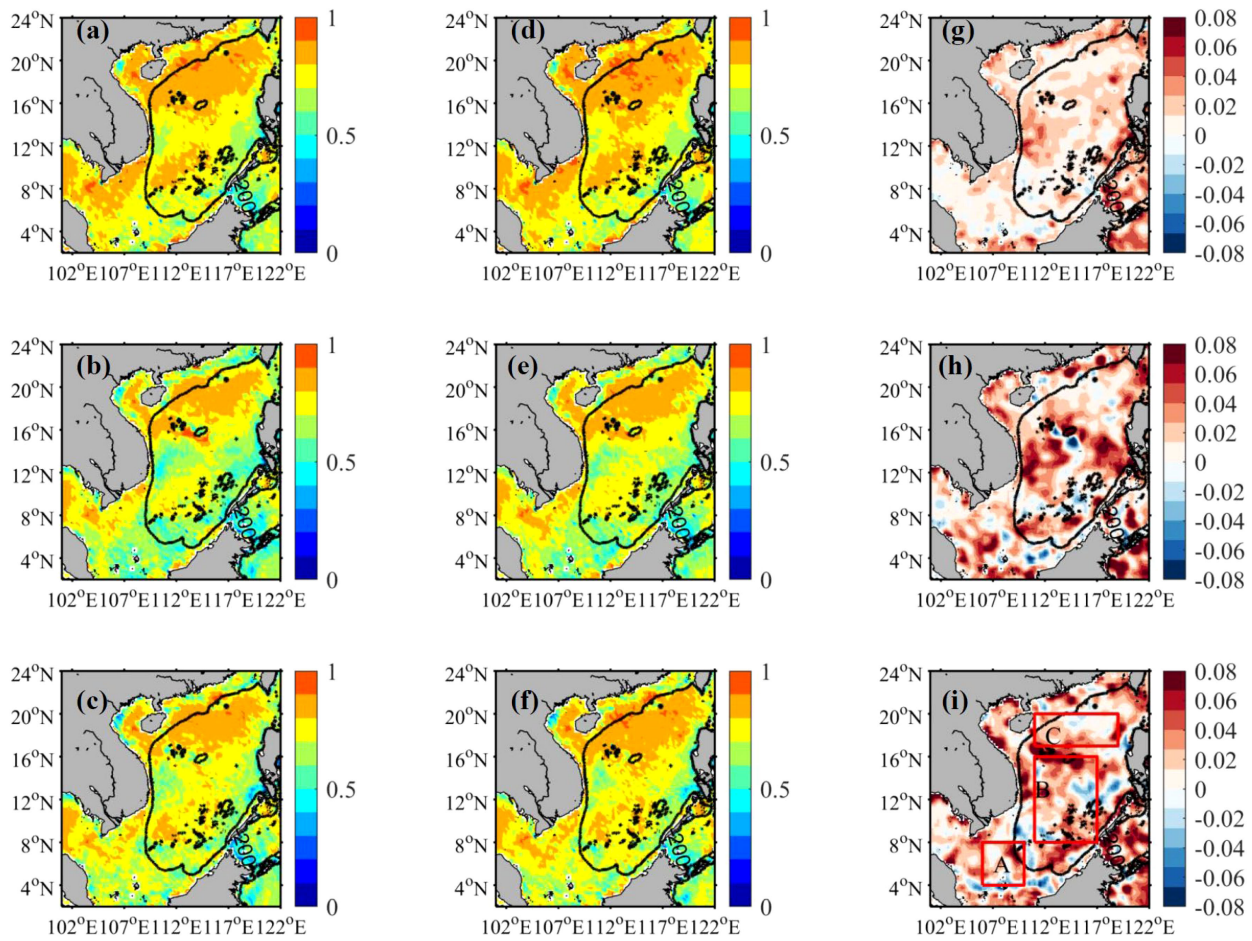


FIGURE 6
Spatial distributions of Pearson correlation coefficients (with $p < 0.05$) in training **(a, d)**, testing **(b, e)**, and validation **(c, f)** datasets. The **(g–i)** were, Δr, calculated by **(d)** minus **(a)**, **(e)** minus **(b)**, and **(f)** minus **(c)**, respectively. The first column represents Exp1 (Exp2). Boxes A, B, and C in **(i)** covered the NSCS, central SCS, and Sunda Shelf.

was not as high as in Figure 7a. However, Exp2 performed better in simulating this patch with higher Chl-a concentration closed to the 0.4 contour (Figure 7c), although it was still lower than that in Exp1. In addition, to the northwest of the high Chl-a patch, the Chl-a concentration was higher than in the True Chl-a (Figures 7a, b). Nonetheless, Exp2 provided a better prediction of Chl-a distribution (Figure 7c) in this area as True Chl-a (Figure 7a). Similarly, the high Chl-a patches near 112°E, 16°N, predicted by the Exp1 and Exp2, were different (Figures 7e, f). The Chl-a concentration in Exp1 was higher than in the True Chl-a (Figure 7d) and Exp2 (Figure 7f). The high Chl-a derived from Exp2 was more comparable to that in the true Chl-a (Figures 7d, f). East of Vietnam, high surface Chl-a is generally induced by upwelling and a southwesterly wind-driven jet (Qiu et al., 2011; Liu et al., 2012; Gao et al., 2013; Chen et al., 2014, 2021). A snapshot of high Chl-a extending from the coast to the east of Vietnam, aligned with the jet (indicated by the strengthened velocity), was shown in Figure 7g. Our model successfully reproduced the high Chl-a along the jet (Figures 7h, i), although the concentrations were not as pronounced as those in the true Chl-a (see red arrow in Figure 7g). Exp2 demonstrated a better prediction of Chl-a along the jet, with higher Chl-a concentrations (see red circles in Figure 7h, i).

This comparison between Exp1 and Exp2 demonstrated that additional variables, SSH and currents, are beneficial to predict the details of the Chl-a distribution. To some extent, the spatial distribution of SSH reflects vertical information, such as the thermocline. Approximately 28.7 cyclonic eddies and 27.9 anticyclonic eddies occur annually in the SCS, which significantly influence the ecosystem of the SCS (Xiu et al., 2010). Mesoscale eddies played a significant role in modulating surface Chl-a through eddy advection, eddy pumping, eddy trapping, and eddy-induced Ekman pumping in the SCS (Gaube et al., 2014; Xiu et al., 2016). Eddy pumping played an important role in controlling surface Chl-a variability to the west of the Luzon Strait and northwest of Luzon Island (Xiu et al., 2016). Yu et al. (2019) found that sea level anomalies are highly correlated with surface Chl-a. Meanwhile, Xiu et al. (2016) revealed that horizontal eddy advection highly influenced the Chl-a off the Vietnam coast. Therefore, including SSH and advection as model inputs enabled the predicted data to more effectively reproduce surface Chl-a.

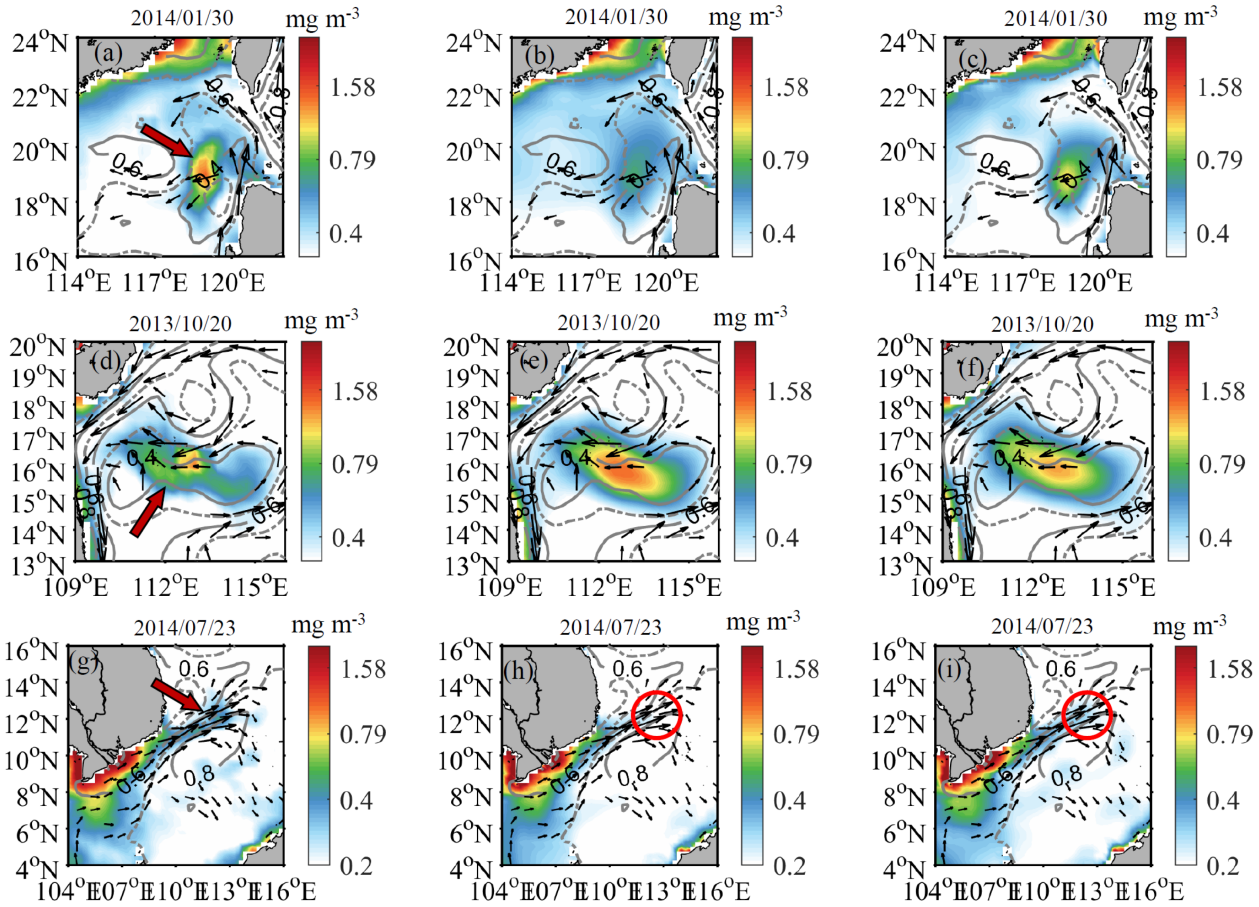


FIGURE 7
Spatial distributions of Chl-a snapshots in the True **(a, d, g)** and Validation datasets **(b, c, e, f, h, i)**. The second and third columns show surface Chl-a in Exp1 and Exp2, respectively. The gray contours represent SSH (0.1 m interval between solid and dashed contours), and the black arrows indicate velocity, exceeding 0.4 m s^{-1}, vectors in HYCOM.

3.5 Application of the model in 2017

The model trained in Exp2 was further applied to predict surface Chl-a in 2017. Based on model performance in the NSCS and SSCS (Figure 6), spatially averaged Chl-a in Boxes A, B, and C was used to assess temporal variability. The predicted Chl-a largely captured the magnitude and temporal variability of surface Chl-a across Boxes A, B, and C (Figures 8a-1, b-1, c-1). Model performance, as measured by correlation coefficients, was highest in the NSCS, followed by the Sunda Shelf and the central SCS (Figures 8a-2, b-2, c-2). Although the model effectively reproduced the temporal variability of surface Chl-a, particularly the seasonal cycle, its performance was relatively less accurate for daily-scale Chl-a variations, as indicated by the distribution of observed Chl-a (Figures 8a-1, b-1, c-1). To improve model validation, we further calculated 8-day averaged surface Chl-a and compared predicted values with observed Chl-a. On the 8-day scale, correlation coefficients between predicted and observed Chl-a were higher than those on the daily scale (Figures 8d-2, e-2, f-2). Observed Chl-a data aligned more closely with predicted values, and both RMSE and MAE indicated reduced errors in 8-day averaged results (Figures 8d-1, e-1, f-1).

One possible reason for the reduced daily-scale accuracy was that daily variations in surface Chl-a were more complex than those on longer timescales. Small-scale dynamic processes, such as fronts and submesoscale eddies, played an essential role in vertical nutrient transport (Callbeck et al., 2017; Jing et al., 2021; Zheng and Jing, 2022). However, the horizontal resolution of model inputs may limit the model's ability to capture these small-scale features, affecting day-scale performance. Additionally, surface Chl-a is often associated with vertical nutrients distribution (Geng et al., 2019; Liu et al., 2020), but obtaining continuous, widespread data on nutrient distribution in the vertical direction remains challenging. These factors constrain the model's precision in predicting daily-scale Chl-a variability.

4 Conclusion

In this study, we developed a statistical model based on the ResUnet architecture to predict daily Chl-a in the SCS through atmospheric and oceanic physical data. The strong correlation between the model-

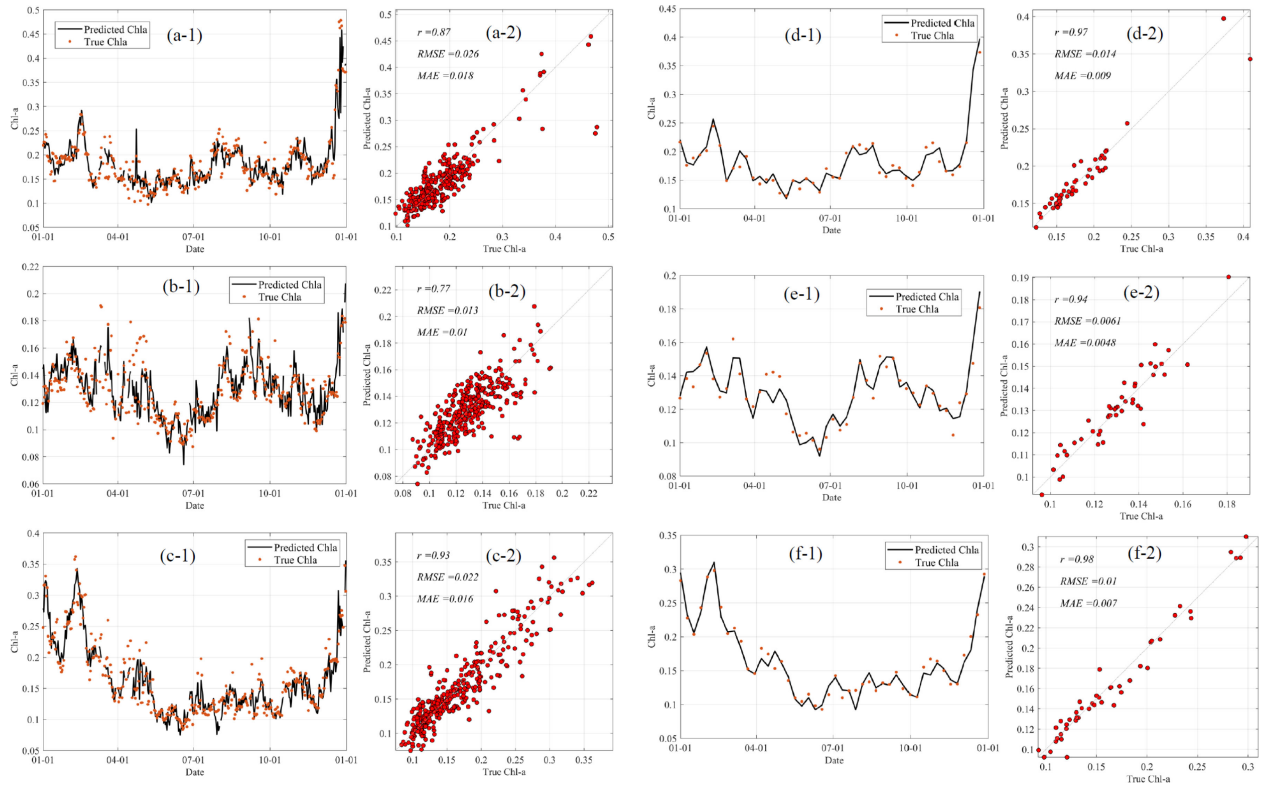


FIGURE 8
Time series of daily spatially averaged Chl-a (units: mg m^{-3}) in Box A **(a-1)**, B **(b-1)**, and C **(c-1)**, along with their corresponding scatter plots of true versus predicted Chl-a in Box A **(a-2)**, B **(b-2)**, and C **(c-2)**. Panels **(d-1)**, **(e-1)**, and **(f-1)** represent the 8-day averaged Chl-a and their corresponding scatter plots of true versus predicted Chl-a [**(d-2)**, **(e-2)**, and **(f-2)**] in Boxes A, B, and C, respectively. The location of boxes is in Figure 6i.

predicted and true Chl-a demonstrates that the model performed well in estimating surface Chl-a. It supported the feasibility of predicting surface Chl-a based on atmospheric and oceanic data.

The model performed better in the NSCS than in the SSCS. In the NSCS, the combination of atmospheric factors and SST was sufficient to reproduce the temporal variability in Chl-a. This superior performance can likely be attributed to the strong correlation between SST and surface Chl-a in this region. In the SSCS, the model-predicted variability of Chl-a had better performance in Exp2, which denoted that the oceanic dynamic factors, such as surface currents and SSH, played a vital role in estimating the Chl-a in the SSCS using deep learning methods.

While the model moderately captured the spatial distribution features in Chl-a when considering only wind-related variables and SST, its performance improved significantly when oceanic dynamic data were included. The addition of surface currents and SSH enabled the model to accurately represent areas with elevated Chl-a due to eddies, particularly around the Luzon Strait and the southeastern side of Hainan Island. The SSH is generally associated with eddies, which enhances the ability of model to predict elevated Chl-a resulting from eddies. In conclusion, the incorporation of ocean dynamics into ecological prediction models based on deep learning technology offers effectively ways and enhances the accuracy of Chl-a predictions in the SCS.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: ERA-5: https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5; HYCOM: https://www.hycom.org/; Chla: https://doi.org/10.5281/zenodo.10478524.

Author contributions

WF: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. AL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. HJ: Supervision, Writing – review & editing. CS: Conceptualization, Supervision, Validation, Writing – review & editing. PX: Supervision, Funding acquisition, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work is financially supported by the National Key Research and Development Program of China (2023YFF0805004, 2022YFE0136600 and 2022YFC3105303), the National Natural Science Foundation of China (42376154), Huanggang Normal University (No. 2042023053), Guangdong Basic and Applied Basic Research Foundation (2022A1515240069 and 2024A1515012032). The work was supported by the Outstanding Postdoctoral Scholarship, State Key Laboratory of Marine Environmental Science at Xiamen University. The work would not have been possible without the free and open access to Chlorophyll-a data, and we sincerely thank the Shilin Tang Team at the South China Sea Institute of Oceanography, Chinese Academy of Sciences, for their invaluable assistance with this work. Also, we gratefully thank the HYCOM consortium for providing the HYCOM data, and the European Centre for Medium-Range Weather Forecasts (ECMWF) for the ERA data, which were invaluable to this research. Special thanks are due to the reviewers of the manuscript.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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

Aleshin, M., Illarionova, S., Shadrin, D., Ivanov, V., Vanovskiy, V., and Burnaev, E. (2024). Machine learning-based modeling of chl-a concentration in Northern marine regions using oceanic and atmospheric data. *Front. Mar. Sci.* 11. doi: 10.3389/fmars.2024.1412883

Boyce, D. G., Lewis, M., and Worm, B. (2012). Integrating global chlorophyll data from 1890 to 2010. *Limnol. Oceanogr. Methods* 10, 840–852. doi: 10.4319/lom.2012.10.840

Bygate, M., and Ahmed, M. (2024). Monitoring water quality indicators over Matagorda Bay, Texas, using Landsat-8. *Remote Sens.* 16, 1120. doi: 10.3390/rs16071120

Cai, Z., Gan, J., Liu, Z., Hui, C. R., and Li, J. (2020). Progress on the formation dynamics of the layered circulation in the South China Sea. *Prog. Oceanogr.* 181, 102246. doi: 10.1016/j.pocean.2019.102246

Callbeck, C. M., Lavik, G., Stramma, L., Kuypers, M. M. M., and Bristow, L. A. (2017). Enhanced nitrogen loss by Eddy-induced vertical transport in the offshore Peruvian oxygen minimum zone. *PloS One* 12, e0170059. doi: 10.1371/journal.pone.0170059

Chang, Y., Shih, Y.-Y., Tsai, Y.-C., Lu, Y.-H., Liu, J. T., Hsu, T.-Y., et al. (2022). Decreasing trend of kuroshio intrusion and its effect on the chlorophyll-a concentration in the Luzon Strait, South China Sea. *GISci. Remote Sens.* 59, 633–647. doi: 10.1080/15481603.2022.2051384

Chen, Y. L. (2005). Spatial and seasonal variations of nitrate-based new production and primary production in the South China Sea. *Deep Sea Res. Part I* 52, 319–340. doi: 10.1016/j.dsr.2004.11.001

Chen, Y., Shi, H., and Zhao, H. (2021). Summer phytoplankton blooms induced by upwelling in the Western South China Sea. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.740130

Chen, G., Xiu, P., and Chai, F. (2014). Physical and biological controls on the summer chlorophyll bloom to the east of Vietnam. *J. Oceanogr.* 70, 323–328. doi: 10.1007/s10872-014-0232-x

Chu, P. C., Edmons, N. L., and Fan, C. (1999). Dynamical mechanisms for the South China Sea seasonal circulation and thermohaline variabilities. *J. Phys. Oceanogr.* 29, 2971–2989. doi: 10.1175/1520-0485(1999)029<2971:DMFTSC>2.0.CO;2

Dai, M., Su, J., Zhao, Y., Hofmann, E. E., Cao, Z., Cai, W.-J., et al. (2022). Carbon fluxes in the coastal ocean: synthesis, boundary processes, and future trends. *Annu. Rev. Earth Planet. Sci.* 50, 593–626. doi: 10.1146/annurev-earth-032320-090746

Diakogiannis, F. I., Waldner, F., Caccetta, P., and Wu, C. (2020). ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS J. Photogramm. Remote Sens.* 162, 94–114. doi: 10.1016/j.isprsjprs.2020.01.013

Dierssen, H. M. (2010). Perspectives on empirical approaches for ocean color remote sensing of chlorophyll in a changing climate. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17073–17078. doi: 10.1073/pnas.0913800107

Ding, W., and Li, C. (2024). Algal blooms forecasting with hybrid deep learning models from satellite data in the Zhoushan fishery. *Ecol. Inf.* 82, 102664. doi: 10.1016/j.ecoinf.2024.102664

Elfwing, S., Uchibe, E., and Doya, K. (2017).Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. Available online at: http://arxiv.org/abs/1702.03118 (Accessed November 3, 2024).

Fang, W., Fang, G., Shi, P., Huang, Q., and Xie, Q. (2002). Seasonal structures of upper layer circulation in the southern South China Sea from *in situ* observations. *J. Geophys. Res.* 107, 21-23. doi: 10.1029/2002JC001343

Fang, M., Ju, W., Liu, X., Yu, Z., and Qiu, F. (2014). Surface chlorophyll-a concentration spatio-temporal variations in the northern South China Sea detected using MODIS data. *Terr. Atmos. Ocean. Sci.* 26, 319–329. doi: 10.3319/TAO.2014.11.14.01(Oc

Fang, G., Wang, G., Fang, Y., and Fang, W. (2012). A review on the South China Sea western boundary current. *Acta Oceanol. Sin.* 31, 1–10. doi: 10.1007/s13131-012-0231-y

Fang, G., Wang, Y., Wei, Z., Fang, Y., Qiao, F., and Hu, X. (2009). Interocean circulation and heat and freshwater budgets of the South China Sea based on a numerical model. *Dynam. Atmos. Oceans* 47, 55–72. doi: 10.1016/j.dynatmoce.2008.09.003

Fernández-González, C., Tarran, G. A., Schuback, N., Woodward, E. M. S., Arístegui, J., and Marañón, E. (2022). Phytoplankton responses to changing temperature and nutrient availability are consistent across the tropical and subtropical Atlantic. *Commun. Biol.* 5, 1035. doi: 10.1038/s42003-022-03971-z

Gan, J., Li, H., Curchitser, E. N., and Haidvogel, D. B. (2006). Modeling South China Sea circulation: Response to seasonal forcing regimes. *J. Geophys. Res.* 111, C06034. doi: 10.1029/2005JC003298

Gao, S., Wang, H., Liu, G., and Li, H. (2013). Spatio-temporal variability of chlorophyll a and its responses to sea surface temperature, winds and height anomaly in the western South China Sea. *Acta Oceanol. Sin.* 2, 48–58. doi: 10.1007/s13131-013-0266-8

Gaube, P., McGillicuddy, D. J., Chelton, D. B., Behrenfeld, M. J., and Strutton, P. G. (2014). Regional variations in the influence of mesoscale eddies on near-surface chlorophyll. *J. Geophys. Res. Oceans* 119, 8195–8220. doi: 10.1002/2014JC010111

Geng, B., Xiu, P., Shu, C., Zhang, W., Chai, F., Li, S., et al. (2019). Evaluating the roles of wind- and buoyancy flux-induced mixing on phytoplankton dynamics in the northern and central South China Sea. *J. Geophys. Res. Oceans* 124, 680–702. doi: 10.1029/2018JC014170

Guo, L., Xiu, P., Chai, F., Xue, H., Wang, D., and Sun, J. (2017). Enhanced chlorophyll concentrations induced by kuroshio intrusion fronts in the northern South China Sea. *Geophys. Res. Lett.* 44, 11,565–11,572. doi: 10.1002/2017GL075336

Huang, Z., Zhuang, W., Hu, J., and Huang, B. (2019). Observations of the Luzon cold Eddy in the northeastern South China Sea in May 2017. *J. Oceanogr.* 75, 415–422. doi: 10.1007/s10872-019-00510-z

Jing, Z., Fox-Kemper, B., Cao, H., Zheng, R., and Du, Y. (2021). Submesoscale fronts and their dynamical processes associated with symmetric instability in the northwest Pacific subtropical ocean. *J. Phys. Oceanogr.* 51, 83–100. doi: 10.1175/JPO-D-20-0076.1

Jouini, M., Lévy, M., Crépon, M., and Thiria, S. (2013). Reconstruction of satellite chlorophyll images under heavy cloud coverage using a neural classification method. *Remote Sens. Environ.* 131, 232–246. doi: 10.1016/j.rse.2012.11.025

Kishino, M., Ishizaka, J., Saitoh, S., Senga, Y., and Utashima, M. (1997). Verification plan of ocean color and temperature scanner atmospheric correction and phytoplankton pigment by moored optical buoy system. *J. Geophys. Res.* 102, 17197–17207. doi: 10.1029/96JD04008

Krestenitis, M., Androulidakis, Y., and Krestenitis, Y. (2024). Deep learning-based forecasting of sea surface temperature in the interim future: application over the Aegean, Ionian, and Cretan Seas (NE Mediterranean Sea). *Ocean Dyn.* 74, 149–168. doi: 10.1007/s10236-023-01595-3

Kuo, N. (2000). Satellite observation of upwelling along the western coast of the South China Sea. *Remote Sens. Environ.* 74, 463–470. doi: 10.1016/S0034-4257(00)00138-3

Lao, Q., Liu, S., Wang, C., and Chen, F. (2023). Global warming weakens the ocean front and phytoplankton blooms in the Luzon Strait over the past 40 years. *J. Geophys. Res. Biogeo.* 128, e2023JG007726. doi: 10.1029/2023JG007726

Large, W. G., and Pond, S. (1981). Open ocean momentum flux measurements in moderate to strong winds. *J. Phys. Oceanogr.* 11, 324–336. doi: 10.1175/1520-0485(1981)011<0324:OOMFMI>2.0.CO;2

Li, A., Shao, T., Zhang, Z., Fang, W., Li, W., Xu, J., et al. (2023). Improvement in spatiotemporal Chl-a data in the South China Sea using the random-forest-based geo-imputation method and ocean dynamics data. *J. Mar. Sci. Eng.* 12, 13. doi: 10.3390/jmse12010013

Li, Q. P., Zhou, W., Chen, Y., and Wu, Z. (2018). Phytoplankton response to a plume front in the northern South China Sea. *Biogeosciences* 15, 2551–2563. doi: 10.5194/bg-15-2551-2018

Liang, W., Tang, D., and Luo, X. (2018). Phytoplankton size structure in the western South China Sea under the influence of a 'jet-eddy system. *J. Marine. Syst.* 187, 82–95. doi: 10.1016/j.jmarsys.2018.07.001

Lin, I.-I., Wong, G. T. F., Lien, C., Chien, C., Huang, C., and Chen, J. (2009). Aerosol impact on the South China Sea biogeochemistry: An early assessment from remote sensing. *Geophys. Res. Lett.* 36, 2009GL037484. doi: 10.1029/2009GL037484

Liu, K.-K., Chao, S.-Y., Shaw, P.-T., Gong, G.-C., Chen, C.-C., and Tang, T. Y. (2002). Monsoon-forced chlorophyll distribution and primary production in the South China Sea: observations and a numerical study. *Deep-Sea Res. I* 49, 1387–1412. doi: 10.1016/S0967-0637(02)00035-3

Liu, Q., Kaneko, A., and Jilan, S. (2008). Recent progress in studies of the South China Sea circulation. *J. Oceanogr.* 64, 753–762. doi: 10.1007/s10872-008-0063-8

Liu, Y., and Li, X. (2023). Impact of surface and subsurface-intensified eddies on sea surface temperature and chlorophyll a in the northern Indian Ocean utilizing deep learning. *Ocean Sci.* 19, 1579–1593. doi: 10.5194/os-19-1579-2023

Liu, X., Wang, J., Cheng, X., and Du, Y. (2012). Abnormal upwelling and chlorophyll-a concentration off South Vietnam in summer 2007. *J. Geophys. Res.* 117, 2012JC008052. doi: 10.1029/2012JC008052

Liu, J., Wang, Y., Yuan, Y., and Xu, D. (2020). The response of surface chlorophyll to mesoscale eddies generated in the eastern South China Sea. *J. Oceanogr.* 76, 211–226. doi: 10.1007/s10872-020-00540-y

Liu, F., Zhang, T., Ye, H., and Tang, S. (2021). Using satellite remote sensing to study the effect of sand excavation on the suspended sediment in the Hong Kong-Zhuhai-Macau Bridge region. *Water* 13, 435. doi: 10.3390/w13040435

Loshchilov, I., and Hutter, F. (2019).Decoupled weight decay regularization. Available online at: http://arxiv.org/abs/1711.05101 (Accessed November 3, 2024).

Lu, W., Yan, X., and Jiang, Y. (2015). Winter bloom and associated upwelling northwest of the Luzon Island: A coupled physical-biological modeling approach. *J. Geophys. Res.: Oceans* 120, 533–546. doi: 10.1002/2014JC010218

Lu, Z., and Gan, J. (2015). Controls of seasonal variability of phytoplankton blooms in the Pearl River Estuary. *Deep Sea Research Part II: Topical Studies in Oceanography.* 117, 86–96. doi: 10.1016/j.dsr2.2013.12.011

Ma, J., Liu, H., Zhan, H., Lin, P., and Du, Y. (2012). Effects of chlorophyll on upper ocean temperature and circulation in the upwelling regions of the South China Sea. *Aquat. Ecosyst. Health Manage.* 15, 127–134. doi: 10.1080/14634988.2012.687663

Ning, X., Chai, F., Xue, H., Cai, Y., Liu, C., and Shi, J. (2004). Physical-biological oceanographic coupling influencing phytoplankton and primary production in the South China Sea. *J. Geophys. Res.* 109, 2004JC002365. doi: 10.1029/2004JC002365

Palacz, A. P., Xue, H., Armbrecht, C., Zhang, C., and Chai, F. (2011). Seasonal and inter-annual changes in the surface chlorophyll of the South China Sea. *J. Geophys. Res.* 116, C09015. doi: 10.1029/2011JC007064

Peñaflor, E. L., Villanoy, C. L., Liu, C.-T., and David, L. T. (2007). Detection of monsoonal phytoplankton blooms in Luzon Strait with MODIS data. *Remote Sens. Environ.* 109, 443–450. doi: 10.1016/j.rse.2007.01.019

Qian, S., Wei, H., Xiao, J., and Nie, H. (2018). Impacts of the Kuroshio intrusion on the two eddies in the northern South China Sea in late spring 2016. *Ocean Dynam.* 68, 1695–1709. doi: 10.1007/s10236-018-1224-y

Qiu, F., Fang, W., and Fang, G. (2011). Seasonal-to-interannual variability of chlorophyll in central western South China Sea extracted from SeaWiFS. Chin. *J. Ocean. Limnol.* 29, 18–25. doi: 10.1007/s00343-011-9931-y

Qu, T. (2000). Upper-layer circulation in the South China Sea. *J. Phys. Oceanogr.* 30, 1450–1460. doi: 10.1175/1520-0485(2000)030<1450:ULCITS>2.0.CO;2

Ronneberger, O., Fischer, P., and Brox, T. (2015). *"U-Net: Convolutional Networks for Biomedical Image Segmentation," in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015.* Eds. N. Navab, J. Hornegger, W. M. Wells and A. F. Frangi (Cham: Springer International Publishing), 234–241. doi: 10.1007/978-3-319-24574-4_28

Roussillon, J., Fablet, R., Gorgues, T., Drumetz, L., Littaye, J., and Martinez, E. (2023). A Multi-Mode Convolutional Neural Network to reconstruct satellite-derived chlorophyll-a time series in the global ocean from physical drivers. *Front. Mar. Sci.* 10. doi: 10.3389/fmars.2023.1077623

Rykaczewski, R. R., and Dunne, J. P. (2011). A measured look at ocean chlorophyll trends. *Nature* 472, E5–E6. doi: 10.1038/nature09952

Shang, S., Li, L., Li, J., Li, Y., Lin, G., and Sun, J. (2012). Phytoplankton bloom during the northeast monsoon in the Luzon Strait bordering the Kuroshio. *Remote Sens. Environ.* 124, 38–48. doi: 10.1016/j.rse.2012.04.022

Shen, S., Leptoukh, G. G., Acker, J. G., Zuojun, Y., and Kempler, S. J. (2008). Seasonal variations of chlorophyll a concentration in the northern South China Sea. *IEEE Geosci. Remote Sens. Lett.* 5, 315–319. doi: 10.1109/LGRS.2008.915932

Shropshire, T., Li, Y., and He, R. (2016). Storm impact on sea surface temperature and chlorophyll a in the Gulf of Mexico and Sargasso Sea based on daily cloud-free satellite data reconstructions. *Geophys. Res. Lett.* 43, 12199-12207. doi: 10.1002/2016GL071178

Shu, Y., Wang, Q., and Zu, T. (2018). Progress on shelf and slope circulation in the northern South China Sea. *Sci. China Earth Sci.* 61, 560–571. doi: 10.1007/s11430-017-9152-y

Song, Y., and Jiang, H. (2023). A deep learning–based approach for empirical modeling of single-point wave spectra in open oceans. *J. Phys. Oceanogr.* 53, 2089–2103. doi: 10.1175/JPO-D-22-0198.1

Sun, R., Li, P., Gu, Y., Zhou, C., Liu, C., and Zhang, L. (2023). Seasonal variation of the shape and location of the Luzon cold eddy. *Acta Oceanol. Sin.* 42, 14–24. doi: 10.1007/s13131-022-2084-3

Tang, D., Kawamura, H., Lee, M.-A., and Van Dien, T. (2003). Seasonal and spatial distribution of chlorophyll-a concentrations and water conditions in the Gulf of Tonkin, South China Sea. *Remote Sens. Environ.* 85, 475–483. doi: 10.1016/S0034-4257(03)00049-X

Tang, S., Liu, F., and Chen, C. (2014). Seasonal and intraseasonal variability of surface chlorophyll a concentration in the South China Sea. *Aquat. Ecosyst. Health Manage.* 17, 242–251. doi: 10.1080/14634988.2014.942590

Wang, X., Du, Y., Zhang, Y., and Wang, T. (2023). Effects of multiple dynamic processes on chlorophyll variation in the Luzon Strait in summer 2019 based on glider observation. *J. Ocean. Limnol.* 41, 469–481. doi: 10.1007/s00343-022-1416-7

Wang, J., Tang, D., and Sui, Y. (2010). Winter phytoplankton bloom induced by subsurface upwelling and mixed layer entrainment southwest of Luzon Strait. *J. Marine. Syst.* 83, 141–149. doi: 10.1016/j.jmarsys.2010.05.006

Wernand, M. R., van der Woerd, H. J., and Gieskes, W. W. C. (2013). Trends in ocean colour and chlorophyll concentration from 1889 to 2000, worldwide. *PloS One* 8, e63766. doi: 10.1371/journal.pone.0063766

Wright, P. N. (1997). "Real-time chlorophyll and nutrient data from a new marine data buoy in Southampton Water, UK," in *Seventh International Conference on Electronic Engineering in Oceanography - Technology Transfer from Research to Industry* (IEE, Southampton, UK), 73–78. doi: 10.1049/cp:19970665

Xian, T., Sun, L., Yang, Y.-J., and Fu, Y.-F. (2012). Monsoon and eddy forcing of chlorophyll-a variation in the northeast South China Sea. *Int. J. Remote Sens.* 33, 7431–7443. doi: 10.1080/01431161.2012.685970

Xie, S., Xie, Q., Wang, D., and Liu, W. T. (2003). Summer upwelling in the South China Sea and its role in regional climate variations. *J. Geophys. Res.* 108, 2003JC001867. doi: 10.1029/2003JC001867

Xiu, P., Chai, F., Shi, L., Xue, H., and Chao, Y. (2010). A census of eddy activities in the South China Sea during 1993–2007. *J. Geophys. Res.* 115, 2009JC005657. doi: 10.1029/2009JC005657

Xiu, P., Guo, M., Zeng, L., Liu, N., and Chai, F. (2016). Seasonal and spatial variability of surface chlorophyll inside mesoscale eddies in the South China Sea. *Aquat. Ecosyst. Health Manage.* 19, 250–259. doi: 10.1080/14634988.2016.1217118

Xue, H., Chai, F., Pettigrew, N., Xu, D., Shi, M., and Xu, J. (2004). Kuroshio intrusion and the circulation in the South China Sea. *J. Geophys. Res.* 109, 2002JC001724. doi: 10.1029/2002JC001724

Yang, Y., Li, X., and Li, X. (2024). A self-attention-based deep learning model for estimating global phytoplankton pigment profiles. *IEEE Trans. Geosci. Remote Sens.* 62, 1–15. doi: 10.1109/TGRS.2024.3435044

Yang, Y. J., Xian, T., Sun, L., and Fu, Y. F. (2012). Summer monsoon impacts on chlorophyll-a concentration in the middle of the South China Sea: climatological mean and annual variability. *Atmos. Oceanic Sci. Lett.* 5, 15–19. doi: 10.1080/16742834.2012.11446961

Ye, H., Yang, C., Dong, Y., Tang, S., and Chen, C. (2024). A daily reconstructed chlorophyll- a dataset in the South China Sea from MODIS using OI-SwinUnet. *Earth Syst. Sci. Data* 16, 3125–3147. doi: 10.5194/essd-16-3125-2024

Yu, Y., Xing, X., Liu, H., Yuan, Y., Wang, Y., and Chai, F. (2019). The variability of chlorophyll-a and its relationship with dynamic factors in the basin of the South China Sea. *J. Marine. Syst.* 200, 103230. doi: 10.1016/j.jmarsys.2019.103230

Zhao, Q., Peng, S., Wang, J., Li, S., Hou, Z., and Zhong, G. (2024). Applications of deep learning in physical oceanography: a comprehensive review. *Front. Mar. Sci.* 11. doi: 10.3389/fmars.2024.1396322

Zheng, R., and Jing, Z. (2022). Submesoscale-enhanced filaments and frontogenetic mechanism within mesoscale eddies of the South China Sea. *Acta Oceanol. Sin.* 41, 42–53. doi: 10.1007/s13131-021-1971-3

Zhou, G., Chen, J., Liu, M., and Ma, L. (2024). A spatiotemporal attention-augmented ConvLSTM model for ocean remote sensing reflectance prediction. *Int. J. Appl. Earth Observ. Geoinform.* 129, 103815. doi: 10.1016/j.jag.2024.103815

Appendix A. terms in deep learning method used in this study

- Feature: In the context of deep learning, a feature represents an individual measurable attribute or characteristic that can be used to describe and analyze an observation or phenomenon.
- Batch Normalization (BatchNorm) Layer: This layer standardizes the inputs of each minibatch, which enhances the stability and efficiency of the training process by reducing internal covariate shift.
- Convolutional Layer: The convolutional layer applies a set of filters to the input data, producing feature maps that capture spatial hierarchies and patterns. This layer performs the convolution operation by sliding the filters over the input and computing the dot product between the filter and the input data, which is fundamental for feature extraction in convolutional neural networks.
- Max Pooling Layer: This layer decreases the spatial dimensions of the input feature maps by extracting the maximum value from each sub-region. Max pooling aids in minimizing computational complexity and mitigating overfitting.
- Sigmoid Linear Unit (SiLU) Activation Function: The SiLU activation function, also known as the Swish function, is defined as:

$$\mathrm{SiLU}(x) = x \cdot \sigma(x)$$

where $\sigma(x)$ is the sigmoid function, given by:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

It combines the properties of linear and sigmoid functions, allowing for smooth, non-linear transformations that can improve the training dynamics of neural networks. The SiLU function has been shown to perform well in various deep learning tasks due to its ability to enhance gradient flow and adaptively control the output.

- Residual Connection: A residual connection bypasses one or more intermediate layers, directly feeding the output of one layer to subsequent layers. This technique aids in training deeper networks by alleviating the vanishing gradient problem.
- Skip Connection: A skip connection, also known as a shortcut connection, involves bypassing one or more layers in the neural network and directly passing the output from an earlier layer to a deeper layer.
- Up-Sampling: In the UNet architecture, up-sampling is employed in the expansive path to restore the resolution of the feature maps. This step is essential for reconstructing high-resolution outputs from lower-resolution feature representations.
- Down-Sampling: Down-sampling decreases the spatial dimensions of the input feature maps, commonly used in the contracting path of the UNet. This process simplifies the information, enabling the model to capture more global features in the earlier layers.
- AdamW Optimizer: The AdamW optimizer is an extension of the Adam optimization algorithm that incorporates weight decay directly into the optimization process. Unlike traditional Adam, which applies weight decay as part of the regularization term added to the loss, AdamW decouples weight decay from the optimization steps, leading to better regularization and improved training dynamics.

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 04 April 2025
DOI 10.3389/fmars.2025.1555286

OPEN ACCESS

EDITED BY
Chao Chen,
Suzhou University of Science and Technology, China

REVIEWED BY
Xiaobo Li,
The Chinese University of Hong Kong, China
Hao Wang,
China University of Petroleum, China

*CORRESPONDENCE
Weitao Chen
wtchen@cug.edu.cn
Xuwen Qin
qinxuwen@163.com

RECEIVED 04 January 2025
ACCEPTED 17 March 2025
PUBLISHED 04 April 2025

CITATION
Shen S, Wang H, Chen W, Wang P, Liang Q and Qin X (2025) A novel edge-feature attention fusion framework for underwater image enhancement.
Front. Mar. Sci. 12:1555286.
doi: 10.3389/fmars.2025.1555286

A novel edge-feature attention fusion framework for underwater image enhancement

Shuai Shen[1], Haoyi Wang[1], Weitao Chen[1]*, Pingkang Wang[2], Qianyong Liang[3] and Xuwen Qin[4,5]*

[1]Faculty of Computer-Science, China University of Geosciences, Wuhan, Hubei, China, [2]Department of Fundamental Investigations, China Geological Survey, Beijing, China, [3]Guangzhou Marine Geological Survey, China Geological Survey, Guangzhou, Guangdong, China, [4]Key Laboratory of Geological Survey and Evaluation of Ministry of Education, China University of Geosciences, Wuhan, Hubei, China, [5]China Aero Geophysical Survey and Remote Sensing Centre for Natural Resources, China Geological Survey, Beijing, China

Underwater images captured by Remotely Operated Vehicles are critical for marine research, ocean engineering, and national defense, but challenges such as blurriness and color distortion necessitate advanced enhancement techniques. To address these issues, this paper presents the CUG-UIEF algorithm, an underwater image enhancement framework leveraging edge feature attention fusion. The method comprises three modules: 1) an Attention-Guided Edge Feature Fusion Module that extracts edge information via edge operators and enhances object detail through multi-scale feature integration with channel-cross attention to resolve edge blurring; 2) a Spatial Information Enhancement Module that employs spatial-cross attention to capture spatial interrelationships and improve semantic representation, mitigating low signal-to-noise ratio; and 3) Multi-Dimensional Perception Optimization integrating perceptual, structural, and anomaly optimizations to address detail blurring and low contrast. Experimental results demonstrate that CUG-UIEF achieves an average peak signal-to-noise ratio of 24.49 dB, an 8.41% improvement over six mainstream algorithms, and a structural similarity index of 0.92, a 1.09% increase. These findings highlight the model's effectiveness in balancing edge preservation, spatial semantics, and perceptual quality, offering promising applications in marine science and related fields.

KEYWORDS

underwater image enhancement, edge feature attention fusion, spatial crossattention, multidimensional perception optimization, attention-guided edge feature fusion

1 Introduction

Underwater images, captured in aquatic environments using remotely operated vehicles (ROVs), are crucial for marine exploration, underwater archaeology, and fishery monitoring, providing visual representations of underwater scenes and objects. However, underwater imaging environments are complex. The images obtained by ROVs are limited

by aggravated color distortion, objects with the same color in the background, and difficulty in edge distinction.

Underwater image enhancement (UIE) improves the quality of underwater images by mitigating their characteristic degradation features and bringing images closer to their true color and clarity, as observed in normal lighting environments. This enables more effective extraction and utilization of valuable features (Alsakar et al., 2024). High-quality underwater image data help reveal unknown marine life and geological features in the deep sea and provide critical information for biodiversity protection (Nazir and Kaleem, 2021), marine environmental monitoring (Wang et al., 2007), and resource sample collection (Mazzeo et al., 2022).

UIE techniques can be divided into two categories: traditional and deep learning-based methods. Traditional UIE techniques include color correction and image restoration methods. Color correction methods such as color balancing can improve color distortion but cannot address blurring and detail loss. Image restoration methods that incorporate physical models, such as light transmission or dehazing models, improve image clarity and optical effects more effectively (Hu et al., 2022). Common color correction methods often perform pixel-level restoration of image colors. For instance, Banik et al. (2018) used gamma correction in the value channel of the hue, saturation, value space to enhance low-light image contrast but introduced problems such as over-enhancement and halos. Garg et al. (2018) applied CLAHE and percentile methods to enhance underwater images and obtained good results in specific scenes but limited improvement in certain water environments. Image-restoration methods typically integrate physical models. Zhu (2023) proposed an enhancement algorithm based on graph theory that improves contrast and color using CIELab and red, green, blue (RGB) spaces combined with CLAHE. However, owing to the independent operations in each color space, the method lacks robustness in complex scenes. Drews et al. (2016) enhanced blue-green channels using a light propagation model but introduced red color distortion. Xiong et al. (2020) applied a linear model and nonlinear adaptive weighting strategy based on the Beer–Lambert law (Swinehart, 1962) to adjust underwater image colors. Recent studies have developed enhanced methods based on conventional algorithmic frameworks to address imaging degradation in specific scenarios. Zhang et al. (2025) proposes a cascaded restoration algorithm grounded in quadtree search-guided background region classification and cross-domain synergy, which integrates dynamic channel discrepancy compensation, S-curve-optimized homomorphic filtering, and chromatic space fusion, thereby significantly improving underwater image fidelity and object recognition robustness. Li et al. (2025) proposes a cascaded restoration algorithm integrating quadtree search-guided background region classification and a cross-domain collaboration mechanism, which effectively addresses color distortion and detail blurring in underwater optical imaging through dynamic channel discrepancy compensation and S-curve-optimized homomorphic filtering, thereby significantly enhancing object detection robustness and visual task performance. However, the methods do not perform well with foggy and low-light underwater images. In general, traditional methods based on fixed underwater priors perform well in specific scenes but are limited by the unpredictability of underwater environments and thus lack general applicability.

Deep learning-based UIE methods use large datasets to train models that adaptively handle various problems, such as color distortion, blurring, and low contrast. These methods can restore image details more accurately and adapt to diverse underwater scenarios. Among the deep learning methods, generative adversarial networks (GANs) have gained prominence in the early stages of UIE for their ability to address limited data availability (Goodfellow et al., 2014). Li et al. (2017) proposed WaterGAN, which corrects underwater image colors by training on both aerial and underwater real images. However, the aerial image model introduces unrealistic background colors. Fabbri et al. (2018) proposed UGAN, which uses CycleGAN generated paired datasets and a Pix2Pix-like structure for UIE. However, CycleGAN generates artifacts under certain scenarios. Despite the requirement of high-quality training data, their proposed method struggles with low-quality underwater images. These methods effectively restore color but often face challenges such as over-enhanced contrast, information loss, instability, and convergence difficulties.

Convolutional neural network (CNN)-based methods (Wang et al., 2021; Lyu et al., 2022; Yang et al., 2023) are particularly effective for UIE tasks owing to their strong feature extraction capabilities and nonlinear feature mapping, which enable them to adapt to various underwater scenes. Wang et al. (2017) designed an end-to-end CNN-based network for color correction and deblurring by employing a pixel disturbance strategy to improve model convergence speed and accuracy. However, their method overfocuses on local features while neglecting the overall semantic information, global color, and light–shadow relationships in the image. Li et al. (2019) developed a paired underwater image enhancement benchmark (UIEB) dataset and proposed Water-Net, a CNN-based model that serves as a benchmark for CNN applications in UIE. Li et al. (2020) trained their proposed UWCNN on synthetic underwater images of various scenes, which resulted in different model parameters. However, owing to the singularity of the training data scenes, the model is overly sensitive to subtle changes in underwater environments and thus, performs poorly. Islam et al. (2020) proposed the UFO-120 dataset and a residual nested CNN called Deep SESR, which has a multimodal objective function for both enhancement and super-resolution of images. However, the shared feature space in this model can cause significant features from the super-resolution task to interfere with the color performance of image enhancement. The aforementioned CNN-based models have powerful feature-learning capabilities and can adapt to complex underwater environments; however, CNNs primarily extract features through local receptive fields, which renders fully capturing global information challenging. Consequently, enhanced images often show a marked locality with coordination problems among objects in complex underwater environments.

To address this limitation, several studies have used Swin Transformers (Liu et al., 2021) for UIE. Sun et al. (2022) enhanced the underwater image contrast by inputting images into

a Swin Transformer following gamma and white balance corrections. However, white balance and gamma correction cannot fully resolve the complex problems of underwater images, particularly in foggy and blurry scenes. Peng et al. (2023) constructed a large-scale underwater image dataset and proposed a channel multi-scale fusion transformer and spatial global feature transformer to enhance severely attenuated color channels and spatial regions. However, the sensitivity of different color spaces to various colors varies, which degrades model stability in scenarios with strong color contrast. Transformer architecture, with its unique mechanisms and processing methods, has tremendous potential and value as a primary framework for UIE. Zhu et al. (2024) proposed an adaptive multi-scale image fusion cascaded neural network that integrates polarization-based multi-dimensional features to improve image enhancement quality under low-quality imaging conditions. Concurrently, the team establishing a standardized evaluation framework for polarization-aware visual restoration algorithms. Zhu et al. (2025) proposed a Fourier-guided dual-channel diffusion network, enhances underwater images via phase-based edge refinement and amplitude mapping, coupled with a lightweight transformer denoiser, outperforming leading methods in generalization and visual quality on real underwater datasets. Wang et al. (2025) proposed a SAM-powered framework for underwater image enhancement, integrating precise foreground-background segmentation, region-specific color correction, adaptive contrast enhancement, and high-frequency detail reconstruction to mitigate crosstalk and blurring, thereby significantly improving restoration fidelity and visual quality. Considering that underwater images exhibit inconsistent attenuation characteristics across different color channels and spatial regions and that the object edges in these images degrade, the proposed network focuses on these characteristics to restore underwater image information and achieve high-quality underwater image data.

The main contributions of this paper are as follows:

1. We propose a network model, CUG-UIEF, based on U-Net and a multi-feature cross-fusion module, which greatly improves the quality of underwater images.
2. We introduce a multi-feature cross-fusion module that enhances the feature representation of images at different scales, thereby improving the overall quality and accuracy of the final output.
3. We evaluate the proposed CUG-UIEF model on the UIEB, low-light and super-resolution underwater image (LSUI), and U45 datasets and compared its performance with that of six other mainstream models. The experimental results show that CUG-UIEF achieves substantial improvements in the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). The results also demonstrated excellent performance in both underwater image quality metrics and underwater color image quality assessments, indicating that the CUG-UIEF effectively overcomes underwater environmental interference and can be applied in related fields.

2 Proposed method

2.1 Network structure

The overall structure of the CUG-UIEF is shown in Figure 1; it can be divided into three parts: an encoder, a multi-feature cross-fusion module (DDEM), and decoder. The encoder converts the input image into a deep feature representation. The decoder gradually fuses the features and performs upsampling to reconstruct an underwater image. In this study, the multi-scale features extracted by the encoder were input into the DDEM, and its output was fused with the upsampling results at each stage of the decoder. An enhanced underwater image was obtained after the final upsampling step.

Encoder stage: This module extracts multi-scale features through the Swin Transformer layer and performs downsampling to capture the details and global information in the image. The deep feature representation provides rich semantic information for the subsequent DDEM module and decoder, which facilitates the final image reconstruction and enhancement.

DDEM module: Uses the Sobel operator to extract edge information from the multi-scale features extracted in the four stages of the encoder and inputs the edge and multi-scale features together into the channel cross-attention (CCA) module to fuse the feature information across channels. Subsequently, the output of the CCA is passed to the spatial cross-attention module to capture the long-distance dependencies among the multi-scale features. Following layer normalization and GeLU activation, the final features are sent to the decoder to gradually restore the spatial resolution and reconstruct the enhanced image.

Decoder stage: The decoder first upsamples the output of the final stage of the encoder and inputs it into the Swin Transformer block. Subsequently, the output of the DDEM is fused with the upsampled results of each decoder stage. The decoder restores the spatial resolution through gradual upsampling to reconstruct an enhanced image. The parameters of the Swin Transformer layer are adjusted at this stage to maintain the integrity of the features, whereas the upsampling layer is used to restore the size of the feature maps. The final upsampling restores the features to the resolution of the original input and projects them onto the RGB channels through the convolutional layer to generate an enhanced underwater image.

2.2 Multi-feature cross-fusion module (DDEM)

The proposed multi-feature cross-fusion module fuses the features extracted from the four multi-scale encoder stages (Figure 2). It generates enhanced feature representations and connects these enhanced features to the corresponding decoder stages. The module can be further divided into attention-guided edge fusion and spatial information enhancement modules.

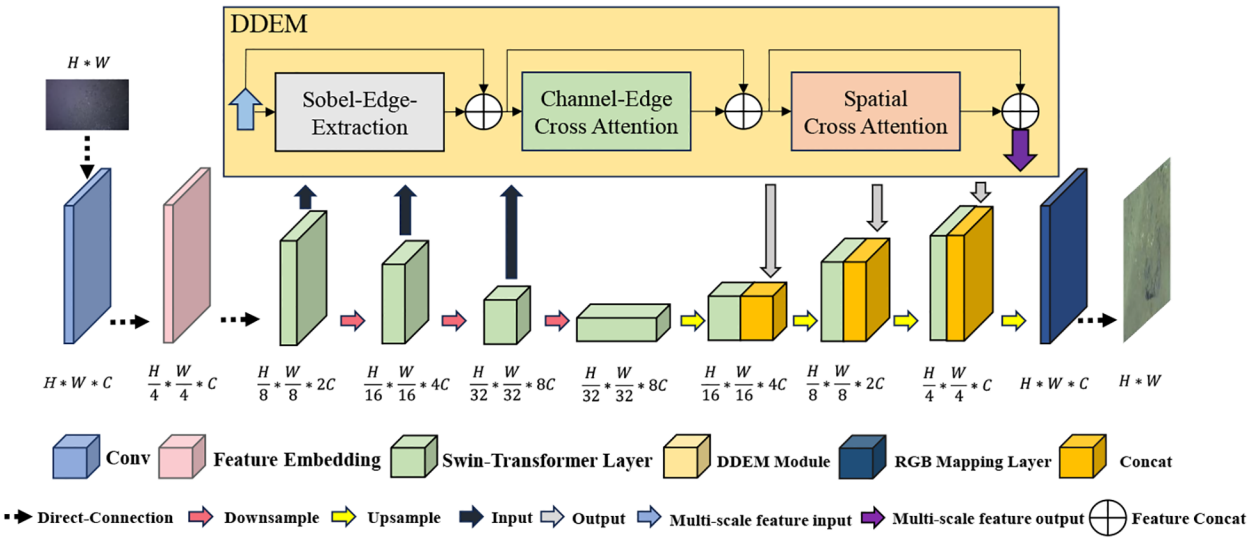


FIGURE 1
Architecture of the proposed network, featuring an encoder–decoder structure enhanced with the addition of a multi-feature cross-fusion module.

The specific operations performed by the module are as follows: The edge features are extracted from the multi-scale features output by the encoder and then input layer by layer together with the multi-scale features into the edge fusion and spatial information enhancement modules. Attention maps are constructed by fusing the features of the multi-scale encoder, enabling them to capture long-distance dependencies across different stages to achieve more accurate and comprehensive modeling of complex scenes and dynamic changes.

Through this series of operations, the output results are processed by layer normalization and subjected to nonlinear mapping via the GeLU activation function to establish dynamic correlations between the feature maps at different levels and edge feature maps.

2.2.1 Attention-guided edge fusion module

This module promotes information interactions between features at different levels in the channel dimension. In this study, the edge features are gradually fused in multiple stages. The weights of the weighted edge features are adjusted using CCA to ensure that detailed information, such as colors and textures, can be accurately transmitted. Upon being output to the decoder stage, as the decoding process proceeds, the weighting coefficients are dynamically adjusted based on the local information of the image; thus, the edge features are enhanced in detailed areas while minimizing interference in the background or smooth regions. In this manner, the edge information is strengthened in key areas (such as object boundaries and detailed parts), while the global consistency and natural appearance of the image are preserved.

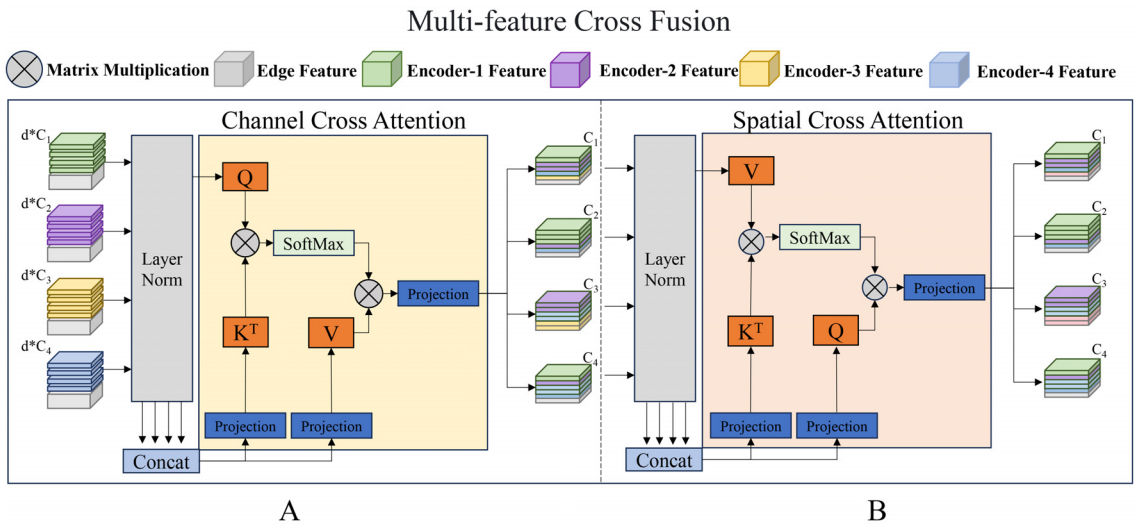


FIGURE 2
Proposed multi-feature cross-fusion module. **(A)** Attention-guided edge fusion module; **(B)** Spatial information enhancement module.

The Sobel operator is applied for edge feature extraction. It identifies edge information by calculating the gray gradient of the area around each pixel in the image. The core of this algorithm lies in its elaborately designed convolution kernels, which perform convolution operations on the horizontal and vertical features of the image, thereby effectively capturing the edge changes in the image in different directions.

The change in the x-axis direction in the Sobel operator is

$$\mathbf{G_x} = \begin{pmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{pmatrix}$$

The change in the y-axis direction is

$$\mathbf{G_y} = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{pmatrix}$$

Approximate gradient values of the image in the horizontal and vertical directions can be obtained by performing convolution operations on the image using these two sets of convolution kernels. The gradient magnitude of each pixel point can then be obtained by calculating the square root of the sum of the squares of these two gradient values (or the sum of their absolute values) to determine the intensity of the edge.

$$G = \sqrt{G_x^2 + G_y^2}$$

The Sobel operator extracts edge information across different scales, performs layer normalization along the channel dimension, and conducts weighted fusion with the multi-scale feature output from the encoder stage.

$$\mathbf{fused_{feature}} = \alpha \cdot \mathbf{edge_{feature}} + (1 - \alpha) \cdot \mathbf{encoder_{feature}}$$

Subsequently, the fused features automatically adjust the attention distribution by calculating the similarity between channels to strengthen key features in the image. At this stage, layer normalization is first performed on each token to stabilize the training process. Subsequently, all tokens are concatenated along the channel dimension to create unified keys and values while retaining each token as an independent query. The linear projection in the self-attention mechanism is replaced with a 1 × 1 depthwise convolutional projection. This enables cross-channel information integration and interaction and enhances its nonlinear characteristics. The process formula is as follows:

$$\mathbf{K, V} = \mathbf{conv1D}(\mathbf{concat}(\mathbf{T}_1, \mathbf{T}_2 \ldots \mathbf{T_i}))$$

$$\mathbf{Q_i} = \mathbf{conv1D}(\mathbf{T}_1, \mathbf{T}_2 \ldots \mathbf{T_i})$$

$$\mathbf{CCA}(\mathbf{Q_i, K, V}) = \mathbf{Softmax}(\mathbf{Q_i^T K S})\mathbf{V^T}$$

Q_i, K, and V are matrices that represent the queries, keys, and values, respectively, which are obtained by concatenating the tokens along the channel dimension. S is the scaling factor. Once the output of the CCA is connected to the original tokens, the enhanced features are input into the spatial information enhancement module.

2.2.2 Spatial Information Enhancement Module

This module dynamically adjusts the feature weights of different regions of the image by calculating the correlations among different spatial positions, thereby enhancing the key features in the image. Edge features provide essential structural cues for the image, enabling the model to focus more on the detailed areas of the image while reducing attention allocation to smooth areas, thus avoiding excessive enhancement. Combined with enhanced multi-scale features, the module can capture the details of the image at different levels, effectively restoring the detail loss in underwater images caused by light attenuation and blurring. At this stage, all the tokens are first subjected to layer normalization along the channel dimensions and then concatenated. In contrast to the edge-fusion module, this module uses concatenated tokens as queries and keys; each token is used as a value. Moreover, 1 × 1 depthwise convolutions are also used for projection onto the queries, keys, and values. This design enables the spatial information enhancement module to focus on information integration in the spatial dimension, thereby complementing the edge fusion module to collaboratively establish a comprehensive and enhanced feature representation. The process is defined by the following formulas:

$$\mathbf{K, Q} = \mathbf{conv1D}(\mathbf{concat}(\mathbf{T}_1, \mathbf{T}_2 \ldots \mathbf{T_i}))$$

$$\mathbf{V_i} = \mathbf{conv1D}(\mathbf{T}_1, \mathbf{T}_2 \ldots \mathbf{T_i})$$

$$\mathbf{SCA}(\mathbf{Q_i, K, V}) = \mathbf{Softmax}(\mathbf{Q^T K S})\mathbf{V_i^T}$$

Q, K, and V_i are matrices that represent queries, keys, and values, respectively. S is the scaling factor.

To ensure that the generated enhanced features can effectively serve the decoder, the following processing steps are adopted. First, layer normalization and the GeLU activation function are applied to the output to stabilize the features and introduce nonlinear transformations. Subsequently, through a combined sequence of an upsampling layer, a 1 × 1 convolution, batch normalization, and GeLU activation function, necessary size adjustments and enhancements are made to the features, which are fused with the features in the decoder stage. The upsampling layer is used to restore the spatial resolution of the feature maps, ensuring that the details of the image can be better reconstructed in the decoding stage. The 1 × 1 convolution is used for channel compression and feature fusion, enhancing the expressive ability of the model, while batch normalization ensures the consistency of features among different layers. The GeLU activation function introduces nonlinear transformations that aid in handling complex feature relationships. This method ensures the continuity and consistency of information and greatly improves the decoding efficiency and performance of the entire network.

2.3 Loss function

We propose a multi-dimensional perceptual loss function for training the CUG-UIEF to align the enhanced images with human visual perception and improve detail reconstruction.

1) Perceptual Loss.

Deep features capture high-level semantic information from images. By comparing the feature maps of the two images in a pretrained network, their perceptual similarity can be evaluated.

$$\mathbf{l_{per}} = |\mathbf{x} - \mathbf{y}|,$$

x represents the predicted image, and y represents the real image.

2) Multi-scale Structural Similarity Loss.

Multi-Scale Structural Similarity (MSSSIM) is an image quality assessment metric that evaluates brightness, contrast, and structural features across multiple scales, providing a measure more aligned with human visual perception.

$$\mathbf{l_{ms-ssim}} = 1 - \prod_{\mathbf{m}=1}^{\mathbf{M}} \left(\frac{2\mathbf{u_p u_g} + \mathbf{c}_1}{\mathbf{u_g^2} + \mathbf{u_p^2} + \mathbf{c}_1} \right)^{\beta_m} \left(\frac{2\boldsymbol{\sigma}_{\mathbf{pg}} + \mathbf{c}_2}{\boldsymbol{\sigma}_{\mathbf{p}}^2 + \boldsymbol{\sigma}_{\mathbf{g}}^2 + \mathbf{c}_2} \right)^{\gamma^m}$$

Here, M represents different scales. u_g and u_p represent the means of the predicted image and ground truth, respectively. σ_p and σ_g represent the standard deviations between the predicted and real images. σ_{pg} represents the covariance between the predicted and real images. β_m and γ_m represent the relative importance constants between the two items. c_1 and c_2 are constants.

3) Charbonnier Loss.

The Charbonnier loss function is a variant of the L1 loss function. It prevents the denominator from reducing to zero by introducing a small positive number ε and ensures smoother changes when the gradient is large. It maintains the sharpness of an image while reducing noise.

$$\mathbf{l_{charbonnier}} = \sqrt{\mathbf{x}^2 + \in^2}$$

X represents the difference between the predicted image and ground truth. ε is a small positive number used for numerical stability.

Finally, the loss function is expressed as

$$\mathbf{l} = \boldsymbol{\lambda}_1 \mathbf{l_{per}} + \boldsymbol{\lambda}_2 \mathbf{l_{ms-ssim}} + \boldsymbol{\lambda}_3 \mathbf{l_{charbonnier}} .$$

Hyperparameters λ_1, λ_2, and λ_3 determine the balance between the overall performance and the local texture details. Following experimental analysis the parameters were set to 1, 2, and 1, respectively.

3 Experiments and analyses

3.1 Experimental environment and parameter settings

The proposed model was implemented using PyTorch 2.4.0. It was trained on an NVIDIA RTX 2080Ti GPU without a pretrained network. During the training process, the Adam optimizer was adopted, and the initial learning rate was set to 0.0005, with the β parameter pair being (0.9, 0.999). Training was performed for 700 epochs, and the number of samples in each batch was four.

3.2 Datasets

This study uses three datasets.

1. UIEB Dataset (Li et al., 2019): This dataset included 950 real underwater images. Among them, 890 images had corresponding reference images, and another 60 underwater images without reference images were used as the challenging data. In this study, 90 pairs of challenging images in multiple scenes with corresponding reference images from the UIEB were selected as the test set Test-U90, and 60 images without reference images were used as the test set Test-C60. The remaining images were divided into training and validation sets at a 8:2 ratio. The training data were enhanced using random cropping, size adjustment, and random rotation.
2. LSUI Dataset (Peng et al., 2023): This is a large-scale underwater image dataset that contains 5,004 underwater images with reference images. It contains richer underwater scenes. Forty-five images were selected from this dataset as the test set, Test-L45. The remaining images were divided into training and validation sets at a 8:2 ratio. The training data were enhanced through random cropping, size adjustment, and random rotation.
3. U45 Dataset: The U45 dataset is a publicly available underwater image test dataset that contains 45 underwater images in different scenes and involves underwater degradation phenomena, such as color shift, low contrast, and fogginess. Forty-five images were used as the test set, Test-U45.

3.3 Evaluation metrics and comparative algorithms

Reference Evaluation Metrics: To quantify the performance of each model on the dataset with reference images, this study adopted two measurement standards: PSNR and SSIM. These two indicators help measure the similarity between the restored and reference images. PSNR is an objective quality metric calculated based on the mean squared error between the original image and the enhanced image, with the unit of decibel (dB). In UIE, a higher PSNR value indicates that the enhanced image has a smaller error than the original image and, thus, better quality. SSIM is an index used to measure the similarity between two images. It considers luminance, contrast, and structural information, and its value ranges from –1 to 1. In UIE, the closer the SSIM value is to one, the more similar the enhanced image to the original image in terms of structure, luminance, and contrast, suggesting a higher image quality.

No-reference Evaluation Metrics: For the test sets of images without reference images, we adopted three evaluation methods: underwater color image quality evaluation (UCIQE), underwater image quality measure (UIQM) and Underwater Ranker(URanker)

(Guo et al., 2023). UCIQE focuses on the color density, saturation, and contrast of images and uses a linear combination of these three aspects as the quantitative form of color cast, blurring, and low contrast. UIQM includes color (UICM), sharpness (UISM), and contrast measurements (UIConM). As the scores of these methods increase, the image processing results become more aligned with the visual perception preferences of human beings.

Comparative Algorithms: The comparative algorithms adopted in this experiment are representative algorithms among traditional UIE methods and deep-learning-based UIE methods, which can verify the effectiveness and advancement of the proposed method, including the UIE algorithm based on color correction: Fusion (Ancuti et al., 2012); UIE algorithms based on image restoration: IBLA (Wang et al., 2013); HL (Berman et al., 2021); WWPF (Zhang et al., 2023); CBLA (Jha and Bhandari, 2024); UIE algorithms based on deep learning: UWCNN (Li et al., 2020), Shallow-UWnet (Naik et al., 2021), USUIR (Fu et al., 2022), URSCT (Ren et al., 2022), DiffWater (Guan et al., 2023).

TABLE 1 PSNR and SSIM scores of different methods on the test set Test-U90.

	Method	TEST-U90	
		PSNR	SSIM
Traditional Method	HL	14.8429	0.6497
	IBLA	14.9395	0.6742
	Fusion	21.1843	0.8639
	CBLA	15.2359	0.6614
	WWPF	18.5371	0.7062
Deep-Learning Method	Sha-UWnet	17.4575	0.7174
	UWCNN	15.4532	0.7560
	USUIR	20.5514	0.8544
	URSCT	22.5976	0.9171
	Diff-Water	20.1567	0.8391
	Ours	**24.4952**	**0.9262**

In the results, boldface indicates the best data and blue denotes the suboptimal data.

3.4 Experimental results

All experimental results are presented with the best outcomes bolded and the second-best outcomes highlighted in blue font. This section first presents the test results of the model based on the UIEB training set on the Test-U90 dataset. As indicated in Table 1, CUG-UIEF outperformed the other algorithms in terms of the PSNR and SSIM. Moreover, compared to the second-best performance, CUG-UIEF achieved percentage gains of 8.41% and 0.1% in PSNR and SSIM, respectively. This study also conducted a no-reference evaluation comparison of Test-C60 and Test-U45. Table 2 presents all the statistical results. Both UIQM and UCIQE have specific feature biases and are relatively sensitive to the contrast of images. Therefore, results based on visual priors and physical models can yield higher scores. Our experimental results align with this conclusion. And the proposed method achieves the best performance on the URanker evaluation metric, with an average improvement of 12.21% over the second-best model. Therefore, the results cannot indicate whether the processed images are the best in all aspects. However, by combining the results of the two parameters, the images performed well in terms of contrast and color. CUG-UIEF obtained the second-best result among the models that were used in the experiment, only lower than that of the fusion method. Combined with the previous results, this shows that the generalization ability and actual performance of the CUG-UIEF are the best.

3.5 Comparative mechanism analysis of algorithms

The fusion algorithm addresses underwater color cast and low-contrast degradation through adaptive weight mapping, yet exhibits critical limitations when confronting specific technical challenges. Its edge restoration capability deteriorates in low signal-to-noise ratio (SNR) regions, producing blurred textures and artificial transitions around fine structural details, while contrast optimization remains suboptimal under non-uniform illumination caused by suspended particulates. Furthermore, the water-quality-dependent input generation mechanism demonstrates unstable color correction performance across chromatic water types, particularly failing to compensate for wavelength-specific absorption in turbid greenish waters where waterborne noise amplifies color inconsistency along depth gradients. These limitations stem from the algorithm's inherent constraints in decoupling overlapping degradation patterns and adapting to spatially variant underwater optical conditions.

The IBLA algorithm decomposes images via luminance-ordering error metrics and bright-pass filtering to separately regulate reflectance and illumination, dynamically adjusting their weights through dual-logarithmic transformations. While effective for uniform scenes, the framework suffers from edge-texture mismatches in areas with overlapping illumination-reflectance gradients, where low-SNR conditions exacerbate erroneous boundary segmentation and nonlinear illumination transitions degrade fine details. The logarithmic weight adaptation further struggles to resolve high dynamic range conflicts, causing halo artifacts near specular highlights and contextual inconsistency in shadowed low-contrast regions. These limitations arise from inadequate noise-robust disentanglement of radiometrically coupled components under complex degradation patterns.

The HL algorithm frames color restoration as a single-image dehazing task by estimating attenuation ratios for the blue-red and blue-green color channels, with a color distribution screening mechanism to identify optimal parameter combinations. However, this approach faces three critical limitations in addressing underwater-specific degradation: Its unified attenuation coefficient oversimplifies spectral interactions, failing to resolve edge blurriness

TABLE 2 UIQM and UCIQE scores of different methods on test sets C60 and U45.

	Method	Test-C60			Test-U45		
		UCIQE	UIQM	URanker	UCIQE	UIQM	URanker
Traditional Method	HL	0.5311	2.8774	0.094	0.5126	1.9423	0.751
	IBLA	0.5642	**3.3236**	0.815	0.4612	1.2768	0.945
	Fusion	**0.5848**	2.8092	0.745	**0.6473**	1.6984	0.726
	CBLA	0.4781	2.4273	1.285	0.5139	1.7141	1.392
	WWPF	0.5135	2.4861	1.348	0.5641	1.7311`	1.351
Deep-Learning Method	Sha-UWnet	0.4198	2.2751	0.921	0.4595	1.6893	1.257
	UWCNN	0.4894	2.4523	1.687	0.4524	1.4338	1.582
	USUIR	0.5673	2.3234	1.618	0.5131	1.8952	1.685
	URSCT	0.5529	2.7453	1.713	0.5729	2.1861	1.724
	Diff-Water	0.5372	2.5894	1.632	0.5338	2.0142	1.583
	Ours	0.5737	2.8168	**1.982**	0.5937	**2.3247**	**1.859**

In the results, boldface indicates the best data and blue denotes the suboptimal data.

caused by wavelength-dependent scattering anisotropy. The channel-agnostic model amplifies noise in low-SNR scenarios, particularly in red-dominated deep-water regions where backscatter varies disproportionately. Linear color compensation ignores depth-related contrast attenuation gradients, leading to inaccurate recovery in shaded seabed areas with nonlinear illumination decay. These simplifications fundamentally disregard the photometric complexity of real underwater environments, where multi-band light refraction and particulate scattering create spatially varying attenuation patterns.

The color-balanced locally adjustable (CBLA) algorithm targets underwater color distortion and contrast degradation through dual-space hierarchical enhancement, yet reveals critical vulnerabilities when addressing complex photometric interactions. Its RGB-space color restoration mechanism struggles to decouple chromatic shifts from suspended particulate backscattering in high-turbidity environments, occasionally overcompensating blue-green dominance while neglecting wavelength-specific absorption residuals. The CIELAB-space contrast optimization demonstrates limited adaptivity to illumination gradients across depth-varying scenes, where aggressive luminosity adjustments in localized regions may amplify noise patterns and induce halo artifacts near high-frequency textures. Furthermore, the separate processing pipelines for color correction and contrast enhancement fail to maintain spectral consistency in transitional zones between adjusted and unprocessed areas, particularly under abrupt optical density changes caused by marine snow or biological layers. These deficiencies originate from the method's sequential processing framework that insufficiently models the nonlinear coupling between wavelength attenuation and turbidity-induced light diffusion.

The weighted wavelet visual perception fusion (WWPF) method tackles underwater color distortion and contrast degradation through multi-strategy hierarchical optimization, yet reveals critical constraints when handling complex photonic interactions. Its attenuation-map-guided color correction exhibits incomplete spectral separation in high-turbidity greenish waters, where particulate backscattering interferes with wavelength-specific absorption estimation, occasionally preserving residual cyan dominance while overcompensating red-channel artifacts. The maximum entropy optimized global contrast enhancement demonstrates limited dynamic range adaptation across depth-varying illumination fields, where uniform intensity stretching may amplify noise in low-transmission regions while compressing texture details in high-clarity zones. Furthermore, the wavelet-based multi-scale fusion mechanism shows inadequate edge preservation at high-frequency subbands when processing particulate-laden scenes, as directional filter banks struggle to differentiate between authentic structural contours and suspended particle clusters, resulting in oversmoothed textures near marine snow interfaces. These limitations stem from the method's implicit assumption of linear degradations and insufficient modeling of nonlinear light-particle-camera interactions in turbid aquatic environments.

The UWCNN algorithm constructs a synthetic degradation dataset using spectral-attenuation priors to train a lightweight CNN for direct underwater image restoration, thereby reducing error propagation. While effective for general color cast correction, its wavelength-agnostic framework introduces spectral bias by oversimplifying depth-dependent chromatic shifts and angular illumination variations inherent in real underwater environments. Specifically, the model fails to address nonlinear wavelength absorption caused by suspended particulates and depth-varying water types, leading to color channel imbalance in scenes with multi-spectral artificial lighting or bioluminescent interference. Furthermore, its static prior integration neglects photometric divergence between shallow and deep-water zones, resulting in inconsistent color constancy when reconstructing red-depleted regions or high-turbidity sediments. These limitations stem from inadequate modeling of spectrally asymmetric degradation and

cross-domain generalization across heterogeneous underwater optical conditions.

The Sha-UWnet employs a parameter-efficient architecture to optimize underwater image enhancement, leveraging prioritized feature extraction to balance computational cost and restoration quality. While its streamlined design effectively addresses global color shifts, the constrained network depth impedes hierarchical abstraction of multi-scale edge contexts, resulting in blurred boundary delineation and textural discontinuities in low-contrast turbid waters. Specifically, the shallow structure fails to resolve edge-texture conflicts caused by suspended particle scattering, often miscalculating gradient magnitudes in regions with overlapping foreground-background chromaticity. Furthermore, its limited receptive field struggles to suppress waterborne noise while preserving high-frequency details, leading to artificial sharpening artifacts near bioluminescent features or sediment-rich zones. These limitations highlight the inherent trade-off between model efficiency and multi-scale degradation disentanglement in underwater optical environments.

The USUIR algorithm reformulates unsupervised restoration through homology-driven cycle consistency between original and synthetically re-degraded images, theoretically circumventing the need for paired training data. While effective for global error minimization, the framework exhibits edge gradient confusion in low signal-to-noise ratio regions, failing to resolve sub-pixel boundary discontinuities caused by suspended particle scattering or nonlinear light attenuation. This manifests as blurred bio-structural contours and textural oversmoothing in turbid waters where foreground-background chromatic similarity exacerbates edge ambiguity. Furthermore, its spectrally insensitive homology constraints inadequately model wavelength-dependent absorption, inducing color channel crosstalk that amplifies greenish hue bias in deep pelagic zones and artificial saturation spikes under multi-spectral artificial lighting. These limitations stem from insufficient physical priors to disentangle spatially coupled degradation patterns across heterogeneous underwater optical domains.

The URSCT algorithm integrates Swin Transformer into a U-Net framework to enhance global context modeling for structural and chromatic restoration, while its RSCTB module employs convolutional layers to refine local features. Although this hybrid design improves cross-scale feature aggregation in uniform underwater scenes, the global attention mechanism in Swin Transformer induces boundary erosion when processing low-contrast edges or suspended particle-induced textures, where multi-scale edge ambiguity arises from nonlinear light scattering. Concurrently, the convolutional RSCTB module exhibits limited texture-edge decoupling capacity, failing to recover high-frequency boundary cues lost during transformer-based global smoothing, particularly in high-turbidity regions with overlapping bio-optical signals. This synergistic deficiency manifests as gradient reversal artifacts along complex seabed contours and chromatic offsets in shadowed areas, highlighting the algorithm's inadequate fusion of spectral-spatial priors to address depth-variant degradation patterns in dynamic underwater environments.

The DiffWater method addresses underwater color distortion and quality degradation through conditional diffusion modeling, yet demonstrates critical vulnerabilities when confronting nonlinear photometric interactions in complex aquatic scenarios. Its channelwise color compensation mechanism in RGB space shows incomplete chromatic separation in turbid greenish waters, where wavelength-dependent scattering interferes with particulate density estimation, occasionally preserving blue-green dominance while introducing artificial magenta casts in shadow regions. The conditional DDPM framework exhibits unstable noise prediction capabilities under dynamic illumination fields, where conditional guidance from color-compensated inputs may misdirect the denoising trajectory, generating texture-inconsistent hallucinated details near high-particle-concentration zones. Furthermore, the sequential integration of color correction and diffusion processes demonstrates spectral incoherence in transitional depth layers, particularly failing to preserve wavelength attenuation gradients when processing scenes with abrupt optical density changes caused by algal blooms or sediment plumes. These limitations stem from the method's simplified assumption of additive degradation patterns and insufficient physical modeling of the nonlinear correlation between waterborne light scattering and depth-dependent chromatic absorption.

The proposed UIE algorithm in this study employs edge feature attention fusion to address critical problems, such as edge blurriness, low SNR, and low contrast in underwater images. It integrates three innovative modules: (1) Edge operators extract edge information through gradient-sensitive feature learning, while CCA fuses multi-scale features using cross-channel coherence analysis, restoring object edge details by jointly optimizing high-frequency components and improving visual performance.(2) A spatial cross-attention mechanism strengthens spatial structure information via edge-guided attention propagation, preserving details under low signal-to-noise ratio conditions through noise-adaptive feature reinforcement.(3) A multi-dimensional perception optimization method enhances semantic understanding, structural integrity, and local contrast using frequency-aware adversarial learning, while mitigating the effects of outliers through multi-scale degradation disentanglement. Collectively, these modules establish hierarchical edge-texture synchronization, where edge restoration and feature fusion are systematically coordinated to resolve cross-scale degradation conflicts in turbid underwater environments.

3.6 Component ablation and fusion validation

The excellent performance of the CUG-UIEF proposed in this study for UIE mainly benefits from the multi-feature cross-fusion module and the redesigned loss function. To verify the effectiveness of the modules proposed in this study, we conducted ablation studies using the UIEB dataset as the training set on Test-U90 and by selecting 45 challenging images from the LSUI dataset as Test-L45.

The original model selected in this study has been described in the literature (Ren et al., 2022). The specific experimental settings were consistent with those used in a previous experiment. Table 3 presents the results of this index. DDEM-1 represents CUG-UIEF with the edge fusion module removed, and DDEM-2 represents CUG-UIEF with the spatial information enhancement module removed.

As indicated in Table 3, the proposed model achieved the best quantitative performance on the two test datasets, reflecting the effectiveness of the combination of multi-feature cross-fusion and multidimensional perceptual loss function modules.

As shown in Figure 3, compared with the original model, the added multi-feature cross-fusion module better address the problem of cyan-green color casts. When processing images, the cross-attention mechanism can adaptively focus on the interactions between cyan-green channels and other channels, avoiding excessive or insufficient utilization of cyan-green channel information. It emphasizes local details than on the convolution in the original model. The color distribution in real scenes was better matched adjusting the weights and contributions of the cyan-green channels in the image to a more reasonable level, thus effectively correcting the cyan-green color cast and improving the accuracy and naturalness of the image colors. Moreover, after adding edge features, the attention mechanism can focus on the structural information in the image and avoid wasting resources in unimportant areas. Following the addition of edge features, the details of the stones and creatures in the two comparison images became clearer. As we can observe in Figure 4, owing to the multidimensional perceptual loss function, the obtained images exhibit enhanced details, improved color restoration, vivid object edges, high contrast, and clear boundaries.

3.7 Qualitative comparison through visualization

First, the image comparison results of the UIEB dataset are presented. A comprehensive training was conducted using the UIEB dataset. The test data selected for this study were sampled according to six scenes with distinct characteristics: shadow, texture, blur, blue, yellow, and green. The images that best represented each type of scene and were to some extent challenging were chosen. Figure 5 presents the enhancement results of the different methods.

TABLE 3 Statistical results of the ablation study on the modules and loss functions.

Module	Test-U90		Test-L45	
	PSNR	SSIM	PSNR	SSIM
Origin	23.2074	0.9178	21.9878	0.9164
DDEM-1	23.4076	0.9183	22.3455	0.9142
DDEM-2	25.5647	0.9195	23.2346	0.9234
CUG-UIEF	**26.4693**	**0.9286**	**24.5212**	**0.9276**
Loss	**PSNR**	**SSIM**	**PSNR**	**SSIM**
l_p	24.3572	0.9142	22.0478	0.9123
l_p,l_m	25.7823	0.9212	23.5689	0.9201
l_p,l_m,l_c	**26.4693**	**0.9286**	**24.5212**	**0.9276**

In the results, boldface indicates the best data.

In the green scene, color casts occurred in the results of HL, Sha-UWnet, and UWCNN. Only Fusion, USUIR, URSCT, and CUG-UIEF showed color restoration similar to that of the real image. However, CUG-UIEF achieved the closest color restoration effect to the real image, and the details in the shadowed parts were the clearest and most distinguishable. In the blue scene, except for USUIR, URSCT, and CUG-UIEF, none of the methods restored the real illumination effect, and only CUG-UIEF truly restored the color texture of the fish in the upper left corner. In foggy and textured scenes, only the proposed URSCT and CUG-UIEF achieved good effects in defogging and enhancing textures and object edges. However, compared to real situations, both methods had some deficiencies. In the final yellow scene, only CUG-UIEF retained delicate edge information during defogging. Overall, CUG-UIEF was visually superior to the other methods.

This section also presents the image results of CUG-UIEF on no-reference datasets. The tests were performed on two test sets, Test-C60 and Test-U45.

Test-C60 includes five underwater environments—red, yellow, green, blue, and foggy scenes—all of which were affected by high backscattering and color deviation. The most representative images of each type were selected for visual comparison. As shown in Figure 6, HL, CBLA, WWPF, Sha-UWnet, UWCNN, and URSCT exhibited obvious color deviations in some cases. In the yellow scene, HL, IBLA, and USUIR restored the paddle blade to purple, whereas UWCNN and CUG-UIEF restored it to yellow, which is closer to the normal visual perception of humans. Moreover, CUG-UIEF can better restore blurred details in the original image. In the green and blue scenes, only URSCT and CUG-UIEF achieved good restoration of the background and surfaces of the creatures. In the foggy scene, URSCT had a significant defogging effect but overly enhanced the red color in the original image. CUG-UIEF attempted to retain the information of the original image while defogging, and the color restoration at the bottom background was more in line with normal perception. In the shadow and texture scenes, all the methods except USUIR, URSCT, and CUG-UIEF, exhibited color restoration deviations. These three methods could restore the details in the shaded parts while retaining the natural illumination, but only CUG-UIEF could retain sufficient light–dark contrast and object details while providing improved color for the seawater background.

TEST-U45 contains multiple scenes, such as color deviation and foggy scenes. Multiple scenes were selected for the experiments, and representative scenes were selected for display. As is shown Figure 7, except for HL, UWCNN, CBLA, WWPF and Sha-UWnet, the methods exhibited a lower degree of color deviation. In the shadow and texture scenes, Fusion, USUIR, URSCT, and CUG-UIEF performed well in color restoration and texture information preservation. However, in the blue scene, only the URSCT and CUG-UIEF restored colors that were more in line with normal visual perception and preserved the texture information of the objects well. In the green scene, only CUG-UIEF could better reflect the natural illumination environment and delicate details.

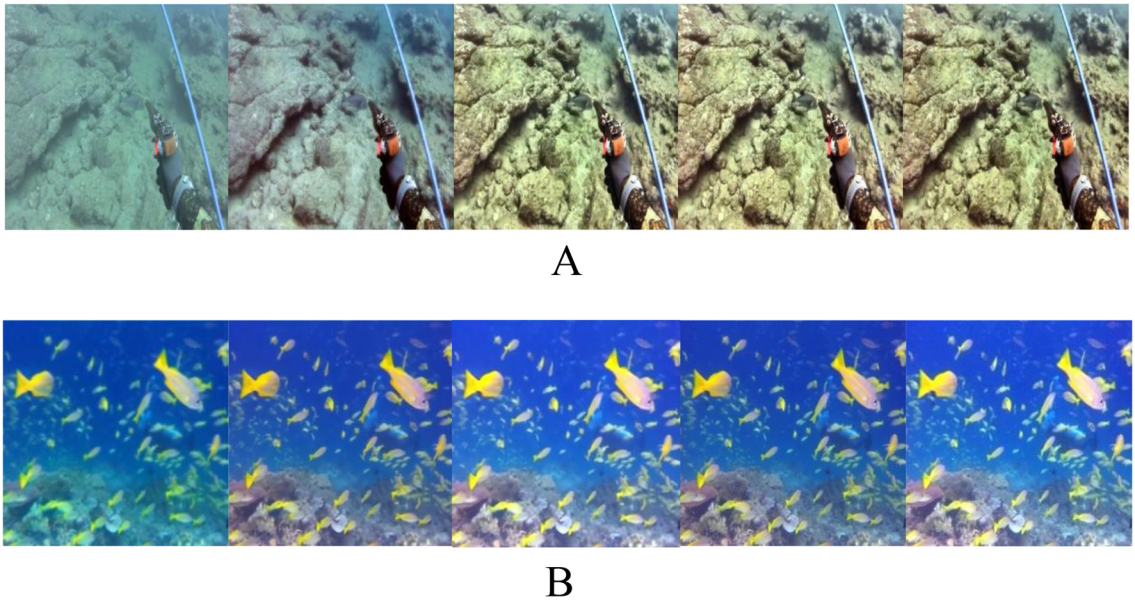


FIGURE 3
Ablation study on the contribution of cross-fusion. Each panel includes the original image (RAW), the results of the original method, the results of DDEM-1, the results of DDEM-2, and the results of CUG-UIEF. **(A)** Test-U90, **(B)** Test-L45.

In summary, HL, UWCNN, WWPF and Sha-UWnet were prone to color-cast phenomena. The IBLA improved the quality of underwater images using local adaptive methods but performed poorly in yellow, foggy, and some blue scenes. Fusion greatly increased artificial colors to enhance contrast but could not adapt well to the changes caused by foggy scenes. USUIR performed well in most scenes but often exhibited a red-shift phenomenon in blue scenes. URSCT had good robustness and strong defogging ability; however, when restoring objects in foggy scenes, it was prone to red-shift phenomena. CUG-UIEF had good robustness and

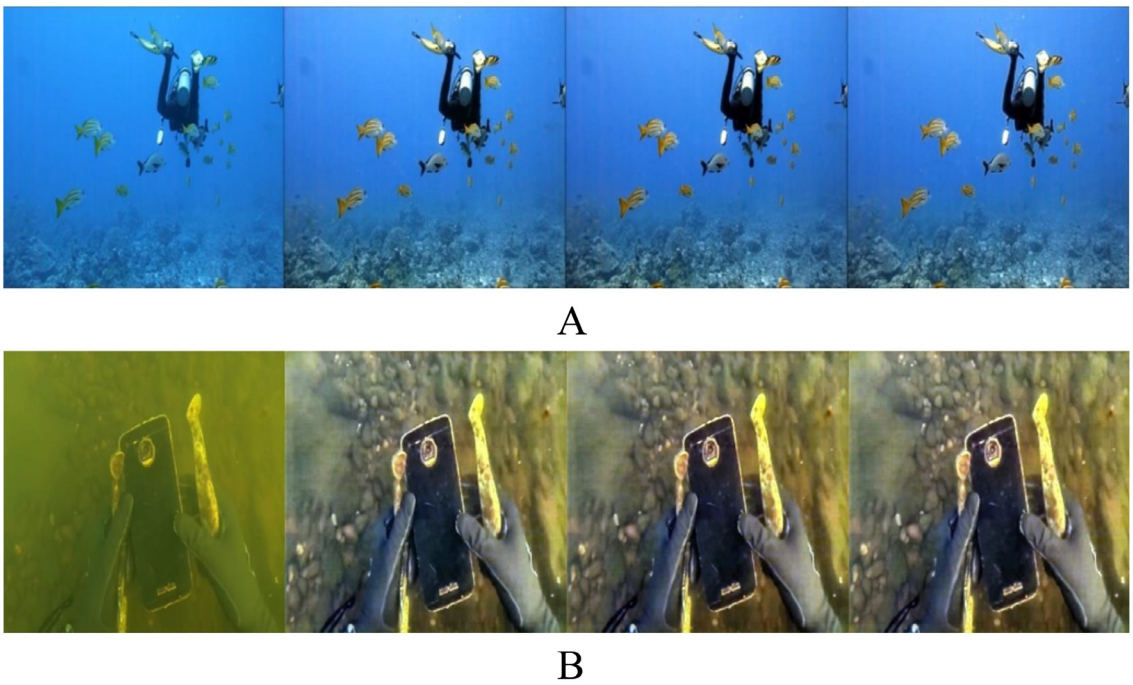


FIGURE 4
Multi-dimensional ablation study. Each panel includes the original image (RAW) and the results of the original method. The results of using only L_p, the results of using both L_p and L_m, and the results of using L_p, L_m, and L_c simultaneously. **(A)** Test-U90, **(B)** Test-L45.

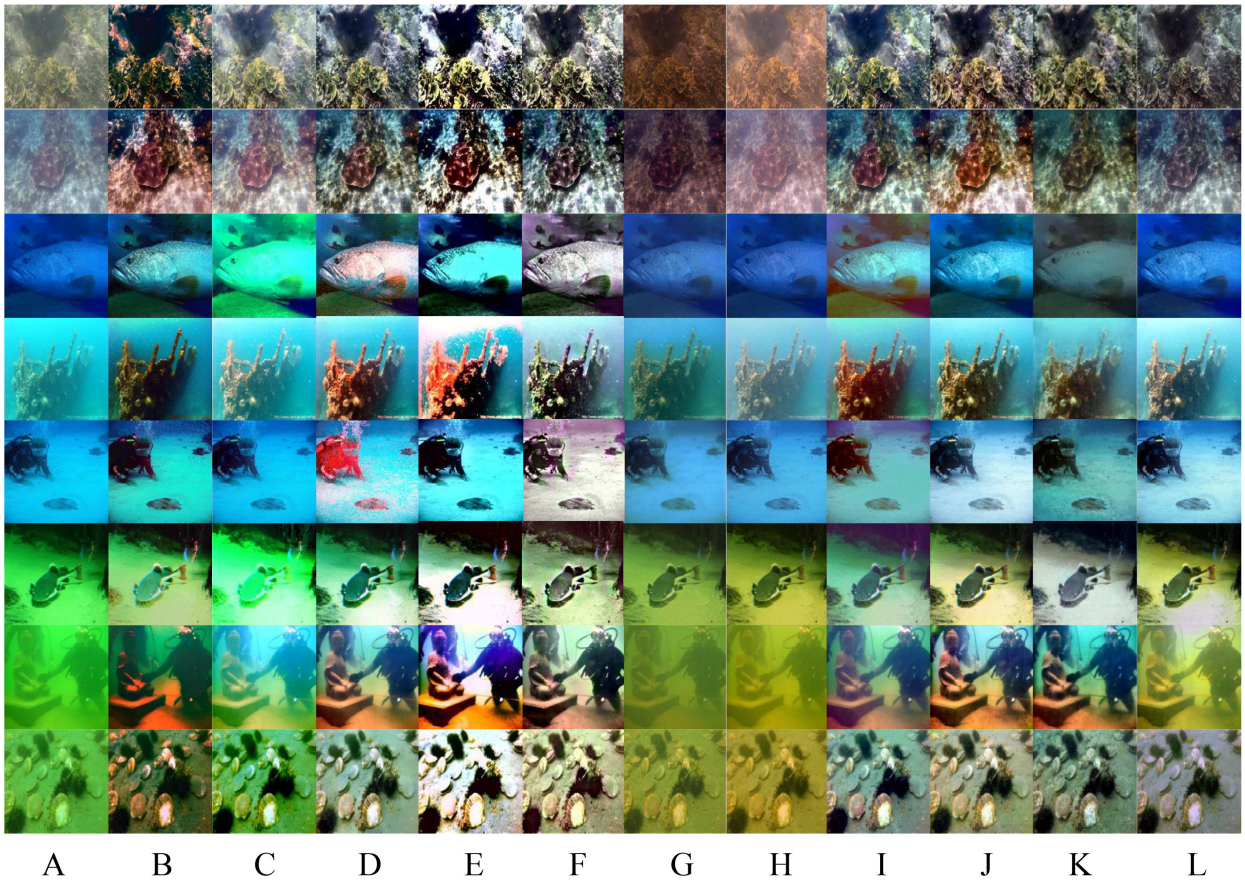


FIGURE 5
Comparison of the underwater images sampled from the Test-U90 dataset. **(A)** RAW **(B)** HL **(C)** IBLA **(D)** Fusion **(E)** CBLA **(F)** WWPE **(G)** Sha-UWnet **(H)** UWCNN **(I)** USUIR **(J)** URSCT **(K)** Diff-Water **(L)** Ours **(M)** Ground truth.

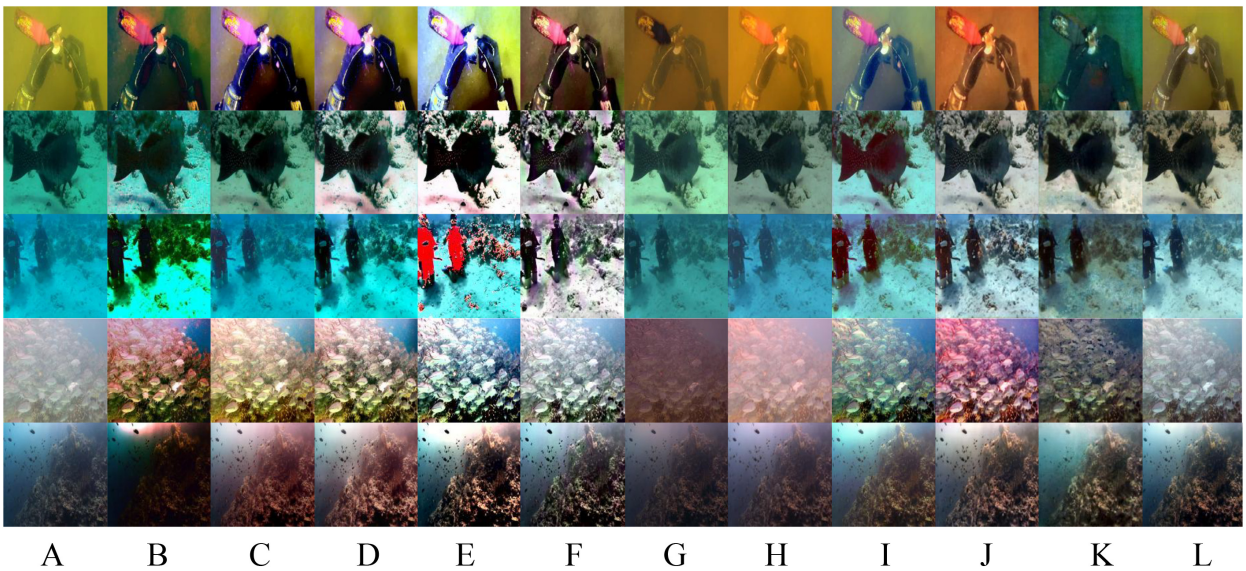


FIGURE 6
Visual comparison of the underwater images sampled from the Test-C60 dataset. **(A)** RAW **(B)** HL **(C)** IBLA **(D)** Fusion **(E)** CBLA **(F)** WWPE **(G)** Sha-UWnet **(H)** UWCNN **(I)** USUIR **(J)** URSCT **(K)** Diff-Water **(L)** Ours.

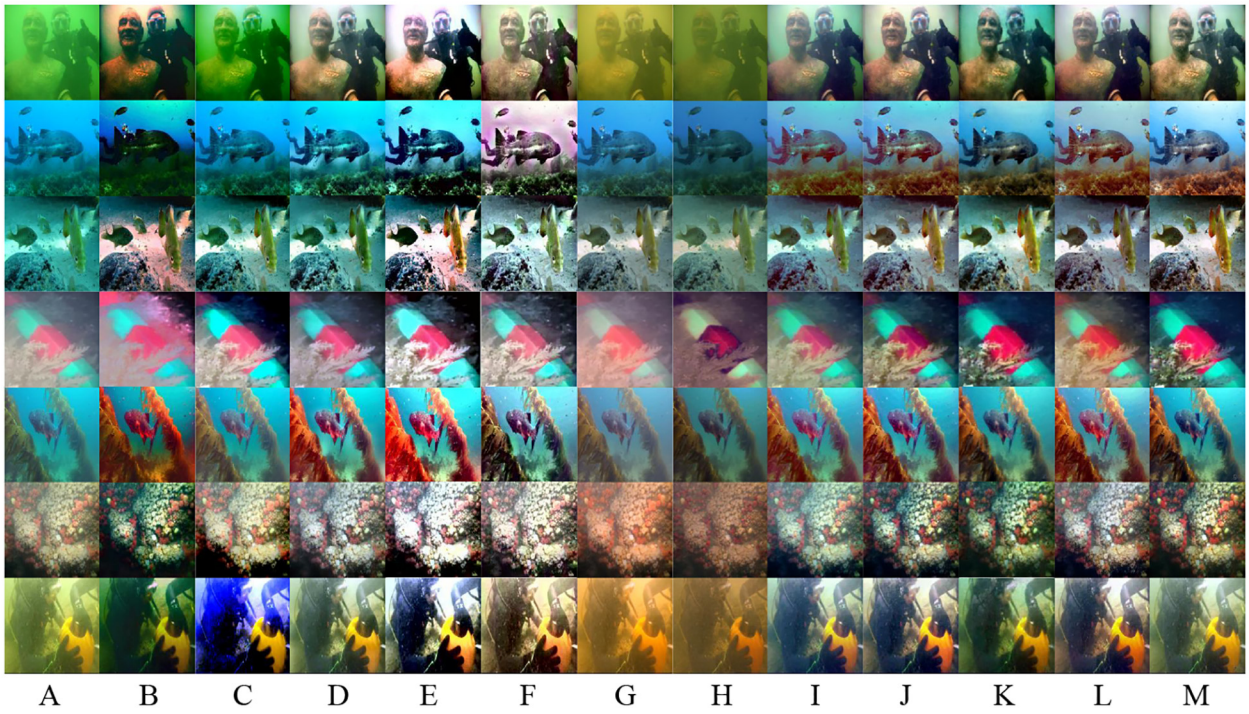


FIGURE 7
Visual comparison of the underwater images sampled from the Test-U45 dataset. **(A)** RAW **(B)** HL **(C)** IBLA **(D)** Fusion **(E)** CBLA **(F)** WWPE **(G)** Sha-UWnet **(H)** UWCNN **(I)** USUIR **(J)** URSCT **(K)** Diff-Water **(L)** Ours.

performed well in yellow, blue, green, foggy, shadowed, and textured scenes.

3.8 Application and inference efficiency

As feature extraction-matching and edge detection constitute core technical pillars in underwater image analysis, this study systematically validated the necessity of the proposed method as a preprocessing module for feature matching and edge detection. In feature matching tasks, the SIFT algorithm was utilized to extract feature points, complemented by the RANSAC algorithm for false match elimination. Feature matching was performed on preprocessed 256×256 pixels underwater stereo image pairs from the SQUID dataset. Figure 8 revealed that the proposed approach significantly optimized matching performance while concurrently improving visual quality. Table 4 demonstrates that compared to baseline methods, our scheme ranked second in both initial and valid matches, yet achieved the highest matching precision. Integrative qualitative-quantitative analyses corroborated the critical utility of this method for underwater feature matching tasks.

Regarding edge detection tasks, all images in the Test-C60 and Test-U45 datasets underwent enhancement prior to edge extraction and evaluation via the Canny operator. Detection performance was quantified using three metrics: Precision, F1 (harmonic mean of precision and recall), and Edge Pixel Ratio (EPR). Table 5 indicated that our method ranked first in accuracy and second in EPR relative to state-of-the-art approaches. Figure 9 demonstrates the experimental findings: In Test-U45 fish samples, the enhanced edge detection preserves intact morphological contours while precisely discriminating target-background depth disparities, revealing underwater spatial hierarchy. In Test-C60 columnar targets, the algorithm achieves complete extraction of artificial structures' geometric edges with enhanced low-light gradient responses, where continuous seagrass blade edges further validate optical attenuation compensation. Convergent qualitative and quantitative evidence validated the significant contribution of this method to underwater edge detection tasks.

To evaluate the practical applicability of underwater image enhancement algorithms, we conducted a systematic comparison of inference efficiency among competing methods. The experiments were performed using the UIEB dataset as the benchmark, with average inference times calculated across all test samples. Traditional algorithms were executed in batch processing mode, while deep learning approaches employed pre-trained models on the UIEB training set for inference. Owing to significant architectural variations among deep learning algorithms, substantial discrepancies in inference times were observed across different models. As demonstrated in Table 6, conventional algorithms maintain absolute superiority in computational speed, whereas the proposed framework achieves the second-highest efficiency among deep learning methods while demonstrating a competitive advantage over structurally complex traditional approaches. These findings validate the proposed method's significant advantages in balancing computational complexity with practical deployment feasibility.

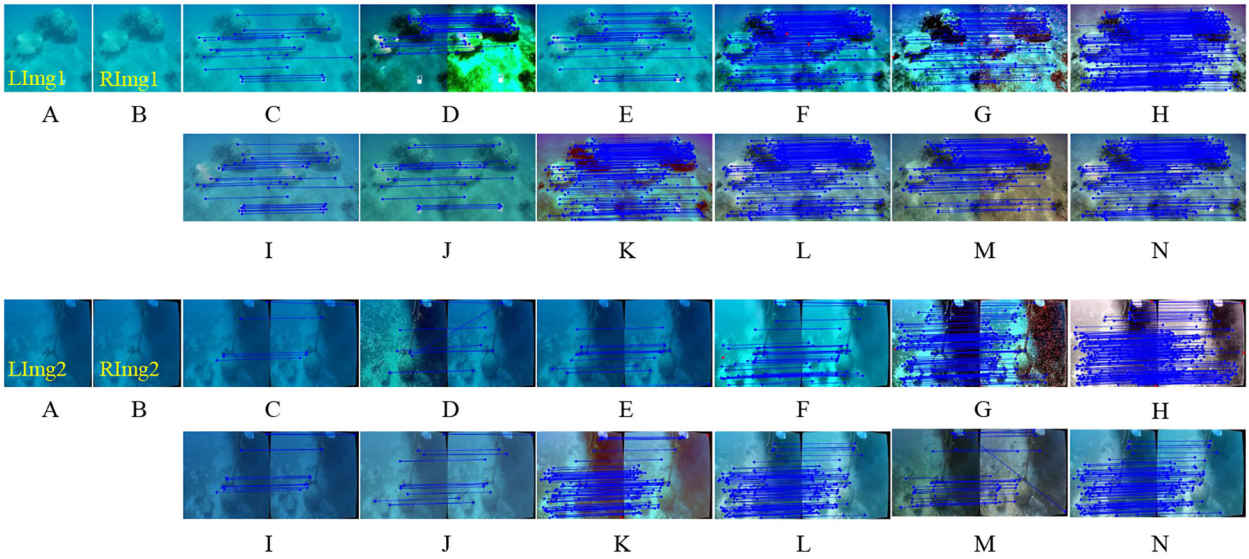


FIGURE 8
Application examples of underwater feature matching. **(A)** RAW-Left **(B)** RAW-right **(C)** RAW **(D)** HL **(E)** IBLA **(F)** Fusion **(G)** CBLA **(H)** WWPE **(I)** Sha-UWnet **(J)** UWCNN **(K)** USUIR **(L)** URSCT **(M)** Diff-Water **(N)** Ours.

TABLE 4 Mean evaluation results of underwater feature matching on the SQUID dataset.

	Method	Initial matches	Valid matches	Precision
Traditional Method	HL	44.32	38.67	87.25%
	IBLA	37.42	31.73	84.79%
	Fusion	46.56	39.14	84.06%
	CBLA	87.35	78.15	89.46%
	WWPF	**198.58**	**166.21**	83.70%
Deep-Learning Method	Sha-UWnet	36.74	31.27	85.12%
	UWCNN	39.55	32.52	82.29%
	USUIR	152.46	135.14	88.64%
	URSCT	163.24	145.81	89.24%
	Diff-Water	108.47	97.09	89.51%
	Ours	172.68	155.81	**90.23%**

In the results, boldface indicates the best data and blue denotes the suboptimal data.

TABLE 5 Mean evaluation results of underwater feature matching on the Test-C60 and Test-U45 dataset.

	Method	Precision	F1	EPR
Traditional Method	HL	0.6011	0.1406	0.0303
	IBLA	0.6208	0.4789	0.1492
	Fusion	0.6149	0.3311	0.0926
	CBLA	0.6215	**0.5435**	**0.2021**
	WWPF	0.6571	0.5252	0.1888

(Continued)

TABLE 5 Continued

	Method	Precision	F1	EPR
Deep-Learning Method	Sha-UWnet	0.6666	0.0295	0.0057
	UWCNN	0.6523	0.3145	0.0649
	USUIR	0.6314	0.2227	0.0542
	URSCT	0.6289	0.4445	0.1436
	Diff-Water	0.6403	0.4474	0.1381
	Ours	**0.6719**	0.4761	0.1989

In the results, boldface indicates the best data and blue denotes the suboptimal data.

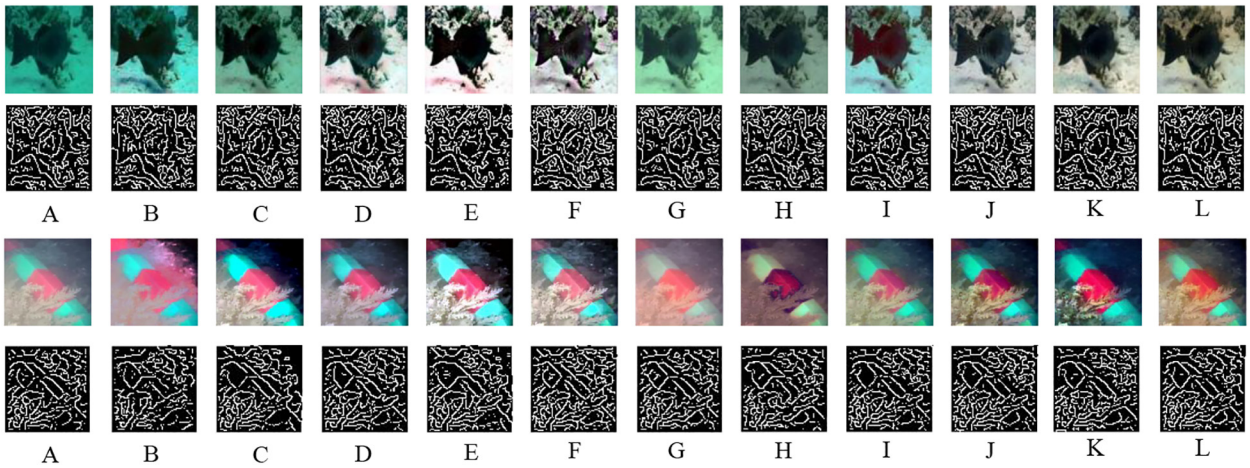


FIGURE 9
Application examples of canny edge detection. **(A)** RAW **(B)** HL **(C)** IBLA **(D)** Fusion **(E)** CBLA **(F)** WWPE **(G)** Sha-UWnet **(H)** UWCNN **(I)** USUIR **(J)** URSCT **(K)** Diff-Water **(L)** Ours.

TABLE 6 Inference Efficiency Comparison.

	Method	Per-image inference time
Traditional Method	HL	0.284s
	IBLA	0.622s
	Fusion	1.756s
	CBLA	**0.199s**
	WWPF	0.652s
Deep-Learning Method	Sha-UWnet	3.643s
	UWCNN	2.911s
	USUIR	1.876s
	URSCT	1.061s
	Diff-Water	44.322s
	Ours	1.083s

In the results, boldface indicates the best data and blue denotes the suboptimal data.

4 Conclusion

This study presented a deep learning model for UIE that improves blurring and color distortion caused by light scattering and attenuation. The proposed model integrates a multi-feature cross-fusion module, which combines edge features with encoder features and utilizes a channel-cross attention mechanism, effectively enhancing the clarity of blurred areas and improving edge detail capture. Additionally, the spatial information enhancement module strengthens feature interactions across different locations, enabling more natural restoration of color-distorted regions, thereby bringing the image closer to true colors and clarity. Through multi-dimensional perception optimization, the model further improves clarity, color accuracy, and edge details. Experimental results confirm the superior ability of the model to restore image details and correct color distortion. Ablation studies highlight the effectiveness of both the multi-feature cross-fusion module and multi-dimensional perception optimization in enhancing detail and overall color consistency. However, the

dehazing performance of the model in large-scale foggy underwater images requires further improvement. Future work will incorporate multispectral data to address the limitations, enhance dehazing performance, and improve the overall robustness and generalizability of the model in complex scenarios, ultimately providing more reliable image enhancement solutions for practical underwater operations and 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 authors.

Author contributions

SS: Data curation, Methodology, Writing – original draft. WC: Supervision, Writing – review & editing, Methodology. HW: Data curation, Validation, Writing – original draft. PW: Data curation, Validation, Writing – review & editing. QL: Validation, Visualization, Writing – review & editing. XQ: Funding acquisition, Methodology, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was jointly supported by the China Geology Survey Project (NO. 20241910244) and the Opening Fund of Key Laboratory of Geological Survey and Evaluation of Ministry of Education (No. GLAB2024ZR01).

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

Alsakar, Y. M., Sakr, N. A., El-Sappagh, S., Abuhmed, T., and Elmogy, M. (2024). Underwater image restoration and enhancement: A comprehensive review of recent trends, challenges, and applications. *Vis. Comput.* 1–49. doi: 10.1007/s00371-024-03630-w

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 (CVPR)*. (Providence, RI, USA: IEEE), p. 81–88. doi: 10.1109/CVPR.2012.6247661

Banik, P. P., Saha, R., and Kim, K.-D. (2018). "Contrast enhancement of low-light image using histogram equalization and illumination adjustment," in *2018 international conference on electronics, information, and Communication (ICEIC)*, Vol. 2018. (Hawaii, USA: IEEE), p. 1–4. doi: 10.23919/ELINFOCOM.2018.8330564

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

Drews, P. L. J., Nascimento, E. R., Botelho, S. S. C., and Campos, M. F. M. (2016). Underwater depth estimation and image restoration based on single images. *IEEE Comput. Graph. Appl.* 36, 24–35. doi: 10.1109/MCG.2016.26

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)*. (Brisbane, Australia: IEEE), p. 7159–7165. doi: 10.1109/ICRA.2018.8460552

Fu, Z., Lin, H., Yang, Y., Chai, S., Sun, L., Huang, Y., et al. (2022). "Unsupervised underwater image restoration: From a homology perspective," in *Proceedings of the AAAI conference on artificial intelligence (AAAI)*, Vol. 36. (AAAI), 643–651. doi: 10.1609/aaai.v36i1.19944

Garg, D., Garg, N. K., and Kumar, M. (2018). Underwater image enhancement using blending of CLAHE and percentile methodologies. *Multimedia Tool. Appl.* 77, 26545–26561. doi: 10.1007/s11042-018-5878-8

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial nets. in Advances in neural information processing systems. Montreal, Quebec, Canada: 27(NIPS). doi: 10.5555/2969033.2969125.

Guan, M., Xu, H., Jiang, G., Yu, M., Chen, Y., Luo, T., et al. (2023). DiffWater: Underwater image enhancement based on conditional denoising diffusion probabilistic model. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 17, 2319–2335. doi: 10.1109/JSTARS.2023.3344453

Guo, C., Wu, R., Jin, X., Han, L., Zhang, W., Chai, Z., et al. (2023). "Underwater ranker: Learn which is better and how to be better," in *Proceedings of the AAAI conference on artificial intelligence.(AAAI)*, Vol. 1. (Washington, DC, USA: AAAI), 702–709. doi: 10.1609/aaai.v37i1.25147

Hu, K., Weng, C., Zhang, Y., Jin, J., and Xia, Q. (2022). An overview of underwater vision enhancement: From traditional methods to recent deep learning. *J. Mar. Sci. Eng.* 10, 241. doi: 10.3390/jmse10020241

Islam, M. J., Luo, P., and Sattar, J. (2020). Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception. In *16th Robotics: Science and Systems* (Corvalis, OR, USA: MIT Press Journals), 18–28. doi: 10.15607/RSS.2020.XVI.018

Jha, M., and Bhandari, A. K. (2024). CBLA: color-balanced locally adjustable underwater image enhancement. *IEEE Trans. Instrum. Meas.* 73, 3396850. doi: 10.1109/TIM.2024.3396850

Li, B., Chen, Z., Lu, L., Qi, P., Zhang, L., Ma, Q., et al. (2025). Cascaded frameworks in underwater optical image restoration. *Inform. Fusion.* 117, 102809. doi: 10.1016/j.inffus.2024.102809

Li, C., Anwar, S., and Porikli, F. J. P. R. (2020). Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognit.* 98, 107038. doi: 10.1016/j.patcog.2019.107038

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, 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.* 3, 1–1. doi: 10.1109/LRA.2017.2730363

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021). "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*. (New York: IEEE), 9992–10002. doi: 10.1109/ICCV48922.2021.00986

Lyu, Z., Peng, A., Wang, Q., and Ding, D. (2022). An efficient learning-based method for underwater image enhancement. *Displays* 74. doi: 10.1016/j.displa.2022.102174

Mazzeo, A., Aguzzi, J., Calisti, M., Canese, S., Vecchi, F., Stefanni, S., et al. (2022). Marine robotics for deep-sea specimen collection: A systematic review of underwater grippers. *Sensors (Basel).* 22, 648. doi: 10.3390/s22020648

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 Online* Vol. 35. (AAAI), 15853–15854. doi: 10.1609/aaai.v35i18.17923

Nazir, S., and Kaleem, M. (2021). Advances in image acquisition and processing technologies transforming animal ecological studies. *Ecol. Inf.* 61, 101212. doi: 10.1016/J.ECOINF.2021.101212

Peng, L., Zhu, C., and Bian, L. (2023). U-shape transformer for underwater image enhancement. *IEEE Trans. Image Process.* 32, 3066–3079. doi: 10.1109/TIP.2023.3276332

Ren, T., Xu, H., Jiang, G., Yu, M., and Luo, T. (2022). Reinforced swin-convs transformer for underwater image enhancement. *IEEE Trans. Geosci. Remote. Sens.* 60, 1–16. doi: 10.1109/TGRS.2022.3205061

Sun, J., Dong, J., and Lv, Q. (2022). "2022. Swin transformer and fusion for underwater image enhancement," in *International Workshop on Advanced Imaging Technology (IWAIT)*. (Hong Kong, China: SPIE) Vol. 12177, pp. 627–631. doi: 10.1117/12.2626075

Swinehart, D. F. (1962). The beer-lambert law. *J. Chem. Educ.* 39, 333. doi: 10.1021/ed039p333

Wang, Y. D., Guo, J., Gao, H., and Yue, H. (2021). UIEC^2-Net: CNN-based underwater image enhancement using two color space. *Signal Process. Image Commun.* 96. doi: 10.1016/j.image.2021.116250

Wang, C. B., Ji, K., Huang, Q., and Xia, Y. (2007). "Generation of multi-spectral scene images for ocean environment," in *MIPPR 2007: Multispectral Image Processing*, Vol. 6787. (Wuhan, China: SPIE), p. 596–603. doi: 10.1117/12.751757

Wang, H., Köser, K., and Ren, P. (2025). Large foundation model empowered discriminative underwater image enhancement. *IEEE Trans. Geosci. Remote. Sens* 63, 1–17. doi: 10.1109/TGRS.2025.3525962

Wang, Y., Zhang, J., Cao, Y., and Wang, Z. (2017). "A deep CNN method for underwater image enhancement," in *24th IEEE International Conference on Image Processing (ICIP)*. (IEEE), p. 1382–1386. doi: 10.1109/ICIP.2017.8296508

Wang, S., Zheng, J., Hu, H. M., and Li, B. (2013). Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Trans. Image Process.* 22, 3538–3548. doi: 10.1109/TIP.2013.2261309

Xiong, J., Zhuang, P., and Zhang, Y. (2020). "An efficient underwater image enhancement model with extensive Beer-Lambert law," in *IEEE International Conference on Image Processing (ICIP)*. (Abu Dhabi, United Arab Emirates: IEEE), p. 893–897. doi: 10.1109/ICIP40778.2020.9191131

Yang, T., Yang, T., Gao, W., Wang, P., Li, X., Lv, Z., et al. (2023). Underwater target detection algorithm based on automatic color level and bidirectional feature fusion. *Laser Optoelectron. Prog.* 60, 11–18. doi: 10.3788/LOP213139

Zhang, W., Li, X., Huang, Y., Xu, S., Tang, J., and Hu, H. (2025). Underwater image enhancement via frequency and spatial domains fusion. *Opt. Laser. Eng.* 186, 108826. doi: 10.1016/j.optlaseng.2025.108826

Zhang, W., Zhou, L., Zhuang, P., Li, G., Pan, X., Zhao, W., et al. (2023). Underwater image enhancement via weighted wavelet visual perception fusion. *IEEE Trans. Circuits Syst. Video Technol.* 34.4, 2469–2483. doi: 10.1109/TCSVT.2023.3299314

Zhu, D. (2023). Underwater image enhancement based on the improved algorithm of dark channel. *Mathematics.* 11, 1382. doi: 10.3390/math11061382

Zhu, Z., Li, X., Ma, Q., Zhai, J., and Hu, H. (2025). FDNet: Fourier transform guided dual-channel underwater image enhancement diffusion network. *Sci. China. Technol. Sc.* 68.1, 1100403. doi: 10.1007/s11431-024-2824-x

Zhu, Z., Li, X., Zhai, J., and Hu, H. (2024). PODB: A learning-based polarimetric object detection benchmark for road scenes in adverse weather conditions. *Inform. Fusion.* 108, 102385. doi: 10.1016/j.inffus.2024.102385

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 15 April 2025
DOI 10.3389/fmars.2025.1542832

OPEN ACCESS

EDITED BY
Muhammad Yasir,
China University of Petroleum (East China), China

REVIEWED BY
Nuno Pessanha Santos,
Portuguese Military Academy, Portugal
Martin Aubard,
University of Porto, Portugal

*CORRESPONDENCE
Jing Han
hanj@nwpu.edu.cn

RECEIVED 10 December 2024
ACCEPTED 11 March 2025
PUBLISHED 15 April 2025

CITATION
Cui X, Zhang J, Zhang L, Zhang Q and Han J (2025) Small object detection in side-scan sonar images based on SOCA-YOLO and image restoration. *Front. Mar. Sci.* 12:1542832. doi: 10.3389/fmars.2025.1542832

Small object detection in side-scan sonar images based on SOCA-YOLO and image restoration

Xiaodong Cui, Jiale Zhang, Lingling Zhang, Qunfei Zhang and Jing Han*

School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an, China

Although side-scan sonar can provide wide and high-resolution views of submarine terrain and objects, it suffers from severe interference due to complex environmental noise, variations in sonar configuration (such as frequency, beam pattern, etc.), and the small scale of targets, leading to a high misdetection rate. These challenges highlight the need for advanced detection models that can effectively address these limitations. Here, this paper introduces an enhanced YOLOv9(You Only Look Once v9) model named SOCA-YOLO, which integrates a Small Object focused Convolution module and an Attention mechanism to improve detection performance to tackle the challenges. The SOCA-YOLO framework first constructs a high-resolution SSS (sidescan sonar image) enhancement pipeline through image restoration techniques to extract fine-grained features of micro-scale targets. Subsequently, the SPDConv (Space-to-Depth Convolution) module is incorporated to optimize the feature extraction network, effectively preserving discriminative characteristics of small targets. Furthermore, the model integrates the standardized CBAM (Convolutional Block Attention Module) attention mechanism, enabling adaptive focus on salient regions of small targets in sonar images, thereby significantly improving detection robustness in complex underwater environments. Finally, the model is verified on a public side-scan sonar image dataset Cylinder2. Experiment results indicate that SOCA-YOLO achieves Precision and Recall at 71.8% and 72.7%, with an mAP50 of 74.3%. It outperforms the current state-of-the-art object detection method, YOLO11, as well as the original YOLOv9. Specifically, our model surpasses YOLO11 and YOLOv9 by 2.3% and 6.5% in terms of mAP50, respectively. Therefore, the SOCA-YOLO model provides a new and effective approach for small underwater object detection in side-scan sonar images.

KEYWORDS

side-scan sonar, image restoration, YOLOv9, attention mechanism, Space-to-Depth Convolution

1 Introduction

Side-scan sonar (何勇光, 2020) is an extensively utilized underwater sensing technology, mainly applied in underwater terrain mapping, object detection, and exploration tasks. In contrast to conventional downward-looking sonar, side-scan sonar transmits acoustic waves at horizontal or inclined angles, thereby covering a larger area of seabed features and improving detection performance. As a result, side-scan sonar is widely utilized in areas such as maritime archaeology, submerged pipeline monitoring, and wreck exploration (Gomes et al., 2020; Tian et al., 2007; Fengchun et al., 2002; Sun et al., 2021; Jinhua et al., 2016). Nevertheless, the intricate underwater environment often introduces multiple sources of noise and blurring in side-scan sonar images, including scattering noise, multipath artifacts, noise streaks, and acoustic shadow distortions. Furthermore, instrumental noise arises from the sensor's inherent electronic noise and the transducer's non-ideal properties, potentially leading to image signal degradation. The interaction of these noise factors results in considerable difficulties in processing side-scan sonar images for real-world applications.

The unique properties of side-scan sonar images introduce significant difficulties in target detection. Firstly, sonar imagery often exhibits considerable background noise and spurious objects, including natural seabed formations and acoustic backscatter from sediment particles, which frequently resemble real targets and result in an elevated false alarm rate in detection models. Secondly, targets in sonar images generally manifest as small, diffuse high-intensity reflections with vague edges and uneven signals, making them indistinguishable from surrounding textures and increasing the difficulty of segmentation from the background. Furthermore, side-scan sonar image data exhibit substantial distribution discrepancies across different scenarios. Given the high cost and inefficiency of underwater data acquisition, labeled datasets are often scarce. This non-uniformity and data insufficiency severely hinder the generalization capability of algorithms, posing a formidable challenge for achieving accurate target detection in complex underwater settings. The rapid progress in artificial intelligence and machine learning has facilitated the fusion of advanced image processing techniques with target detection models, substantially enhancing side-scan sonar image quality and improving the precision of seabed target detection (Yasir et al., 2024; Cheng et al., 2023; Wen et al., 2024; Fan et al., 2022; Yu et al., 2021; Fayaz et al., 2022).

Among existing underwater target detection methods for side-scan sonar images, some object detection models have become relatively outdated and struggle to meet the current diversified underwater application requirements. Although some studies have improved traditional deep learning models, these enhancements often fail to adequately consider the inherent structural characteristics of side-scan sonar images. This neglect of sonar image characteristics makes targeted model optimization challenging, resulting in subpar detection performance in practical applications. Furthermore, while some modified models have enhanced detection capabilities to some extent, their parameter counts have also increased substantially, leading to higher computational costs. Therefore, developing an effective underwater target detection method tailored to the specific requirements of side-scan sonar images is particularly crucial. As shown in Figure 1, by improving existing object detection models with greater emphasis on the structure and characteristics of side-scan sonar images, we can significantly enhance detection performance while effectively controlling model parameters and computational complexity, thereby providing more reliable metrics for underwater detection tasks.

This paper is structured into four main sections: The first section provides a literature review, systematically summarizing the current research status in underwater side-scan sonar image target detection. The second section focuses on methodology, providing a detailed explanation of the proposed detection model and its theoretical framework. The third section presents experimental validation, where multiple comparative and ablation experiments empirically analyze the performance advantages of the proposed model. The fourth section provides conclusions and future perspectives, discussing in depth the future research directions and trends in underwater side-scan sonar image processing based on an evaluation of the model's practical performance.

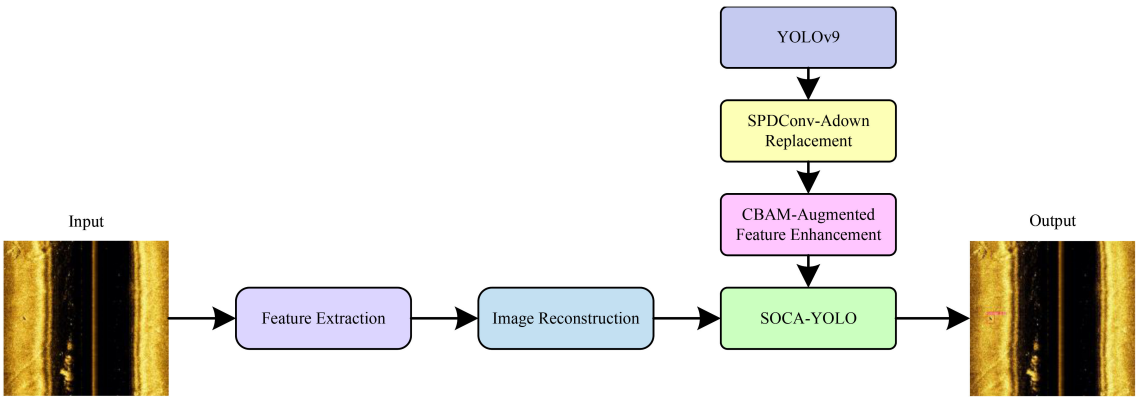


FIGURE 1
SSS object detection architecture.

The main contributions of this paper are as follows:

1. SwinIR-based sonar image enhancement method: To address the issues of low quality and high noise interference in traditional sonar images, the SwinIR super-resolution reconstruction network is introduced into the field of sonar image preprocessing. This method can more effectively enhance image quality, providing clearer input features for subsequent target detection.
2. Optimal model selection for small target detection in images: In the task of small target detection in side-scan sonar images, a comparison of existing object detection network models reveals that YOLOv9, through its auxiliary reversible branch, retains critical feature information, significantly enhancing the model's ability to detect small targets, particularly improving target recognition accuracy in complex backgrounds.
3. CBAM-enhanced detection model: Building upon the standard YOLOv9 network, the convolutional block attention module (CBAM) is innovatively incorporated. Unlike the original model's reliance solely on convolutional feature extraction, this method adaptively focuses on key target features, significantly improving target detection accuracy in complex underwater environments.
4. SPDConv replacement for ADown downsampling scheme: To address the challenges of small target detection in sonar images, the original ADown module in YOLOv9 is replaced with the SPDConv (Space-to-Depth Convolution) module. Compared to traditional downsampling methods, this improvement effectively mitigates the issue of small target feature loss.
5. Sonar image dataset reconstruction and evaluation: Existing public sonar datasets are systematically restructured, and a data partitioning standard more aligned with practical application scenarios is proposed. Experimental results demonstrate that the proposed improvements outperform traditional methods across all metrics.

2 Related work

As deep learning technology advances, various effective approaches have been introduced in image enhancement. The goal of image enhancement is to enhance image visual quality and interpretability using different algorithms, spanning from basic filtering to sophisticated color adjustment and detail refinement. Methods based on Convolutional Neural Networks (CNNs), such as Du (Du et al., 2023), employ four conventional CNN models for training and predicting on the same submarine SSS dataset. A comparative analysis was conducted on the predictive accuracy and computational efficiency of the four CNN models. Generative Adversarial Networks (GANs) employ adversarial learning between a generator and a discriminator to produce highly detailed images. Jiang (Jiang et al., 2020), for example, introduced a GAN-based semantic image synthesis model that can efficiently generate high-quality SSS images with reduced computational cost and time. Swin Transformer (Liu et al., 2021) serves as a versatile vision model designed mainly for image classification, object detection, and semantic segmentation (Lin et al., 2022; Gao et al., 2022; He et al., 2022; Jannat and Willis, 2022), with potential applications in image enhancement and video processing. It is specifically designed for efficient high-resolution image processing and has demonstrated superior performance in multiple visual tasks. SwinIR (Liang et al., 2021), built upon Swin Transformer, is a deep learning framework tailored for image restoration, encompassing super-resolution, noise reduction, and deblurring, among other tasks. Retaining the strengths of Swin Transformer, it integrates task-specific optimizations for image restoration, leading to improved processing efficiency and output quality. SwinIR has demonstrated significant performance improvements across various fields. For instance, in medical imaging, its application in low-dose PET/MRI restoration achieves a mean SSIM of 0.91 at a 6.25% dose level, substantially enhancing image quality (Wang et al., 2023b). In the domain of remote sensing, experiments on benchmark datasets show that SwinIR can enhance the resolution of satellite and aerial images—at a 2× scaling factor, its PSNR reaches 35.367dB and its SSIM increases to 0.9449, thereby facilitating more accurate topographic monitoring and mapping (Ali et al., 2023). Moreover, in video enhancement and facial recognition (Zheng et al., 2022; Lin, 2023), SwinIR's robust feature extraction and reconstruction capabilities significantly improve detail recovery and overall performance, as evidenced by its competitive results in multiple top-tier challenges. These advancements in deep learning have propelled significant innovations in image enhancement techniques.

In the field of computer vision, object detection and image enhancement are two complementary and important research directions. Image enhancement techniques aim to improve image quality, providing more accurate inputs for object detection, while object detection techniques focus on identifying and localizing objects of interest within images. Deep learning-based object detection methods are primarily divided into two categories: one-stage methods and two-stage methods. One-stage detection models directly predict target locations and categories through a single network forward pass, offering faster speed but potentially slightly lower accuracy. Representative works include the SSD (Single Shot Detector) series (Liu et al., 2016) and the YOLO (You Only Look Once) family (Redmon, 2016; Redmon and Farhadi, 2017; Redmon, 2018; Bochkovskiy et al., 2020; Li et al., 2022; Wang et al., 2023a, 2025, 2024). Two-stage detection models first generate candidate regions and then classify and regress these regions, achieving higher accuracy but at a relatively slower speed. Representative works include the R-CNN family (Girshick et al., 2014; Ren et al., 2016; He et al., 2017). Currently, these methods have been widely applied in underwater object detection tasks using sonar images and have achieved significant results (Heng et al., 2024; Yang et al., 2025; Ma et al., 2024; Yulin et al., 2020; Polap et al., 2022).

Deep learning-based side-scan sonar image enhancement and object detection technologies have achieved significant progress in both theoretical research and practical applications. Burguera et al

(Burguera and Oliver, 2016) employed a probability model-based high-resolution seabed mapping method, correcting sonar data using physical models to generate high-precision images surpassing the device's resolution, laying the foundation for scientific applications. Tang et al. (2023) proposed a deep learning-based real-time object detection method, incorporating lightweight network design to address the challenges of detection efficiency and accuracy in complex underwater terrains. Li et al. (2024) designed an image generation algorithm for zero-shot and few-shot scenarios by combining UA-CycleGAN and StyleGAN3, significantly enhancing the generalization performance of deep learning-based object detection models. Yang et al. (2023) employed diffusion models to generate high-fidelity sonar images and validated the effectiveness of these enhanced data in practical object detection tasks. Zhu et al. (2024) significantly improved the stability and global information extraction capabilities of generative models by introducing CC-WGAN and CBAM modules, while also enhancing the accuracy of object detection. Yang et al. (2024) generated full-category sonar image samples using diffusion models combined with transfer learning, and trained object detection and semantic segmentation models with these samples, significantly improving model performance and data diversity. Aubard et al. (2024) proposed the YOLOX-ViT model, effectively compressing the model size using knowledge distillation while maintaining high detection performance, particularly reducing false alarm rates in underwater environments. Peng et al. (2024) designed a single-image enhancement method based on the CBL-sinGAN network, incorporating CBAM modules and L1 loss functions to enhance the construction capability of small-sample object detection models while preserving sonar image style.

3 Method

This section introduces the proposed SOCA-YOLO model, which integrates the image restoration model SwinIR, the CBAM (Woo et al., 2018) attention mechanism, the SPDConv (Sunkara and Luo, 2022) convolution module, and the YOLOv9 object detection model.

3.1 SwinIR

Image restoration is the process of transforming low-quality images into high-quality versions. SwinIR, a model based on the Swin Transformer, is primarily used for image super-resolution, denoising, and JPEG compression artifact reduction.

SwinIR combines the strengths of both Transformers and CNNs, outperforming traditional CNNs in handling large images due to its local attention mechanism. SwinIR employs a sliding window approach, dividing the input image into multiple small windows and processing each window separately, while retaining the Transformer's ability to manage long-range pixel relationships within the image. As illustrated in Figure 2, SwinIR is designed based on the Swin Transformer and comprises three modules: Shallow Feature Extraction, Deep Feature Extraction, and High-Quality Image Reconstruction.

The Shallow Feature Extraction module extracts initial features through convolutional layers, preserving lowfrequency information and passing it to the reconstruction module. The Deep Feature Extraction module incorporates Residual Swin Transformer Blocks (RSTB), which achieve local attention and cross-window interactions through multiple Swin Transformer layers. Residual connections provide a shortcut for feature aggregation, and convolutional layers further enhance the features. Finally, the High-Quality Image Reconstruction module combines shallow and deep features to produce high-quality images. Each module is detailed below.

Shallow Feature Extraction Module: This module uses a 3×3 convolution to extract shallow features. The main purpose of this process is to retain low-frequency information, leading to better and more stable results. A low-quality imageis $I_L$input at the input stage, and after passing through the shallow feature extraction module H_S, the shallow feature F_0 is obtained as shown in Equation 1:

$$F_0 = H_S(I_L) \tag{1}$$

Deep Feature Extraction Module: This module consists of several RSTBs (Residual Swin Transformer Blocks) and a 3×3 convolution. Each RST is composed of an even number of Swin Transformer Layers (STL) and a convolution layer. This module further processes the shallow features, resulting in its deep feature F_D, as shown in Equation 2.

$$F_D = H_D(F_0) \tag{2}$$

Here, H_D represents the deep feature extraction module.

High-Quality Image Reconstruction: The shallow and deep features are aggregated, transferring both the lowfrequency and high-frequency information of the image to the reconstruction layer. The high-quality image reconstruction module uses a sub-pixel convolution layer to upsample the feature map, resulting in the reconstructed high-quality image I_H, as shown in Equation 3:

$$I_H = H_{RE}(F_0 + F_D) \tag{3}$$

Here, H_{RE} represents the high-quality image reconstruction module.

3.2 CBAM

The Convolutional Block Attention Module (CBAM) is an efficient attention module for feedforward convolutional neural networks, proposed by Sanghyun Woo et al, as illustrated in Figure 3a. CBAM enhances the model's perceptive capability by incorporating a Channel Attention Module (CAM) (Figure 3b) and a Spatial Attention Module (SAM) (Figure 3c) into CNNs, thereby improving performance without adding significant network complexity. As a lightweight and versatile module, CBAM can be seamlessly integrated into any CNN architecture, adding minimal parameters and enabling end-to-end training with YOLOv9 models.

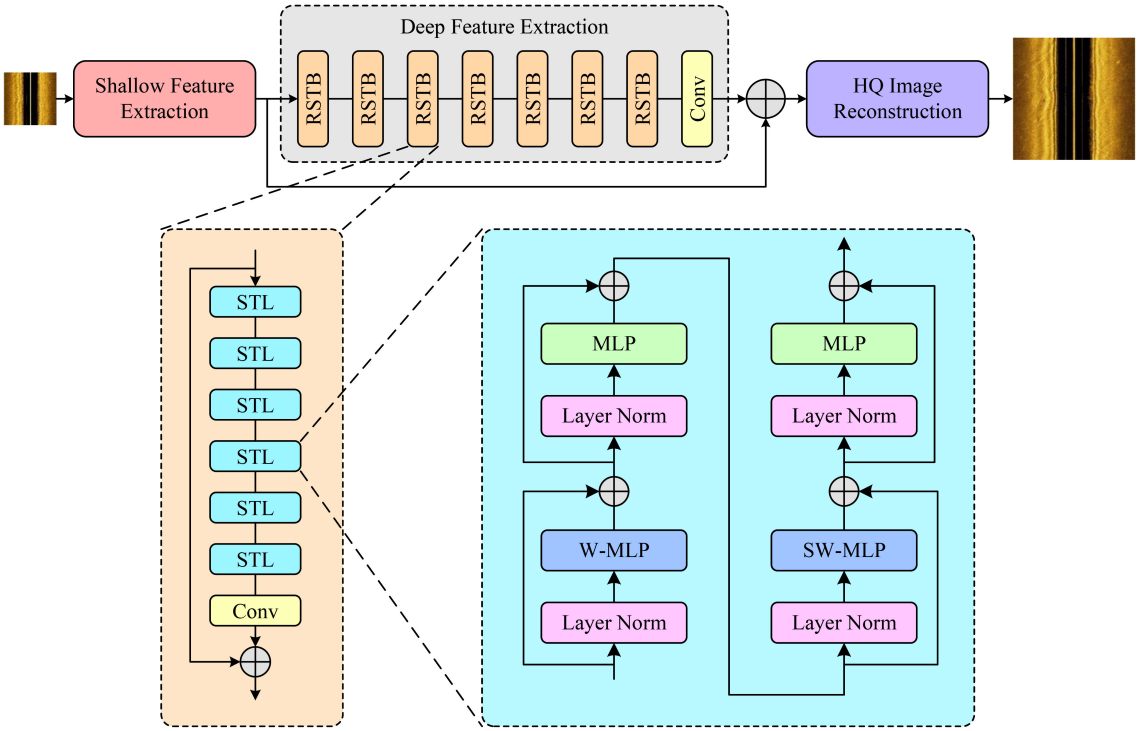


FIGURE 2
SwinIR transformer architecture.

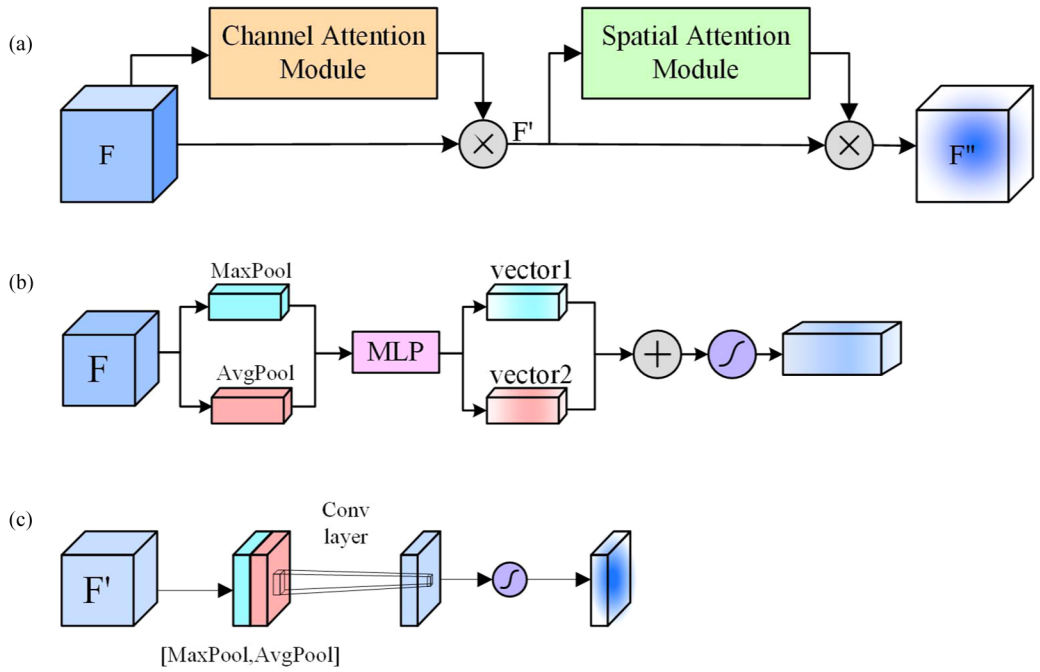


FIGURE 3
(a) Convolutional Block Attention Module (CBAM) architecture, **(b)** Channel Attention Module (CAM) architecture, **(c)** Spatial Attention Module (SAM) architecture.

The input feature map F first passes through the CAM, where the channel weights are multiplied with the input feature map to produce F'. Then, F' is fed into the SAM, where the normalized spatial weights are multiplied with the input feature map of the spatial attention mechanism, resulting in the final weighted feature map F''.

3.3 Space-to-Depth Convolution

The fundamental principle of SPDConv (Space-to-Depth Convolution) is to enhance the performance of convolutional neural networks (CNNs) when processing low-resolution images and small objects, as illustrated in Figure 4. This improvement is achieved through the following key steps:

1. Replacing Strided Convolutions and Pooling Layers: SPDConv is designed to replace traditional strided convolution and pooling layers, which often cause the loss of fine-grained information when dealing with low-resolution images or small objects.
2. Space-to-Depth (SPD) Layer: This transformation layer converts the spatial dimensions of the input image into the depth dimension, increasing the feature map depth without information loss. The SPD layer is critical for retaining spatial information, especially when processing low-resolution images and small objects. By converting spatial information into the depth dimension, the SPD layer mitigates the information loss typically associated with traditional strided convolutions and pooling operations.
3. Non-strided Convolution Layer: A convolutional layer with a stride of 1, applied after the SPD transformation, preserves fine-grained information by avoiding size reduction of the feature map. This non-strided convolution enables feature extraction while maintaining the full resolution of the feature map, which is essential for enhancing recognition performance on low-resolution images and small objects.

SPDConv effectively processes low-resolution images and small objects by combining space-to-depth transformations with non-strided convolutions. This method addresses the fine-grained information loss commonly caused by traditional strided convolutions and pooling layers during downsampling. By preserving spatial information through the SPD layer and converting it into depth features, combined with non-strided convolutions to capture finer details, SPDConv excels in small object detection tasks. It significantly enhances detection accuracy and adaptability to low-resolution images, offering a novel solution for small object detection and related tasks.

3.4 YOLOv9

Proposed in 2024, YOLOv9 is an object detection network that excels in both detection accuracy and processing speed. The model

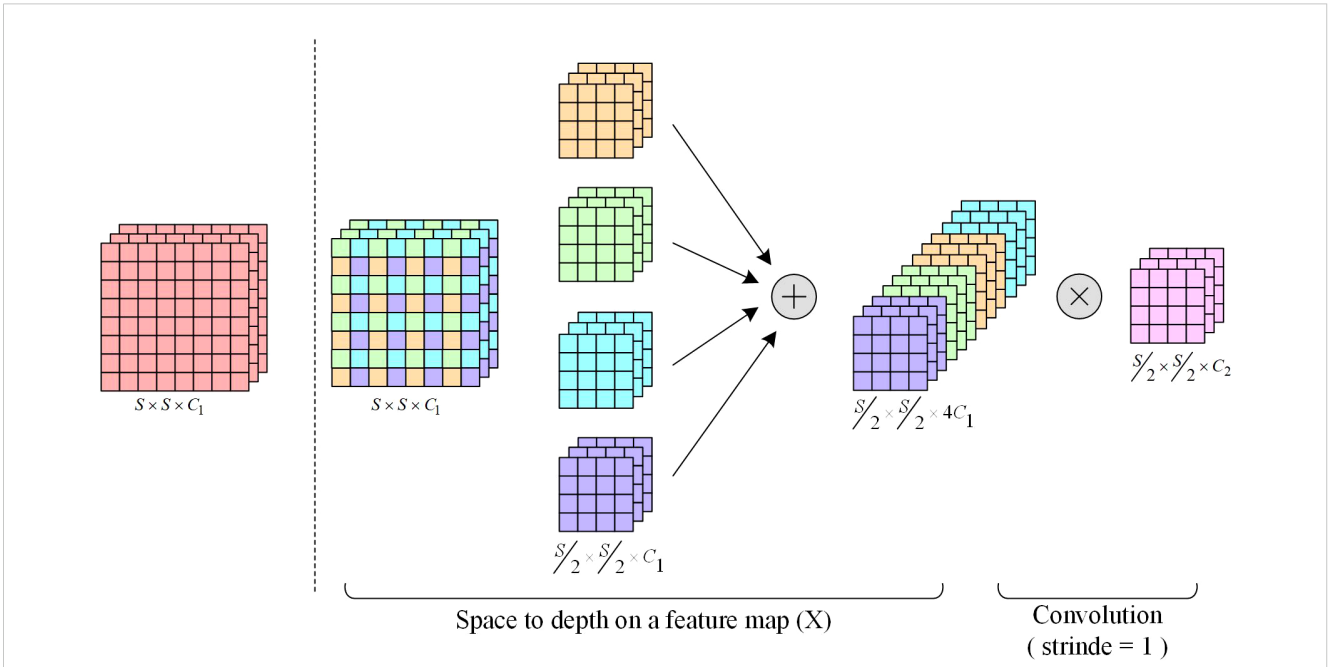


FIGURE 4
SPDConv architecture.

introduces Programmable Gradient Information (PGI), as illustrated in Figure 5. Through auxiliary reversible branches, PGI allows deep features to retain essential object characteristics, enabling the network to preserve crucial visual features of the target without sacrificing important information. This approach enhances YOLOv9's ability to maintain high performance even in complex detection scenarios.

PGI consists of three components: the main branch, multi-level auxiliary information, and the auxiliary reversible branch. Each component is detailed below:

Main Branch: The main branch includes the backbone network, neck network, and head network, which are common components in the YOLO series. The backbone network primarily uses Conv and RepNCSPELAN4 layers for feature extraction. The neck network comprises Upsample, Conv, and RepNCSPELAN4 layers, utilizing an FPN+PAN structure for multi-scale target detection. The head network processes features from the neck network to predict and classify large, medium, and small objects.

Auxiliary Reversible Branch: This branch addresses information loss that occurs as network depth increases, leading to information bottlenecks that hinder reliable gradient generation from the loss function. It introduces an additional network between the feature pyramid layers and the main branch to integrate gradient information from multiple prediction heads.

Multi-level Auxiliary Information: Multi-level auxiliary information involves inserting an integrated network between the feature pyramid's sub-layers and the main branch under auxiliary supervision. This network aggregates gradient information from various prediction heads and passes it to the main branch for parameter updates. Consequently, the feature pyramid in the main branch is not dominated by specific objects, enabling the main branch to retain comprehensive information necessary for learning target features.

3.5 SOCA-YOLO

In this study, we have improved upon the YOLOv9 object detection framework to address challenges such as noise interference, small target size, and edge blurring in side-scan sonar images. Due to the unique imaging mechanism of side-scan sonar, the images often exhibit high noise and low contrast, which can hinder traditional detection models from effectively extracting fine-grained features. Although YOLOv9 demonstrates notable advantages in real-time performance and multi-scale feature fusion, its standard convolutional layers and global feature extraction strategies still exhibit certain limitations when handling such specialized scenarios. Therefore, we propose two main improvements: the introduction of the CBAM attention mechanism into the model and the replacement of some standard convolutional layers with SPDConv modules, thereby achieving more precise feature extraction and fusion for small targets. The modified network model is illustrated in Figure 6.

In our improved model, the overall architecture still adheres to the core design principles of YOLOv9, divided into three components: Backbone, Neck, and Head. However, novel

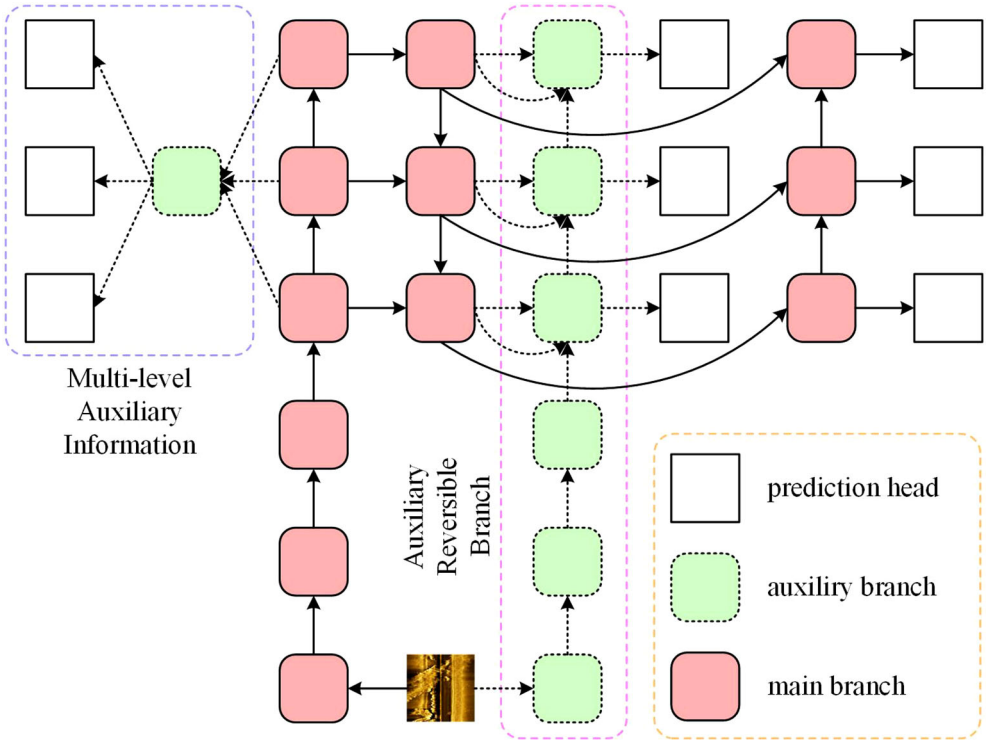


FIGURE 5
YOLOv9 architecture.

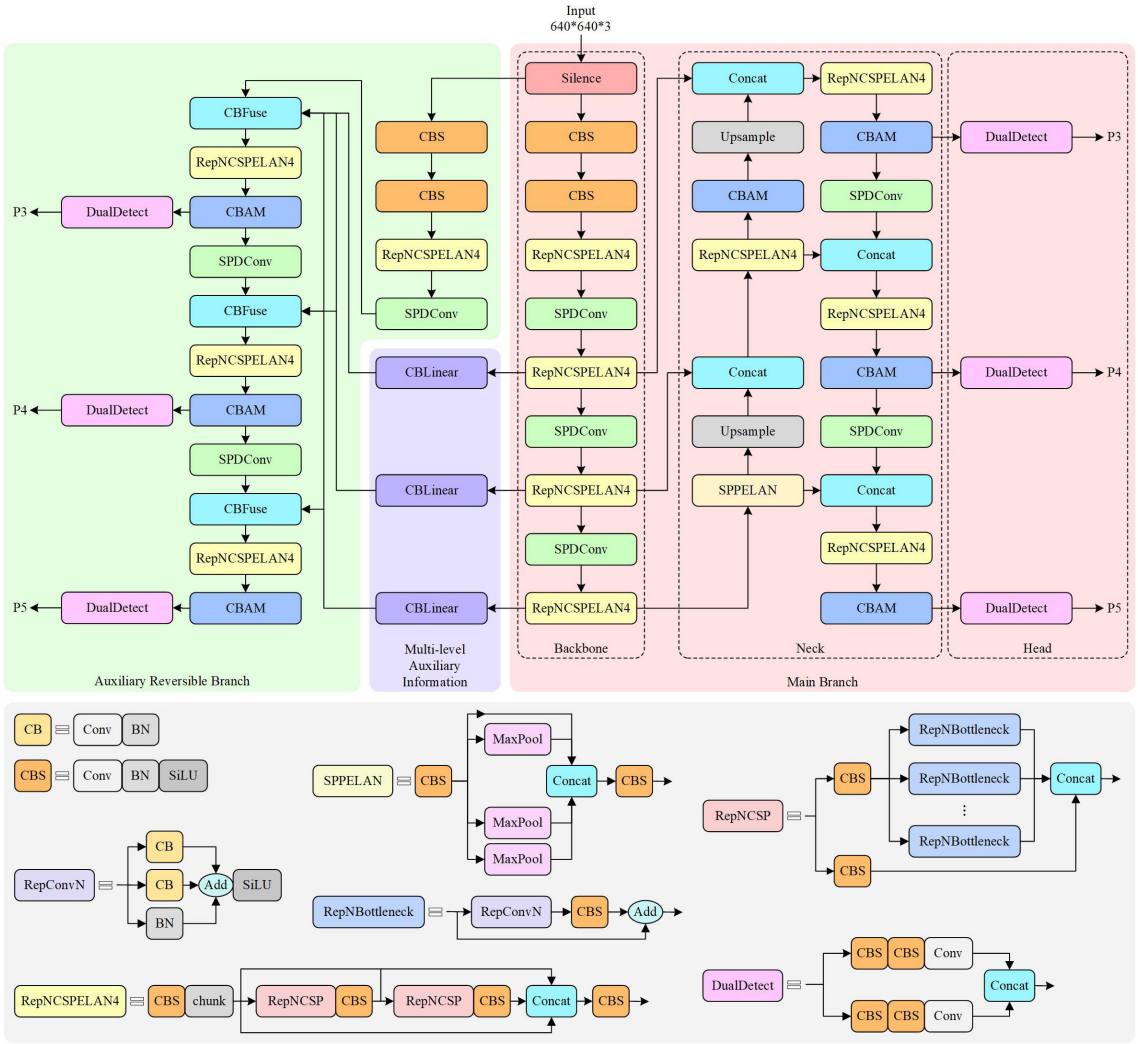


FIGURE 6
SOCA-YOLO architecture.

modules have been strategically incorporated at each stage to adapt to the characteristics of side-scan sonar images. First, the Backbone section integrates SPDConv modules alongside traditional convolutional layers to enhance multi-scale representation capabilities in feature extraction. Specifically, the SPDConv module performs spatial reorganization of input feature maps. This operation can be formally described as follows: let the input feature map be defined in Equation 4.

$$x \in \mathbb{R}^{C\times H\times W} \tag{4}$$

Initially, SPDConv samples x to derive four sub-regions, as shown in Equation Equation 5.

$$\begin{aligned} x_1 &= x[\ldots, :2, :2], \quad x_2 = x[\ldots, 1:2, :2], \\ x_3 &= x[\ldots, :2, 1:2], \quad x_4 = x[\ldots, 1:2, 1:2] \end{aligned} \tag{5}$$

The four sub-features are concatenated in the channel dimension, resulting in a new feature map, as shown in Equation 6.

$$x_{\text{SPD}} = \text{Concat}\{x_1, x_2, x_3, x_4\} \in \mathbb{R}^{4C\times\frac{H}{2}\times\frac{W}{2}}, \tag{6}$$

Subsequently, a 3×3 convolutional layer (denoted as $\text{Conv}_{3\times3}$) is employed for fusion, producing the output features, as shown in Equation 7:

$$y = \text{Conv}_{3\times3}(x_{\text{SPD}})\,. \tag{7}$$

This spatial reorganization and downsampling strategy not only reduces the size of the feature maps and computational load but also effectively captures fine-grained information through channel expansion, offering significant advantages for detecting small, blurry targets in side-scan sonar images.

In the Backbone and some Head modules, we also embed the CBAM to apply dual attention weighting to the features. Specifically, let the input feature be $F \in \mathbb{R}^{C\times H\times W}$, and first, channel statistics are computed through global average pooling and max pooling along the channel dimension, as shown in Equation 8:

$$f_{\text{avg}}(c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F(c,i,j), \quad f_{\max}(c) = \max_{i,j} F(c,i,j). \tag{8}$$

These two sets of statistics are processed through a shared multi-layer perceptron (MLP) and a Sigmoid activation to generate the channel attention vector $M_c \in \mathbb{R}^C$, which is then multiplied with the original features on a per-channel basis to obtain the intermediate feature $F' = M_c \otimes F$. Next, average and max pooling are applied along the channel dimension of F', followed by concatenation, a 7×7 convolution, and Sigmoid activation to generate the spatial attention map $M_s \in \mathbb{R}^{H \times W}$, which is then used to output the spatially weighted feature, as shown in Euqation 9:

$$F_{\text{att}} = M_s \otimes F'. \tag{9}$$

This process allows the network to automatically focus on the target regions, effectively suppress background noise, and further enhance the discriminative ability for small target features.

In the overall architecture, the multi-scale features extracted by the Backbone are strengthened by the SPDConv and CBAM modules and then passed to the Neck section. The Neck employs an FPN and PAN-style multi-scale feature fusion strategy, merging features from different levels in an abstract formulation, as shown in Euqation 10:

$$F_{\text{neck}} = \sum_{i=1}^{N} w_i \cdot f_i(F_{\text{att}}), \tag{10}$$

Here, $f_i(\cdot)$ denotes the feature transformation function for each scale branch, and w_i represents the corresponding weight. This fusion not only retains fine-grained information from each layer but also enriches the global semantics, making it particularly suitable for detecting small targets in side-scan sonar images.

In the Head section, the improved features are processed through a series of modules such as SPPELAN, RepNCSPELAN4, and CBAM, and then further integrated using upsampling and cross-layer concatenation (Concat) to merge multi-scale information. It is worth mentioning that in the subsequent design of the Head, we also introduce multi-level reversible auxiliary branches (through CBLinear and CBFuse modules), which re-fuse features from different levels of the Backbone, providing stronger discriminative signals for final target detection. Finally, after passing through the DualDDetect module, the network outputs detection results containing target categories, bounding box coordinates, and confidence scores, as shown in Equation 11:

$$\hat{Y} = f_{\text{head}}(F_{\text{neck}}), \tag{11}$$

The network is then trained end-to-end using a multi-task loss function, composed of localization loss, classification loss, and confidence loss, as shown in Equation 12:

$$L = \lambda_{\text{loc}} L_{\text{loc}} + \lambda_{\text{cls}} L_{\text{cls}} + \lambda_{\text{conf}} L_{\text{conf}}. \tag{12}$$

This improvement strategy fully integrates the advantages of SPDConv for spatial reorganization and downsampling, CBAM's dual attention weighting ability for features, and the overall design of multi-scale fusion. It significantly enhances the model's detection performance for small targets in side-scan sonar images, while balancing real-time processing and efficiency, providing a solid theoretical and technical foundation for future practical deployment.

During model training, the original images are first uniformly resized to a standard dimension of 640×640×3. This standardization ensures consistency in input data. Subsequently, the images undergo a series of convolution and pooling operations, through which the network generates feature maps of varying scales. Shallow feature maps retain finer details for detecting small targets, while deep feature maps capture global information for large target detection. This multi-scale feature extraction mechanism effectively enhances the network's capability to detect targets of varying sizes.

4 Experiments and analysis

To validate that our SOCA-YOLO network achieves superior results on public side-scan sonar images compared to other methods, we designed the following experiments. First, we applied SwinIR to preprocess the original dataset, generating a high-resolution dataset. We then compared various object detection models, demonstrating that our network exhibits a certain level of superiority. Additionally, we conducted comparative experiments using different convolution modules and attention mechanisms to verify the effectiveness of the SPDConv module and the CBAM attention mechanism. Finally, ablation experiments confirmed that each of our proposed improvements contributes positively to the overall performance.

4.1 Environment and dataset

To comprehensively assess the effectiveness of the proposed method, we conduct experiments in a high-performance computing environment and evaluate the model on a publicly available side-scan sonar image dataset. This section provides a detailed description of the experimental setup and dataset used in our study.

4.1.1 Environment

To ensure the reproducibility of experiments and the efficiency of computational performance, the experimental environment in this study is built on the mainstream deep learning framework PyTorch, fully meeting the computational requirements for model training and inference. Detailed configuration information is presented in Table 1.

4.1.2 Dataset

The experimental dataset used in this paper is the publicly available Cylinder2 ([Dataset] yeesonmin@naver.com, 2022), utilized to evaluate the model's performance. Released in 2022, this dataset contains 478 side-scan sonar images categorized into two classes: cylinders and manta rays, with each image containing exactly one object. Each object occupies a relatively small pixel area compared to the full image, making this dataset suitable for

TABLE 1 System configuration.

Name	Configuration
Python	3.9.18
PyTorch	1.12.0
CUDA	11.3
CPU	Intel(R) Core(TM) i5-13600KF@3.50GHz
GPU	NVIDIA GeForce RTX 4070Ti (12GB)

underwater small object detection tasks. We excluded the portion containing manta rays (140 images), retaining only the 338 cylinder images. The dataset was subsequently split into training, validation, and test sets in an 8:1:1 ratio, which was then used to train the network. The basic configuration of the dataset is shown in Table 2.

4.2 Evaluation metrics

During the image restoration stage using SwinIR, the image quality was evaluated using standard metrics, including Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM).

PSNR: Given a clean image and a noisy image of size m×n, the Mean Squared Error (MSE) is defined, as shown in Equation 13:

$$\mathrm{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \tag{13}$$

At this point, PSNR is defined as shown in Equation 14:

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{\mathrm{MAX}_I^2}{\mathrm{MSE}}\right) \tag{14}$$

Here, MAX_I^2 represents the maximum possible pixel value in the image. If each pixel is represented by 8-bit binary, then the maximum value is 255. Typically, if the pixel value is represented in B-bit binary, then $\mathrm{MAX}_I = 2^B - 1$.

SSIM: The SSIM formula is based on three comparison measures between samples x and y: luminance (Equation 15), contrast (Equation 16), and structure (Equation 17).

$$l(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \tag{15}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \tag{16}$$

TABLE 2 Dataset split settings.

Dataset	Images
Train	270
Val	34
Test	34

$$s(x,y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3} \tag{17}$$

Typically, $c_3 = \frac{c_2}{2}$, where μ_x represents the mean of x, μ_y represents the mean of y, σ_x^2 is the variance of x, σ_y^2 is the variance of y, and σ_{xy} is the covariance between x and y. Thus SSIM can be expressed, as shown in Equation 18:

$$\mathrm{SSIM}(x,y) = [l(x,y)^\alpha \cdot c(x,y)^\beta \cdot s(x,y)^\gamma] \tag{18}$$

Setting, $\alpha = \beta = \gamma = 1$ we obtain Equation 19:

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{19}$$

During the training and testing phases, the model's performance is evaluated according to the PASCAL VOC 2010 standard, which includes Precision (P), Recall (R), and Mean Average Precision (mAP). P represents the proportion of samples correctly predicted as positive out of all samples predicted as positive by the model. R represents the proportion of correctly predicted positive samples out of all true positive samples. mAP is used to comprehensively assess the model's performance across all categories by calculating the average precision at various recall thresholds. Since this paper focuses on detecting a single target type, the AP value is equivalent to the mAP value. Ideally, a higher mAP value indicates better model performance. The formulas for calculating P, R, and mAP are provided in equations Equations 20–23.

$$P = \frac{TP}{TP + FP} \tag{20}$$

$$R = \frac{TP}{TP + FN} \tag{21}$$

$$AP = \int_0^1 P(R)\,dR \tag{22}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{23}$$

Here, TP represents true positives, where positive samples are correctly predicted as positive; FP represents false positives, where negative samples are incorrectly predicted as positive; and FN represents false negatives, where positive samples are incorrectly predicted as negative.

4.3 Experimental results

To validate the effectiveness of the proposed method, we conduct a series of comparative experiments. First, we apply SwinIR for image restoration and analyze its impact on the quality of side-scan sonar images. Then, we perform multiple comparative studies, including object detection model comparison, attention mechanism comparison, and convolution module comparison. These experiments provide a comprehensive evaluation of the contributions of different components to the overall detection performance.

4.3.1 Using SwinIR for image processing

In this paper, we employ the SwinIR model as a preprocessing step to enhance the quality of original side-scan sonar images. The enhanced images are subsequently used to train and validate the SOCA-YOLO model, which is designed for small object detection. Pretrained weights from the official SwinIR GitHub repository (Liang et al., 2021) are utilized to leverage the architecture's robust super-resolution capabilities. The application of SwinIR results in processed side-scan sonar images with sharper edges, reduced noise, and improved fine details—key factors for accurate detection. Figure 7 presents comparative examples of the original and enhanced images, illustrating the effectiveness of this preprocessing step.

To intuitively assess the effectiveness of SwinIR in enhancing image clarity, we used PSNR and SSIM to compare the experimental results. The findings indicate that, compared to the original images, the processed images achieved average PSNR and SSIM values of 36.14 and 0.9807, respectively. These results demonstrate that SwinIR not only improves the visual quality and resolution of the images but also yields higher PSNR and SSIM values. Consequently, this enhancement facilitates more accurate detection of small objects, with notable improvements across various detection metrics.

4.3.2 Comparative experiment

1. Comparison of SOCA-YOLO with mainstream object detection networks.

To verify the performance of this method, we conducted comparative experiments with several mainstream object detection models, including SSD, Faster R-CNN, and various YOLO series models. Table 3 presents the experimental results of each model on the side-scan sonar dataset.

As shown in Table 3, the proposed method outperforms the original YOLOv9 and other object detection algorithms across multiple metrics. Specifically, compared to the original YOLOv9, P increases by 4.2%, R by 7.2%, and mAP50 by 6.5%. In comparison with SSD, Faster R-CNN, and the latest YOLO models, the proposed algorithm demonstrates superior performance in metrics such as P, R, and mAP. Although YOLO11 achieves a higher P of 75.8% compared to SOCA-YOLO's 71.8%, SOCA-YOLO surpasses YOLO11 in both recall and mAP50, highlighting its balanced and robust detection capabilities.

These results indicate that the algorithm significantly enhances the detection capability for small underwater targets. Figure 8 displays sample results of SOCA-YOLO target detection, illustrating noticeable improvements in both detection metrics and practical detection outcomes. However, some instances of missed and false detections remain in the detection process.

Furthermore, to provide a more comprehensive comparison of our model's superiority, we also compared the P-R curves. Figure 9 presents the P-R curve of the original YOLOv9 and the P-R curve of SOCA-YOLO.

In summary, for small object detection in underwater side-scan sonar images, the proposed method significantly outperforms mainstream object detection algorithms. Figure 10 compares the detection results of SOCA-YOLO with other models for the same target. As shown in the Figure 10, while other models produce false positives and missed detections, SOCA-YOLO accurately identifies the target, demonstrating its robustness and precision.

2. Comparison of SPDConv with other convolutional methods.

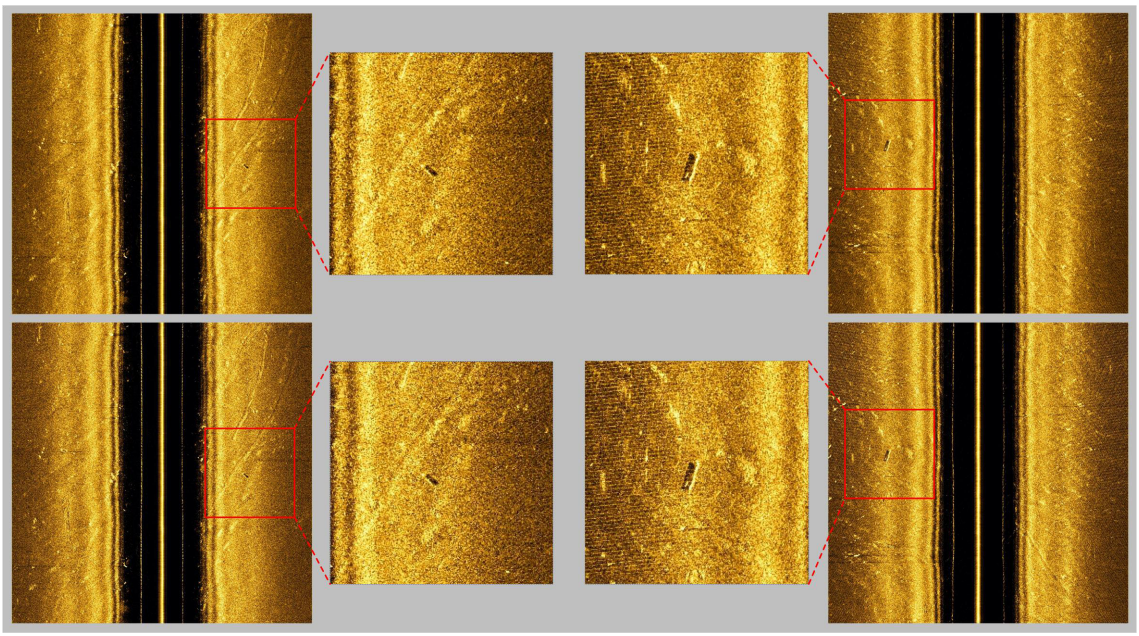


FIGURE 7
Partial results of SwinIR preprocessing, with the first row showing the original images, the second row showing the restored images, and the red boxes indicating a zoomed-in view of the target region.

TABLE 3 Comparison of SOCA-YOLO with mainstream object detection networks.

Methods	Precision / %	Recall / %	mAP50 / %
SSD	48.6	51.5	44.8
Faster-RCNN	42.4	52.9	45.5
YOLOv9	42.4	52.9	45.5
YOLOv10	70.6	65.3	71.4
YOLO11	75.8	66.7	72.0
SOCA-YOLO	71.8	72.7	74.3

To verify the contribution of the introduced convolution module SPDConv to our model's improvements, we replaced the original YOLOv9 convolution module ADown with AConv, AKConv, and SPDConv, respectively. ADown is the default convolution module in YOLOv9; AConv is a standard convolution module consisting of a pooling layer and a convolution layer; AKConv (Zhang et al., 2023) is a variable kernel convolution module that allows the kernel to dynamically adjust its shape and size based on target characteristics; SPDConv is the proposed convolution module in our SOCA-YOLO network, designed for superior detection capability on low-resolution images and small objects. We tested each module replacement on side-scan sonar images without SwinIR preprocessing. The experimental results are shown in Table 4.

As show in Table 4, SPDConv demonstrates significant advantages in object detection tasks, outperforming other convolutional modules across all key metrics. Specifically, SPDConv achieves a P of 70.4%, a R of 71%, and a mAP50 of 72.6%. These results represent improvements of 2.8%, 5.5%, and 4.8%, respectively, compared to those obtained using the original YOLOv9 convolution module, ADown. Compared to traditional convolutional modules, these improvements are particularly important for enhancing the overall performance of the YOLOv9 network. SPDConv not only improves precision but also significantly enhances the network's detection consistency (i.e., the balanced performance of P and R), making it especially suitable for small object detection in side-scan sonar images.

3. Comparison of CBAM with other attention mechanisms.

To validate the effectiveness of the attention mechanism in our network model, we conducted comparative experiments incorporating various popular attention modules, including the SE module (Hu et al., 2018), CA module (Hou et al., 2021), ECA module (Wang et al., 2020), CBAM module, and the baseline YOLOv9 network without any attention mechanism. Each attention module was integrated into the same position within the YOLOv9 network to ensure the comparability of results. Consistent training and validation datasets were used throughout the experiments to maintain fairness. The experimental results are presented in Table 5.

The results demonstrate that the performance improvements provided by attention mechanisms depend on the specific module design. Among these, CBAM achieved the best performance, significantly enhancing both detection P and R. This outcome highlights the effectiveness of CBAM's dual-branch design in capturing feature correlations at multiple levels, thereby

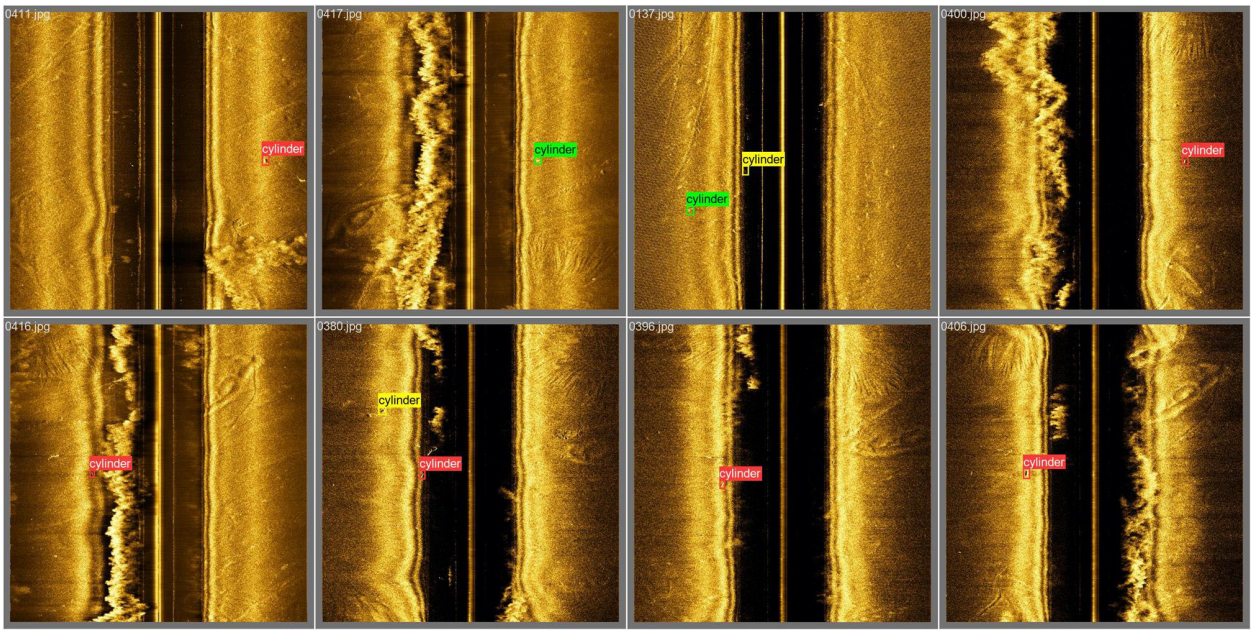


FIGURE 8
Partial results of SOCA-YOLO detection, with red boxes representing correctly detected targets, yellow boxes representing false detections, and green boxes representing missed detections.

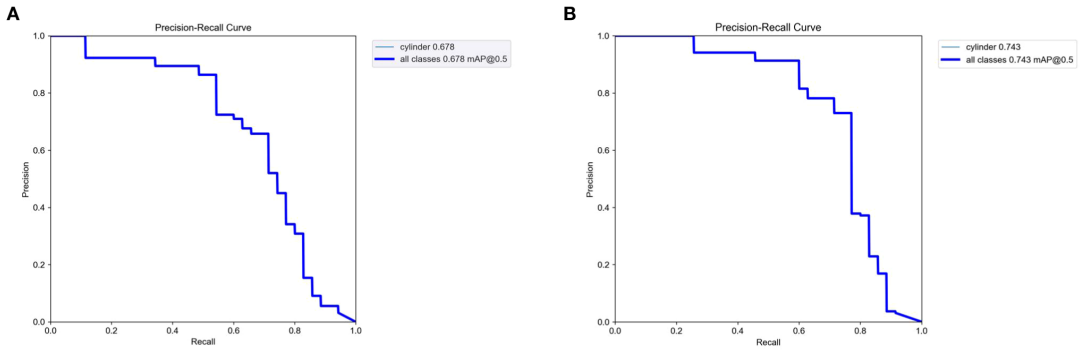


FIGURE 9
P-R curves. **(A)** shows the P-R curve of the original YOLOv9. **(B)** shows the P-R curve of SOCA-YOLO.

improving the model's ability to locate and classify targets. In comparison, the SE module, which focuses on channel attention, shows notable classification improvements in specific scenarios but offers relatively limited gains in complex environments. The CA module, by incorporating coordinate information, improves the locality of feature representations and performs well in scenarios with targets of varying aspect ratios. The ECA module strikes a balance by reducing the computational cost of attention but delivers limited improvements in small object detection.

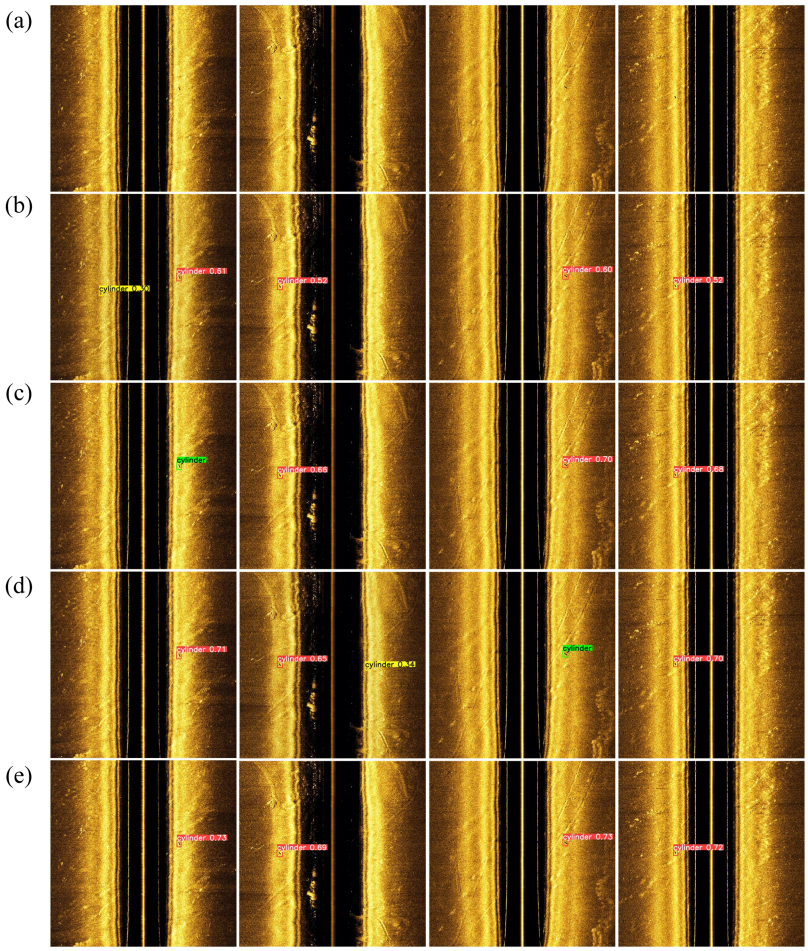


FIGURE 10
Detection results from different object detection models for four targets, with each image containing exactly one target: **(a)** Original image, **(b)** Faster-RCNN, **(c)** YOLOv9, **(d)** YOLOv10, **(e)** SOCA-YOLO. Red boxes indicate correctly detected targets, yellow boxes indicate false detections, and green boxes indicate missed detections.

TABLE 4 Comparison of SPDConv with other convolutional methods.

Model	Precision / %	Recall / %	mAP50 / %
ADwon	67.6	65.5	67.8
AConv	66.7	65.7	67.8
AKConv	68.1	69.4	69.1
SPDConv	70.4	71.0	72.6

Table 5 shows that the CBAM module achieved the best performance, with a P of 69.9%, R of 72.2%, and mAP50 of 73.3%. These values represent improvements of 2%, 6.7%, and 5.5%, respectively, compared to the baseline YOLOv9 network. However, the results also indicate that while certain attention modules provide performance enhancements, not all attention mechanisms positively impact object detection tasks. The selection and design of attention modules should be carefully adjusted and optimized to align with the specific characteristics of the task.

4.4 Ablation study

To evaluate the impact of each proposed innovation on network performance, we conducted ablation experiments on different modules. This study primarily examines the effects of using SwinIR for preprocessing the original images, replacing the original YOLOv9 convolution module with SPDConv, and adding the CBAM attention mechanism. These three enhancements were gradually incorporated into the YOLOv9 network. The experiments were conducted on the side-scan sonar image dataset, and the experimental outcomes are presented in Table 6.

TABLE 5 Comparison of CBAM with other attention mechanisms.

Model	Precision / %	Recall / %	mAP50 / %
YOLOv9	67.6	65.5	67.8
YOLOv9+SE	65.6	67.4	67.2
YOLOv9+CA	66.4	70.7	68.8
YOLOv9+ECA	69.5	70.3	66.3
YOLOv9+CBAM	69.9	72.2	73.3

As shown in Table 6, preprocessing the original dataset using SwinIR and applying the resulting high-quality images for SOCA-YOLO training and testing increased the mAP50 by 0.3%. Replacing the convolution module in the original YOLOv9 network resulted in a 5.6% increase in mAP50 compared to the original YOLOv9 results. Finally, adding the CBAM module to the YOLOv9 network with the replaced convolution module further increased the mAP50 by 0.6%. These experimental results demonstrate that each improvement is meaningful. Compared to the original network, the cumulative mAP50 increase of 6.5% significantly reduces missed detections and false detections of small objects in the original YOLOv9 network.

4.5 Generalization experiment

To validate the generalization capability of the object detection method proposed in this paper under different data distributions, we selected another publicly available side-scan sonar image dataset as the test platform (Santos et al., 2024). This dataset differs significantly from the data used during training, with marked variations in the capture environment, target types, and noise interference, thereby thoroughly assessing the model's adaptability and robustness in new scenarios. The dataset primarily comprises 1170 high-resolution side-scan sonar images and includes two types of targets—NOn-Mine-like BOttom Objects (NOMBO) and MIne-Like COntacts (MILCO)—with varying sizes and shapes. The experimental results are presented in Table 7. It can be seen that the method proposed in this paper outperforms traditional detection approaches across evaluation metrics, demonstrating strong generalization ability.

Additionally, to further analyze the detection performance across different target categories, the P-R curves for each category were plotted, as shown in Figure 11.

From the above experimental results, it is evident that the proposed method effectively adapts to noise and interference issues in public side-scan sonar image data across different marine environments, achieving high detection accuracy and recall.

5 Conclusions

In this paper, we introduced the object detection algorithm YOLOv9 with several modifications. The specific improvements are as follows: (1) Using the SwinIR model to preprocess the original

TABLE 6 Ablation study.

YOLOv9	SwinIR	SPDConv	CBAM	Precision / %	Recall / %	mAP50 / %
✓	×	×	×	67.6	65.5	67.8
✓	✓	×	×	73.5	63.4	68.1
✓	✓	✓	×	69.6	71.6	73.7
✓	✓	✓	✓	71.8	72.7	74.3

The symbol "✓" indicates that the condition was included in the experiment, while "×" signifies that the condition was not incorporated into the experimental setup.

TABLE 7 Performance of YOLO9 and SOCA-YOLO.

Method	Precision / %	Recall / %	mAP50 / %
YOLO9	82.1	65.3	74.3
SOCA-YOLO	93.7	76.2	83.7

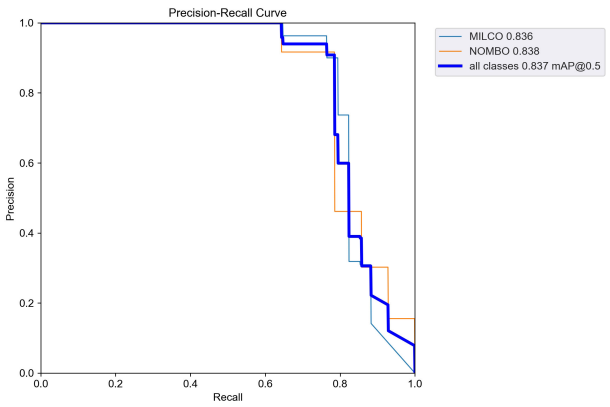


FIGURE 11
Performance of different target categories.

dataset and generate a re-divided high-resolution image dataset. (2) Adding the CBAM attention mechanism to the original YOLOv9 model to enhance focus on regions of interest. (3) Replacing the original ADown module with the SPDConv convolution module, which is more effective for small object detection. The resulting SOCA-YOLO model was applied for small object detection in underwater side-scan sonar images, achieving a Precision of 71.8%, Recall of 72.7%, and mAP50 of 74.3% on the enhanced dataset. These results indicate that the method significantly improves target detection performance in side-scan sonar images.

In future work, expanding the dataset is a crucial research direction. Although the current dataset has demonstrated the feasibility of our method, its limited scope may constrain the model's robustness and generalization ability. By incorporating additional datasets from different environments, operational conditions, and various sonar devices, we can capture a broader range of image features and noise characteristics. Such dataset expansion not only enables more comprehensive model training but also allows fine-tuning and validation of the model across various real-world scenarios. Furthermore, given the inherent unique noise characteristics of side-scan sonar images, developing specialized image processing techniques is particularly crucial. Future research can focus on designing denoising and image enhancement algorithms tailored to issues such as speckle noise and signal interference in sonar data. Exploring the integration of multimodal data is also a highly promising direction. For example, combining side-scan sonar data with optical or hyperspectral imaging data can provide complementary information, thereby improving the overall performance of detection and classification tasks. Such data fusion is expected to lead to the development of more robust and accurate models, ultimately driving new methodologies and applications in underwater imaging and analysis.

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

XC: Conceptualization, Methodology, Software, Validation, Writing – review & editing, Project administration, Resources, Supervision. JZ: Validation, Writing – review & editing, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft. LZ: Resources, Validation, Writing – review & editing, Supervision. QZ: Project administration, Validation, Writing – review & editing. JH: Resources, Validation, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This project was supported by the National Natural Science Foundation of China (Grant No.62271397 and Grant No.62171384).

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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

[Dataset] yeesonmin@naver.com (2022). *cylider2 dataset*. Available online at: https://universe.roboflow.com/yeesonmin-naver-com/cylider2 (Accessed December 2, 2024).

何勇光 (2020). 海洋侧扫声呐探测技术的现状及发展[J]. 工程建设与设计. 2020 (04), 275–276. doi: 10.13616/j.cnki.gcjsysj.2020.02.328

Ali, A. M., Benjdira, B., Koubaa, A., Boulila, W., and El-Shafai, W. (2023). Tesr: two-stage approach for enhancement and super-resolution of remote sensing images. *Remote Sens.* 15, 2346. doi: 10.3390/rs15092346

Aubard, M., Antal, L., Madureira, A., and Ábrahám, E. (2024). Knowledge distillation in yolox-vit for side-scan sonar object detection. *arXiv preprint arXiv:2403.09313*. doi: 10.48550/arXiv.2403.09313

Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*. doi: 10.48550/arXiv.2004.10934

Burguera, A., and Oliver, G. (2016). High-resolution underwater mapping using side-scan sonar. *PloS One* 11, e0146396. doi: 10.1371/journal.pone.0146396

Cheng, C., Hou, X., Wen, X., Liu, W., and Zhang, F. (2023). Small-sample underwater target detection: a joint approach utilizing diffusion and yolov7 model. *Remote Sens.* 15, 4772. doi: 10.3390/rs15194772

Du, X., Sun, Y., Song, Y., Sun, H., and Yang, L. (2023). A comparative study of different cnn models and transfer learning effect for underwater object classification in side-scan sonar images. *Remote Sens.* 15, 593. doi: 10.3390/rs15030593

Fan, X., Lu, L., Shi, P., and Zhang, X. (2022). A novel sonar target detection and classification algorithm. *Multimedia Tools Appl.* 81, 10091–10106. doi: 10.1007/s11042-022-12054-4

Fayaz, S., Parah, S. A., and Qureshi, G. (2022). Underwater object detection: architectures and algorithms–a comprehensive review. *Multimedia Tools Appl.* 81, 20871–20916. doi: 10.1007/s11042-022-12502-1

Fengchun, L., Dianlun, Z., and Haitao, G. (2002). Image segmentation based upon bounded histogram and its application to sonar image segmentation. *J. Harbin Eng. Univ.* 2002, 1–3.

Gao, L., Zhang, J., Yang, C., and Zhou, Y. (2022). Cas-vswin transformer: A variant swin transformer for surface-defect detection. *Comput. Industry* 140, 103689. doi: 10.1016/j.compind.2022.103689

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA. 2014, 580–587. doi: 10.1109/CVPR.2014.81

Gomes, D., Saif, A. S., and Nandi, D. (2020). "Robust underwater object detection with autonomous underwater vehicle: A comprehensive study," in *Proceedings of the International Conference on Computing Advancements*, Dhaka, Bangladesh. (New York, NY, USA: Association for Computing Machinery), 1–10. doi: 10.1145/3377049.3377052

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask R-CNN," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2980–2988. doi: 10.1109/ICCV.2017.322

He, X., Zhou, Y., Zhao, J., Zhang, D., Yao, R., and Xue, Y. (2022). Swin transformer embedding unet for remote sensing image semantic segmentation. *IEEE Trans. Geosci. Remote Sens.* 60, 1–15. doi: 10.1109/TGRS.2022.3230846

Heng, Z., Shuping, H., Jiaying, G., Yubo, H., and Honggang, L. (2024). "Research on the automatic detection method of side-scan sonar image of small underwater target," in *2024 IEEE 7th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, Chongqing, China, 1567–1573. doi: 10.1109/ITNEC60942.2024.10733067

Hou, Q., Zhou, D., and Feng, J. (2021). "Coordinate attention for efficient mobile network design," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, 13708–13717. doi: 10.1109/CVPR46437.2021.01350

Hu, J., Shen, L., and Sun, G. (2018). "Squeeze-and-excitation networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1 Aug. 2020, 42, pp. 2011–2023. doi: 10.1109/TPAMI.2019.2913372

Jannat, F.-E., and Willis, A. R. (2022). Improving classification of remotely sensed images with the swin transformer. *SoutheastCon* 2022, 611–618. doi: 10.1109/SoutheastCon48659.2022.9764016

Jiang, Y., Ku, B., Kim, W., and Ko, H. (2020). Side-scan sonar image synthesis based on generative adversarial network for images in multiple frequencies. *IEEE Geosci. Remote Sens. Lett.* 18, 1505–1509. doi: 10.1109/LGRS.2020.3005679

Jinhua, L., Jinpeng, J., and Peimin, Z. (2016). Automatic extraction of the side-scan sonar imagery outlines based on mathematical morphology. 海洋学报 38, 150–157. doi: 10.3969/j.issn.0253-4193.2016.05.014

Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., et al. (2022). Yolov6: A single-stage object detection framework for industrial applications. *arXiv preprint arXiv:2209.02976*. doi: 10.48550/arXiv.2209.02976

Li, L., Li, Y., Wang, H., Yue, C., Gao, P., Wang, Y., et al. (2024). Side-scan sonar image generation under zero and few samples for underwater target detection. *Remote Sens.* 16, 4134. doi: 10.3390/rs16224134

Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., and Timofte, R. (2021). "Swinir: Image restoration using swin transformer," in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Montreal, BC, Canada, 2021, pp. 1833–1844. doi: 10.1109/ICCVW54120.2021.00210

Lin, H. (2023). Adversarial training of swinir model for face super-resolution processing. *Front. Computing Intelligent Syst.* 5, 87–90. doi: 10.54097/fcis.v5i1.11846

Lin, A., Chen, B., Xu, J., Zhang, Z., Lu, G., and Zhang, D. (2022). Ds-transunet: Dual swin transformer u-net for medical image segmentation. *IEEE Trans. Instrumentation Measurement* 71, 1–15. doi: 10.1109/TIM.2022.3178991

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). SSD: Single Shot MultiBox Detector. In: B. Leibe, J. Matas, N. Sebe and M. Welling (eds) *Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science*, 9905. Springer, Cham. doi: 10.1007/978-3-319-46448-0_2

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021). "Swin transformer: Hierarchical vision transformer using shifted windows," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, QC, Canada., 9992–10002. doi: 10.1109/ICCV48922.2021.00986

Ma, Q., Jin, S., Bian, G., and Cui, Y. (2024). Multi-scale marine object detection in side-scan sonar images based on bes-yolo. *Sensors* 24, 4428. doi: 10.3390/s24144428

Peng, C., Jin, S., Bian, G., Cui, Y., and Wang, M. (2024). Sample augmentation method for side-scan sonar underwater target images based on cbl-singan. *J. Mar. Sci. Eng.* 12, 467. doi: 10.3390/jmse12030467

Polap, D., Wawrzyniak, N., and Włodarczyk-Sielicka, M. (2022). Side-scan sonar analysis using roi analysis and deep neural networks. *IEEE Trans. Geosci. Remote Sens.* 60, 1–8. doi: 10.1109/TGRS.2022.3147367

Redmon, J. (2016). "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 779–788. doi: 10.1109/CVPR.2016.91

Redmon, J. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*. doi: 10.48550/arXiv.1804.02767

Redmon, J., and Farhadi, A. (2017). "Yolo9000: better, faster, stronger," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 6517–6525. doi: 10.1109/CVPR.2017.690

Ren, S., He, K., Girshick, R., and Sun, J. (2016). 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

Santos, N. P., Moura, R., Torgal, G. S., Lobo, V., and de Castro Neto, M. (2024). Side-scan sonar imaging data of underwater vehicles for mine detection. *Data Brief* 53, 110132. doi: 10.1016/j.dib.2024.110132

Sun, C., Wang, L., Wang, N., and Jin, S. (2021). Image recognition technology in texture identification of marine sediment sonar image. *Complexity* 2021, 6646187. doi: 10.1155/2021/6646187

Sunkara, R., and Luo, T. (2022). No more strided convolutions or pooling: A new cnn building block for low-resolution images and small objects. In: M. R. Amini, S. Canu, A. Fischer, T. Guns, P. Kralj Novak and G. Tsoumakas (eds) *Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2022. Lecture Notes in Computer Science*. (Cham: Springer). 13715. doi: 10.1007/978-3-031-26409-2_27

Tang, Y., Wang, L., Jin, S., Zhao, J., Huang, C., and Yu, Y. (2023). Auv-based side-scan sonar real-time method for underwater-target detection. *J. Mar. Sci. Eng.* 11, 690. doi: 10.3390/jmse11040690

Tian, X., Liu, Z., and Zhou, D. (2007). Mine target recognition algorithm of sonar image. *Sys. Eng. Electron* 7, 1049–1052.

Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M. (2023a). "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, 7464–7475. doi: 10.1109/CVPR52729.2023.00721

Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., et al. (2024). Yolov10: Real-time end-to-end object detection. *Advances in Neural Information Processing Systems* 37, 107984–108011.

Wang, Y.-R., Wang, P., Adams, L. C., Sheybani, N. D., Qu, L., Sarrami, A. H., et al. (2023b). Low-count whole-body pet/mri restoration: an evaluation of dose reduction spectrum and five state-of-the-art artificial intelligence models. *Eur. J. Nucl. Med. Mol. Imaging* 50, 1337–1350. doi: 10.1007/s00259-022-06097-w

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., 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)*, Seattle, WA, USA, 11534–11542. doi: 10.1109/CVPR42600.2020.01155

Wang, C.-Y., Yeh, I.-H., and Mark Liao, H.-Y. (2025). Yolov9: Learning what you want to learn using programmable gradient information. In: A. Leonardis, E. Ricci, S.

Roth, O. Russakovsky, T. Sattler and G. Varol (eds) Computer Vision – ECCV 2024. ECCV 2024. Lecture Notes in Computer Science. (Cham: Springer), 15089. doi: 10.1007/978-3-031-72751-1_1

Wen, X., Zhang, F., Cheng, C., Hou, X., and Pan, G. (2024). Side-scan sonar underwater target detection: Combining the diffusion model with an improved yolov7 model. *IEEE J. Oceanic. Eng.* 49, 976–991. doi: 10.1109/JOE.2024.3379481

Woo, S., Park, J., Lee, J.-Y., and Kweon, I. S. (2018). CBAM: Convolutional Block Attention Module. In: V. Ferrari, M. Hebert, C. Sminchisescu and Y. Weiss (eds) *Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science*, (Cham: Springer). vol 11211. doi: 10.1007/978-3-030-01234-2_1

Yang, N., Li, G., Wang, S., Wei, Z., Ren, H., Zhang, X., et al. (2025). Ss-yolo: A lightweight deep learning model focused on side-scan sonar target detection. *J. Mar. Sci. Eng.* 13, 66. doi: 10.3390/jmse13010066

Yang, Z., Zhao, J., Yu, Y., and Huang, C. (2024). A sample augmentation method for side-scan sonar full-class images that can be used for detection and segmentation. *IEEE Trans. Geosci. Remote Sens.* 62, 1–11. doi: 10.1109/TGRS.2024.3371051

Yang, Z., Zhao, J., Zhang, H., Yu, Y., and Huang, C. (2023). A side-scan sonar image synthesis method based on a diffusion model. *J. Mar. Sci. Eng.* 11, 1103. doi: 10.3390/jmse11061103

Yasir, M., Liu, S., Pirasteh, S., Xu, M., Sheng, H., Wan, J., et al. (2024). Yoloshiptracker: Tracking ships in sar images using lightweight yolov8. *Int. J. Appl. Earth Observation Geoinformation* 134, 104137. doi: 10.1016/j.jag.2024.104137

Yu, Y., Zhao, J., Gong, Q., Huang, C., Zheng, G., and Ma, J. (2021). Real-time underwater maritime object detection in side-scan sonar images based on transformer-yolov5. *Remote Sens.* 13, 3555. doi: 10.3390/rs13183555

Yulin, T., Shaohua, J., Gang, B., Yonzhou, Z., and Fan, L. (2020). "Wreckage target recognition in side-scan sonar images based on an improved faster r-cnn model," in *2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)*, Bangkok, Thailand, 348–354. doi: 10.1109/ICBASE51474.2020.00080

Zhang, X., Song, Y., Song, T., Yang, D., Ye, Y., Zhou, J., et al. (2023). Akconv: Convolutional kernel with arbitrary sampled shapes and arbitrary number of parameters. *arXiv preprint arXiv:2311.11587*. doi: 10.48550/arXiv.2311.11587

Zheng, M., Xing, Q., Qiao, M., Xu, M., Jiang, L., Liu, H., et al. (2022). "Progressive training of a two-stage framework for video restoration," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, New Orleans, LA, USA, 1023–1030. doi: 10.1109/CVPRW56347.2022.00115

Zhu, J., Li, H., Qing, P., Hou, J., and Peng, Y. (2024). Side-scan sonar image augmentation method based on cc-wgan. *Appl. Sci.* 14, 8031. doi: 10.3390/app14178031

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 16 April 2025
DOI 10.3389/fmars.2025.1539371

OPEN ACCESS

EDITED BY
Muhammad Yasir,
China University of Petroleum (East China), China

REVIEWED BY
Tao Xu,
Henan Institute of Science and Technology, China
Chen Congcong,
Southeast University, China

*CORRESPONDENCE
Shanwen Zhang
wjdw716@163.com

RECEIVED 04 December 2024
ACCEPTED 18 March 2025
PUBLISHED 16 April 2025

CITATION
Wang Z, Guo J, Zhang S and Zhang Y (2025) Sonar-based object detection for autonomous underwater vehicles in marine environments.
Front. Mar. Sci. 12:1539371.
doi: 10.3389/fmars.2025.1539371

Sonar-based object detection for autonomous underwater vehicles in marine environments

Zhen Wang[1,2], Jianxin Guo[1], Shanwen Zhang[1]* and Yucheng Zhang[3]

[1]College of Electronic Information, Xijing University, Xi'an, China, [2]College of Computer Science, Northwestern Polytechnical University, Xi'an, China, [3]College of Computer Science, Xijing University, Xi'an, China

Sonar image object detection plays a crucial role in obstacle detection, target recognition, and environmental perception in autonomous underwater vehicles (AUVs). However, the complex underwater acoustic environment introduces various interferences, such as noise, scattering, and echo, which hinder the effectiveness of existing object detection methods in achieving satisfactory accuracy and robustness. To address these challenges in forward-looking sonar (FLS) images, we propose a novel multi-level feature aggregation network (MLFANet). Specifically, to mitigate the impact of seabed reverberation noise, we designed a low-level feature aggregation module (LFAM), which enhances key low-level image features, such as texture, edges, and contours in the object regions. Given the common presence of shadow interference in sonar images, we introduce the discriminative feature extraction module (DFEM) to suppress redundant features in the shadow regions and emphasize the object region features. To tackle the issue of object scale variation, we designed a multi-scale feature refinement module (MFRM) to improve both classification accuracy and positional precision by refining the feature representations of objects at different scales. Additionally, the CIoU-DFL loss optimization function was constructed to address the class imbalance in sonar data and reduce model computational complexity. Extensive experimental results demonstrate that our method outperforms state-of-the-art detectors on the Underwater Acoustic Target Detection (UATD) dataset. Specifically, our approach achieves a mean average precision (mAP) of 81.86%, an improvement of 7.85% compared to the best-performing existing model. These results highlight the superior performance of our method in marine environments.

KEYWORDS

autonomous underwater vehicles, forward-looking sonar, marine object detection, feature extraction, feature fusion

1 Introduction

As an important underwater exploration means, sonar technology is widely used in the field of marine resource development (Zhang et al., 2022b), marine scientific studies (Grzadziel, 2020), and national defense security (Hansen et al., 2011). A forward-looking sonar (FLS) system can realize the positioning, imaging, and recognition of underwater targets by transmitting sound waves and receiving echo information (Liu et al., 2015), so it has significant advantages in underwater object detection and monitoring tasks. FLS image object detection (Karimanzira et al., 2020) refers to using computer vision and signal processing technology to perform object detection and recognition on the image data obtained by sonar devices to achieve the classification, positioning, and tracking of underwater objects. Different from natural scene images, sonar images are affected by the underwater environment and terrain. As shown in Figure 1, there are serious interferences, such as seabed reverberation noise, sediment shadow region, and background clutter information, in the sonar image. Moreover, FLS images commonly contain underwater objects with different scales and weak feature information, which presents great challenges for sonar object detection.

Compared to object detection in natural scene images, sonar image object detection faces unique challenges due to severe noise interference, complex environments, substantial variations in object scales, and weak saliency of object features. These factors often lead to low detection accuracy, missed detections, and false positives. To address these issues, many methods based on hand-crafted feature extraction combined with classifiers have been proposed. These approaches rely on algorithms for extracting features such as edges, contours, and textures from sonar image regions of interest, followed by classifiers such as support vector machine (SVM) (Chandra and Bedi, 2021), AdaBoost (Collins et al., 2002), and K-nearest neighbors (KNN) (Zhang and Zhou, 2007) for object recognition. For example, Abu and Diamant (2019) developed an object detection framework for synthetic aperture sonar (SAS) images based on unsupervised statistical learning. In the context of FLS images, Zhou et al. (2022b) combined fuzzy C-means and K-means clustering to extract target features through global clustering. Kim and Yu (2017) employed multi-scale feature extraction to obtain Haar-like features from sonar target regions, leveraging AdaBoost to cascade weak classifiers for detection. In efforts to address noise interference, Xinyu et al. (2017) applied fast curve transforms to filter noise and K-means clustering for object region pixel extraction. Zhang et al. (2023) used non-local mean filtering to remove speckle noise and applied super-pixel segmentation to delineate object contours. Although these hand-crafted feature-based methods combined with classifiers have been widely used in sonar object detection, they are limited by their applicability to simple underwater scenes or single-object detection. In more complex underwater acoustic environments and multi-class object detection scenarios, these methods exhibit shortcomings such as insufficient robustness, poor real-time performance, and limited ability to meet high-precision detection requirements.

Benefiting from the robust feature extraction and representation capabilities of convolutional neural networks (CNNs) (Gu et al., 2018), CNN-based methods have gained widespread use in object detection tasks, achieving significantly improved detection performance (Li et al., 2021). These methods leverage frameworks similar to those used in natural scene object detection, such as Faster R-CNN (Ren et al., 2016), You Only Look Once, Version 3 (YOLOv3) (Redmon and Farhadi, 2018), and FPN (Lin et al., 2017a), to detect various types of sonar images, including forward-looking sonar, side-scan sonar, and synthetic aperture sonar. For example, based on the FPN framework, Li et al. (2024) proposed a dual spatial attention network that utilizes a multi-layer convolutional structure to extract features at different scales, with the attention mechanism enhancing feature representation to improve sonar object detection accuracy. To address sonar object detection in complex underwater acoustic environments, Zhao et al. (2023) introduced a composite backbone network that extracts multi-level feature information. Their method uses the shuffle convolution

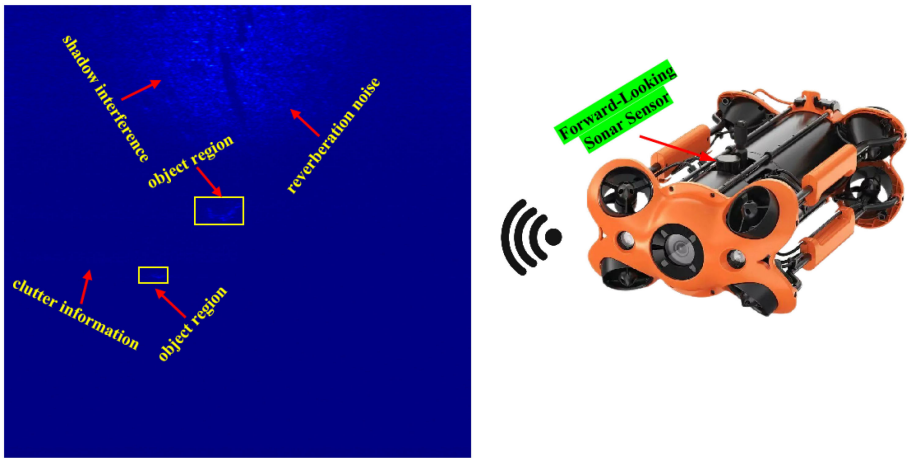


FIGURE 1
Example of a forward-looking sonar image containing object region, seabed reverberation noise, clutter information, and shadow interference.

block attention mechanism and multi-scale feature fusion module to suppress redundant feature interference. Inspired by the two-stage object detection network architecture, Wang et al. (2022d) developed the sonar object detection model, which includes multi-level feature extraction and fusion modules to handle both forward-looking and side-scan sonar detection challenges. Building on the YOLO series of detectors, Zhang et al. (2022a) incorporated the coordinate attention mechanism to extract spatial position features from sonar image regions. They also employed model pruning and compression techniques to enhance the real-time performance of their detector. Yasir et al. (2024) proposed the YOLOShipTracker for ship detection, which has achieved better results in tiny object detection in complex scenes. For tiny object detection, Wang et al. (2022c) introduced the multi-branch shuffle module to reconstruct features at different scales, along with a mixture attention mechanism to strengthen feature representation of small object regions and mitigate clutter interference. Combining CNNs with transformer models, Yuanzi et al. (2022) proposed the TransYOLO detector, which integrates a cascade structure to capture texture and contour features from sonar images, utilizing the attention mechanism for multi-scale feature fusion. Kong et al. (2019) developed the YOLOv3-DPFIN, which achieves effective sonar object detection in complex underwater environments. Their approach employs dense connections for multi-scale feature transmission and the cross-attention mechanism to enhance object region features while reducing reverberation noise interference.

Although CNN-based sonar object detection methods have shown significant improvements over traditional hand-crafted feature extraction techniques, they still face challenges in certain difficult scenarios, such as seabed reverberation noise, shadow interference, object scale variation, and tiny object detection. It is well established that CNN-based object detection methods achieve excellent performance primarily due to their powerful feature extraction capabilities. However, the inherent characteristics of sonar images, such as noise and interference, significantly hinder the feature extraction process of CNN models, making it difficult to fully capture the valuable information necessary for effective sonar image object detection. As illustrated in Figure 2, we provide visualization results of convolution feature heatmaps in challenging scenarios involving seabed reverberation noise interference, shadow interference, clutter, and multi-scale object transformations. These visualizations clearly demonstrate how these interference factors disrupt the feature extraction process of CNN models, leading to a notable decline in detection accuracy across different categories of sonar objects. To address the challenge of sonar image object detection in complex marine acoustic environments, we propose a multi-level feature aggregation network (MLFANet) for FLS image detection. Different from traditional CNN-based methods, MLFANet is specifically designed for challenging sonar detection tasks. The main contributions of this study are as follows:

- **Low-Level Feature Aggregation Module (LFAM):** We introduce the LFAM, a novel module that enhances low-level features and suppresses the impact of seabed reverberation noise, improving feature extraction and object detection in noisy underwater environments. The LFAM significantly enhances the robustness of sonar object detection in the presence of acoustic interference.
- **Discriminative Feature Extraction Module (DFEM):** To handle large-scale shadow regions, we designed the DFEM, which filters redundant features and refines object region representations. The DFEM improves the accuracy of object localization and classification, making MLFANet more efficient in detecting objects even in highly cluttered or shadowed regions.
- **Multi-Scale Feature Refinement Module (MFRM):** We developed the MFRM to address the challenge of object

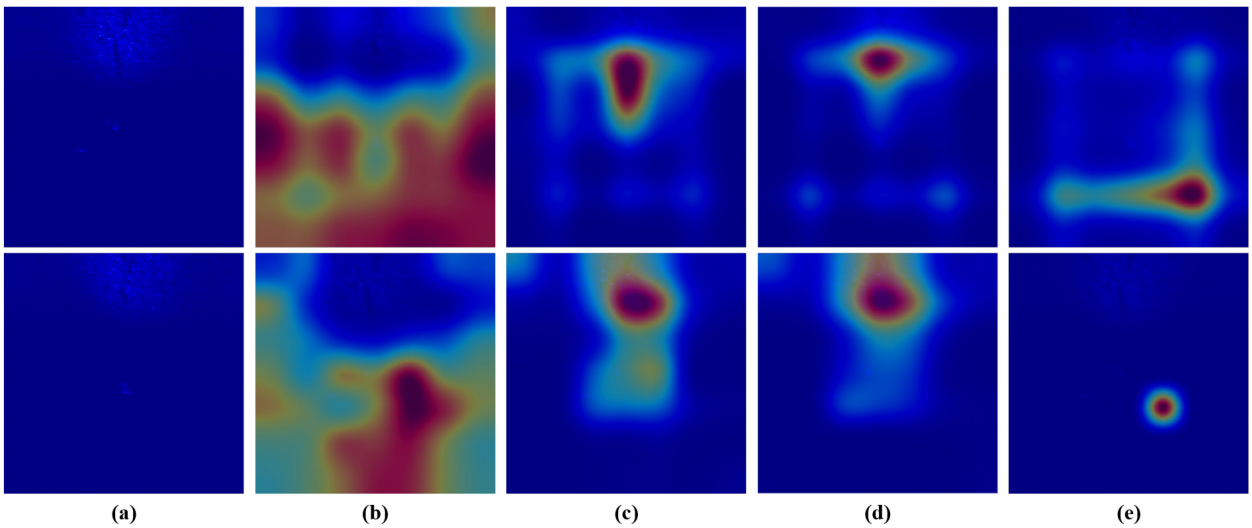


FIGURE 2
Visualization of convolution feature heat map under different interference scenes. **(a)** Two original FLS images. **(b)** Seabed reverberation noise interference. **(c)** Shadow interference. **(d)** Clutter information interference. **(e)** Multi-scale object transformation.

scale variation. The MFRM extracts and fuses fine-grained multi-scale features, enabling the network to handle objects of various sizes more effectively, ensuring that small, medium, and large objects are all accurately detected.

- **CIoU-DFL Loss Function:** To address the issue of object category imbalance in sonar datasets, we propose the CIoU-DFL loss function. This loss function optimizes the model by improving the accuracy of bounding box predictions and reducing computational complexity, particularly for challenging sonar image datasets with skewed category distributions.
- **Extensive Experimental Validation:** We perform extensive experiments on the Underwater Acoustic Target Detection (UATD) dataset, demonstrating that MLFANet outperforms existing state-of-the-art methods in terms of both efficiency and accuracy. Our results highlight the effectiveness of MLFANet in real-world sonar object detection tasks, particularly in complex underwater environments.

The article is organized as follows. Section 2 presents an overview of related works. Section 3 introduces the proposed MLFANet framework and related components. Section 4 presents the experimental results and analysis. Finally, the conclusion is drawn in Section 5.

2 Related works

2.1 Multi-scale feature extraction

For CNN-based object detection methods, multi-scale features play an important function in improving model detection accuracy, fusing global context information, and enhancing model robustness and generalization. Currently, widely used multi-scale feature extraction methods include constructing multi-scale convolution structures (Mustafa et al., 2019), using feature pyramid networks (Lin et al., 2017a), and designing adaptive extraction strategies (Zhou et al., 2022a). Guo et al. (2020) constructed AugFPN to obtain semantic multi-scale features and used residual feature augmentation to highlight the object region feature information. Ma et al. (2020) used the cascade structure to extract multi-scale context information and used feature parameter sharing to establish the correlation of different scale features. To reduce the detail information loss in the multi-scale feature extraction process, Kim et al. (2018b) achieved feature restoration by constructing the global relationship between channel and spatial features. Jiang et al. (2024) used the dense feature pyramid network for small object detection, which uses the multi-scale parallel structure to obtain different scale feature information of the multi-scale object region. MFEFNet (Zhou et al., 2024) uses the efficient spatial feature extraction module to obtain context semantic information and uses a progressive feature extraction strategy to obtain multi-scale features of context information. Tang et al. (2022) constructed a scale-aware feature pyramid structure to obtain multi-scale feature information of the object deformation region and used the feature alignment module to solve the feature offset problem. However, these multi-scale feature extraction methods focus on the extraction of spatial and semantic features, ignoring the important contribution of low-level feature information. Especially for FLS image object detection, low-level features can effectively improve the positioning precision of the object detection model. In this article, we construct the LFAM to obtain low-level multi-scale feature information of the FLS image to improve positioning and recognition accuracy for the sonar object region.

2.2 Contextual feature fusion

Since the contextual information can provide more object region and background information, it can effectively improve the detection accuracy of the object detection model for small object categories. FLS image object detection is a typical small object detection scene, so it is essential to fully mine and fuse the global context feature information. Currently, the commonly used context feature fusion methods include the context feature pyramid (Kim et al., 2018a), global context model (Du et al., 2023), and multi-scale context structure (Wang et al., 2022a). Liang et al. (2019) used the feature pyramid structure to obtain multi-scale context feature information and performed context feature fusion using a spatial-channel reconstruction strategy. Cheng et al. (2020) constructed a cross-scale feature fusion framework to extract local context features and used the region feature aggregation module to achieve context feature fusion. Lu et al. (2021) used the multi-layer feature fusion module to obtain context feature information and introduced a dual-path attention mechanism and multi-scale receptive field module for context feature fine-grained fusion. CANet (Chen et al., 2021) uses a patch attention mechanism to obtain context patch spatial feature information and uses feature mapping and semantic enhancement modules to filter the valuable information of context features. Dong et al. (2022) used deformable convolution and feature pyramids to obtain multi-scale global information and the multi-level feature fusion module is used to fuse local-global context features. These aforementioned context feature fusion methods can effectively fuse feature information of different scales to improve the feature representation for the object region. However, for FLS image object detection, due to the interference of shadow region and clutter information, the existing context feature fusion method cannot solve the feature redundancy problem. In this article, we design the DFEM to suppress redundant feature representation and achieve context feature fusion.

2.3 Visual attention mechanism

An important component of an object detection model, the visual attention mechanism enhances feature representation, solving object deformation and feature correlation modeling. Currently, the attention modules widely used in object detection models include the spatial attention mechanism (Zhu et al., 2019),

channel attention mechanism (Wang et al., 2020), and self-attention mechanism (Shaw et al., 2018). Gong et al. (2022) used the self-attention mechanism to obtain the robust invariant feature information of the object region to enhance the small object region feature representation. Wang et al. (Wang and Wang, 2023) constructed a pooling and global feature fusion self-attention mechanism to obtain the feature correlation and used the feature adaption module for fine-grained feature fusion. Zhu et al. (2018) constructed a cascade attention mechanism to obtain global receptive field information and used dual encoder-decoder attention to reduce feature information loss. Miao et al. (2022) used cross-context attention to obtain local-global feature information and used a spatial-channel attention module to enhance different scale features. To accurately detect multi-scale objects with complex backgrounds, Xiao et al. (2022) designed a pixel attention mechanism to model the pixel correlation information of different object regions and used the self-attention mechanism to enhance the pixel region feature representation. Although the existing visual attention mechanism can effectively enhance the model feature representation and solve the object scale variable problem, for FLS image object detection, due to the serious interference of clutter information and underwater terrain in the object region, the existing attention mechanism struggles to fine-grain enhance the object region feature information, so it cannot obtain satisfactory detection results for small object categories. To solve this problem, inspired by the deformable convolution and attention mechanism, we construct the MFRM to improve the detection accuracy for multi-scale sonar objects by extracting the robust invariant feature information of the object region.

3 Methodology

To solve the problem of object detection in FLS sonar images, based on the YOLOX (Ge et al., 2021) detector, we constructed MLFANet to detect different object categories in sonar images. As shown in Figure 3, the proposed MLFANet introduces the LFAM, DFEM, and MFRM on the basis of the YOLOX detector. Specifically, to improve the object detection performance in complex seabed reverberation noise interference scenes, the LFAM is used to enhance the shallow feature information (*C*1, *C*2, and *C*3) of the backbone network, so that the model can obtain more feature information that is conducive to improving the object positioning precision. Then, to suppress redundant feature representation in deep feature information (*C*4 and *C*5), the DFEM is used to obtain valuable information on deep features to optimize the sonar object detection effect under shadow interference conditions. Moreover, to improve the recognition accuracy of the detector for different categories of sonar objects, we introduce the DFEM into the neck structure, which performs fine-grained fusion of multi-scale feature maps by generating attention weights to further enhance the representation ability of the feature maps and alleviate clutter noise information interference. For the model parameter optimization, we combine CIoU (Zheng et al., 2020) and the DLF (Li et al., 2020) loss function to solve the problem of sample category imbalance and model computational complexity.

3.1 Low-level feature aggregation module

Since the interference of signal intensity difference and reverberation noise, there are many dark areas in sonar images, which makes it difficult for the existing feature extraction network (Elharrouss et al., 2022) to obtain low-level feature information such as texture, edge, and contour of sonar object regions, and the obtained low-level features lead to serious loss in the process of convolutional feature transmission. To solve this problem, we designed the LFAM and embedded it into the backbone network to compensate for the feature information loss of deep convolution by mining the low-level features obtained in the shallow

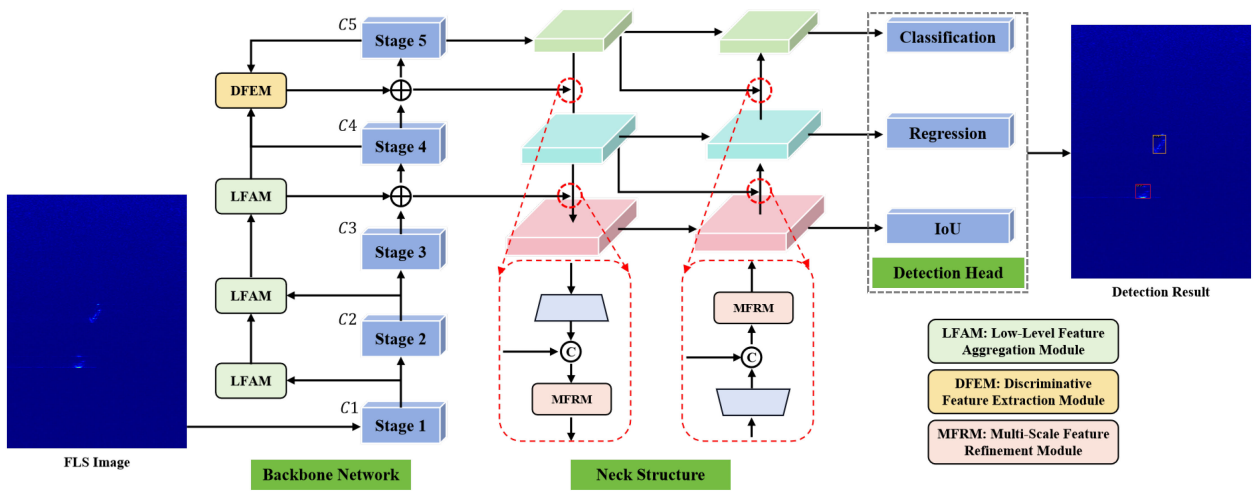


FIGURE 3
The overall architecture of the proposed multi-level feature aggregation network (MLFANet), including the low-level feature aggregation module (LFAM), discriminative feature extraction module (DFEM), and multi-scale feature refinement module (MFRM).

convolution stages. The LFAM is designed to enhance low-level feature information, such as texture, edges, and contours, while suppressing seabed reverberation noise that commonly disrupts the feature extraction process.

The specific structure of the LFAM is shown in Figure 4, where the backbone network consists of five convolution stages, and denotes the feature map obtained in the *lth* convolution stage and $l \in [1, 5]$. The proposed LFAM takes the feature maps C_1, C_2, and C_3 obtained in the shallow convolution stage as input features, and performs feature fusion in turn to generate the aggregation feature map $\boldsymbol{G} \in \mathbb{R}^{C\times H\times W}$, so that it can retain more low-level feature information. The specific fusion process is as follows:

$$\boldsymbol{G} = K_{3\times3}(K_{3\times3}(C_1) \oplus C_2) \oplus C_3 \tag{1}$$

where $K_{3\times3}(\cdot)$ represents the 3 × 3 convolution function for feature map resolution and feature channel adjustment, and $\oplus$ denotes the element-by-element summation operation. The aggregate feature map is used as the output of parallel pooling, which uses different pooling layers to obtain the context information of the aggregate feature map to extract more discriminative low-level features. The parallel pooling consists of different pooling functions, namely $1 \times W$ strip pooling, $H \times 1$ strip pooling, and $S \times S$ spatial pooling and residual connection. For the aggregate feature map $\boldsymbol{G}$ with a size of $H \times W$, the feature map is averaged using strip pooling with a pooling range of $(1, W)$ and $(H, 1)$, which compresses the feature map and encodes feature information along the vertical and horizontal directions. Furthermore, the use of strip pooling establishes long-distance dependencies between discretely distributed feature regions for spatial dimension information in the vertical and horizontal directions and obtains low-level feature information such as edges and contours of the object region in the global dimension. The calculation of strip pooling is as follows:

$$y_w = \frac{1}{H}\sum\nolimits_{0\le i<H} \boldsymbol{G}(i, W) \tag{2}$$

$$y_h = \frac{1}{H}\sum\nolimits_{0\le j<W} \boldsymbol{G}(H, j) \tag{3}$$

where $y_w \in \mathbb{R}^{C\times1\times W}$ and $y_h \in \mathbb{R}^{C\times H\times1}$ represent the feature tensors obtained by strip pooling with sizes of 1 × 1 and 3 × 3, respectively. The one-dimensional convolution is used to integrate the adjacent feature information inside the feature tensor, and the bilinear interpolation operation is used to recover the spatial information of feature tensor y_w and y_h. To generate low-level features with rich edges and contours, the feature tensor is fused by using the element-by-element multiplication operation. The calculation process is as follows:

$$z_1 = \mathcal{F}_{ex}(f_{3\times1}(y_w)) \oplus \mathcal{F}_{ex}(f_{1\times3}(y_h)) \tag{4}$$

where $\mathcal{F}_{ex}(\cdot)$ represents the bilinear interpolation operation, and $f_{3\times1}(\cdot)$ and $f_{1\times3}(\cdot)$ represent the one-dimensional convolution operation with the size of 3 × 1 and 1 × 3, respectively. Moreover, the parallel pooling introduces spatial pooling with a range of $S \times S$, which can use rectangular pooling windows to detect densely distributed object region feature information and obtain texture feature information of sonar objects in the local receptive field range. The residual connection is used to preserve the original spatial information of the aggregate feature map $\boldsymbol{G}$, and it is fused with the spatial pooling feature to generate low-level texture feature tensor z_2. The specific calculation process is as follows:

$$z_2 = P_{s\times s}(\boldsymbol{G}) \oplus \boldsymbol{G} \tag{5}$$

where $P_{s\times s}(\cdot)$ denotes the spatial pooling with a size of 1 × 1. For feature tensors z_1 and z_2, the 3 × 3 convolution is used to further extract detailed information, and the feature stitching operation is used to generate feature map $z_3 \in \mathbb{R}^{C\times H\times W}$ with

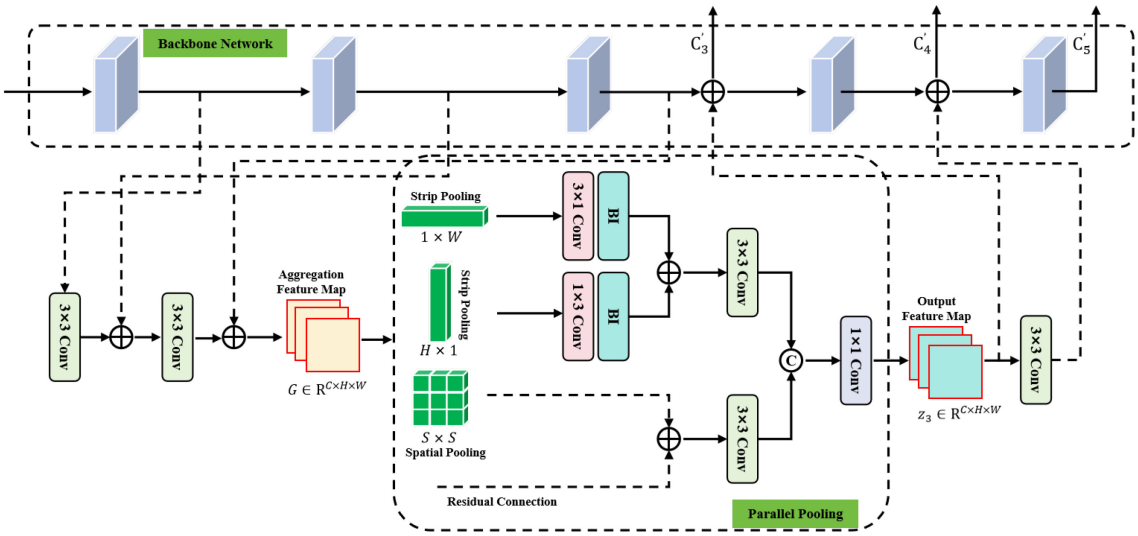


FIGURE 4
The specific structure of the low-level feature aggregation module (LFAM) includes 3×3 convolution, 1 × 1 convolution, 3 × 1, 1 × 3 one-dimensional convolution, element-by-element summation, channel stitching, and bilinear interpolation operation.

more discriminative information. The calculation is as follows:

$$z_3 = K_{1\times1}([K_{3\times3}(z_1); K_{3\times3}(z_2)]) \tag{6}$$

where $K_{1\times1}(\cdot)$ and $K_{3\times3}(\cdot)$ represent convolution operations with sizes of 1×1 and 3×3, respectively, and $[\cdot;\cdot]$ denotes the feature stitching operation on the channel dimension. The feature map z_3 is fused with the features C_3 and C_4 in the deep convolution stage of the backbone network, and input to the subsequent convolution stage to compensate for low-level feature information loss. The feature maps $C_3^{'}$, $C_4^{'}$, and $C_5^{'}$ generated by the fuse operation can retain more effective edge, contour, and texture feature information, which is beneficial for improving the positioning precision for different object categories. The generation process of feature maps $C_3^{'}$, $C_4^{'}$, and $C_5^{'}$ is calculated as follows:

$$C_3^{'} = C_3 \oplus z_3 \tag{7}$$

$$C_4^{'} = K_{3\times3}(z_3) \oplus \mathcal{F}_{\text{conv}}^{4}(C_3^{'}) \tag{8}$$

$$C_5^{'} = \mathcal{F}_{\text{conv}}^{5}(C_4^{'}) \tag{9}$$

where $K_{3\times3}(\cdot)$ represents the convolution operation with a size of 3 × 3, and $\mathcal{F}_{\text{conv}}^{l}(\cdot)$ denotes the *lth* convolution stage. The LFAM leverages feature aggregation and parallel pooling operations to extract discriminative low-level feature information. By preserving key spatial details and reducing noise interference, LFAM enhances the model's ability to detect object boundaries and localization precision.

3.2 Discriminative feature extraction module

Due to the redundant feature interference in the feature extraction process of the convolution operation (Qin et al., 2020), it is difficult to retain valuable tiny object region information. To solve this problem, we propose the DFEM, as shown in Figure 5. The DFEM improves the robustness of feature extraction in shadowed and cluttered regions by suppressing redundant features and enhancing salient object features. For the deep feature information (C_4 and C_5) obtained by the backbone network, given the specific feature mapping $\boldsymbol{X} \in \mathbb{R}^{C\times W\times H}$, where C, H, and W represent the number of channels, width, and height of the feature map, respectively. To mine the local regions with discriminative attributes in convolution features, the obtained deep features are divided into k regions along the W dimension, where each region feature is defined as $\boldsymbol{X}_i \in \mathbb{R}^{C\times W/k\times H}$. The feature description importance factor corresponding to each region is calculated as

$$\boldsymbol{a}_i = \text{SoftMax}(\mathcal{F}_{\text{GAP}}(K_{1\times1}(\boldsymbol{X}_i))) \tag{10}$$

where $K_{1\times1}(\cdot)$ represents the convolution operation with a size of 1 × 1, $\mathcal{F}_{\text{GAP}}(\cdot)$ denotes the global average pooling function, and the softmax function is used for feature normalization. The high importance factor indicates that the region feature significance is strong. By comparing the importance factor of different regions, the region with strong discrimination feature description in W dimension can be located. We use the descriptor $\boldsymbol{Y}$ to denote the positioning region and separate it from the initial feature $\boldsymbol{X}$. The region $\boldsymbol{Y}$ is uniformly split into n sub-regions along the H dimension, and $\boldsymbol{Y}_j \in \mathbb{R}^{C\times W/k\times H/n}$ is used to denote the feature information of each sub-region, where $j \in [1, 2, \ldots, n]$. The calculation of the importance factor for sub-region feature description is as follows:

$$\boldsymbol{b}_j = \text{SoftMax}\left(\mathcal{F}_{\text{GPA}}(K_{1\times1}(\boldsymbol{Y}_j))\right) \tag{11}$$

The normalized importance factor of each sub-region can be used to discriminate the sub-region $\boldsymbol{Y}_J^{'} \in \mathbb{R}^{C\times W/k\times H/n}$ with important feature information in the feature mapping $\boldsymbol{X}$. By using the above feature discrimination process, it can effectively solve the deviation problem of feature extraction and enhance the localization ability for the discriminant feature region. To further mine the valuable information in the feature map, we use the discriminative feature enhancement-suppression strategy to preprocess the sub-region feature $\boldsymbol{Y}_J^{'}$, and obtain the feature maps $\boldsymbol{Y}_e \in \mathbb{R}^{C\times W/k\times H}$ and $\boldsymbol{Y}_s \in \mathbb{R}^{C\times W/k\times H}$. The calculation is as follows:

$$\boldsymbol{Y}_e = \boldsymbol{Y} + \alpha \times (\boldsymbol{E} \otimes \boldsymbol{Y}) \tag{12}$$

$$\boldsymbol{Y}_s = \boldsymbol{S} \otimes \boldsymbol{Y} \tag{13}$$

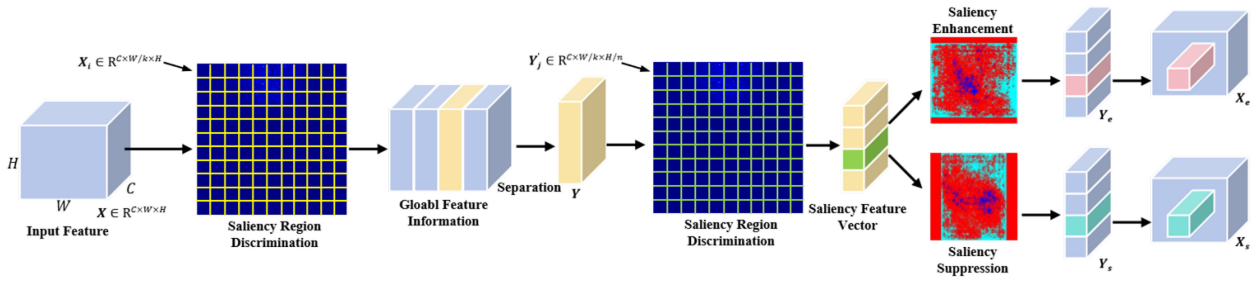


FIGURE 5
The specific structure of the discriminative feature extraction module (DFEM) includes saliency region discrimination, global feature information extraction, saliency region selection, and saliency feature enhancement and suppression.

where $\otimes$ represents the element-by-element multiplication, and the specific calculation for features $\boldsymbol{E}$ and $\boldsymbol{S}$ are as follows:

$$\begin{cases} \boldsymbol{E} = (\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n)^{\mathrm{T}} \\ \boldsymbol{e}_j = \begin{cases} \boldsymbol{b}_j, \text{if} \boldsymbol{b}_j = \max(\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n) \\ 0, \text{otherwise} \end{cases} \end{cases} \tag{14}$$

$$\begin{cases} \boldsymbol{S} = (\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n)^{\mathrm{T}} \\ \boldsymbol{s}_j = \begin{cases} 1-\beta, \text{if} \boldsymbol{b}_j = \max(\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n) \\ 1, \text{otherwise} \end{cases} \end{cases} \tag{15}$$

where α and β denote the coefficients used to control feature enhancement and suppression, respectively. The original feature $\boldsymbol{Y}$ is replaced by feature maps $\boldsymbol{Y}_e$ and $\boldsymbol{Y}_s$, and fused with feature $\boldsymbol{X}_i$ along the $\boldsymbol{W}$ dimension to generate the discriminative enhancement feature $\boldsymbol{X}_e \in \mathbb{R}^{C\times W\times H}$ and the discriminative suppression feature $\boldsymbol{X}_s \in \mathbb{R}^{C\times W\times H}$, respectively. By using a discriminative enhancement operation, it can effectively suppress redundant feature representation to improve the detection accuracy for tiny object categories in sonar images. The DFEM improves the robustness of feature extraction in shadowed and cluttered regions by suppressing redundant features and enhancing salient object features.

3.3 Multi-scale feature refinement module

Due to interference in underwater environments, FLS images contain serious object deformation problems, which makes it difficult for the object detection network to extract fine-grained feature information from the object region, and it is prone to lose the valuable feature information in the shadow region. To solve this problem, we constructed the MFRM and embedded it into the neck structure of the detector to enhance the feature extraction capacity for the deformation object regions. The MFRM consists of region location branch and feature refinement branch, and the specific structure is shown in Figure 6. The MFRM addresses the challenge of detecting objects at varying scales by extracting robust, scale-invariant features and refining multi-scale feature representations. The region location branch is used to position the range of object region, which uses 7 × 7 convolution to obtain local feature information and extract the valuable feature region information for the input feature map $\boldsymbol{X} \in \mathbb{R}^{C\times W\times H}$. The 7 × 7 convolution kernel provides a larger receptive field compared to smaller kernels (e.g., 3 × 3 or 5 × 5), enabling the extraction of richer local feature information. Parallel dilated convolution with different dilation coefficients is used to expand the range of receptive fields and stitch the dilated convolution features to aggregate fine-grained context information. To generate the region attention map, the 3 × 3 convolution is used to encode the context information to obtain the object region features. The calculation is as follows:

$$\boldsymbol{U}_1 = \mathcal{K}_{3\times3}([\mathcal{F}^{6}_{\mathrm{atr}}(K_{7\times7}(\boldsymbol{X})); \mathcal{F}^{12}_{\mathrm{atr}}(K_{7\times7}(\boldsymbol{X}))]) \tag{16}$$

where $K_{3\times3}(\cdot)$ and $K_{7\times7}(\cdot)$ represent convolution operations with sizes of 3 × 3 and 7 × 7, respectively; $\mathcal{F}^{6}_{\mathrm{atr}}(\cdot)$ and $\mathcal{F}^{12}_{\mathrm{atr}}(\cdot)$ denote the dilation coefficients of 6 and 12; $[\cdot;\cdot]$ represents the feature splicing operation on the spatial dimension. The feature refinement branch obtains the fine-grained feature information of the object region through the feature cross-dimensional interaction. This branch performs different global adaptive pooling operations on the input feature map $\boldsymbol{X} \in \mathbb{R}^{C\times H\times W}$ to obtain global spatial feature information and perform feature space compression. Specifically, 1 × 1 global adaptive average pooling is used to compress the global feature spatial information, 3×3 global adaptive average pooling is used to enhance the global feature representation, and 2 × 2 global adaptive maximum pooling is used to enhance the feature structure information and refine the global feature information obtained by the global adaptive average pooling. The feature tensor obtained by the different pooling operations is converted into vector representation using feature reconstruction to achieve a cross-dimensional interaction of feature

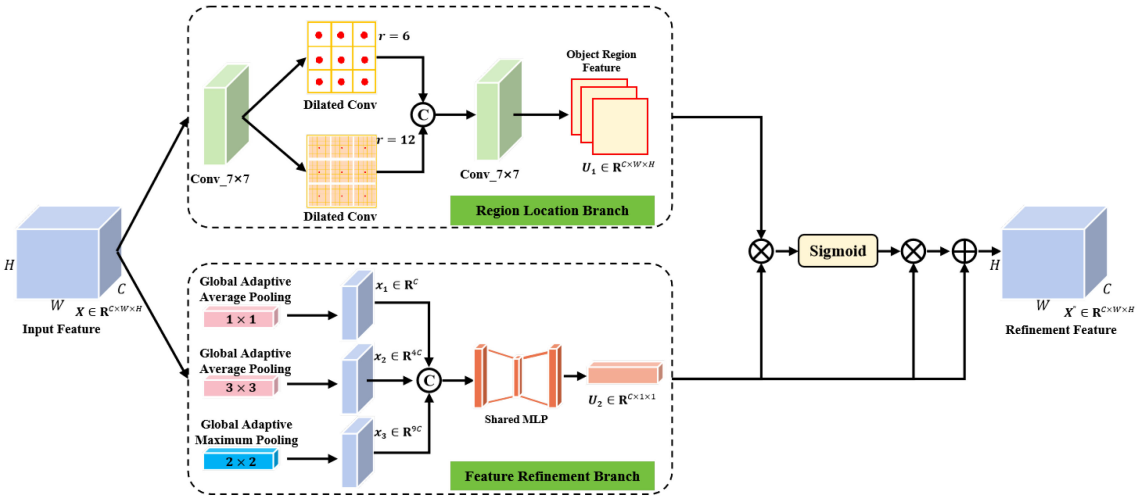


FIGURE 6
The specific structure of the multi-scale feature refinement module (LFAM) includes a region location branch and feature refinement branch.

information on the spatial dimension and fuse it with the object region features retained on the channel dimension to generate one-dimensional feature vectors $x_1 \in \mathbb{R}^C$, $x_2 \in \mathbb{R}^{4C}$ and $x_3 \in \mathbb{R}^{9C}$. The one-dimensional feature vector is spliced to obtain the feature vector $x_4 \in \mathbb{R}^{14C}$ that aggregates rich cross-dimensional interaction feature information. The specific calculation of this process is as follows:

$$\boldsymbol{x}_1 = \mathcal{F}_{\text{resize}}\left(\boldsymbol{P}^1_{\text{avg}}(\boldsymbol{X})\right) \tag{17}$$

$$\boldsymbol{x}_2 = \mathcal{F}_{\text{resize}}\left(\boldsymbol{P}^2_{\text{avg}}(\boldsymbol{X})\right) \tag{18}$$

$$\boldsymbol{x}_3 = \mathcal{F}_{\text{resize}}\left(\boldsymbol{P}^3_{\text{max}}(\boldsymbol{X})\right) \tag{19}$$

$$\boldsymbol{X}_c = [\boldsymbol{x}_1; \boldsymbol{x}_2; \boldsymbol{x}_3] \tag{20}$$

where $P^n_{\text{avg}}(\cdot)$ represents the global adaptive average pooling function with a size of $n \times n$, $P^n_{\text{max}}(\cdot)$ represents the global adaptive maximum pooling function with a size of $n \times n$, and $\mathcal{F}_{\text{resize}}(\cdot)$ feature reconstruction operation. The multi-layer perceptron composed of the fully connected layer and non-linear activation function is used to encode the feature vector $\boldsymbol{X}_c$ to generate the feature descriptor $\boldsymbol{U}_2 \in \mathbb{R}^{C\times1\times1}$. The specific calculation process is as follows:

$$\boldsymbol{U}_2 = \text{MLP}(\boldsymbol{X}_c) = \mathcal{F}_1(\delta(\mathcal{F}_2(\boldsymbol{X}_c))) \tag{21}$$

where $\mathcal{F}_1 \in \mathbb{R}^{C/r\times C}$ and $\mathcal{F}_2 \in \mathbb{R}^{C\times C/r}$ represent different fully connected functions, and set $r = 32$; δ denotes the ReLU activation function. Element-by-element multiplication is used to fuse the region attention mapping $\boldsymbol{U}_1$ and the feature descriptor $\boldsymbol{U}_2$, and the Sigmoid function is used to normalize the feature values to the range of $(0, 1)$ to generate the attention weight $\boldsymbol{M}$. The original feature map $\boldsymbol{X}$ is weighted to achieve object feature adaptive optimization to highlight the object region feature information and reduce the seabed reverberation noise interference. The specific calculation is as follows:

$$\boldsymbol{M} = \sigma(\boldsymbol{U}_1 \otimes \boldsymbol{U}_2) \tag{22}$$

$$\boldsymbol{Y} = \boldsymbol{X} \oplus (\boldsymbol{X} \otimes \boldsymbol{M}) \tag{23}$$

where $\otimes$ represents element-by-element multiplication, σ denotes the Sigmoid activation function, $\oplus$ denotes element-by-element summation, and $\boldsymbol{Y}$ represents the multi-scale refinement feature map. The MFRM uses a dual-branch architecture to effectively model object regions at different scales. The region location branch focuses on coarse object localization, while the feature refinement branch enhances fine-grained feature details through cross-dimensional feature interactions. This ensures that objects of different sizes, from small to large, are accurately detected and classified.

3.4 Loss function optimization

To optimize the proposed MLFANet detector, we combined CIoU (Zheng et al., 2020) and DLF (Li et al., 2020) to calculate the regression loss of the bounding box. The constructed loss function uses DLF loss to obtain the loss probability of the bounding box and object label by calculating the cross-entropy function. The distribution probability of the bounding box is restored as the prediction box, and CIoU is used to calculate the loss value of the prediction box and truth box to achieve the optimization of the prediction box generation process. The calculation of CIoU is as follows:

$$\mathcal{L}_{\text{CIoU}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{gt})}{(c_w)^2 + (c_h)^2} + \frac{4}{\pi}\left(\arctan\frac{w_{gt}}{h_{gt}} - \arctan\frac{w}{h}\right)^2 \tag{24}$$

where IoU represents the intersection in the union of the prediction bounding box and truth bounding box; $\rho^2(b, b^{gt})$ denotes the Euclidean distance between the prediction box and the truth box; h and w represent the height and width of the prediction box; h_{gt} and w_{gt} represent the height and width of the truth box; c_h and c_w denote the height and width of the minimum bounding box consisting of the prediction box and truth box. Since CIoU ignores the problem of sample imbalance, smaller positional offsets lead to significant decreases in IoU values for small object regions in sonar images, while large size object regions will produce an IoU difference. Moreover, since the calculation process involves the solution of inverse trigonometric function, it increases the model computational complexity. To solve this problem, we introduce the normalized Wasserstein distance (NWD) position regression loss function, which uses the two-dimensional Gaussian distribution to calculate the similarity between the prediction box and truth box. The loss calculation process can reflect the true distance between the prediction box and object region distribution, and it has strong robustness to the object scale scaling, so it is more suitable for solving the tiny object detection problem. The specific calculation of the NWD position loss function is as follows:

$$N_a = [cx_a, cy_a, w_a/2, h_a/2]^T \tag{25}$$

$$N_b = [cx_b, cy_b, w_a/2, h_a/2] \tag{26}$$

$$W_2^2(N_a, N_b) = \| (N_a, N_b) \|_2^2 \tag{27}$$

$$\mathcal{L}_{\text{NWD}}(N_a, N_b) = \exp\left(-\sqrt{W_2^2(N_a, N_b)}/C\right) \tag{28}$$

where C denotes the number of object categories; $W_2^2(N_a, N_b)$ denotes the distance measure; N_a and N_b denote the Gaussian distributions modeled by $A = (cx_a, cy_a, w_a, h_a)$ and $B = (cx_b, cy_b, w_b, h_b)$, respectively. Since CIoU is suitable for large size object categories, we combine CIoU and NWD to construct the loss optimization function. The specific calculation is as follows:

$$\mathcal{L}_{\text{CIoU\_NWD}} = \alpha \cdot \mathcal{L}_{\text{CIoU}} + (1 - \alpha) \cdot \mathcal{L}_{\text{NWD}} \tag{29}$$

where α represents the adaptive weight adjustment coefficient, $\mathcal{L}_{\text{CIoU}}$ and $\mathcal{L}_{\text{NWD}}$ denote the CIoU loss function and the NWD loss function, respectively.

4 Experiments and analysis

In this section, we present a detailed description of the forward-looking sonar image dataset, model training strategy, experimental parameter setting, evaluation metrics, ablation studies, and robustness analysis.

4.1 FLS image dataset

To verify the effectiveness and feasibility of the proposed method, we conducted experimental verification on the UATD dataset (Qin et al., 2020) in a real-scene underwater acoustic environment. The dataset was released in 2022 and was provided by Peng Cheng Laboratory, Shenzhen, China. It used Tritech Gemini 1200ik multi-beam forward-looking sonar for image collection. The sonar operates at two acoustic frequencies, 720kHZ for lone-range object detection, and 1,200kHz for enhanced high-resolution imaging at shorter ranges. The data collection sites were located in Golden Pebble Beach in Dalian (39.0904292°N, 122.0071952°E) and Haoxin Lake in Maoming (21.7011602°N, 110.8641811°E).

The dataset contains 9,200 high-resolution original forward-looking sonar images and corresponding manual annotation information. To improve the readability of the sonar images, we performed Gaussian filtering and pseudo-color enhancement on the original images, as shown in Figure 7. The annotation object categories provided by the dataset contain a cube, ball, cylinder, human body model, tire, circle cage, square cage, metal bucket, plane model, and ROV, and the corresponding physical entities and sizes are shown in Figure 8. We present the statistical information of the number of different object categories in Figure 9a, from which it can be seen that the dataset has a serious category imbalance problem. To further analyze the dataset, we calculated the area and aspect ratio of the rectangular label boxes of different object categories, and drew the corresponding histogram, as shown in Figures 9b, c. It can be seen that the different object category sizes were diverse, as the minimum area covered 12 pixels, and the maximum area included 38,272 pixels; the rectangle minimum ratio of length/width was 0.22, and the maximum ratio was 7.95. From the above statistical information, it can be shown that the dataset poses a great challenge to the sonar image object detection task.

4.2 Training strategies and implementation details

The specific details of the dataset and hyperparameters in the experiment are described as follows.

4.2.1 Dataset setting

For the 9,200 forward-looking sonar images contained in the UATD dataset, we randomly split them into the training, verification, and testing sets based on the ratio of 7:2:1. Specifically, the training set contained 6,440 images, the verification set contained 1,840 images, and the testing set contained 920 images. To further improve the model robustness and generalization performance, data augmentation methods including random rotation, image deformation, brightness transformation, image sharpening, and adding noise were used to supplement the number of training set samples. The use of data augmentation can also alleviate the overfitting problem in the

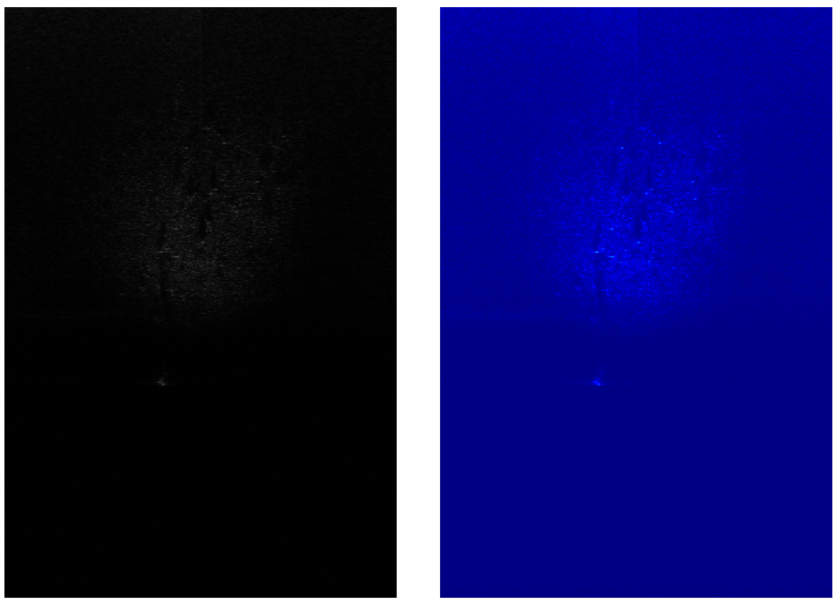


FIGURE 7
The original forward-looking sonar image and the preprocessed image from the UATD dataset. **(a)** original image. **(b)** preprocessed image.

FIGURE 8
The physical sonar target entities and their corresponding size in the UATD dataset. The size is measured in meters.

model training process. Moreover, limited by the device memory, we uniformly scaled the original sonar image to 512×512 pixels in the training process and maintained the original image size for the verification and testing sets.

4.2.2 Training strategies

The experiments were conducted on a workstation equipped with an Intel i9-12900T CPU, 64GB RAM, an NVIDIA GeForce RTX 4090 GPU, and the Ubuntu 18.04 operating system. The code was implemented using the PyTorch 2.1.0 and MMDetection 3.2.0 frameworks. All models were trained and evaluated on the UATD dataset using the same training, validation, and testing splits to ensure fairness. During training, input images were resized to 512×512 pixels, and data augmentation techniques, including random horizontal flipping, random rotation, and color jittering, were applied equally to all models to improve robustness and prevent overfitting. Mixed precision training was employed to enhance training speed and memory efficiency.

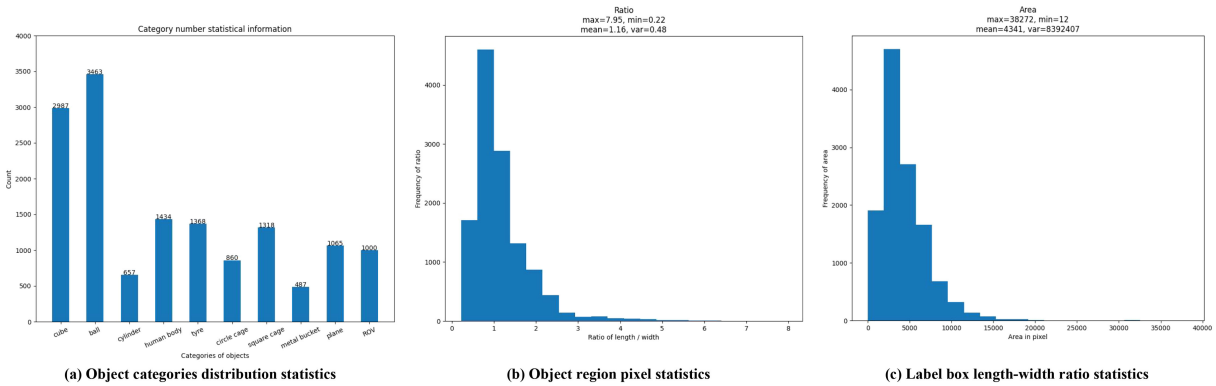


FIGURE 9
An overview of the detailed statistical information of the UATD dataset. **(a)** Object categories distribution statistics. **(b)** Object region pixel statistics. **(c)** Label box length-width ratio statistics.

For the proposed MLFANet, we used a ResNet-50 or ResNet-101 backbone pre-trained on ImageNet. The batch size was set to 8, and the optimizer was SGD with momentum (0.9) and a weight decay of 0.0001. The initial learning rate was set to 0.02 and reduced by a factor of 10 at epochs 8 and 11, with a total of 12 training epochs (1× schedule). To further optimize performance, we adopted a three-stage training strategy: (1) pre-training the backbone on ImageNet with a batch size of 32 and an initial learning rate of 0.001, decayed every 1,000 iterations; (2) fine-tuning the pre-trained backbone on the sonar image dataset with a batch size of 8, an initial learning rate of 0.001, and decay applied every 500 iterations; and (3) training the entire model with a batch size of 16, an initial learning rate of 0.0001, and decay applied every 2,000 iterations. This staged strategy ensured optimal parameter learning and mitigated overfitting.

For the baseline models, we used their standard configurations as described in their original implementations. For example, Faster R-CNN, RetinaNet, Cascade R-CNN, Dynamic R-CNN, and DH R-CNN were trained with a ResNet-50 backbone, a batch size of 8, an initial learning rate of 0.02 (reduced by a factor of 10 at epochs 8 and 11), and 12 training epochs. CenterNet was trained with a ResNet-101 backbone, a batch size of 16, an initial learning rate of 0.01 (reduced at epochs 30 and 45), and 50 training epochs. The DETR-based models (e.g., DETR, DAB-DETR, Sparse R-CNN, and CO-DETR) used AdamW optimizers, with a batch size of 4 and an initial learning rate of 0.0001 for the transformer and 0.00001 for the backbone. These models were trained for 50 epochs, with learning rate reductions at epoch 40. ViTDet used a ViT-B backbone, a batch size of 8, an initial learning rate of 0.0001, and was trained for 36 epochs, with learning rate reductions at epochs 24 and 30. By using consistent preprocessing, training splits, and hyperparameters tailored to each model, we ensured a fair and comprehensive comparison across all methods.

4.3 Evaluation metrics

To quantitatively evaluate the effectiveness and advantages of the proposed sonar object detection model, we used the precision, recall, average precision (AP), false alarm rate (FAR), F1 score, and frames-per-second (FPS) metrics commonly used in natural scene image object detection tasks as the evaluation metrics. First, we defined TP, FP, TN, and FN as true positive, false positive, true negative, and false negative. Specifically, TP indicates the model correctly detects the sonar object, FP denotes a non-object is falsely detected as the object region, TN indicates the model correctly predicts the non-object category, and FN denotes the object region is mistakenly predicted as a non-object. The calculation of different evaluation metrics is as follows.

1) The precision is defined as the proportion of the model's correct object detection to overall detection results.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{30}$$

2) The recall is defined as the proportion of model correct object detection to the truth annotation object.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{31}$$

3) The AP is defined as the area under the precision-recall (PR) curve used to evaluate the model performance.

$$\text{AP}_{\text{IoU}} = \int_0^1 P(R)\, d(R) \tag{32}$$

where IoU denotes the intersection-over-union threshold used to determine whether the detection result belongs to TP or FP. For the sonar object detection task, we set the IoU to 0.5. Additionally, the evaluation metrics AP^s, AP^m, and AP^l of the Microsoft COCO dataset (Lin et al., 2014) were used to further refine the evaluation and analyze model performance.

4) The FAR evaluates the prediction result credibility by calculating the proportion of FP in all the results.

$$\text{FAR} = \frac{\text{FP}}{\text{TP} + \text{FP}} \tag{33}$$

5) The F1 score is defined as the harmonic mean of precision and recall and can assess the comprehensive performance of the object detection model.

$$\text{F1\_score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{34}$$

6) The FPS represents the speed of the object detection model to process a single frame image per second.

$$\text{FPS} = 1/T_{\text{single}} \tag{35}$$

where T_{single} denotes the time taken to process a single forward-looking sonar image.

4.4 Comparison experiments and analysis

To demonstrate the advantages of the proposed forward-looking sonar object detector MLFANet, we compared it with 11 state-of-the-art object detection models on the UATD dataset. The compared methods can be classified into CNN-based methods and Transformer-based methods. Specifically, the CNN-based methods included Faster R-CNN (Girshick, 2015), RetinaNet (Lin et al., 2017b), Cascade R-CNN (Cai and Vasconcelos, 2019), CenterNet (Duan et al., 2019), Dynamic R-CNN (Zhang et al., 2020), DH R-CNN (Wang et al., 2022b), and Spare R-CNN (Sun et al., 2023); the Transformer-based methods included DETR (Carion et al., 2020), ViTDet (Li et al., 2022), DAB-DETR (Liu et al., 2022) and CO-DETR (Zong et al., 2023). To ensure experiment fairness, the compared methods were retrained on the UATD dataset and used the same training strategy and parameter settings as the proposed methods.

The comparative analysis included quantitative comparison, qualitative comparison, and model complexity analysis. The details are as follows:

4.4.1 Quantitative analysis

The quantitative comparison of different object detection methods was performed on the testing set of the UATD dataset. The performance quantitative analysis results of different methods are shown in Table 1. From the analysis results, compared with other object detection models, the proposed MLFANet obtained the optimal results on multiple evaluation metrics. Additionally, for metrics AP^l, AP^m, and AP^s, the proposed method reached 62.79%, 58.24%, and 45.36%, respectively, which further explains the comprehensive performance advantages of our MLFANet. Specifically, compared with the CNN-based optimal model CenterNet (Duan et al., 2019) and Transformer-based optimal model CO-DETR (Zong et al., 2023), the proposed method was 6.53% and 2.85% higher for the AP metric, respectively. For the CNN-based methods, such as Faster R-CNN (Girshick, 2015), RetinaNet (Lin et al., 2017b), and Cascade R-CNN (Cai and Vasconcelos, 2019), the AP values only reached 32.53%, 29.75%, and 34.97%, respectively, and were accompanied by higher FAR values. The reason for this phenomenon is that the seabed reverberation noise and clutter information contained in the sonar image seriously interfere with the feature extraction process of the CNN model, and the use of a simple convolution operation cannot fully extract the valuable feature information. Moreover, the weak and dark light characteristics of the sonar image object region diminish the positioning and recognition of the CNN-based methods, so they cannot achieve the ideal detection accuracy. Since the Transformer model has better global feature extraction and modeling effect, compared with the CNN-based method, the Transformer-based method has a slight advantage for the sonar image object detection task. For example, compared with Dynamic R-CNN (Girshick, 2015), ViTDet (Li et al., 2022) was 8.40% and 6.86% higher for the AP and F1 score, respectively. Furthermore, for the metrics AP^l and AP^m, the optimal Transformer-based model CO-DETR (Zong et al., 2023) reached 58.93% and 54.68%, indicating that the method can accurately detect large/medium size objects in sonar images. However, the imaging characteristics of sonar images cause redundant information interference in the global information correlation modeling process of the Transformer-based method, which makes it difficult to achieve satisfactory results for small object detection. For instance, the AP^s values of ViDet (Li et al., 2022), DAB-DETR (Liu et al., 2022), and CO-DETR (Zong et al., 2023) were only 41.32%, 39.76%, and 42.18%, and these methods have high false alarm rates. The reason for this problem is that the Transformer model only focuses on global feature information extraction, ignoring the important value of local feature information, resulting in false discrimination of small object region features as background information features. To verify the detection accuracy of different object detection models for different object categories in sonar images, we randomly selected 1,200 images from the UATD dataset as experimental data. As shown in Table 2, the mean AP (mAP) value of the proposed MLFANet was 81.86%, which is better than all the compared methods. The quantitative results further illustrate the superior detection performance of the proposed method compared to other object detection models. For the AP value of each sonar object category, we can conclude that for the tiny object categories, i.e. the ball, circle cage, and tire, the optimal CNN-based model CenterNet (Duan et al., 2019) only reached 61.28%, 39.78%, and 30.12%, and the optimal Transformer-based model CO-DETR (Zong et al., 2023) only reached 62.85%, 45.63%, and 35.92%. For the large-size object categories, i.e., the cube, plane, and metal bucket, the experimental results in Table 2 show that

TABLE 1 Performance comparison of different object detection methods on the testing set of the UATD dataset, where the score in bold is the highest score.

Model	Backbone	Precision	Recall	F1 score	AP	AP50	AP75	AP^l	AP^m	AP^s	FAR
Faster R-CNN	ResNet-50	0.8245	0.8547	0.8393	0.3253	0.8013	0.2179	0.4768	0.4312	0.3147	0.1755
RetinaNet	ResNet-50	0.7852	0.8165	0.8005	0.2975	0.7928	0.1852	0.4573	0.4127	0.3052	0.2148
Cascade R-CNN	ResNet-50	0.8564	0.8872	0.8715	0.3497	0.8417	0.2368	0.4892	0.4562	0.3387	0.1436
CenterNet	ResNet-101	0.8864	0.8953	0.8908	0.3958	0.8736	0.2873	0.5579	0.5124	0.3865	0.1136
Dynamic R-CNN	ResNet-50	0.8426	0.8692	0.8557	0.3375	0.8327	0.2295	0.4936	0.4457	0.3249	0.1574
DH R-CNN	ResNet-50	0.8647	0.8873	0.8758	0.3589	0.8562	0.2674	0.5183	0.4618	0.3621	0.1353
DETR	ResNet-50	0.8958	0.9267	0.9110	0.4122	0.8893	0.3275	0.5724	0.5218	0.4018	0.1042
Sparse R-CNN	ResNet-101	0.8782	0.8879	0.8830	0.3624	0.8624	0.2587	0.5276	0.4835	0.3512	0.1218
ViTDet	ViT-B	0.9128	0.9385	0.9255	0.4215	0.9032	0.3386	0.5597	0.5197	0.4132	0.0872
DAB-DETR	ResNet-50	0.9067	0.9249	0.9157	0.4037	0.8924	0.3194	0.5482	0.4973	0.3976	0.0933
CO-DETR	Swin-L	0.9273	0.9486	0.9378	0.4326	0.9162	0.3417	0.5893	0.5468	0.4218	0.0727
MLFANet (Ours)	ResNet-50	0.9438	0.9652	0.9543	0.4583	0.9548	0.3578	0.6142	0.5679	0.4427	0.0562
	ResNet-101	**0.9521**	**0.9716**	**0.9617**	**0.4611**	**0.9602**	**0.3792**	**0.6279**	**0.5824**	**0.4536**	**0.0479**

TABLE 2 Comparison of category detection accuracy of different object detection methods, where the score in bold is the highest score.

Model	Backbone	Cube	Ball	Cylinder	HB	Plane	CC	SC	MB	Tire	ROV	mAP
Faster R-CNN	ResNet-50	0.8126	0.5247	0.7582	0.6978	0.8668	0.3576	0.6632	0.8345	0.2864	0.7238	0.6516
RetinaNet	ResNet-50	0.7834	0.4873	0.7763	0.6504	0.8174	0.2865	0.5983	0.7853	0.2981	0.7124	0.6195
Cascade R-CNN	ResNet-50	0.8345	0.5562	0.7956	0.7126	0.8972	0.4128	0.6895	0.8672	0.2573	0.7559	0.6779
CenterNet	ResNet-101	0.8872	0.6128	0.8325	0.7382	0.9203	0.3978	0.7326	0.8763	0.3012	0.8057	0.7105
Dynamic R-CNN	ResNet-50	0.8257	0.5369	0.8154	0.6893	0.8438	0.3736	0.6782	0.8325	0.2297	0.7354	0.6561
DH R-CNN	ResNet-50	0.8536	0.5738	0.7862	0.7025	0.9056	0.4265	0.7024	0.8614	0.2895	0.7564	0.6858
DETR	ResNet-50	0.8842	0.6297	0.8537	0.7458	0.9159	0.4758	0.7253	0.8713	0.3158	0.8126	0.7230
Sparse R-CNN	ResNet-101	0.8264	0.5261	0.7436	0.6951	0.8397	0.3695	0.6915	0.8427	0.2697	0.7327	0.6537
ViTDet	ViT-B	0.8976	0.6385	0.8423	0.7626	0.9234	0.4425	0.7456	0.9057	0.3387	0.8051	0.7302
DAB-DETR	ResNet-50	0.8653	0.5642	0.8535	0.7715	0.9386	0.3871	0.7167	0.8976	0.3254	0.8385	0.7158
CO-DETR	Swin-L	0.8946	0.6285	0.8761	0.7869	0.9527	0.4563	0.7315	0.9143	0.3592	0.8274	0.7428
MLFANet (Ours)	ResNet-50	0.9473	0.7942	0.9182	0.8213	0.9738	0.5138	0.7826	0.9485	0.4895	0.8573	0.8046
	ResNet-101	**0.9584**	**0.8157**	**0.9203**	**0.8306**	**0.9814**	**0.5264**	**0.8014**	**0.9526**	**0.5123**	**0.8615**	**0.8161**

HB, CC, SC, and MB denote the human body, circle cage, square cage, and metal bucket.

these compared object detection models still cannot achieve satisfactory detection accuracy. In contrast, the proposed MLFANet obtained AP values of 95.84%, 98.14%, and 95.26% for the large-size object categories, respectively. Additionally, for the other object categories such as cylinder, human body, square cage, and ROV, the proposed method achieved AP values of 92.03%, 83.06%, 80.14%, and 86.15%, which are the optimal results for all compared methods. The quantitative analysis results in Tables 1 and 2 show that the proposed method has significant advantages in solving sonar image object detection tasks. The reason is that MLFANet fully considers the interference of seabed reverberation noise, shadow region, and clutter information in the sonar images, and proposes corresponding solutions, so it can obtain better object detection accuracy. To further intuitively compare the performance of different object detection models, we drew the PR curve of different object detection models for comparison. The PR curve in Figure 10 demonstrates the performance of MLFANet compared to baseline models across various classification thresholds. The PR curve of MLFANet exhibits a higher AUC, indicating its ability to achieve both high precision and high recall. This is particularly

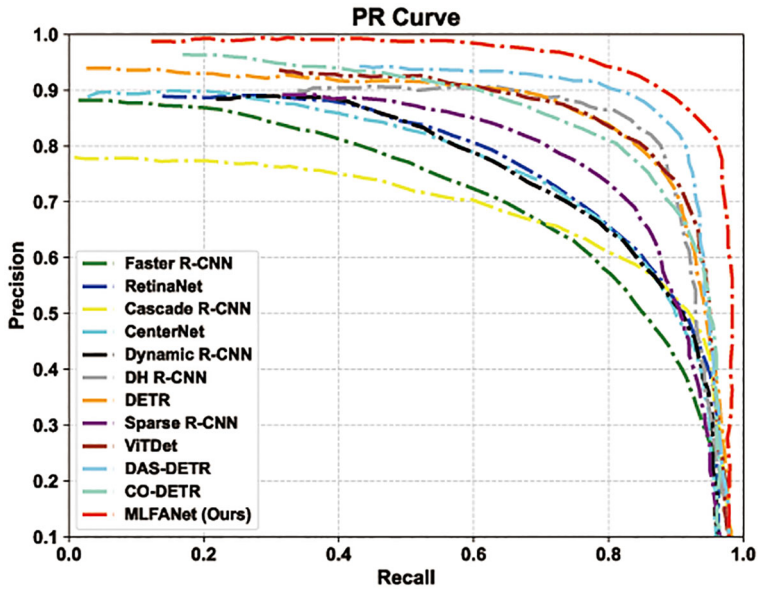


FIGURE 10
Comparison of PR curves for different object detection models.

important for FLS images, where the presence of noise, shadows, and reverberation can lead to false positives or missed detections. Compared to baseline models, MLFANet maintains a more gradual decline in precision as recall increases, reflecting its robustness to challenging underwater conditions. This is attributed to the integration of the LFAM, DFEM, and MFRM, which together enhance feature representation and reduce noise interference. Additionally, the CIoU-DFL loss function contributes to this improved performance by addressing class imbalance and refining object localization and classification. The precision value of MLFANet was the highest among all models, further supporting the superior performance of the proposed framework. This analysis highlights the effectiveness of MLFANet in achieving a favorable precision-recall trade-off, making it well-suited for underwater object detection.

4.4.2 Qualitative analysis

To further demonstrate the effectiveness of the proposed MLFANet, we visualized the prediction results of sonar images under different scene conditions contained in the UATD dataset. As shown in Figures 11–13, these scenes include seabed reverberation noise interference, shadow region interference, and object scale variation. It can be seen from the prediction results that the proposed method can accurately locate and recognize the different categories of sonar objects in the test images with high confidence scores. In contrast, the compared methods suffer from location deviation, high false alarm rate, and recognition failures. Additionally, as shown in Table 3, we present the confidence scores of different object detection models for the object categories in the test images. Following this, we present a detailed analysis of the different object detection model prediction results under three underwater scene conditions and the advantages of the constructed sonar object detector. The qualitative comparison results effectively illustrate the advantages of the proposed method for sonar object detection.

4.4.2.1 Superiority in scenes with seabed reverberation noise interference

The irregularity of underwater terrain seriously affects the propagation and reflection of sound waves on the seabed, so a forward-looking sonar image is disturbed by seabed reverberation noise. As shown in Figure 11, under the interference of seabed reverberation noise, it is difficult for the compared object detection models to obtain satisfactory detection results. For example, for CNN-based object detection models, Faster R-CNN (Girshick, 2015) and RetinaNet (Lin et al., 2017b) could not correctly detect all object categories in sonar images, resulting in false detection and missing detection. The reason is that the non-linear characteristics of seabed reverberation noise interfere with the detection and recognition process of CNN-based methods. For the Transformer-based object detection models, ViTDet (Li et al., 2022) and CO-DETR (Zong et al., 2023) obtained relatively better detection results. However, the results in Figure 11 show that these methods still struggle to accurately detect small-size object categories. In contrast, MLFANet effectively suppress the seabed reverberation noise interference on the feature extraction process,

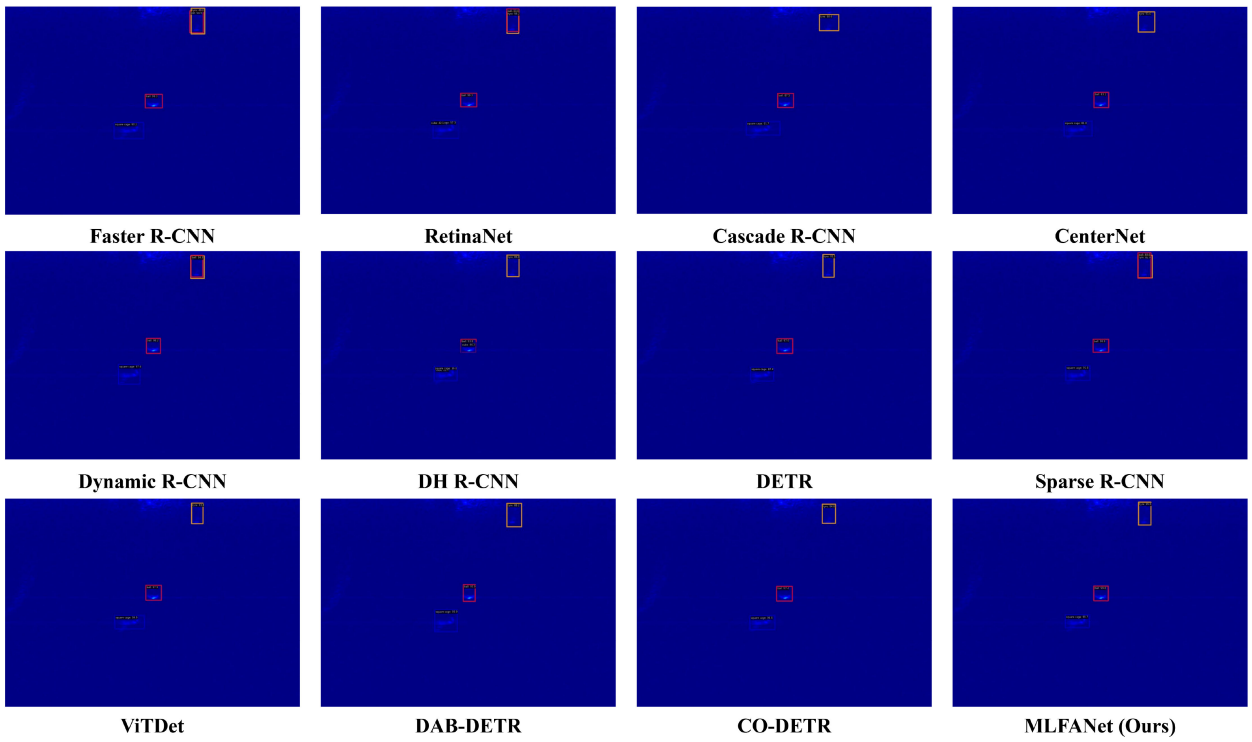


FIGURE 11
Visualization detection results of different object detection models in seabed reverberation noise interference scenes.

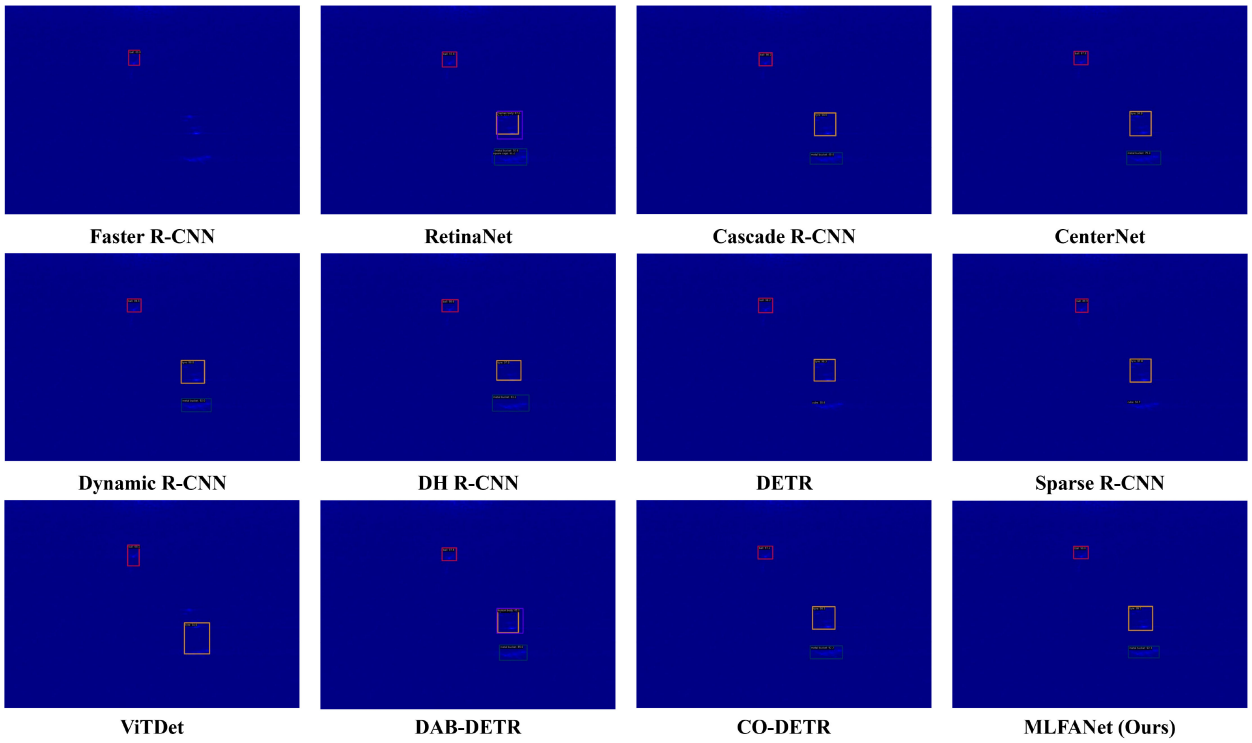


FIGURE 12
Visualization detection results of different object detection models on shadow region interference scenes.

successfully detects different object categories and obtains the higher confidence score. Moreover, in environments with strong seabed reverberation noise, MLFANet occasionally misclassifies noise patterns as objects due to their similar intensity and texture. Future work could focus on integrating advanced noise suppression or training with adversarial noise augmentation to mitigate this issue.

4.4.2.2 Superiority in scenes with shadow region interference

Since the underwater object has the characteristics of absorption, reflection, and scattering of sonar signal, it is difficult for the acoustic wave to directly penetrate the object entity, so the shadow interference region is formed in the reverse of the object region. The existence of the shadow region causes object occlusion, so it is difficult for the object detection model to accurately extract the edge, contour, and detail feature information. As shown in Figure 12, in the shadow interference scene, the compared sonar object detection models struggled to accurately locate and identify the object category and obtained a lower confidence score. Among the competitors, for CNN-based methods, CenterNet (Duan et al., 2019) obtained relatively better detection results. The reason is that the model uses a center point detection strategy to locate the object region, which can effectively alleviate the shadow region interference on the object feature extraction process. For the Transformer-based methods, CO-DETR obtained the optimal detection results. The reason is that it suppresses the representation of redundant feature information in the shadow region through global context modeling, and uses the position encoder mechanism to improve the object positioning accuracy. The proposed method obtains the optimal detection effect, which suppresses and filters the shadow feature interference by focusing on the discriminative feature information of the object region to improve the location and recognition accuracy. In addition, the objects located in regions with strong shadow interference are sometimes missed due to low contrast and insufficient discriminative features. Introducing adaptive contrast enhancement or attention mechanisms could help improve detection in such regions.

4.4.2.3 Superiority in scenes with object multi-scale transformation

Due to the influence of different object entities, object distance transformation, sonar beam angle, and object motion state, there are complex object scale transformation phenomena in the forward-looking sonar image. The variable object scale puts forward higher requirements for the multi-scale feature extraction capability of the object detection model. However, the existing object detection methods can only solve the multi-scale feature extraction problem of natural scene images, while multi-scale feature extraction for sonar images still cannot achieve satisfactory performance. As shown in Figure 13, for sonar images with different scale objects, the compared methods had false alarms and missing detection problems. Among competitors, Cascade R-CNN (Cai and Vasconcelos, 2019) and Dynamic R-CNN (Zhang et al., 2020), which use multi-scale feature extraction strategies, achieved relatively better results. The reason is that these methods

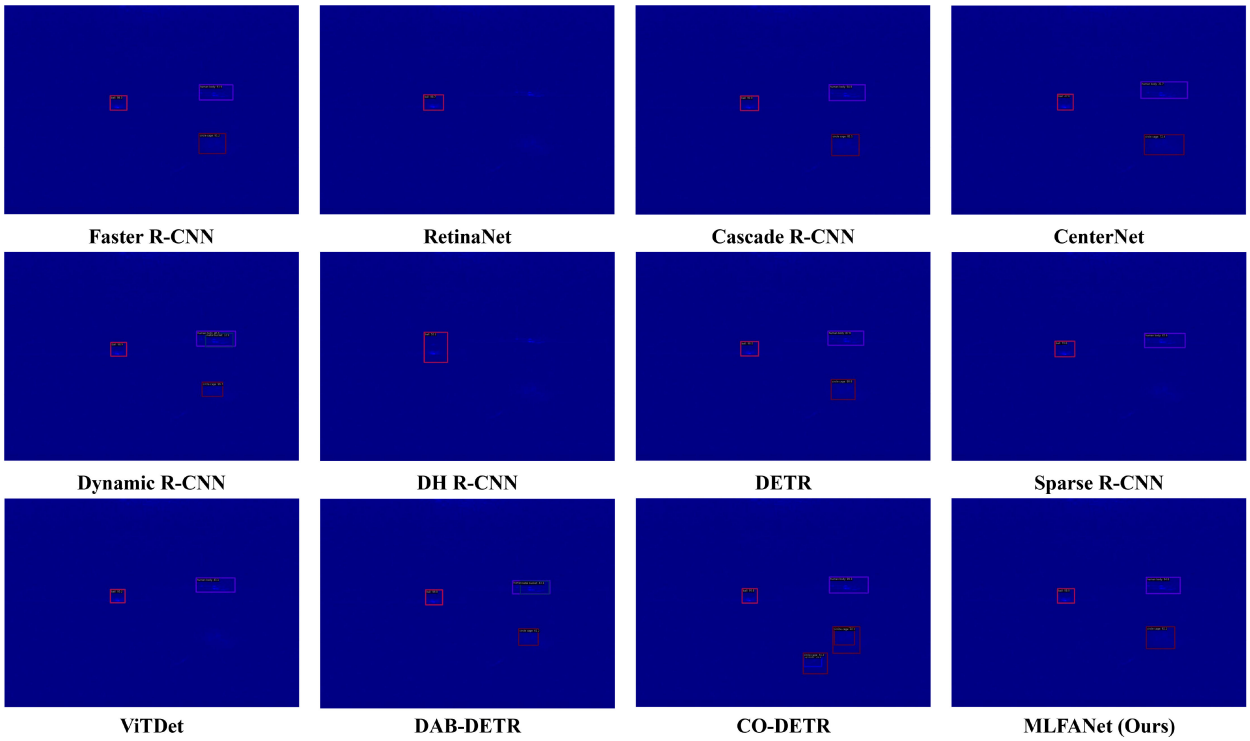


FIGURE 13
Visualization detection results of different object detection models on object multi-scale transformation scenes.

TABLE 3 Comparison of the confidence scores of different object detection methods.

Models	Reverberation noise scenes			Shadow interference scenes			Object scale transformation		
	NA	NO	Confidence	NA	NO	Confidence	NA	NO	Confidence
Faster R-CNN	3	4	80.2%, 99.3%, 64.9%, 49.0%	3	2	68.1%, 34.6%	3	2	95.2%, 45.2%
RetinaNet	3	5	42.1%, 57.3%, 99.3%, 53.6%, 96.2%	3	5	93.8%, 67.0%, 58.3%, 30.6%, 81.2%	3	1	98.7%
Cascade R-CNN	3	3	61.7%, 87.5%, 68.6%	3	3	86.2%, 94.0%, 49.4%	3	3	99.0%, 94.8%, 81.1%
CenterNet	3	3	81.0%, 83.2%, 51.0%	3	3	97.9%, 94.8%, 75.3%	3	3	47.0%, 31.7%, 72.4%
Dynamic R-CNN	3	4	97.6%, 98.2%, 34.3%, 28.6%	3	1	69.0%	3	3	98.3%, 93.9%, 91.2%
DH R-CNN	3	5	36.0%, 61.7%, 44.3%, 93.8%, 86.8%	3	3	98.6%, 97.3%, 61.2%	3	1	52.2%
DETR	3	3	87.4%, 97.0%, 95.5%	3	3	98.2%, 98.3%, 38.4%	3	3	98.0%, 87.6%, 88.8%
Sparse R-CNN	3	4	91.6%, 98.9%, 48.6%, 82.0%	3	3	89.9%, 97.8%, 50.7%	3	2	70.6%, 45.6%
ViDet	3	3	84.9%, 97.9%, 93.8%	3	3	98.3%, 98.8%, 92.0%	3	4	98.9%, 32.9%, 46.6%, 96.7%
DAB-DETR	3	3	93.9%, 93.9%, 88.2%	3	4	97.8%, 45.7%, 78.3%, 85.0%	3	4	98.0%, 43.4%, 54.2%, 61.2%
CO-DETR	3	3	96.6%, 97.4%, 94.8%	3	3	97.2%, 98.3%, 92.2%	3	6	99.8%, 99.3%, 40.2%, 83.9%, 31.2%
MLFANet	3	3	99.7%, 99.8%, 99.6%	3	3	99.0%, 99.5%, 92.3%	3	3	99.0%, 94.8%, 81.1%

HB, CC, SC, and MB denote the human body, circle cage, square cage, and metal bucket.

construct a multi-scale feature extraction structure, which can alleviate the influence of object scale transformation. In contrast, we can observe from Figure 12 that the Transformer-based object detection models were less effective for object scale variable scenarios. Taking the DAB-DETR (Liu et al., 2022) detector as an example, it only focuses on the efficient modeling of global information and ignores the extraction of scale-invariant features, which leads to missing detection and false alarm problems. The proposed MLFANet can effectively detect the different scale object categories in sonar images and obtain a higher confidence score. The reason is that the multi-scale feature refinement module can accurately locate sonar objects with different scales and obtain the robust invariant feature information in sonar image. Moreover, MLFANet struggles with extreme scale variations, leading to missed detections of very small objects or fragmented detections of very large targets. Developing more robust multi-scale feature fusion techniques or scale-invariant detection mechanisms could address this limitation.

4.4.3 Performance in small sample scenarios

To evaluate the potential of MLFANet for small sample learning, we conducted experiments by reducing the training dataset size to simulate limited data conditions. Specifically, 50%, 25%, and 10% of the original training data were used, while the test set remained unchanged. The performance of MLFANet and four representative baseline models (Faster R-CNN, RetinaNet, DETR, and CO-DETR) under these conditions is summarized in Table 4. The results in Table 4 demonstrate that MLFANet consistently outperforms the baseline models across all training data fractions. Notably, in extremely small sample conditions (10% training data), MLFANet achieved an AP of 30.85%, significantly surpassing Faster R-CNN (18.57%), RetinaNet (16.42%), DETR (24.17%), and CO-DETR (26.58%). This highlights the robustness and effectiveness of MLFANet in low-data conditions.

4.4.4 Computational complexity analysis

Since the sonar image object detection task has high requirements for algorithm real-time performance, we compared and analyzed the computational complexity of different object detection models, and the specific results are shown in Figure 14. Table 5 presents the number of parameters, FLOPs (Floating Point Operations), and FPS for each model on a workstation equipped with an NVIDIA RTX 4090 GPU. It can be seen from the comparison results that the CNN-based methods have advantages in computational complexity and real-time performance compared with the Transformer-based object detection models. To take the Transformer-based method ViTDet (Li et al., 2022) as an example, its calculation parameter reached 5,792 MB, and the inference speed

TABLE 4 Model performance verification under small sample conditions.

Data fraction	Model	AP (%)	AP50 (%)	AP75 (%)	AP^l (%)	AP^m (%)	AP^s (%)	FAR (%)
100%	Faster R-CNN	32.53	80.13	21.79	47.68	43.12	31.47	17.55
	RetinaNet	29.75	79.52	18.52	45.73	41.27	30.52	21.48
	DETR	41.22	88.92	32.75	57.24	52.18	40.18	10.42
	CO-DETR	43.26	91.62	34.17	58.92	54.68	42.18	7.27
	MLFANet (Ours)	46.11	96.02	37.92	62.79	58.24	45.36	4.79
50%	Faster R-CNN	29.20	76.80	18.50	43.12	39.84	28.74	19.80
	RetinaNet	26.85	74.30	15.67	41.45	37.20	26.32	23.67
	DETR	36.80	85.20	28.90	50.72	46.10	36.40	12.80
	CO-DETR	39.40	88.10	30.70	53.34	49.36	39.20	9.72
	MLFANet (Ours)	42.50	93.80	34.80	57.30	53.60	42.10	6.30
25%	Faster R-CNN	24.72	63.18	14.52	37.11	32.84	23.45	23.42
	RetinaNet	22.17	64.82	12.33	33.84	30.11	21.52	26.64
	DETR	31.25	76.41	23.74	43.55	39.62	30.18	15.63
	CO-DETR	33.84	80.12	26.24	46.85	42.62	33.51	12.92
	MLFANet (Ours)	37.58	86.74	30.66	49.25	46.13	36.35	8.87
10%	Faster R-CNN	18.57	55.42	9.83	25.17	22.52	15.68	27.92
	RetinaNet	16.42	51.27	8.28	21.84	19.43	14.36	30.65
	DETR	24.17	61.74	17.98	33.65	30.24	24.06	18.73
	CO-DETR	26.58	65.97	20.42	36.84	33.41	26.17	15.83
	MLFANet (Ours)	30.85	74.26	24.74	41.58	38.92	31.74	10.48

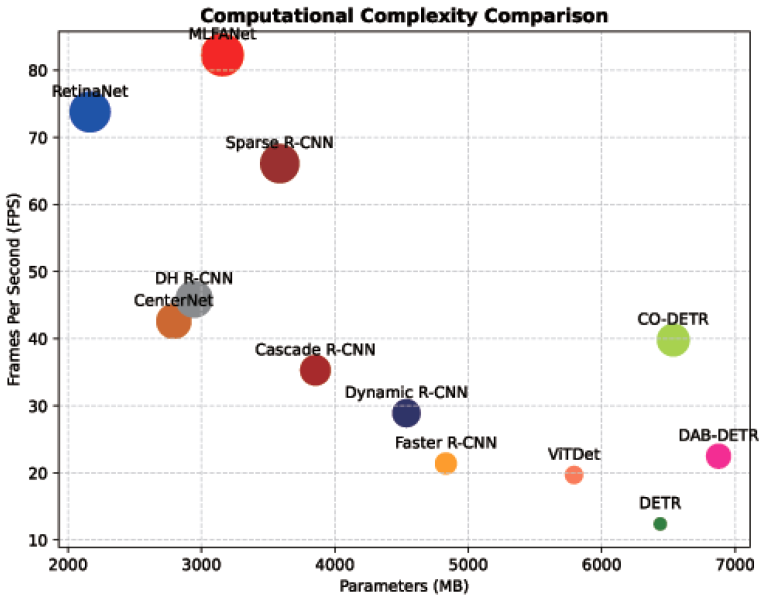


FIGURE 14
Comparison of our method with other object detection models in calculation parameters and inference speed.

was only 19.7 FPS. The reason is that the self-attention mechanism used in the Transformer model requires the calculation of the correlation of each pixel spatial position information, which increases the model inference time and calculation parameters. For the CNN-based methods, to take the Cascade R-CNN (Cai and Vasconcelos, 2019) as an example, the number of calculation parameters was 3,854 MB, and the inference speed reached 35.3 FPS. Although this method outperforms several Transformer-based object detection models, it still fails to address the real-time requirements of the sonar object detection task. In contrast, the computational parameter of the proposed MLFANet reached 3,157 MB, and the inference speed was 82.3 FPS, which was significantly better than the other object detection models. The proposed method can achieve an advantage because it constructs the corresponding feature extraction and fusion module for the sonar image, and effectively alleviates the influence of redundant feature and noise information on the inference process of the object detection model. To further validate the feasibility of MLFANet for deployment on embedded devices, experiments were conducted on an NVIDIA Jetson Xavier NX. The model was optimized using quantization techniques to reduce memory consumption and computational overhead. After optimization, MLFANet achieved an inference speed of 27.4 FPS with a memory footprint of 2.60 GB on the Jetson Xavier NX. These results demonstrate that MLFANet meets the real-time requirements of embedded systems, making it practical for AUV applications such as obstacle avoidance and object tracking.

TABLE 5 Comparison of the computational complexity of different object detection models.

Model	Parameter (MB)	FLOPs (G)	FPS (512×512 image size)
Faster R-CNN	4,138	207.1	32.6
RetinaNet	3,815	198.7	35.9
Cascade R-CNN	3,854	223.2	35.3
CenterNet	3,384	189.2	41.2
Dynamic R-CNN	4,052	204.7	30.1
DH R-CNN	4,327	212.3	29.6
DETR	5,748	350.2	19.4
Sparse R-CNN	4,877	298.1	22.3
ViTDet	5,792	420.5	17.8
DAB-DETR	6,015	368.4	18.5
CO-DETR	5,674	324.5	19.7
MLFANet (Ours)	3,157	183.4	82.3

4.5 Ablation study and analysis

To demonstrate the effectiveness of the important components LFAM, DFEM, and MFRM in the constructed MLFANet, we performed an ablation study on the UATD testing set, and the specific quantitative analysis results are shown in Table 6. In the experiment, we used the YOLOX detector (Ge et al., 2021) as the baseline model and verified the detector performance improvement by adding different components. Additionally, since the different constructed components are mainly for feature extraction and fusion of sonar images, we present the feature map visualization results of the different component modules in Figure 15. The specific analysis of the ablation study is as follows.

4.5.1 Effect of the LFAM

The constructed LFAM aims to fully exploit the low-level feature information such as texture, edge, and contour in the sonar image to improve the discriminating ability of the model for object region and background information. As shown in Table 6, when the LFAM was embedded into the baseline model, it achieved 68.74% (18.5% ↑) mAP on the testing set. Additionally, each object category experienced a corresponding increase in AP value, for example, the ball category had an increase of 25.58%, and the circle cage category had an increase of 22.13%. The feature visualization results corresponding to Figure 15 further show that the LFAM can make the model focus on feature extraction in the sonar object region and significantly enhance the model's feature representation ability for low-level feature information.

4.5.2 Effect of the DFEM

To filter the redundant feature information interference in the feature extraction process, the DFEM was constructed, which obtains the discriminative attributes of the object region by enhancing the local feature information representation in deep convolution. As shown in Table 6, when the DFEM was introduced into the baseline model, its mAP on the testing set reached 70.60%. Moreover, the DFEM enhanced the small object region feature representation, so that the AP values of the ball, circle cage and tire small object categories increased by 30.74%, 25.59%, and 9.60% respectively, and the AP values of the cube, plane and metal bucket large-size object categories increased by 15.69%, 11.47%, and 12.87% respectively. Combined with the LFAM and DFEM, the baseline model achieved significant performance improvement. Compared with the initial results, the

TABLE 6 Quantitative evaluation of the ablation study with different components, where the score in bold is the highest score.

Methods	Cube	Ball	Cylinder	HB	Plane	CC	SC	MB	Tire	ROV	mAP
Baseline	0.7153	0.3274	0.5865	0.4317	0.7528	0.2154	0.5171	0.7349	0.2054	0.5366	0.5024
+LFAM	0.8543	0.5832	0.7941	0.7352	0.8759	0.4367	0.7185	0.8437	0.2681	0.7639	0.6874
+DFEM	0.8722	0.6348	0.8364	0.7281	0.8675	0.4713	0.6985	0.8636	0.3014	0.7863	0.7061
+MFRM	0.8657	0.5962	0.8046	0.7524	0.8854	0.3762	0.6735	0.8893	0.3267	0.8127	0.6983
+LFAM+DFEM	0.8875	0.6584	0.8536	0.7782	0.9107	0.4685	0.7639	0.9164	0.4172	0.8311	0.7485
+LFAM+DFEM+MFRM	**0.9318**	**0.7653**	**0.8735**	**0.8094**	**0.9512**	**0.4956**	**0.7855**	**0.9381**	**0.4597**	**0.8535**	**0.7864**

NA denotes the number of actual objects and NO denotes the number of objects detected.

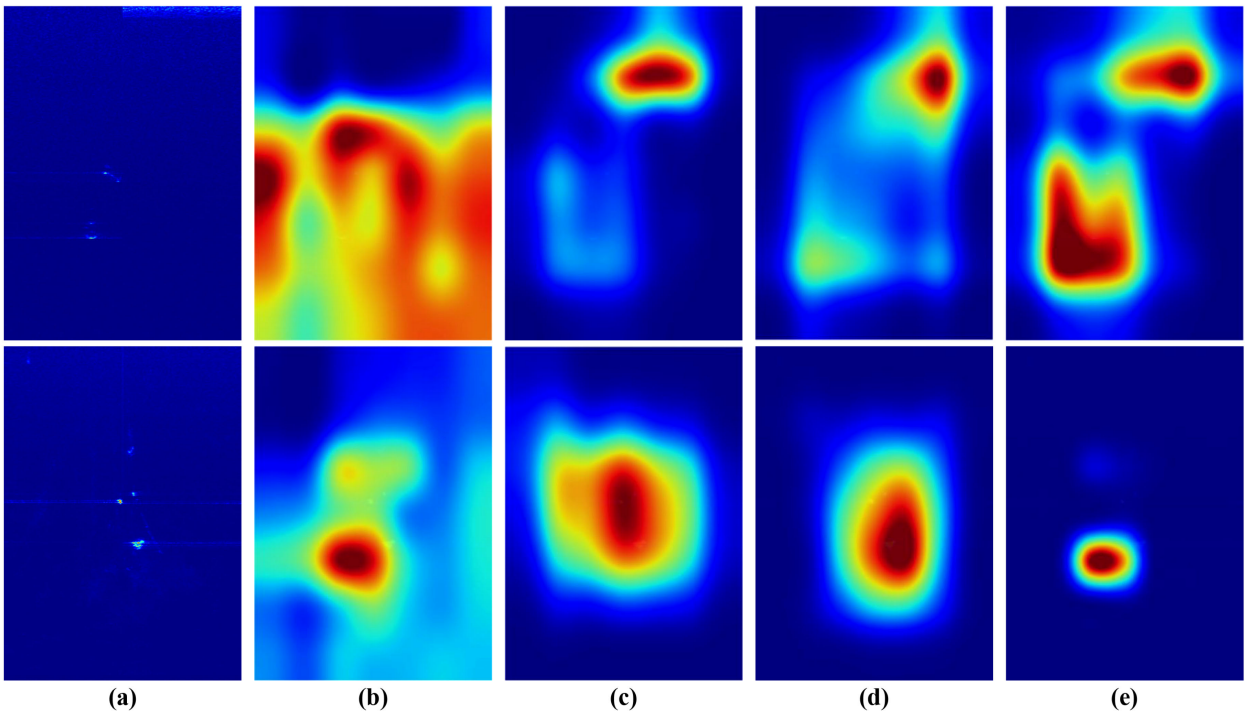


FIGURE 15
Visualizing the feature extraction effects of the LFAM, DFEM, and MFRM. **(a)** original forward-looking sonar images. **(b)** Feature extraction results from baseline. **(c)** Feature extraction results with the LFAM. **(d)** Feature extraction results with the DFEM. **(e)** Feature extraction results with the MFRM.

mAP increased by 24.61%, and the AP values for the cylinder, human body, square cage, and ROV increased by 26.71%, 34.65%, 24.64%, and 29.45%, respectively. The feature visualization results in Figure 15 show that DFEM can effectively filter the redundant feature information interference to improve the sonar object detection accuracy in clutter and shadow information interference scene.

4.5.3 Effect of the MFRM

To solve the problem of multi-scale feature extraction in seabed reverberation noise and shadow region interference scene, the MFRM was constructed, which obtains the scale-invariant features of sonar images by region location branch and feature refinement branch. Different from placing the LFAM and the DFEM in the feature extraction stage, we embedded the MFRM into the neck structure of the detector. As shown in Table 6, when placing the MFRM in the baseline model, it increased the mAP by 19.59%. Additionally, the model obtained a significant boost in AP values for object categories with different scales, for example, it increased by 26.88%, 16.08%, and 12.13% for the ball, circle cage, and tire, respectively. From the results shown in Figure 15, it can be observed that the use of the MFRM effectively improved the model's receptive field deformation ability, so that it could obtain the discriminative feature information of object regions with different scales. Notably, when combining the LFAM, DFEM, and MFRM, the baseline model performance was optimized, and the mAP value on the UATD testing set reached 78.64%, which further demonstrates the effectiveness of the different components in improving the detector performance.

5 Conclusion

To solve the problem of forward-looking sonar image object detection in complex underwater acoustic environment, in this article, we propose a novel multi-level feature aggregation network (MLFANet) to achieve an underwater sonar image object detection task. The proposed MLFANet contains three innovative modules, the LFAM, DFEM, and MFRM. Specifically, the LFAM is used to enhance the low-level feature information representation of sonar images to alleviate the influence of seabed reverberation noise on the feature extraction process. The DFEM enhances the saliency of object region features in deep convolution by constructing the correlation of local-global features to filter shadow and clutter information interference. The MFRM uses the region location and feature refinement branches to extract robust invariant feature information of different scale objects to solve the problem of underwater object multi-scale variation. To demonstrate the effectiveness and advantages of the proposed method, we conducted a series of experiments on a real-scene sonar image dataset, and MLFANet achieved better performance than the existing state-of-the-art methods. The ablation studies further validate the effectiveness and feasibility of the proposed different innovation modules. Although the proposed method can obtain better detection performance, it requires more training samples. Therefore, in future work, we intend to explore the forward-looking sonar image object detection method in small sample conditions.

Data availability statement

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

Author contributions

ZW: Writing – original draft, Writing – review & editing. JG: Writing – review & editing. SZ: Writing – review & editing. YZ: Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work is supported by the National Natural Science Foundation of China (61671465), the National Natural Science Foundation of China (61624931), the Neural Science Foundation of Shaanxi Province (2021JM-537), the Youth Talent Support Program of Shaanxi Science and Technology Association (23JK0701), the Xi'an Science and Technology Planning Projects (20240103), and the China Postdoctoral Science Foundation under Grant (2024M754225).

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

Abu, A., and Diamant, R. (2019). A statistically-based method for the detection of underwater objects in sonar imagery. *IEEE Sensors J.* 19, 6858–6871. doi: 10.1109/JSEN.7361

Cai, Z., and Vasconcelos, N. (2019). Cascade r-cnn: High quality object detection and instance segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 1483–1498. doi: 10.1109/TPAMI.2019.2956516

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., and Zagoruyko, S. (2020). "End-to-end object detection with transformers," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Glasgow, UK. 213–229.

Chandra, M. A., and Bedi, S. (2021). Survey on svm and their application in image classification. *Int. J. Inf. Technol.* 13, 1–11. doi: 10.1007/s41870-017-0080-1

Chen, L., Liu, C., Chang, F., Li, S., and Nie, Z. (2021). Adaptive multi-level feature fusion and attentionbased network for arbitrary-oriented object detection in remote sensing imagery. *Neurocomputing* 451, 67–80. doi: 10.1016/j.neucom.2021.04.011

Cheng, G., Si, Y., Hong, H., Yao, X., and Guo, L. (2020). Cross-scale feature fusion for object detection in optical remote sensing images. *IEEE Geosci. Remote Sens. Lett.* 18, 431–435. doi: 10.1109/LGRS.8859

Collins, M., Schapire, R. E., and Singer, Y. (2002). Logistic regression, adaboost and bregman distances. *Mach. Learn.* 48, 253–285. doi: 10.1023/A:1013912006537

Dong, X., Qin, Y., Gao, Y., Fu, R., Liu, S., and Ye, Y. (2022). Attention-based multi-level feature fusion for object detection in remote sensing images. *Remote Sens.* 14, 3735. doi: 10.3390/rs14153735

Du, B., Huang, Y., Chen, J., and Huang, D. (2023). "Adaptive sparse convolutional networks with global context enhancement for faster object detection on drone images." in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, Vancouver, Canada. 13435–13444. doi: 10.1109/CVPR52729.2023.01291

Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., and Tian, Q. (2019). "Centernet: Keypoint triplets for object detection," in *Proceedings of the IEEE/CVF international conference on computer vision (ICCV)*, Seoul, South Korea. 6569–6578.

Elharrouss, O., Akbari, Y., Almaadeed, N., and Al-Maadeed, S. (2022). Backbones-review: Feature extraction networks for deep learning and deep reinforcement learning approaches. *arXiv preprint arXiv:2206.08016*. doi: 10.48550/arXiv.2206.08016

Ge, Z., Liu, S., Wang, F., Li, Z., and Sun, J. (2021). Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430*. doi: 10.48550/arXiv.2107.08430

Girshick, R. (2015). "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision* (ICCV), Santiago, Chile. 1440–1448.

Gong, H., Mu, T., Li, Q., Dai, H., Li, C., He, Z., et al. (2022). Swin-transformer-enabled yolov5 with attention mechanism for small object detection on satellite images. *Remote Sens.* 14, 2861. doi: 10.3390/rs14122861

Grzadziel, A. (2020). Results from developments in the use of a scanning sonar to support diving operations from a rescue ship. *Remote Sens.* 12, 693. doi: 10.3390/rs12040693

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., et al. (2018). Recent advances in convolutional neural networks. *Pattern recognition* 77, 354–377. doi: 10.1016/j.patcog.2017.10.013

Guo, C., Fan, B., Zhang, Q., Xiang, S., and Pan, C. (2020). "Augfpn: Improving multi-scale feature learning for object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Seattle, WA, USA. 12595–12604.

Hansen, R. E., Callow, H. J., Sabo, T. O., and Synnes, S. A. V. (2011). Challenges in seafloor imaging and mapping with synthetic aperture sonar. *IEEE Trans. Geosci. Remote Sens.* 49, 3677–3687. doi: 10.1109/TGRS.2011.2155071

Jiang, L., Yuan, B., Du, J., Chen, B., Xie, H., Tian, J., et al. (2024). Mffsodnet: Multi-scale feature fusion small object detection network for uav aerial images. *IEEE Trans. Instrumentation Measurement* 73, 1–14. doi: 10.1109/TIM.2024.3381272

Karimanzira, D., Renkewitz, H., Shea, D., and Albiez, J. (2020). Object detection in sonar images. *Electronics* 9, 1180. doi: 10.3390/electronics9071180

Kim, B., and Yu, S.-C. (2017). "Imaging sonar based real-time underwater object detection utilizing adaboost method," in *2017 IEEE Underwater Technology (UT) (IEEE)*, Busan, South Korea. 1–5.

Kim, S.-W., Kook, H.-K., Sun, J.-Y., Kang, M.-C., and Ko, S.-J. (2018a). "Parallel feature pyramid network for object detection," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany. 234–250.

Kim, Y., Kang, B.-N., and Kim, D. (2018b). "San: Learning relationship between convolutional features for multi-scale object detection," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany. 316–331.

Kong, W., Hong, J., Jia, M., Yao, J., Cong, W., Hu, H., et al. (2019). Yolov3-dpfin: A dual-path feature fusion neural network for robust real-time sonar target detection. *IEEE Sensors J.* 20, 3745–3756. doi: 10.1109/JSEN.7361

Li, Z., Liu, F., Yang, W., Peng, S., and Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE Trans. Neural Networks Learn. Syst.* 33, 6999–7019. doi: 10.1109/TNNLS.2021.3084827

Li, Y., Mao, H., Girshick, R., and He, K. (2022). "Exploring plain vision transformer backbones for object detection," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Tel Aviv, Israel. 280–296.

Li, X., Wang, W., Wu, L., Chen, S., Hu, X., Li, J., et al. (2020). Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *Adv. Neural Inf. Process. Syst.* 33, 21002–21012. doi: 10.5555/3495724.3497487

Li, Z., Xie, Z., Duan, P., Kang, X., and Li, S. (2024). Dual spatial attention network for underwater object detection with sonar imagery. *IEEE Sensors J.* 24, 6998–7008. doi: 10.1109/JSEN.2023.3336899

Liang, X., Zhang, J., Zhuo, L., Li, Y., and Tian, Q. (2019). Small object detection in unmanned aerial vehicle images using feature fusion and scaling-based single shot detector with spatial context analysis. *IEEE Trans. Circuits Syst. Video Technol.* 30, 1758–1770. doi: 10.1109/TCSVT.76

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017a). "Feature pyramid networks for object detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Honolulu, HI, USA, 2117–2125.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017b). "Focal loss for dense object detection," in *Proceedings of the IEEE/CVF international conference on computer vision (ICCV)*, Venice, Italy, 2980–2988.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). "Microsoft coco: Common objects in context," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Zurich, Switzerland, 740–755.

Liu, S., Li, F., Zhang, H., Yang, X., Qi, X., Su, H., et al. (2022). Dab-detr: Dynamic anchor boxes are better queries for detr. *arXiv preprint arXiv:2201.12329*. doi: 10.48550/arXiv.2201.12329

Liu, X., Zhou, F., Zhou, H., Tian, X., Jiang, R., and Chen, Y. (2015). A low-complexity real-time 3-d sonar imaging system with a cross array. *IEEE J. Oceanic Eng.* 41, 262–273. doi: 10.1109/JOE.2015.2439851

Lu, X., Ji, J., Xing, Z., and Miao, Q. (2021). Attention and feature fusion ssd for remote sensing object detection. *IEEE Trans. Instrumentation Measurement* 70, 1–9. doi: 10.1109/TIM.2021.3118092

Ma, W., Wu, Y., Cen, F., and Wang, G. (2020). Mdfn: Multi-scale deep feature learning network for object detection. *Pattern Recognition* 100, 107149. doi: 10.1016/j.patcog.2019.107149

Miao, S., Du, S., Feng, R., Zhang, Y., Li, H., Liu, T., et al. (2022). Balanced single-shot object detection using cross-context attention-guided network. *Pattern recognition* 122, 108258. doi: 10.1016/j.patcog.2021.108258

Mustafa, H. T., Yang, J., and Zareapoor, M. (2019). Multi-scale convolutional neural network for multi-focus image fusion. *Image Vision Computing* 85, 26–35. doi: 10.1016/j.imavis.2019.03.001

Qin, Y., Yan, C., Liu, G., Li, Z., and Jiang, C. (2020). Pairwise gaussian loss for convolutional neural networks. *IEEE Trans. Ind. Inf.* 16, 6324–6333. doi: 10.1109/TII.9424

Redmon, J., and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*. doi: 10.48550/arXiv.1804.02767

Ren, S., He, K., Girshick, R., and Sun, J. (2016). 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

Shaw, P., Uszkoreit, J., and Vaswani, A. (2018). Self-attention with relative position representations. *arXiv preprint arXiv:1803.02155*. doi: 10.48550/arXiv.1803.02155

Sun, P., Zhang, R., Jiang, Y., Kong, T., Xu, C., Zhan, W., et al. (2023). Sparse r-cnn: An end-to-end framework for object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 15650–15664. doi: 10.1109/TPAMI.2023.3292030

Tang, L., Tang, W., Qu, X., Han, Y., Wang, W., and Zhao, B. (2022). A scale-aware pyramid network for multi-scale object detection in sar images. *Remote Sens.* 14, 973. doi: 10.3390/rs14040973

Wang, J., Feng, C., Wang, L., Li, G., and He, B. (2022c). Detection of weak and small targets in forward-looking sonar image using multi-branch shuttle neural network. *IEEE Sensors J.* 22, 6772–6783. doi: 10.1109/JSEN.2022.3147234

Wang, Z., Guo, J., Zeng, L., Zhang, C., and Wang, B. (2022d). Mlffnet: Multilevel feature fusion network for object detection in sonar images. *IEEE Trans. Geosci. Remote Sens.* 60, 1–19. doi: 10.1109/TGRS.2022.3214748

Wang, B., Ji, R., Zhang, L., and Wu, Y. (2022a). Bridging multi-scale context-aware representation for object detection. *IEEE Trans. Circuits Syst. Video Technol.* 33, 2317–2329. doi: 10.1109/TCSVT.2022.3221755

Wang, D., Shang, K., Wu, H., and Wang, C. (2022b). Decoupled r-cnn: Sensitivity-specific detector for higher accurate localization. *IEEE Trans. Circuits Syst. Video Technol.* 32, 6324–6336. doi: 10.1109/TCSVT.2022.3167114

Wang, C., and Wang, H. (2023). Cascaded feature fusion with multi-level self-attention mechanism for object detection. *Pattern Recognition* 138, 109377. doi: 10.1016/j.patcog.2023.109377

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., and Hu, Q. (2020). "Eca-net: Efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, Seattle, WA, USA, 11534–11542.

Xiao, J., Guo, H., Yao, Y., Zhang, S., Zhou, J., and Jiang, Z. (2022). Multi-scale object detection with the pixel attention mechanism in a complex background. *Remote Sens.* 14, 3969. doi: 10.3390/rs14163969

Xinyu, T., Xuewu, Z., Xiaolong, X., Jinbao, S., and Yan, X. (2017). "Methods for underwater sonar image processing in objection detection," in *2017 International conference on computer systems, electronics and control (ICCSEC)*, Dalian, China, 941–944.

Yasir, M., Liu, S., Pirasteh, S., Xu, M., Sheng, H., Wan, J., et al. (2024). Yoloshiptracker: Tracking ships in sar images using lightweight yolov8. *Int. J. Appl. Earth Observation Geoinformation* 134, 104137. doi: 10.1016/j.jag.2024.104137

Yuanzi, L., Xiufen, Y., and Weizheng, Z. (2022). Transyolo: high-performance object detector for forward looking sonar images. *IEEE Signal Process. Lett.* 29, 2098–2102. doi: 10.1109/LSP.2022.3210839

Zhang, M., Cai, W., Wang, Y., and Zhu, J. (2023). A level set method with heterogeneity filter for side-scan sonar image segmentation. *IEEE Sensors J.* 24, 584–595. doi: 10.1109/JSEN.2023.3334765

Zhang, H., Chang, H., Ma, B., Wang, N., and Chen, X. (2020). "Dynamic r-cnn: Towards high quality object detection via dynamic training," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Glasgow, UK. 260–275.

Zhang, H., Tian, M., Shao, G., Cheng, J., and Liu, J. (2022a). Target detection of forward-looking sonar image based on improved yolov5. *IEEE Access* 10, 18023–18034. doi: 10.1109/ACCESS.2022.3150339

Zhang, Y., Zhang, H., Liu, J., Zhang, S., Liu, Z., Lyu, E., et al. (2022b). Submarine pipeline tracking technology based on auvs with forward looking sonar. *Appl. Ocean Res.* 122, 103128. doi: 10.1016/j.apor.2022.103128

Zhang, M.-L., and Zhou, Z.-H. (2007). Ml-knn: A lazy learning approach to multi-label learning. *Pattern recognition* 40, 2038–2048. doi: 10.1016/j.patcog.2006.12.019

Zhao, Z., Wang, Z., Wang, B., and Guo, J. (2023). Rmfenet: Refined multi-scale feature enhancement network for arbitrary oriented sonar object detection. *IEEE Sensors J.* 23, 29211–29226. doi: 10.1109/JSEN.2023.3324476

Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., and Ren, D. (2020). "Distance-iou loss: Faster and better learning for bounding box regression," in *Proceedings of the AAAI conference on artificial intelligence*, New York, USA. 34. 12993–13000.

Zhou, T., Si, J., Wang, L., Xu, C., and Yu, X. (2022b). Automatic detection of underwater small targets using forward-looking sonar images. *IEEE Trans. Geosci. Remote Sens.* 60, 1–12. doi: 10.1109/TGRS.2022.3181417

Zhou, K., Zhang, M., Wang, H., and Tan, J. (2022a). Ship detection in sar images based on multi-scale feature extraction and adaptive feature fusion. *Remote Sens.* 14, 755. doi: 10.3390/rs14030755

Zhou, L., Zhao, S., Wan, Z., Liu, Y., Wang, Y., and Zuo, X. (2024). Mfefnet: A multi-scale feature information extraction and fusion network for multi-scale object detection in uav aerial images. *Drones* 8, 186. doi: 10.3390/drones8050186

Zhu, X., Cheng, D., Zhang, Z., Lin, S., and Dai, J. (2019). "An empirical study of spatial attention mechanisms in deep networks," in *Proceedings of the IEEE/CVF international conference on computer vision (ICCV)*, Seoul, South Korea. 6688–6697.

Zhu, Y., Zhao, C., Guo, H., Wang, J., Zhao, X., and Lu, H. (2018). Attention couplenet: Fully convolutional attention coupling network for object detection. *IEEE Trans. Image Process.* 28, 113–126. doi: 10.1109/TIP.2018.2865280

Zong, Z., Song, G., and Liu, Y. (2023). "Detrs with collaborative hybrid assignments training," in *Proceedings of the IEEE/CVF international conference on computer vision (ICCV)*, Paris, France. 6748–6758.

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 20 May 2025
DOI 10.3389/fmars.2025.1549513

OPEN ACCESS

EDITED BY
Muhammad Yasir,
China University of Petroleum (East China), China

REVIEWED BY
Zhiqiang Li,
Guangdong Ocean University, China
Surisetty Kumar,
Space Applications Centre (ISRO), India

*CORRESPONDENCE
Fahim Khan
fkhan4@ucsc.edu;
fkhan19@calpoly.edu

RECEIVED 21 December 2024
ACCEPTED 07 April 2025
PUBLISHED 20 May 2025

CITATION
Khan F, de Silva A, Palinkas A, Dusek G, Davis J and Pang A (2025) RipFinder: real-time rip current detection on mobile devices.
Front. Mar. Sci. 12:1549513.
doi: 10.3389/fmars.2025.1549513

RipFinder: real-time rip current detection on mobile devices

Fahim Khan[1]*, Akila de Silva[2], Ashleigh Palinkas[3], Gregory Dusek[4], James Davis[5] and Alex Pang[5]

[1]Department of Computer Science and Software Engineering, California Polytechnic State University, San Luis Obispo, CA, United States, [2]Department of Computer Science, San Francisco State University, San Francisco, CA, United States, [3]Scripps Institution of Oceanography, University of California, San Diego, San Diego, CA, United States, [4]National Ocean Service, National Oceanic and Atmospheric Administration, Silver Spring, MD, United States, [5]Department of Computer Science and Engineering, University of California, Santa Cruz, Santa Cruz, CA, United States

Rip currents present a significant safety risk to beach tourists and coastal communities, resulting in hundreds of annual drownings all over the world. A key contributing factor to this danger is the lack of awareness among beachgoers about recognizing and avoiding these rip currents. In response to this issue, we introduce RipFinder, a mobile app equipped with machine learning (ML) models trained to detect two types of rip currents. Users can leverage the app's computer vision capabilities to use their phone's camera to identify these hazardous rip currents in real time. The amorphous and ephemeral nature of rip currents makes it challenging to detect them with high accuracy using object detection models. To address this, we propose a client-server ML model-based computer vision system designed specifically to improve rip current detection accuracy. This novel approach enables the app to function with or without internet connectivity, proving particularly beneficial in regions without lifeguards or internet access. Additionally, the app serves as an educational resource, offering in-app information about rip currents. It also promotes citizen science involvement by encouraging users to contribute valuable information on detected rip currents. This paper presents the app's overall design and discusses the challenges inherent to the rip current detection system.

KEYWORDS

rip current detection, data collection, citizen science, coastal observation, computer vision, deep learning, mobile application

1 Introduction

Rip currents are dangerous, strong, fast-moving currents that pull swimmers away from the shore, often leading to drownings and fatalities. They pose a significant hazard to beachgoers and can easily overpower even strong, experienced swimmers. Rip currents are a global issue, affecting coastlines around the world (Zhang et al., 2021; Retnowati et al., 2012; Mucerino et al., 2021). In the United States alone, they account for an estimated 100 drownings a year (Gensini and Ashley, 2010). Rip currents can form suddenly and without obvious signs, which can catch swimmers off guard. While there are general conditions that

can lead to their formation, predicting exactly when and where they will appear is challenging. Furthermore, rip currents are created through various mechanisms and, as a result, exhibit different visual characteristics. This complexity of occurrence and variability in appearance makes them difficult to identify (Castelle et al., 2016). Consequently, many beachgoers lack the essential knowledge and awareness needed to recognize and avoid these perilous currents.

Rip current detection techniques are significantly important because of their potential to save lives. As a public safety issue, the implications extend beyond swimmers. Lifeguards, rescue teams, and even bystanders who try to help can also be put in danger. If rip currents could be detected reliably, then beachgoers and lifeguards could be alerted to the dangers in real-time. This would likely result in a significant decrease in the number of rip current-related incidents and fatalities. By providing more accurate information about rip currents, the general public could make more informed decisions about when it is safe to enter the water, thereby enhancing overall public safety. The development and deployment of tools, such as rip current prediction models (Dusek and Seim, 2013) or mobile apps that can detect and provide real-time alerts and tips about rip currents could be instrumental in these efforts.

While rip currents can often be visually identified by experienced swimmers, surfers, lifeguards, and coastal scientists, traditional detection and data collection methods typically involve *in-situ* instrumentation, such as GPS-equipped drifters and current meters (Leatherman, 2017; MacMahan et al., 2011). However, recent studies have demonstrated that images and video can also be used to detect rip currents. These approaches leverage computer vision and machine learning (ML) models for object detection to spot and identify these potentially dangerous phenomena (de Silva et al., 2021; Silva et al., 2023; Dumitriu et al., 2023; Maryan et al., 2019; Mori et al., 2022; Philip and Pang, 2016; Rampal et al., 2022; Rashid et al., 2021). However, detecting and segmenting rip currents with high accuracy using ML methods presents unique challenges due to their amorphous and ephemeral nature. Given the potentially fatal nature of dangerous rip currents, their detection is a matter of life and death. Thus, high accuracy and reliability are crucial for any rip current detection tool to issue warnings and take preventive actions to decrease the number of rip current-related incidents. Providing such capability for real-world use, i.e., on mobile platforms, adds another layer of technical challenge.

Many object detection ML models can detect rip currents, but the challenge lies in deploying these models in real-time on mobile devices with limited power and computational resources. More accurate yet computationally resource-intensive, ML models cannot run directly on mobile devices. By sending the visual input for object detection to a remote server, it can be achieved on mobile devices. However, this approach is not always feasible, especially in beach locations where server connectivity is unavailable. Alternatively, mobile-optimized ML models can feasibly run using the limited computational resources of portable devices without server connectivity but at the cost of sacrificing accuracy.

To address these challenges, we introduce a mobile application, or app, designed to detect rip currents using ML models for computer vision. Users can identify potential rip currents in real-time by simply aiming their phone's camera toward the ocean. We propose a client-server system of object detection models to balance the trade-off between computational speed and accuracy. Depending on the mobile device's available computational resources and internet connectivity, this app employs one or more ML models to identify rip currents. If the device is relatively new and has adequate computational resources, our app runs two different types of mobile-optimized ML models to enhance the reliability of rip current detection. For older, resource-constrained devices, only one ML model is used. Moreover, when internet connectivity is available, part of the visual data is transmitted to a server for further verification of the detection using a more accurate large model. Our system combines client-server architecture with multiple ML model-based computer vision to enhance the accuracy and reliability of rip current detection. The novelty of our solution lies in its implementation of this combined system, allowing the app to function both with and without internet connectivity. Our app's versatility is especially invaluable in areas where lifeguards are absent or internet access is limited, establishing it as a crucial tool for public safety.

In addition to rip current detection, our app places a strong emphasis on educating users about the dangers of rip currents through informative in-app content and links to additional resources. Our aim is to empower beach enthusiasts with the knowledge necessary to make informed decisions, protecting themselves and others from these hazardous rip currents. Moreover, our app includes a citizen science feature, enabling users to contribute to scientific knowledge. This is done by encouraging them to record and share data, such as geotagged images and videos, along with additional information about detected rip currents. Harnessing the collective power of app users, we can gather valuable data that improves our understanding of rip currents and helps verify existing rip current forecast models. Ultimately, this leads to the development of more effective safety measures and strategies.

The contributions of this paper are as follows:

- Introduction of RipFinder: a mobile app designed for real-time, vision-based rip current detection.
- Development of a client-server system tailored for the ML models utilized in the rip current detection app.
- A comprehensive analysis and comparison of state-of-the-art ML models for rip detection.

2 Related work

2.1 Real-time object detection

Developing a mobile application for effectively and reliably identifying rip currents necessitates real-time object detection capabilities. Deep learning has revolutionized the field of object detection, as well as other computer vision tasks. Convolutional neural networks (CNNs) have become the standard method for

these applications. Numerous large and intricate models, such as Faster R-CNN—a two-stage regionbased detector (Ren et al., 2015)—and DETR (Detection Transformers)—an object detector based on the Transformer architecture (Carion et al., 2020)—offer remarkable accuracy in object detection tasks. For instance, Faster R-CNN has been adeptly used for real-time object detection in drones by connecting to a remote GPU server (Lee et al., 2017). However, these detectors often bear significant computational complexity, rendering them difficult to deploy on mobile or embedded platforms for real-time performance. An earlier server-based system named Glimpse, offering continuous, real-time object recognition for mobile devices, was introduced by Chen et al. (2015). Nonetheless, server-reliant systems prove impractical in locations devoid of internet connectivity.

Achieving accurate and reliable real-time object detection on mobile devices without depending on servers presents inherent challenges. Numerous efforts have been directed toward integrating deep learning methods on mobile devices by creating compact, mobile-optimized ML models. Typically, streamlined architectures, like one-stage CNNs, render the models lightweight, allowing them to function swiftly on mobile devices—making them an ideal choice for real-time object detection. The primary compromise for such efficiency is a minor decrease in accuracy relative to their more elaborate counterparts (Huang et al., 2017). We scrutinized a range of mobile-optimized ML models to ascertain the best fit for our system. SSD-MobileNetV2 (Sandler et al., 2018) stood out as one of the earliest trustworthy models tailored for mobile platforms. Among the contemporary one-stage models refined for mobile devices are variants of RT-DETR (Lv et al., 2023), EfficientDet (Tan et al., 2020), and YOLO (Jocher et al., 2022). Our investigation encompassed a comprehensive evaluation of potential ML models suitable for real-time rip current detection using computer vision on mobile platforms.

2.2 Rip current detection with ML

Given its impact on public safety, the problem of automated rip current detection has been approached using various methods, some of which predate the emergence of deep learning techniques. For example, Philip and Pang (2016) utilized optical flow on video sequences to discern the predominant flow towards the sea, aiding human observers in rip current detection. Maryan et al. (2019) employed modified Haar cascade methods to detect rip currents from time-averaged images. The concept of rip current detection via deep learning-based methods is not entirely new either. de Silva et al. (2021) were among the early adopters of deep learning methods for rip current detection, employing Faster R-CNN, a large model that achieved high accuracy. They introduced a frame aggregation technique that bolstered detection accuracy for fixedposition cameras, but this technique was not suitable for moving cameras. Mori et al. (2022) offered a flow-based method to accentuate and depict rip currents for human observers. However, this approach also demands a stationary camera and serves as a visualization tool rather than an automated detection system. In recent years, there have been several scholarly works about new deep learning model-based rip current detection techniques. For instance, Rashid et al. (2021) and Zhu et al. (2022) presented RipDet and YOLO-Rip, respectively. These lightweight rip current detection models, rooted in Tiny-YOLOv3 and YOLOv5s, belong to the smaller members of the YOLO family and are adept for environments with limited computational power. Rampal et al. (2022) showcased that the mobile-optimized, single-stage model SSD-MobileNetV2 can achieve performance metrics comparable to Faster R-CNN. Furthermore, Dumitriu et al. (2023) explored and compared various iterations of YOLOv8 for rip current segmentation. Silva et al. (2023) unveiled RipViz, an innovation that examines 2D vector fields and interprets pathline behaviors to pinpoint rip currents. Like that of Dumitriu et al. (2023), this method highlights the rip region's shape but identifies currents based on water movement rather than water appearance. Yet, while there is an assortment of effective rip current detection methods employing ML, a real-world application—such as a mobile app—primed for public safety and enhancing awareness for tangible societal impact remains elusive. This work endeavors to fill that void by devising a deployable mobile device-based real-time system for rip current detection.

3 System design and methods

3.1 System architecture

Figure 1 presents an overview of the RipFinder system architecture. Our comprehensive system, designed to effectively identify and alert users of rip currents, is organized into two primary components:

1. The client mobile app serves as the primary user interface. Within this app, we have integrated four ML models, each tailored specifically for mobile devices. As the device processes real-time visual input, these models evaluate the data and issue warnings if rip currents are detected. Depending on the device's processing power, the app can deploy either one or two ML models for detection. More modern devices with substantial resources can utilize two types of mobile-optimized ML models simultaneously, enhancing the reliability of rip current detection. In contrast, older devices with limited resources might default to a single model. Nevertheless, the ultimate decision to use one or two models rests with the user. When feasible, the app suggests users employ two models for optimal detection, but they retain the freedom to choose only one from the available options if preferred.
2. Our system's server-side employs complex ML models that demand significant computational resources and GPU capabilities, ensuring rip currents are detected with high accuracy. When a user captures an image or video via our mobile app, this data is sent to the server for in-depth analysis. After the server-side models process the data, the detection results are relayed back to the mobile app.

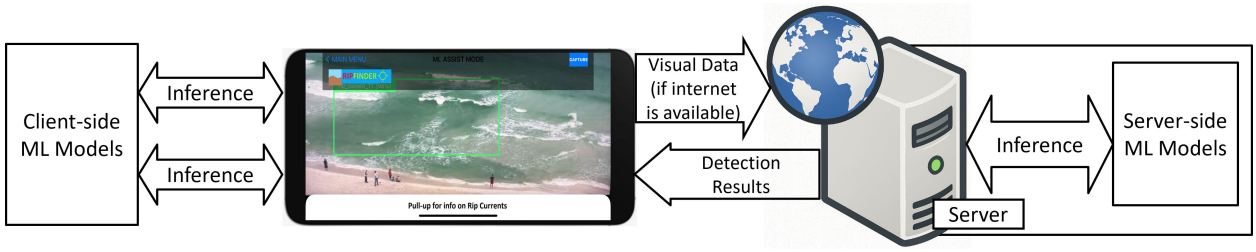


FIGURE 1
The high-level system architecture of RipFinder.

Additionally, we offer the option to execute multiple models on the server, depending on its capabilities (number of CPUs and GPUs, system memory, etc.), enhancing reliability through redundancy.

Our system attempts to improve the reliability of rip current detection in a two-fold way. The use of two models enhances detection reliability on the client app, even though it demands more computational resources. Server-side models, being complex and larger, boast superior accuracy, thus ensuring that server-aided rip current detection is more reliable when internet access is available. The client-side model, meanwhile, operates using the on-device computational resources without the need for an internet connection. The results section further elaborates on the justification behind these two design choices. Thus, our system's design allows it to operate both online and offline.

Training datasets are essential for training both client-side and server-side ML models. We developed our dataset by utilizing the existing dataset from de Silva et al. (2021) and supplementing it with a large amount of our own data. Further details on the dataset and the ML model training process are explained in the implementation section.

3.2 Mobile apps

Figure 2 provides a visual representation of our mobile app's user interface, offering an intuitive, user-friendly environment. We created both Android and iOS versions of the mobile app. The application's design caters to a variety of user needs and includes the following features:

3.2.1 Live camera and visualization tool

The app offers a live camera feature to capture the seashore and serves as a real-time visualizer, placing bounding boxes around detected rip currents in the view, thus acting as an immediate warning system (Figure 2b).

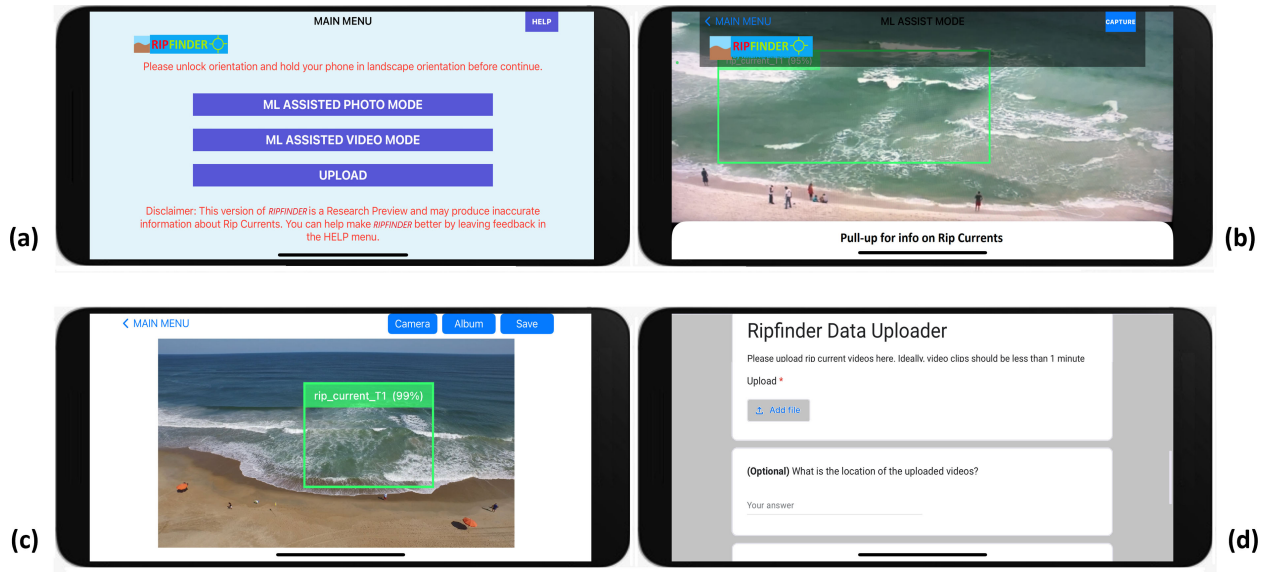


FIGURE 2
GUI of RipFinder App **(a)** Main menu, **(b)** Real-time detection from live camera view, **(c)** Detection from single image, **(d)** Data uploader for citizen science contribution.

3.2.2 ML model selection

From the in-app menu, users can choose the ML model for real-time rip current detection. On devices with higher computational resources, users have the option to turn on or off the use of two models in parallel for increased reliability.

3.2.3 Image and video recording

The app enables real-time rip current detection and the recording of images and videos, letting users document and share potential rip currents with other beachgoers and rip current researchers.

3.2.4 Rip current detection tool for existing images

RipFinder app analyzes existing images on the phone to identify rip currents, offering retrospective insights to users (Figure 2c).

3.2.5 Educational resources

Our app features an educational hub with resources on rip currents, accessible via a pull-up menu and help menu, ensuring users always have information at hand (Figure 2b).

3.2.6 Data upload tool

We integrated a data upload tool (Figure 2d) for users to share geotagged rip current images and observations, fostering community collaboration and enhancing our dataset for improved algorithm refinement.

3.3 Client-side ML models

In our application, RipFinder, we integrate several mobile-optimized ML models, all trained on a rip current dataset for client-side detection. These models have been tailored to ensure swift and efficient performance on mobile devices, which facilitates real-time rip current detection. The current version of RipFinder incorporates the following models:

3.3.1 YOLOv8n and YOLOv8m

YOLOv8, the latest in the YOLO series known for fast object detection (Redmon et al., 2016; Jocher et al., 2023), includes variants like YOLOv8n (nano) and YOLOv8m (medium) optimized for mobile devices. Its architecture facilitates single-pass detections, making it ideal for real-time applications such as rip current detection.

3.3.2 EfficientDet D0 and EfficientDet D2

EfficientDet, known for its object detection prowess (Tan et al., 2020), has a unique scalable architecture that adjusts to computational resources, making it ideal for mobile use; it offers eight variants, D0 to D7, based on image size.

Of the four ML models at our disposal, the app selects one or two mobile-optimized models for rip current detection, contingent upon a device's computational prowess and internet connectivity. Modern, high-end devices employ two models, while the older, resource-constrained devices resort to just one. YOLOv8n and EfficientDet D0, due to their lesser computational demand, are ideally deployed as standalone models or in conjunction with dated or less competent mobile devices. In contrast, YOLOv8m and EfficientDet D2 are better aligned with newer devices boasting significant computational strength.

3.4 Server-side ML models

Server-side, we engage a collection of high-performance ML models tailored for more resource-intensive computations. Given their demanding computational needs, these models are perfectly positioned for server-side deployment, capitalizing on robust hardware resources, including GPUs. For the server side, we've selected:

3.4.1 YOLOv8l and YOLOv8x

The YOLOv8 'l' (large) and 'x' (extra-large) variants (Jocher et al., 2023) are more complex than their mobile-optimized versions, offering higher accuracy but requiring greater computational power, ideal for situations demanding utmost accuracy with ample resources.

3.4.2 Real-time detection transformer

RT-DETR, a real-time adaptation of the DETR transformer-based object detection model (Lv et al., 2023; Carion et al., 2020), maintains DETR's accuracy while ensuring faster performance. We trained its large and extra-large versions, RT-DETR-L and RT-DETR-X, for server-side use.

By leveraging these server-side models that can deliver high accuracy, we bolster the final verification of detected rip currents, reinforcing the reliability of our rip current detection tool.

4 Implementation

Various components of our system were implemented using the latest available technology.

4.1 Dataset

Our training dataset distinguishes between two types of rip currents based on their visual features. The first, termed *bathymetry-controlled rip*, is characterized by areas devoid of breaking waves, presenting as darker and calmer regions flanked by brighter waves. The second, known as *transient rip*, is identified by water discoloration due to sediment plumes that extend beyond the breaking waves. Though both classes represent rip currents, their visual features differ significantly. Detecting one type of rip current with an ML model trained on data from another type is unfeasible. Treating these two types as a single class compromises the effectiveness of the trained model. The label correlograms in the Figure 3 illustrate the distinctions between the two classes based on the labeled regions of images from each class.

For the bathymetry-controlled rip current category, we utilized a dataset consisting of 1780 images made publicly available by de

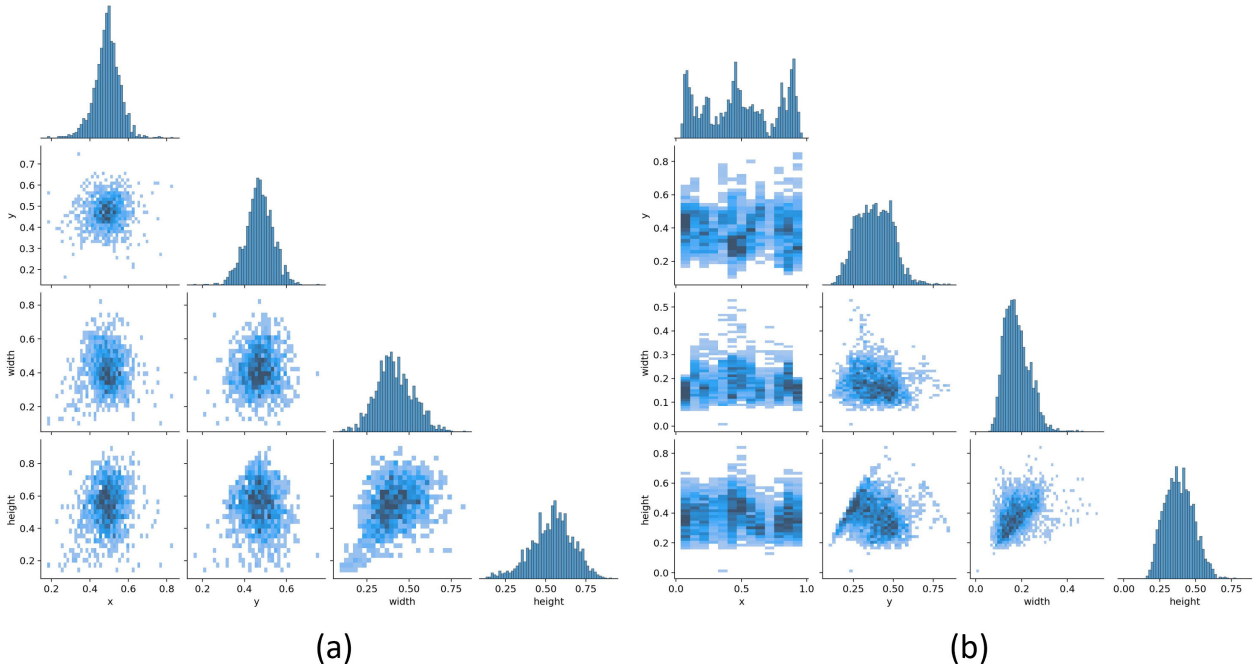


FIGURE 3
The label correlograms for **(a)** bathymetry controller rips and **(b)** transient rips, depicted in the figure, illustrate the distinctions between the two classes based on the labeled regions of images from each class.

Silva et al. (2021). For the transient rip current category, we curated a new dataset comprising 7565 labeled images. These were selectively extracted from videos captured by a drone, which focused on the visual signature of transient rip currents, and a Wi-Fi camera set up specifically for monitoring rip currents. We combined both datasets to train our model in the detection of the two rip current types. This dataset was then divided into an 80:20 split for training and validation, with 80% allocated for training purposes and the remaining 20% used for validation. The efficacy of the trained models was assessed using a series of test videos. Figure 4 showcases a selection of images from our dataset.

It is important to include imagery from diverse geographic regions and environmental conditions to enhance model robustness. Our dataset includes images from publicly available sources, drone footage, and fixed-location cameras. We incorporated the de Silva et al. (2021) dataset, which features satellite imagery from diverse regions. To enhance generalization, we are collaborating with coastal research partners to expand data collection across varied wave conditions, lighting, and water characteristics. Additionally, our citizen science initiative allows users to contribute images, enriching the dataset. While expanding the dataset and refining models is an ongoing effort, it remains independent of RipFinder's core architecture, as the ML models can be continuously updated with improved datasets.

4.2 ML model training and evaluation

We conducted ML model training on an AWS cloud server equipped with eight vCPUs, 61 GB of memory, and an NVIDIA Tesla V100 GPU boasting 16 GB of video memory. The EfficientDet models were trained using the TensorFlow library, while the YOLOv8 and RT-DETR models were trained with the Ultralytics library, which is based on PyTorch. All model trainings were initialized with a maximum of 500 epochs. For all versions of YOLOv8 and RT-DETR, a patience parameter of 50 was set. The patience parameter defines the number of epochs to wait before halting training via early stopping if there's no improvement in performance on a validation dataset. Since the EfficientDet models do not allow for the definition of a patience parameter, we monitored convergence through TensorBoard and manually terminated the training once convergence was observed. All models converged within 300 epochs. We trained all models from scratch, instead of using transfer learning with MS COCO pretrained models from the ML libraries, to prevent negative transfer (Wang et al., 2019). This decision was made because our rip current class data domain is distinct from any of the classes in the MS COCO2017 dataset (Lin et al., 2014).

4.3 Client apps and server

We developed the iOS version of the app in Swift using Xcode, and wrote the Android version in Java with Android Studio. To ensure broad accessibility, we tested the RipFinder app on a wide range of mobile devices, including both high-end and low-end models. While Table 1 presents results from the iPhone 12 Pro (2020) and Google Pixel 6 (2021), which served as our primary development devices, we also validated the app's performance on older and more budget-friendly models such as the Samsung A50 (2019), Samsung S23 (2023), LG G3 (2014), LG G5 (2016), and

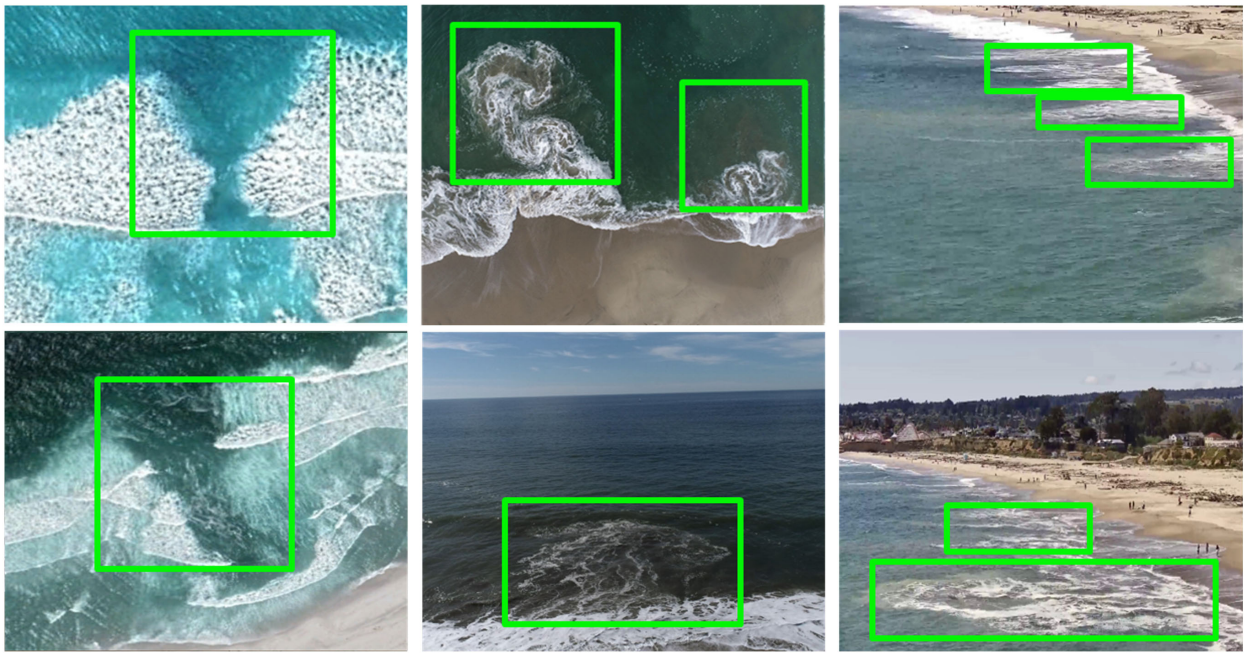


FIGURE 4
Some examples from our training dataset. The images on the first column are from the dataset by de Silva et al. (2021). The images on the second and third columns are from the dataset we collected using a drone and a wireless rip activity monitoring camera, respectively.

Xiaomi Redmi 10A (2022), and older iPhones such as the iPhone XR (2018). This extensive testing confirms that the app performs efficiently across a diverse spectrum of hardware, greatly enhancing its real-world applicability.

The server-side components were programmed in Python. We evaluated the server-side ML models on a desktop server equipped with a 16-core Intel Core i9 3.2 GHz CPU, 30 GB of memory, and an NVIDIA RTX3080 GPU with 10 GB of video memory.

4.4 Data privacy and security measures

Ensuring data privacy and security is a core aspect of RipFinder, particularly for citizen science contributions. All uploaded images, videos, and metadata are encrypted to prevent unauthorized access. Personally identifiable information is anonymized before storage, and location data is collected only with user consent, then obscured or aggregated to prevent tracking. We adhere to institutional ethical guidelines and restrict data access to authorized researchers who validate contributions. Users receive clear terms of use and can request data removal. Our retention policies prevent unnecessary long-term storage, ensuring responsible data handling while supporting rip current research.

4.5 Quality control and validation of user-uploaded data

To ensure the accuracy and reliability of citizen science contributions, RipFinder employs a multi-step validation process combining automated filtering, metadata verification, and expert review.

4.5.1 Automated screening and metadata verification

All user-uploaded images and videos first undergo computer vision-based pre-screening, which filters out irrelevant or low-quality submissions. Additionally, metadata, such as location, timestamp, and environmental conditions, is cross-referenced with rip current forecasts from NOAA and other sources. Any inconsistencies flag submissions for further review.

4.5.2 Expert validation and continuous improvement

Flagged submissions undergo manual review by rip current specialists, including NOAA scientists and coastal researchers, ensuring only verified data is incorporated into the dataset. A continuous feedback loop refines detection accuracy by improving machine learning models over time. Verified contributors may also receive recognition, fostering quality participation.

By integrating automated detection with expert validation, RipFinder ensures that only high-confidence, research-grade data supports scientific analysis and rip current safety efforts.

5 Results and discussion

5.1 Performance analysis of ML models

In this section, we present a performance analysis and comparison of state-of-the-art (SOTA) object detection models tailored for rip current detection. We compared ML models including EfficientDet D0, EfficientDet D1, EfficientDet D2, YOLOv8n, YOLOv8s, YOLOv8m,

10.3389/fmars.2025.1549513

TABLE 1 Comparison of ML models: performance metrics and resource utilization.

ML Model Properties	EfficientDet-D0	EfficientDet-D2	YOLOv8n	YOLOv8m	YOLOv8l	YOLOv8x	RT-DETR-L	RT-DETR-X
Model Size on Server (MB)	13.70	18.50	6.00	49.60	83.60	130.40	63.00	129.00
Avg. FPS on Server	37	21	127	106	86	79	47	35
Model Size on Phone (MB)	4.23	7.04	6.00	49.60	83.60	130.40	63.00	129.00
Avg. FPS on iPhone 12 Pro	48	15	25	17	Not Applicable	Not Applicable	Not Applicable	Not Applicable
Avg. FPS on Pixel 6	26	8	29	18	Not Applicable	Not Applicable	Not Applicable	Not Applicable

YOLOv8l, YOLOv8x, RT-DETR-l, and RT-DETR-x. To gauge the accuracy of these models, we utilized nine test videos annotated with ground truth data. To ensure model generalization and robustness, we validated RipFinder using diverse test videos from independent sources. Four of these videos were selected for their relevance to our rip current detection objectives from the test set introduced by de Silva et al. (2021). Additionally, three videos were drone-captured by us, while the last two originated from a wireless camera at webcoos.org dedicated to rip current monitoring.

While our model validation primarily utilized video data captured from elevated perspectives, we acknowledge that real-world user applications will often involve videos recorded at ground level. However, rip currents exhibit distinct visual characteristics that remain detectable even from a beach-level viewpoint. Lifeguards and experienced swimmers routinely identify rip currents using precisely these visual cues. With appropriate training datasets, ML models can similarly leverage these visual indicators to detect rip currents.

To evaluate the effectiveness of our object detection models, we use Intersection over Union (IoU) as the primary performance metric. Our evaluation methodology follows the object detection benchmarking approach outlined by Padilla et al. (2020), which provides a standardized toolbox for computing IoU. This toolbox calculates the ratio of overlap between the predicted and ground truth bounding boxes, allowing for a precise and objective assessment of detection accuracy.

Unlike classification tasks that rely on confusion matrices, object detection inherently requires spatial accuracy in addition to detection presence. IoU directly accounts for true positives, false positives, and localization precision, making it a more suitable metric for this study. Given the established use of IoU in object detection benchmarks, additional metrics such as precision, recall, and F1 score are not required to support our results and would be redundant in this context.

Our accuracy assessment followed the methodology described by de Silva et al. (2021), where:

$$accuracy = \frac{correct\_labels}{total\_frames}$$

Frames were considered classified as correct if the detected bounding boxes had an Intersection over Union (IoU) score versus ground truth bounding boxes above 0.3. IoU is calculated as:

$$IoU = \frac{area\_of\_intersection}{area\_of\_union}$$

The comparison results are presented in Table 2, and some examples of detected rip currents are shown in Figure 5. Based on these results, we can justify the following two design choices we made.

5.2 Statistical analysis of model performance

While the per-video accuracy results in Table 2 offer a general comparison, we further examined whether the observed accuracy differences among models are statistically meaningful. We treated

TABLE 2 We compared the detection accuracy of the SOTA methods to select the best options for the client and server application.

Test Videos	Client Side Models						Server Side Models			
	EfficientDet			YOLOv8					RT-DETR	
	D0	D1	D2	n	s	m	l	x	L	X
Rip_test_video_1	1.00	1.00	1.00	0.94	0.72	0.99	0.99	0.93	1.00	1.00
Rip_test_video_2	0.99	0.86	1.00	0.01	0.01	0.05	0.20	0.05	1.00	0.99
Rip_test_video_3	0.86	0.84	0.79	0.58	0.30	0.71	0.46	0.53	0.90	0.93
Rip_test_video_4	0.27	0.79	0.72	0.00	0.00	0.04	0.00	0.00	0.85	0.89
Rip_test_video_5	0.73	0.91	1.00	0.76	0.50	1.00	1.00	1.00	1.00	1.00
Rip_test_video_6	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.86	1.00
Rip_test_video_7	0.99	1.00	1.00	0.19	0.35	0.93	1.00	1.00	1.00	1.00
Rip_test_video_8	0.70	0.71	0.71	0.00	0.00	0.00	0.15	0.29	0.76	0.80
Rip_test_video_9	1.00	1.00	1.00	0.21	0.24	0.62	0.71	0.63	1.00	1.00
Average Accuracy	**0.73**	**0.79**	**0.91**	**0.30**	**0.24**	**0.48**	**0.50**	**0.49**	**0.93**	**0.96**

Bold values in the last row represent the average detection accuracy for each model variant across all test videos, used to evaluate overall model performance.

each of the nine test videos as repeated measurements, one "sample" per model, yielding paired accuracy data for each model across the same videos. Below, we illustrate a straightforward method using 95% confidence intervals (CIs) for each model's mean accuracy. These intervals help gauge whether model performance truly differs on average or if apparent differences could be due to sampling variability Rainio et al. (2024).

The mean accuracy of each model was computed by averaging the detection accuracy across all test videos. The standard deviation was calculated to measure the variability in performance across different video samples. The standard error of the mean was calculated as:

$$\mathrm{SE} = \frac{\mathrm{StDev}}{\sqrt{9}},$$

To quantify the uncertainty in these estimates, we determined the 95% confidence interval (CI). We used a two-sided t-distribution (given the small sample size) with 8 degrees of freedom (n − 1 = 9 − 1) and a critical value of ≈ 2.306 for 95% confidence:

$$95\%\ \mathrm{CI} = \bar{x} \pm t_{0.975,\, df=8} \times \mathrm{SE}.$$

A higher-variance model yields a wider interval, possibly overlapping intervals of both stronger and weaker performers. Large differences in means with minimal interval overlap typically point to genuine performance gaps, but borderline cases call for further pairwise statistical testing (e.g., with multiplecomparison corrections).

Table 3 reports the mean accuracy, standard deviation, and 95% CIs for each model. Although the average accuracies match Table 2, the confidence intervals offer insight into the consistency of each model's performance across videos.

5.2.1 Findings and interpretation

From the statistical analysis, the RT-DETR-X model achieved the highest mean accuracy ($\mu = 0.96$) with a very narrow confidence interval ($CI = [0.91,1.00]$), indicating consistent and highly reliable performance. Similarly, RT-DETR-L ($\mu = 0.93$) and EffDet-D2 ($\mu = 0.91$) demonstrated high accuracy with relatively low variability, confirming their robustness for rip current detection. Conversely, YOLOv8n, YOLOv8s, and YOLOv8m exhibited the lowest mean accuracies and the widest confidence intervals, reflecting high variability and inconsistent detection performance. The EffDet-D0 and EffDet-D1 models, while moderately accurate, showed greater performance fluctuations due to their wider confidence intervals.

5.2.2 Implications for model selection

The statistical findings reinforce the rationale behind selecting EffDet-D2 and YOLOv8n for mobile deployment, as they balance accuracy and efficiency. Meanwhile, RT-DETR-L and RT-DETR-X were the most reliable server-side models, offering superior accuracy with minimal variability. These insights confirm that our chosen client-server hybrid approach effectively optimizes both computational efficiency and real-time detection performance. By incorporating statistical validation, we ensure that model selection is based on empirical evidence rather than raw accuracy alone. This strengthens the reliability of RipFinder as a robust and scientifically validated rip current detection tool.

5.3 Other considerations

5.3.1 Running two ML models to increase accuracy

While running multiple models demands more computational resources, it enhances reliability. This design decision stems from the understanding that ML models with varying architectures possess distinct strengths and shortcomings. Research by Mekhalfi et al. (2022) indicates that models from the YOLO family tend to identify more objects, even if their precision varies.

FIGURE 5
Some examples of detected rip currents from our test videos.

In contrast, EfficientDet provides more stable and accurate detection. In many cases, one of the models might not detect specific instances of rip currents, even if they were trained using the same data. For instance, although the rip current in "Rip test video 6" can be detected by EfficientDet D2, it isn't identified by any other mobile models. Thus, deploying two models ensures that a challenging-to-detect rip current is more likely to be detected on a more capable device. Additionally, since rip current detection pertains to safety, minimizing false negatives is more crucial than avoiding excessive false positives. Therefore, while employing two models might seem redundant for general applications, it is beneficial for the purpose of rip current detection.

5.3.2 Two models vs. three or more

The decision to use two models on the client side balanced accuracy, computational demands, and processing time. Although running more than two models could improve detection accuracy through ensemble techniques, the benefits were minimal compared to the significant increase in resource consumption and latency.

Additional models would heavily strain server CPU and GPU resources, leading to higher costs and potential delays during peak usage. Increased latency from more models would compromise real-time detection, critical for user safety. The two-model setup already offers robust redundancy, ensuring reliable detection even if one model underperforms. The combination of YOLOv8l for broad detection and RT-DETR-L for detailed analysis provides a well-rounded solution.

After evaluating various models, EfficientDet-D2 and YOLOv8n were selected for mobile deployment due to their optimal balance of speed, accuracy, and compact size. For server-side operations, YOLOv8l and RT-DETR-L were chosen to maximize accuracy and reliability, enabling effective online and offline functionality. The findings, summarized in Table 1, highlight models that meet both hardware constraints and application needs for proficient rip current detection.

TABLE 3 Mean accuracy, standard deviation, and 95% confidence intervals (CIs) for each model.

Model	Mean	StDev	SE	95% CI
EffDet-D0	0.73	0.36	0.12	(0.45, 1.01)
EffDet-D1	0.79	0.31	0.10	(0.55, 1.03)
EffDet-D2	0.91	0.13	0.04	(0.81, 1.01)
YOLOv8n	0.30	0.37	0.12	(0.02, 0.58)
YOLOv8s	0.24	0.26	0.09	(0.04, 0.44)
YOLOv8m	0.48	0.45	0.15	(0.13, 0.83)
YOLOv8l	0.50	0.43	0.14	(0.17, 0.83)
YOLOv8x	0.49	0.43	0.14	(0.16, 0.82)
RT-DETR-L	0.93	0.09	0.03	(0.86, 1.00)
RT-DETR-X	0.96	0.07	0.02	(0.90, 1.01)

The confidence intervals indicate the range in which each model's true mean accuracy is likely to lie, based on nine test videos.

5.3.3 Running ML models on both the client and server side

More advanced and complex models, such as RT-DETR-L and RT-DETR-X, achieve higher accuracy but are limited to server execution. Thus, when an internet connection is available, server-assisted rip current detection becomes more reliable. The client-side models serve as the primary object detection mechanism, ensuring that rip current detection operates at the highest possible accuracy both with and without internet connectivity.

5.4 Evaluation and model selection

5.4.1 Addressing detection bias

Different machine learning models exhibit varying performance across rip current types, leading to detection bias in some cases. For

example, EfficientDet-D0 struggles with transient rip currents, showing a higher false-negative rate. This discrepancy arises due to differences in model architectures, feature extraction capabilities, and training data distribution. Models optimized for certain visual cues, such as wave breaks in bathymetry-controlled rips, may not generalize as well to transient rips, which often exhibit diffuse, sediment-laden water patterns.

Rather than refining a single model, RipFinder employs a multi-model strategy to balance detection accuracy and computational efficiency. This approach ensures adaptability, allowing the system to leverage mobile-optimized models for real-time detection while utilizing more powerful server-side models when internet access is available. Table 2 compares model performance, highlighting trade-offs between accuracy, speed, and resource constraints.

While this work prioritizes flexibility over single-model optimization, we recognize the importance of improving individual model performance. Future efforts will focus on fine-tuning models using more diverse datasets and reducing false negatives in challenging conditions. By continuously integrating improved architectures and expanded training data, RipFinder will further enhance detection reliability for all rip current types.

5.4.2 Evaluation

Among the ten (10) models highlighted in Table 2, we chose eight (8) for further evaluation. From the less accurate EfficientDet D0 and D1 variants, we selected only D0 because of smaller size. YOLOv8s was similarly excluded due to its poor accuracy. We evaluated the chosen models on a server equipped with a single GPU, an iPhone 12 Pro, and a Google Pixel 6 to determine the best-fit models for each platform (Table 1). Our benchmarking of each model's performance focused on two primary metrics:

1. We evaluated the real-time responsiveness of each model by measuring the frames processed per second (FPS). This metric offers insights into the model's speed and its ability to detect rip currents in real-time scenarios. EfficientDet-D0 and YOLOv8n exhibited higher FPS on mobile devices, marking them as optimal choices for devices with limited computational capabilities. Meanwhile, the enhanced accuracy of EfficientDet-D2 makes it a reliable option while still maintaining real-time performance.
2. Each model's storage footprint needs to be considered for embedding them in a mobile app, given that mobile devices have diverse storage capabilities and may also be running other apps simultaneously. Assessing a model's storage needs ensures that the application remains streamlined and does not overtax the device's memory. While the compactness of EfficientDet-D0 and YOLOv8n makes them as ideal for devices with resource constraints, the relatively small size and superior performance of EfficientDet-D2 make it a trustworthy option.

To further validate the practical applicability of our system, we extended our device testing to include low-end and older Android models. While certain high-end devices demonstrated superior performance, models such as the Xiaomi Redmi 10A and Samsung A50 successfully ran RipFinder, demonstrating that the app is not solely dependent on flagship devices.

5.5 Model performance evaluation

5.5.1 EfficientDet-D0 and D2

EfficientDet-D0 was notable for its high FPS, making it responsive on mobile devices, but it sometimes struggled with detecting transient rip currents in complex backgrounds, leading to occasional false negatives. On the other hand, EfficientDet-D2, while slightly slower, offered higher accuracy in distinguishing rip currents from similar water patterns, making it a more reliable choice for detailed analysis despite its larger storage requirements.

5.5.2 YOLOv8 variants

YOLOv8n excelled in real-time performance due to its compact size and speed, effectively detecting well-defined rip currents but occasionally missing subtler ones. YOLOv8m balanced speed and accuracy, handling both bathymetry-controlled and transient rip currents consistently, making it suitable for mobile deployment. The larger YOLOv8l and YOLOv8x models used server-side provided superior accuracy, detecting even faint rip currents, though their size and computational demands restricted them to server environments. YOLOv8s was excluded due to poor accuracy, particularly in complex scenarios.

5.5.3 RT-DETR variants

RT-DETR-L and RT-DETR-X, designed for server use, offered high accuracy and reliability, excelling in differentiating rip currents from similar patterns like wave shadows and sandbars. Their complex architecture required substantial computational resources, making them suitable only for server-side deployment.

6 Limitations and future work

While RipFinder is designed to improve rip current detection using diverse datasets and a hybrid clientserver architecture, certain limitations remain.

6.1 Dataset scope and generalization

Our dataset includes rip current images from multiple independent sources, such as NOAA, coastal research partners, and public sources (de Silva et al., 2021; Mori et al., 2022). However, we recognize that geographic and environmental variations may still impact model generalization, particularly in detecting rip currents under unique wave conditions or in less studied coastal regions. To mitigate these effects and improve generalization, we are:

- Expanding the dataset by incorporating images from diverse geographic locations and environmental conditions.
- Using data augmentation techniques, such as lighting adjustments, resolution scaling, and viewpoint shifts, to simulate different acquisition conditions.
- Leveraging citizen science contributions to introduce more real-world variability, ensuring models encounter a wider range of rip current appearances.

Future work will include a systematic evaluation of model generalization across different data sources and acquisition methods to further reduce bias and improve detection accuracy in real-world applications.

6.2 Server dependency and offline functionality

The client-server hybrid architecture enhances detection accuracy by leveraging more powerful models on the server. However, we acknowledge that server dependency may limit real-time detection in areas with poor or no internet connectivity. To mitigate this, RipFinder is designed to function independently using on-device models, ensuring continued usability in offline scenarios, although with a trade-off in detection accuracy.

6.3 Potential biases in model training

Training data biases may influence model performance, particularly in detecting less common rip current types. To improve fairness and generalizability, we plan to conduct further bias analysis, integrate domain adaptation techniques, and continuously refine the dataset to address potential imbalances.

6.4 Robustness in complex marine environments

Rip current detection is inherently challenging in extreme conditions, such as strong waves, light variations, and surface reflections. While multi-model detection improves reliability, some edge cases remain difficult to classify. Detection failures often occur when transient rips blend into background wave activity, making them harder to distinguish. As RipFinder is model-agnostic, future iterations can integrate more advanced models specifically trained for challenging marine conditions. Additionally, ongoing data collection through citizen science contributions will help refine model generalization, ensuring greater robustness over time.

6.5 Real-world usability from beach-level perspective

Another key consideration is the real-world usability of the app when deployed by users at beach level rather than from an elevated viewpoint. Although our current dataset primarily includes images captured from drones and other high vantage points, we recognize the importance of validating detections from ground-level perspectives. Future work will involve expanding our dataset to incorporate user-submitted images and videos captured at beach level, enabling the machine learning models to generalize more effectively across various viewing angles. Additionally, we plan to implement citizen-science feedback loops to continuously refine model accuracy based on real-world user data.

7 Conclusion

In this paper, we introduce Ripfinder, a mobile app equipped with an ML-based computer vision tool designed to mitigate the safety hazards associated with rip currents, which are a leading cause of drownings globally. Ripfinder features a sophisticated system that ensures rip current detection even in the absence of internet connectivity, making it indispensable in regions without lifeguards or reliable internet coverage. This capability is crucial for enhancing beach safety in remote and underserved areas.

Beyond its detection capabilities, Ripfinder enriches user knowledge with in-app informational content and videos about rip currents, helping users understand the dangers and how to avoid them. This educational component is vital for raising awareness and promoting safe behaviors at the beach. A standout feature of Ripfinder is its inclusion of citizen science. By inviting users to share data about identified rip currents, the app not only enhances scientific understanding but also fosters community engagement. This participatory approach leverages the collective efforts of users to contribute valuable data that can be used for further research and analysis, ultimately improving the overall understanding of rip current patterns and behaviors.

Ripfinder's integration of public safety, education, and scientific progress underscores its multifaceted approach to ensuring safer beach outings. By combining advanced technology with user engagement and educational resources, Ripfinder aims to create a comprehensive solution that addresses both immediate safety concerns and long-term scientific goals. The app exemplifies how modern technology can be harnessed to address real-world problems, making beaches safer and more enjoyable for everyone.

Data availability statement

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

Author contributions

FK: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. AdS: Conceptualization, Data curation, Validation,

Writing – original draft, Writing – review & editing. AsP: Funding acquisition, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing. GD: Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. JD: Conceptualization, Formal analysis, Investigation, Supervision, Visualization, Writing – original draft, Writing – review & editing. AlP: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work is partially funded by the following grants: Southeast Coastal Ocean Observing Regional Association (SECOORA) sub-award from NOAA award number NA20NOS0120220, the US Coastal Research Program (USCRP) through a Sea Grant award number NA23OAR4170121, and a grant from the UCSC Center for Coastal Climate Resilience.

Acknowledgments

We are grateful for the support and permissions granted by the California State Parks System and the Monterey Bay National Marine Sanctuary to operate our drone for this and related research. We also thank the Santa Cruz Port District, the California State Park Lifeguards and the other beta testing participants for their invaluable assistance in evaluating our system and providing feedback to help us improve RipFinder. We also acknowledge the support from the Google Cloud Research Credits program and AWS Cloud Credit for Research.

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.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. Grammarly and GPT-4 was used in order to improve readability and language during the preparation of this work. After using these tools, we reviewed and edited the content as needed and take full responsibility for the content of the publication.

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.

Author disclaimer

The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of SECOORA, NOAA, or UCSC. The content of the information provided in this publication does not necessarily reflect the position or the policy of the government, and no official endorsement should be inferred.

References

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., and Zagoruyko, S. (2020). "End-to-end object detection with transformers," in *European conference on computer vision* (Glasgow, UK: Springer), 213–229.

Castelle, B., Scott, T., Brander, R., and McCarroll, R. (2016). Rip current types, circulation and hazard. *Earth Sci. Rev.* 163, 1–21. doi: 10.1016/j.earscirev.2016.09.008

Chen, T. Y.-H., Ravindranath, L., Deng, S., Bahl, P., and Balakrishnan, H. (2015). "Glimpse: Continuous, real-time object recognition on mobile devices," in *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems* (Association for Computing Machinery, SenSys '15, New York, NY, USA), 155–168. doi: 10.1145/2809695.2809711

de Silva, A., Mori, I., Dusek, G., Davis, J., and Pang, A. (2021). Automated rip current detection with region based convolutional neural networks. *Coastal Eng.* 166, 103859. doi: 10.1016/j.coastaleng.2021.103859

Dumitriu, A., Tatui, F., Miron, F., Ionescu, R. T., and Timofte, R. (2023). "Rip current segmentation: A novel benchmark and YOLOv8 baseline results," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. (Vancouver, BC, Canada: IEEE) 1261–1271.

Dusek, G., and Seim, H. (2013). A probabilistic rip current forecast model. *J. Coastal Res.* 29, 909–925. doi: 10.2112/JCOASTRES-D-12-00118.1

Gensini, V. A., and Ashley, W. S. (2010). An examination of rip current fatalities in the United States. *Natural Hazards* 54, 159–175. doi: 10.1007/s11069-009-9458-0

Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., et al. (2017). "Speed/accuracy trade-offs for modern convolutional object detectors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (Honolulu, Hawaii: IEEE), 7310–7311.

Jocher, G., Chaurasia, A., and Qiu, J. (2023). *Ultralytics YOLOv8*. (San Francisco, CA: GitHub).

Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., Kwon, Y., Michael, K., et al. (2022). *ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation* (Geneva, Switzerland: Zenodo).

Leatherman, S. B. (2017). Rip current measurements at three south florida beaches. *J. Coastal Res.* 33, 1228–1234. doi: 10.2112/JCOASTRES-D-16-00124.1

Lee, J., Wang, J., Crandall, D., Šabanović, S., and Fox, G. (2017). "Real-time, cloud-based object detection for unmanned aerial vehicles," in *2017 First IEEE International Conference on Robotic Computing (IRC)* (Taichung, Taiwan: IEEE), 36–43. doi: 10.1109/IRC.2017.77

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). "Microsoft coco: Common objects in context," in *European conference on computer vision* (Zurich, Switzerland: Springer), 740–755.

Lv, W., Xu, S., Zhao, Y., Wang, G., Wei, J., Cui, C., et al. (2023). *DETRs beat YOLOs on real-time object detection*. (Seattle, Washington: IEEE).

MacMahan, J., Reniers, A., Brown, J., Brander, R., Thornton, E., Stanton, T., et al. (2011). An introduction to rip currents based on field observations. *J. Coastal Res.* 27, iii–ivi. doi: 10.2112/JCOASTRES-D-11-00024.1

Maryan, C., Hoque, M. T., Michael, C., Ioup, E., and Abdelguerfi, M. (2019). Machine learning applications in detecting rip channels from images. *Appl. Soft Comput.* 78, 84–93. doi: 10.1016/j.asoc.2019.02.017

Mekhalfi, M. L., Nicolò, C., Bazi, Y., Rahhal, M. M. A., Alsharif, N. A., and Maghayreh, E. A. (2022). Contrasting YOLOv5, Transformer, and EfficientDet detectors for crop circle detection in desert. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5. doi: 10.1109/LGRS.2021.3085139

Mori, I., de Silva, A., Dusek, G., Davis, J., and Pang, A. (2022). Flow-based rip current detection and visualization. *IEEE Access* 10, 6483–6495. doi: 10.1109/ACCESS.2022.3140340

Mucerino, L., Carpi, L., Schiaffino, C. F., Pranzini, E., Sessa, E., and Ferrari, M. (2021). Rip current hazard assessment on a sandy beach in liguria, nw mediterranean. *Natural Hazards* 105, 137–156. doi: 10.1007/s11069-020-04299-9

Padilla, R., Netto, S. L., and Da Silva, E. A. (2020). "A survey on performance metrics for object-detection algorithms," in *2020 international conference on systems, signals and image processing (IWSSIP) (IEEE)*. (Niteroi, Brazil: IEEE), 237–242.

Philip, S., and Pang, A. (2016). "Detecting and visualizing rip current using optical flow," in *EuroVis (Short Papers)* (Groningen, The Netherlands: Eurographics Association), 19–23.

Rainio, O., Teuho, J., and Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Sci. Rep.* 14, 6086. doi: 10.1038/s41598-024-56706-x

Rampal, N., Shand, T., Wooler, A., and Rautenbach, C. (2022). Interpretable deep learning applied to rip current detection and localization. *Remote Sens.* 14, 91–9. doi: 10.3390/rs14236048

Rashid, A. H., Razzak, I., Tanveer, M., and Robles-Kelly, A. (2021). "RipDet: A fast and lightweight deep neural network for rip currents detection," in *International Joint Conference on Neural Networks (IJCNN)*. (Shenzhen, China: IEEE), 1–6.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). "You only look once: Unified, real-time object detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (Las Vegas, NV, USA: IEEE), 779–788.

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, 6048–6070. doi: 10.1109/TPAMI.2016.2577031

Retnowati, A., Marfai, M. A., and Sumantyo, J. S. (2012). Rip currents signatures zone detection on alos palsar image at parangtritis beach, Indonesia. *Indonesian J. Geogr.* 43, 12–27.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). "MobileNetV2: Inverted residuals and linear bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE Computer Society, Los Alamitos, CA, USA), 4510–4520. doi: 10.1109/CVPR.2018.00474

Silva, A. d., Zhao, M., Stewart, D., Hasan, F., Dusek, G., Davis, J., et al. (2023). RipViz: Finding rip currents by learning pathline behavior. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 2023 (Seattle, WA, USA: IEEE), 1–13. doi: 10.1109/TVCG.2023.3243834

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 (CVPR)*. (Long Beach, CA, USA: IEEE), 10781–10790.

Wang, Z., Dai, Z., Poczos, B., and Carbonell, J. (2019). "Characterizing and avoiding negative transfer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (Long Beach, CA, USA: IEEE) 11293–11302.

Zhang, Y., Huang, W., Liu, X., Zhang, C., Xu, G., and Wang, B. (2021). Rip current hazard at coastal recreational beaches in China. *Ocean Coastal Manage.* 210, 105734. doi: 10.1016/j.ocecoaman.2021.105734

Zhu, D., Qi, R., Hu, P., Su, Q., Qin, X., and Li, Z. (2022). YOLO-Rip: A modified lightweight network for rip currents detection. *Front. Marine Sci.* 9. doi: 10.3389/fmars.2022.930478

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 13 June 2025
DOI 10.3389/fmars.2025.1586237

OPEN ACCESS

EDITED BY
Chao Chen,
Suzhou University of Science and Technology, China

REVIEWED BY
Jianmin Yang,
Sun Yat-sen University, China
Tingt Lyu,
Ocean University of China, China

*CORRESPONDENCE
Shengwen Gong
gsw780604@126.com

RECEIVED 02 March 2025
ACCEPTED 21 May 2025
PUBLISHED 13 June 2025

CITATION
Jiang J, Cheng W, Gong S and Wang J (2025) A deep learning-based data augmentation method for marine mammal call signals. *Front. Mar. Sci.* 12:1586237. doi: 10.3389/fmars.2025.1586237

A deep learning-based data augmentation method for marine mammal call signals

Jiaming Jiang[1,2], Wanlu Cheng[1,2], Shengwen Gong[1,2]* and Jingjing Wang[1,2]

[1]School of Information Science and Technology, Qingdao University of Science and Technology, Qingdao, China, [2]Shandong Key Laboratory of Deep Sea Equipment Intelligent Networking, Qingdao, Shandong, China

In marine ecology research, it is crucial to accurately identify the marine mammal species active in the target area during the current season, which helps researchers understand the behavioral patterns of different species and their ecological environment. However, the difficulty and high cost of collecting marine mammal calls, coupled with limited publicly available datasets, result in insufficient data for support, making it difficult to obtain accurate and reliable identification results. To address this problem, we propose MarGEN, a deep learning-based augmentation method for marine mammal call signal data. This method processed the call data into Mel spectrograms, then designed a self-attention conditional generative adversarial network to generate new samples of Mel spectrograms that were highly similar to the real data, and finally reconstructed them into call signals using WaveGlow. The classification experiments on the calls of four Marine mammals show that MarGEN significantly enriches the diversity and volume of the data, increasing the classification accuracy of the model by an average of 4.7%. The method proposed in this paper greatly promotes marine ecological protection and sustainable development, while effectively advancing research progress in bionic covert underwater acoustic communication technology.

KEYWORDS

marine ecology, marine mammal call signals, MarGEN, deep learning, data augmentation, self-attention conditional generative adversarial network

1 Introduction

Marine mammal calls serve as important ecological signals, carrying a wealth of behavioral and environmental information. Accurately recognizing marine mammal calls not only contributes to species monitoring and conservation but also facilitates the assessment of the health of the marine environment. At the same time, accurate recognition of marine mammal calls also has important military application value, bionic covert underwater acoustic communication technology embeds secret information into marine mammal calls to improve the security of underwater communication Qiao et al.

(2018); Ma et al. (2024), the working principle diagram of this technology is shown in Figure 1. The prerequisite for realizing this technique is to accurately identify the active marine mammals in the target sea area in the current season, so as to select the appropriate calls for bionic communication. Currently, deep learning-based recognition classification offers the most effective results Shi et al. (2023); Dong et al. (2020), but its training demands a large number of data samples as support Li et al. (2021). However, the current limited availability of marine mammal call data significantly reduces the performance of deep learning-based recognition and classification models. Buda et al. (2018). Therefore, increasing the number and diversity of Marine mammal call data has become the key to improving the recognition accuracy.

Data augmentation is a method to expand the size of datasets Khan et al. (2024), which not only enhances the predictive ability of classification models but also provides diversity-rich call signals for bionic communication. Currently, data augmentation methods have performed well in the field of computer vision, which has attracted researchers to focus on its application in the field of audio Sun et al. (2024); Xu et al. (2024).

The cropping method Garcea et al. (2023) obtains multiple cropped sub-data by sliding the audio sequence over a sliding window. Scaling methods Lie and Chang (2006) are implemented by adjusting the audio amplitude or frequency, amplitude scaling is achieved by multiplying all the elements of the time series by some constant, and frequency scaling is achieved by changing the sampling rate of the audio signal. Adding some random noise to the original data can also increase data diversity Kishk and Dhillon (2017), but inappropriate noise may mask important signal features and lead to degradation of model performance. The random oversampling technique Wei et al. (2022) achieves data augmentation by randomly selecting samples for replication. The Synthetic Minority Oversampling Technique (SMOTE) Azhar et al. (2023) generates new samples by interpolating the minority class sample, which improves the problem of unbalanced data distribution. SpecAugment Kim et al. (2024) is a data augmentation method that operates on the audio spectrum. By distorting or masking the spectrogram of the speech signal, the data diversity during model training is increased. Experiments have demonstrated that this method can significantly reduce the word error rate and improve the robustness of the model in speech recognition tasks. This method performs data augmentation on individual sequences, utilizing only the nature of the sequence itself and not taking the overall distribution of the dataset into account.

In the wake of rapid advancements in artificial intelligence, researchers have started to apply deep learning techniques to data augmentation. Yan employed a convolutional neural network model for data augmentation of music in a rhythm game. He took the first 30 seconds of 16 piano arrangements as input, generated additional material that mimicked the original styles through Jukebox and extended them to 60 seconds for data enhancement. However, this method is time-consuming because it generates only one sample at a time Yan (2024). The adversarial training model of Generative Adversarial Networks (GANs) Goodfellow et al. (2020); Wu et al. (2020) gives them excellent generative results. Significant advancements and outcomes have been achieved in the generation of high-resolution and realistic images, which has a wide range of potentials in the field of computer vision and image generation, which also encourages researchers to apply GANs In the field of audio generation. Some researchers have applied GANs to environmental sounds and footstep signal generation with better results Bahmei et al. (2022); Chakraborty and Kar (2023). At present, among the published methods, no researcher has applied the data augmentation method based on GANs to marine mammal call signals.

We proposed MarGEN, a data augmentation method for marine mammal call signals based on audio transformation and a Self-Attention Conditional Generative Adversarial Network (SACGAN). It can effectively enrich the number and diversity of marine mammal call signals and greatly improve the recognition accuracy of the model. The main contributions of this paper are as follows.

1. We proposed a novel method for generating marine mammal call signals, marking the first application of generative adversarial networks in the field of marine mammal call signal data augmentation.
2. We designed a self-attention conditional generation adversarial network for generating new samples that are highly similar to the Mel spectrograms Hong and Suh (2023); Ustubioglu et al. (2023) of real marine mammal calls. The network innovatively added conditional variables representing marine mammal species and self-attention modules and replaced some of the convolutional layers with improved Inception blocks, which significantly improved the model performance and the quality of the generated samples.
3. In order to analyze the performance of our generated call signals, we performed classification experiments and compared them with baseline datasets, which demonstrated the superiority of our method in terms of prediction accuracy.
4. The proposed method can effectively extend the existing marine mammal sound database. It greatly advances the research progress in marine mammal conservation and bionic covert underwater acoustic communication technology. It also provides a reference method for the generation of other types of sound.

2 Data preprocessing

The dataset used in this study comes from the Watkins Marine Mammal Sound Database Sayigh et al. (2016), which provides a variety of call clips of marine mammals recorded in real marine environments. In this study, four marine mammal calls, which are widely distributed in China's sea area and have a relatively large

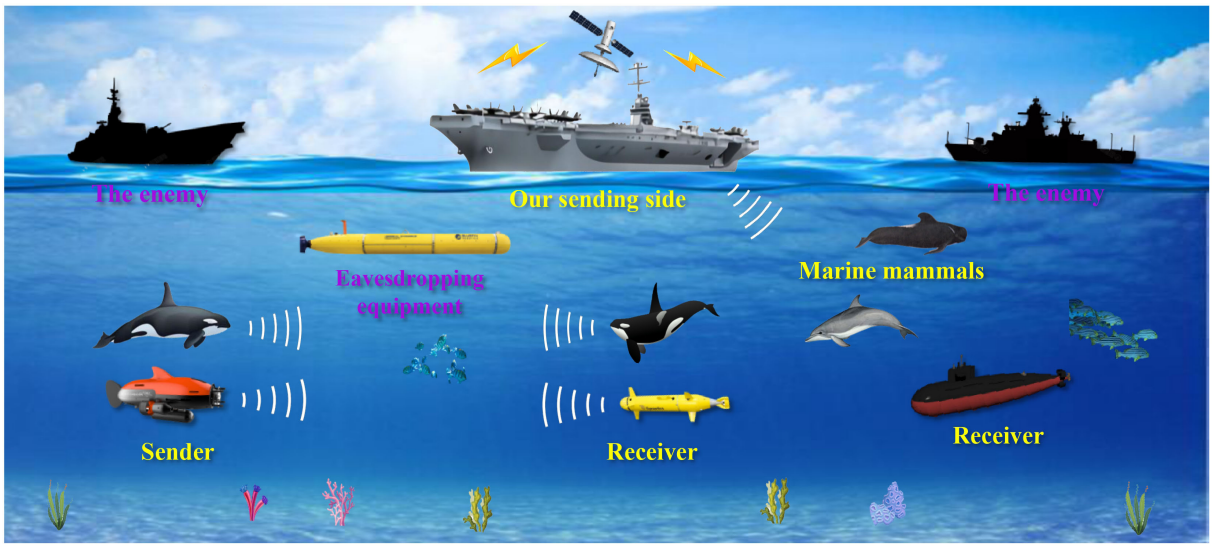


FIGURE 1
Working principle diagram of bionic covert underwater acoustic communication technology.

amount of data, were selected for downloading. After clipping, denoising, resampling, and other operations, 4190 samples with a duration of 1 second are finally obtained and labeled. The distribution and characteristics of the samples are shown in Table 1.

3 MarGEN method

The overall flowchart of the MarGEN method is illustrated in Figure 2, consisting of three main steps. In the first step, due to the large number of audio sampling points of marine mammal calls, resulting in many network parameters and training difficulties, and given that generative adversarial networks are more mature in the image generation domain, we converted the marine mammal call audio files into the form of spectrograms that are more suitable for machine learning to understand the characterization. In the second step, we innovatively designed the SACGAN, whose generator and discriminator engaged in continuous adversarial training until Nash equilibrium Lv et al. (2024) was reached, thereby generating new samples that closely resembled the original images. In the final step, the generated spectrogram was converted into audio signals using WaveGlow Prenger et al. (2019).

TABLE 1 Distribution and characteristics of samples.

Species Name	Abbreviation	Sample Size	Sampling Rate
Killer Whale	KW	1394	48000Hz
Humpback Whale	HW	908	48000Hz
Pilot Whale	PW	1165	48000Hz
Bottlenose Dolphin	BND	723	48000Hz

3.1 Feature extraction

The features of different call samples behave similarly in the time domain but differ significantly in the frequency domain. Therefore, the feature representations chosen in this study were mainly frequency domain features. First, the original signal is analyzed in time-frequency by Short-Time Fourier Transform (STFT) to extract its local frequency domain information. The STFT can effectively capture the spectral changes of the signal in the time dimension. On this basis, Mel frequency cepstrum coefficient (MFCC) based analysis can be further frequency transformed according to the auditory perception of the human ear, thus preserving the key features of the signal. Its accuracy and computational efficiency are better than other representations in the speech recognition task. Therefore, the Mel spectrogram was chosen as the feature representation in this study. The expression for the MEL frequency is shown in Equation 1:

$$M = 2595\log_{10}(1 + \frac{f}{700}) \quad (1)$$

Where M is the frequency in Mel and f is the frequency in Hz, 128 Meier filters are used in this study.

3.2 Self-attention conditional generative adversarial network

GANs consist of a generator and a discriminator. The generator receives random noise and outputs newly generated data samples, while the discriminator is responsible for determining whether the received data is real or generated by the generator. The generator and discriminator engage in adversarial training, which ultimately generates new data that closely resembles real data.

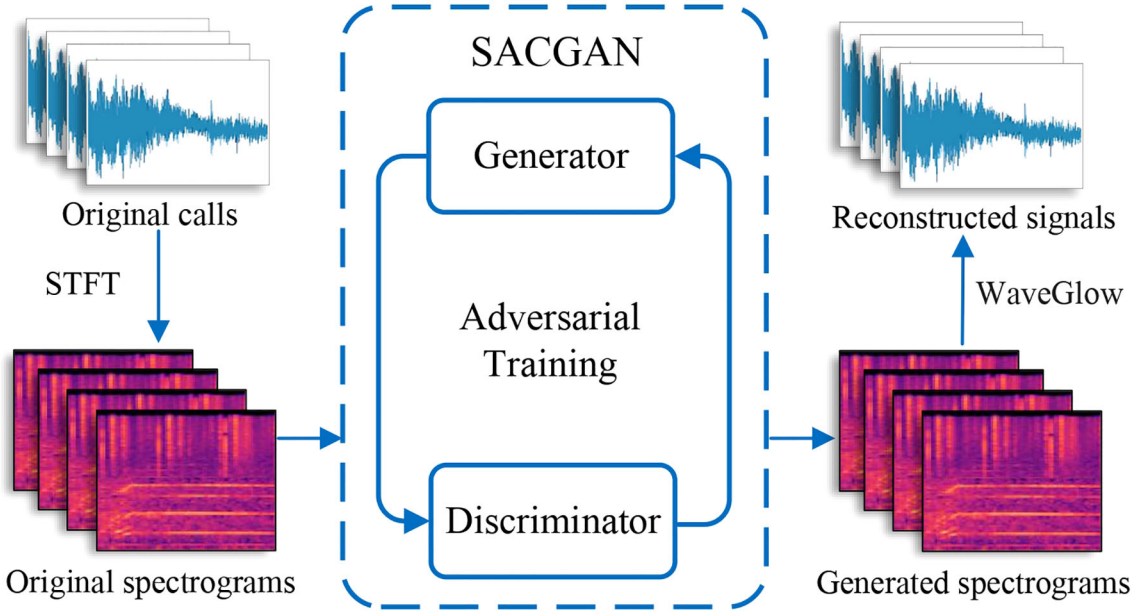


FIGURE 2
Flowchart of MarGEN Method.

We innovatively designed SACGAN, which introduced conditional variables representing marine mammal species and self-attention modules based on generative adversarial networks. Additionally, traditional convolutional neural networks consist of multiple layers of convolutional layers stacked on top of each other, which tends to lead to overfitting as well as difficulty in updating the gradient. The network we designed utilized improved Inception blocks, a structure that combines convolutional kernels of various sizes within the same layer to capture multi-scale information, thereby enhancing the capability of feature extraction.

The specific network structure of SACGAN is shown in Figure 3A. In the generator network structure, the discrete labeled variables were converted to continuous vectors through the Embedding layer, which were spliced with random noise to help the model better understand the input data. The network structure of the Inception block is shown in Figure 3B. We improved its second branch by decomposing a 3x3 convolution into a 1x3 convolution and a 3x1 convolution, further reducing the number of parameters and computational complexity. The residual block consisted of the deconvolution layer, the batch normalization layer, and the activation layer. In the residual block, the gradient information was propagated by means of skip connections to help the generator better recover the image details. A self-attention module was added between two residual blocks to enhance the generator's ability to produce specific content under given conditions, thereby improving generation precision. In the discriminator network structure, the residual block consisted of the convolution layer, the batch normalization layer, and the activation layer. The pooling layer was responsible for reducing the feature dimensions and extracting the main information of the features. We added a self-attention module after the pooling operation to help the model compensate for information loss, ensuring that the model retained some detailed information while capturing the main features. The formula expression of the self-attention mechanism is shown as Equation 2:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

Where Q denotes the query matrix, K denotes the key matrix, V denotes the value matrix, K^T is the transpose matrix of K, and d_k denotes the dimension length.

In addition, the model used the loss function of WGAN-GP Pu et al. (2022); Zhu et al. (2023) to prevent the pattern collapse problem during training. A gradient penalty term was added to the discriminator loss function to ensure that the discriminator function satisfied the Lipschitz continuity constraint, avoiding the problem of gradient explosion or gradient disappearance during the training process and enhancing the convergence speed of the model. The generator loss function is shown as Equation 3:

$$L(G) = -E_{z\sim Pz}[D(G(z|y))] \tag{3}$$

Where Pz denotes the data distribution of samples generated by the generator, z is the randomly sampled noise vector in Pz, and y is the condition variable.

The discriminator loss function is shown as Equation 4:

$$\begin{aligned} L(D) = {} & E_{x\sim pr}[D(x|y)] - E_{z\sim Pz}[D(G(z|y))] \\ & + \lambda E_{\hat{x}\sim P\hat{x}}\left[(\| \nabla_{\hat{x}} D(\hat{x}) \|_2 - 1)^2\right] \end{aligned} \tag{4}$$

Where p_r denotes the data distribution of the real sample, x is the sample in p_r, λ is the gradient penalty term weight, $\lambda E_{\hat{x}\sim P\hat{x}}[(\| \nabla_{\hat{x}} D(\hat{x}) \|_2 - 1)^2]$ is the gradient penalty term, $\hat{x}$ is the stochastic

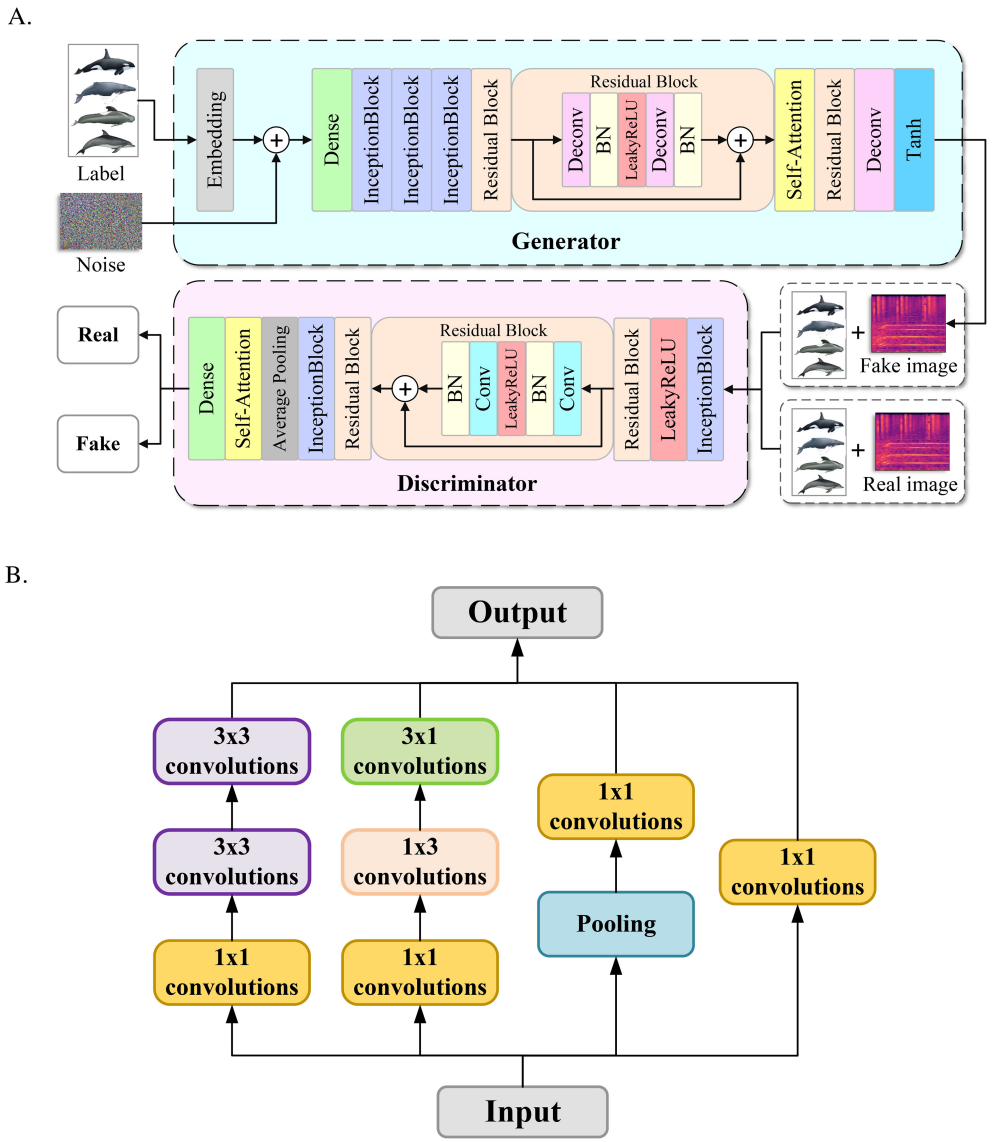


FIGURE 3
(A) Network structure of the self-attention conditional generative adversarial network; **(B)** structure of the inception block.

interpolation between the real sample and the generated sample, $P\hat{x}$ denotes the sampling distribution of the gradient penalty term, and $\| \nabla_{\hat{x}} D(\hat{x}) \|_2$ denotes the gradient parameter of $\hat{x}$, which ensures that the gradient paradigm of the discriminator function is close to 1 and satisfies the Lipschitz constraint.

3.3 Audio signal reconstruction

We used WaveGlow to reconstruct the Mel spectrogram samples generated by SACGAN into audio signals. The model can accurately learn the probability distribution of the audio data and acquire longrange information, resulting in better generation quality and generalization ability. In addition, WaveGlow supports GPU parallel operation, significantly accelerating the audio synthesis speed.

4 Experiment

We designed generation experiments and classification experiments. The generation experiments were used to increase the number and diversity of existing datasets. The classification experiments were used to validate the effectiveness of the MarGEN method.

4.1 Generation experiment

The experimental programming language was Python 3.9, and the network construction was built using Pytorch 1.10 deep learning framework. We trained SACGAN with 4,190 Mel spectrograms of marine mammal calls, setting the labels for killer whale calls to 0, humpback whale calls to 1, pilot whale calls to 2, and bottlenose

dolphin calls to 3. The experimental dataset was divided into a training set and a test set in an 8:2 ratio, with a learning rate set at 1e-4; the batch size was 64; the number of training epochs was 2000. An alternating training strategy was adopted, in which the discriminator was trained six times corresponding to the training of the generator once.

Figure 4 shows an example of Mel spectrograms generated using the SACGAN. As shown, SACGAN can generate high-quality Mel spectrograms. In this experiment, a total of 1755 samples of Mel spectrograms of marine mammal calls were generated using the SACGAN.

4.2 Classification experiment

To verify the effectiveness and superiority of the MarGEN method, this experiment trained the same ResNet classification model on two datasets separately for performance evaluation. Table 2 presents the number of samples in the two datasets and their specific distribution. Among them, OD is a dataset consisting of the original marine mammal call signals. MD is a mixed dataset consisting of the original marine mammal call signals and the call signals obtained using the MarGEN method. The 'Factor' column indicates the ratio between the total number of samples after data enhancement (original samples plus generated samples) and the number of original samples. For example, for bottlenose dolphin, a factor of 2.0 indicates that after augmentation, the dataset contains twice as many samples as the original dataset (original: 723 samples, augmented: 1,446 samples). The dataset was divided using 5-fold cross-validation, in which the entire sample was randomly divided into five non-overlapping subsets, each of which accounted for approximately 20% of the entire dataset. In each round of cross-validation, four of them were selected as the training set. The remaining one as the validation set, and a total of five rounds were executed, with a different validation subset being used in each round. The final results are aggregated by the average of the metrics obtained from the 5 rounds of experiments to ensure the stability and generalization ability of the model. At the same time, it is necessary to make sure that the ratio of original data and generated data in the training and validation sets is consistent. The learning rate for the experiments was set to 1e-4; the batch size was 32; and the training epochs were 150.

Figure 5A illustrates the confusion matrix of the classification model trained using OD, while Figure 5B illustrates the confusion matrix of the classification model trained using MD. In these matrices, the diagonal elements represent the correct classification rate for each category, while the off-diagonal elements reflect the misclassifications between species. Through comparison, it can be found that killer whales showed high classification accuracy in both confusion matrices, probably due to the even spacing between fundamental and harmonic frequencies in their calls, regular frequency bands, often accompanied by high-energy dominant frequency components, and clear transverse stripe structure on Mel spectrograms, which had good discriminability, and thus were easy to be accurately recognized by the model. The classification effect of the bottlenose dolphin was significantly improved after the data enhancement. However, the classification accuracy was still at the lowest level, which may be attributed to the following reasons: on the one hand, broad-snouted dolphin calls are complex and diverse, with a large frequency span, which increases the difficulty of identification; On the other hand, broad-snouted dolphins have the smallest number of original samples among the four categories, and the

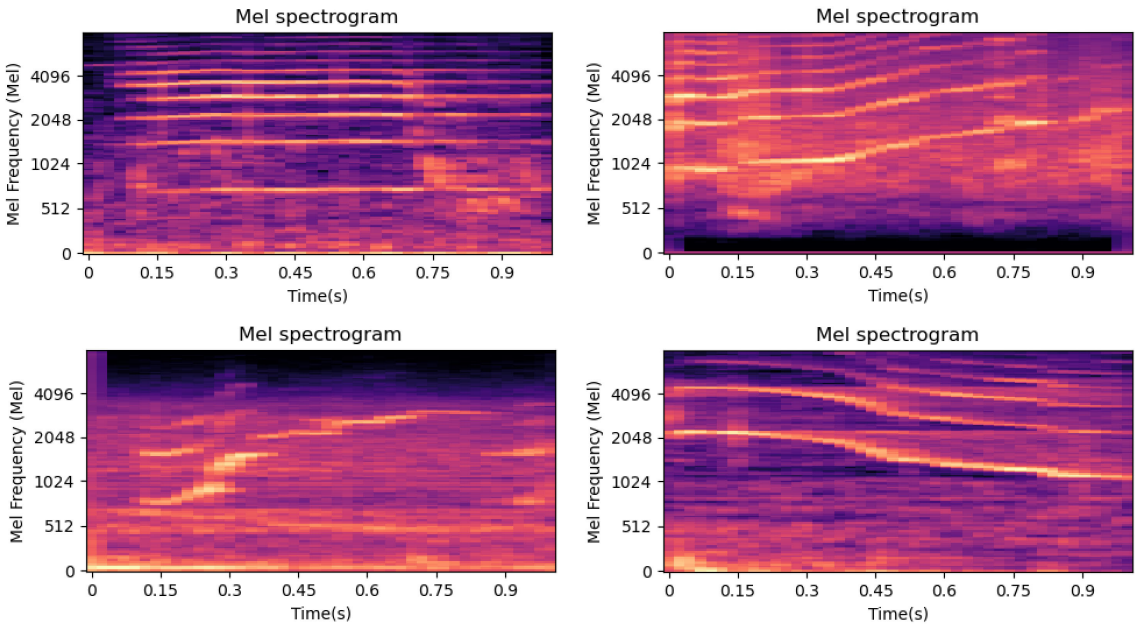


FIGURE 4
Mel spectrograms generated using SACGAN.

TABLE 2 The number of samples and their specific distribution for the two datasets.

Species Name	Abbreviation	OD	Factor	MD
Killer Whale	KW	1394	1.1	1533
Humpback Whale	HW	908	1.6	1452
Pilot Whale	PW	1165	1.3	1514
Bottlenose Dolphin	BND	723	2.0	1446

model does not learn enough of its features at the early stage of training. Although data augmentation greatly mitigates the training bias caused by the uneven samples, there are still some recognition challenges.

In general, the model trained using the MD dataset achieves a higher recognition accuracy for marine mammal calls and exhibits a significantly reduced gap in classification performance between species. These results demonstrate that the proposed MarGEN data augmentation method effectively enhances the model's generalization ability and mitigates the problem of class imbalance.

We selected four classical deep learning models for classification experiments to demonstrate that the MarGEN method can optimize the performance of multiple models. In the experiments, we calculated the Accuracy, Precision, Recall, and F1 Score of the models to comprehensively evaluate their classification performance. We calculated the accuracy, precision, recall, and F1 score of these models in the experiments. The corresponding formulas are as shown in Equations 5–8:

$$Accuracy = \frac{True\ \ Positives + True\ \ Negatives}{True\ \ Positives + False\ \ Positives + True\ \ Negatives + False\ \ Negatives} \tag{5}$$

$$Precision = \frac{True\ \ Positives}{True\ \ Positives + False\ \ Positives} \tag{6}$$

$$Recall = \frac{True\ \ Positives}{True\ \ Positives + FalseNegatives} \tag{7}$$

$$F1\ \ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{8}$$

Where the F1 score is the reconciled average of precision and recall, which can comprehensively evaluate the classification performance.

Table 3 shows that the accuracy of the classification models trained using MD increased by an average of 4.7%, in which the accuracy of the ResNetSE model increased by 5.7% from 90.93% to 96.63%; the F1 score increased by an average of 5.75%, proving that

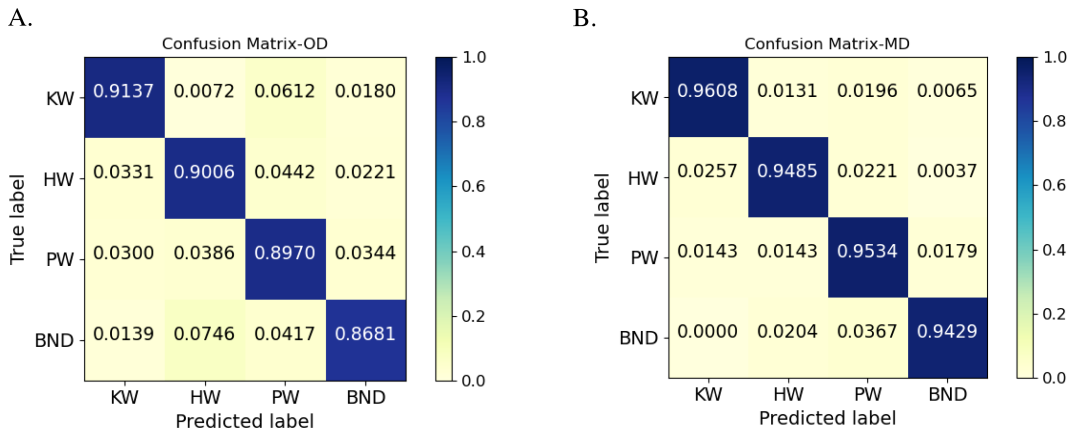


FIGURE 5
(A) Confusion matrix of the classification model trained with OD; **(B)** Confusion matrix of the classification model trained with MD.

TABLE 3 Comparison of performance evaluation indexes for two datasets applied to different classification models.

Model	Accuracy (OD/MD) (%)	Precision (OD/MD) (%)	Recall (OD/MD) (%)	F1 Score (OD/MD) (%)
CNN	89.98/94.37	88.11/95.20	90.06/94.85	89.07/95.02
Res2Net	91.77/95.37	94.42/96.39	91.37/96.08	92.87/96.23
ResNetSE	90.93/96.63	88.03/96.65	86.81/94.29	87.42/95.46
RNN	88.90/94.03	87.08/92.68	89.70/95.34	88.37/93.99

the MarGEN method can significantly improve the performance of multiple deep learning models on the marine mammal call signal recognition task.

5 Conclusion

We have innovatively presented MarGEN, which can effectively realize the high similarity generation of marine mammal call signals and improve their recognition accuracy. First, we designed SACGAN, which can generate Mel spectrograms that are highly similar to the original data, and then we converted the Mel spectrograms into call signals using WaveGlow. The experimental results demonstrated that after using the MarGEN method, the recognition accuracy of different classification models is improved by 4.7% on average, and the F1 score is improved by 5.75% on average. The proposed method in this paper greatly promotes marine ecological protection and sustainable development, and at the same time, it also greatly promotes the research progress of bionic covert hydroacoustic communication technology, which is of great strategic significance. In the future, we will further extend the applicability of the study. On the one hand, we will extend the MarGEN method to more marine species to verify its generalizability in multi-species identification tasks; on the other hand, we will also explore the migration ability of the model under fewer samples to enable the identification and study of data-scarce species.

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

JJ: Software, Writing – original draft, Writing – review & editing, Methodology. WC: Data curation, Investigation, Writing – review & editing. SG: Validation, Visualization, Writing – review & editing. JW: Funding acquisition, Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported in part by the National Natural Science Foundation of China under Grants 62171246 , 62101298 and U24A20215.

Acknowledgments

The authors are grateful for the Watkins Marine Mammal Sound Database website for providing us with the audio of marine mammal calls needed for the experiment, and the support of Python.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

Azhar, N. A., Pozi, M. S. M., Din, A. M., and Jatowt, A. (2023). An investigation of smote based methods for imbalanced datasets with data complexity analysis. *IEEE Trans. Knowledge Data Eng.* 35, 6651–6672. doi: 10.1109/TKDE.2022.3179381

Bahmei, B., Birmingham, E., and Arzanpour, S. (2022). Cnn-rnn and data augmentation using deep convolutional generative adversarial network for environmental sound classification. *IEEE Signal Process. Lett.* 29, 682–686. doi: 10.1109/LSP.2022.3150258

Buda, M., Maki, A., and Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks* 106, 249–259. doi: 10.1016/j.neunet.2018.07.011

Chakraborty, M., and Kar, S. (2023). Enhancing person identification through data augmentation of footstep-based seismic signals. *IEEE Signal Process. Lett.* 30, 1642–1646. doi: 10.1109/LSP.2023.3327650

Dong, S., Zhuang, Y., Yang, Z., Pang, L., Chen, H., and Long, T. (2020). Land cover classification from vhr optical remote sensing images by feature ensemble deep learning network. *IEEE Geosci. Remote Sens. Lett.* 17, 1396–1400. doi: 10.1109/LGRS.2019.2947022

Garcea, F., Serra, A., Lamberti, F., and Morra, L. (2023). Data augmentation for medical imaging: A systematic literature review. *Comput. Biol. Med.* 152, 106391. doi: 10.1016/j.compbiomed.2022.106391

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2020). Generative adversarial networks. *Commun. ACM* 63, 139–144. doi: 10.1145/3422622

Hong, G., and Suh, D. (2023). Mel spectrogram-based advanced deep temporal clustering model with unsupervised data for fault diagnosis. *Expert Syst. Appl.* 217, 119551. doi: 10.1016/j.eswa.2023.119551

Khan, A. A., Chaudhari, O., and Chandra, R. (2024). A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation. *Expert Syst. Appl.* 244, 122778. doi: 10.1016/j.eswa.2023.122778

Kim, K.-H., Oh, K.-H., Ahn, H.-S., and Choi, H.-D. (2024). Time–frequency domain deep convolutional neural network for li-ion battery soc estimation. *IEEE Trans. Power Electron.* 39, 125–134. doi: 10.1109/TPEL.2023.3309934

Kishk, M. A., and Dhillon, H. S. (2017). Stochastic geometry-based comparison of secrecy enhancement techniques in d2d networks. *IEEE Wireless Commun. Lett.* 6, 394–397. doi: 10.1109/LWC.2017.2696537

Li, L., Qiao, G., Liu, S., Qing, X., Zhang, H., Mazhar, S., et al. (2021). Automated classification of tursiops aduncus whistles based on a depth-wise separable convolutional neural network and data augmentation. *J. Acoustical Soc. America* 150, 3861–3873. doi: 10.1121/10.0007291

Lie, W.-N., and Chang, L.-C. (2006). Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification. *IEEE Trans. Multimedia* 8, 46–59. doi: 10.1109/TMM.2005.861292

Lv, Y., Liu, Y.-J., Liu, L., Yu, D., and Chen, Y. (2024). Distributed nash equilibrium searching for multi-agent games under false data injection attacks. *Neurocomputing* 570, 127134. doi: 10.1016/j.neucom.2023.127134

Ma, X., Wang, B., Tian, W., Ding, X., and Han, Z. (2024). Strategic game model for auv-assisted underwater acoustic covert communication in ocean internet of things. *IEEE Internet Things J.* 11, 22508–22520. doi: 10.1109/JIOT.2024.3382649

Prenger, R., Valle, R., and Catanzaro, B. (2019). "Waveglow: A flow-based generative network for speech synthesis," in *ICASSP 2019–2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (Brighton, UK: IEEE), 3617–3621. doi: 10.1109/ICASSP.2019.8683143

Pu, Z., Cabrera, D., Li, C., and de Oliveira, J. V. (2022). Vgan: Generalizing mse gan and wgan-gp for robot fault diagnosis. *IEEE intelligent Syst.* 37, 65–75. doi: 10.1109/MIS.2022.3168356

Qiao, G., Bilal, M., Liu, S., Babar, Z., and Ma, T. (2018). Biologically inspired covert underwater acoustic communication—a review. *Phys. Communication* 30, 107–114. doi: 10.1016/j.phycom.2018.07.007

Sayigh, L., Daher, M. A., Allen, J., Gordon, H., Joyce, K., Stuhlmann, C., et al. (2016). "The watkins marine mammal sound database: an online, freely accessible resource," in *Proceedings of meetings on acoustics*, vol. 27. (Melville, New York, United States: AIP Publishing). doi: 10.1121/2.0000358

Shi, J., Liu, W., Shan, H., Li, E., Li, X., and Zhang, L. (2023). Remote sensing scene classification based on multibranch fusion attention network. *IEEE Geosci. Remote Sens. Lett.* 20, 1–5. doi: 10.1109/LGRS.2023.3262407

Sun, Y., Xu, K., Liu, C., Dou, Y., Wang, H., Ding, B., et al. (2024). Automated data augmentation for audio classification. *IEEE/ACM Trans. Audio Speech Lang. Process.* 32, 2716–2728. doi: 10.1109/TASLP.2024.3402049

Ustubioglu, B., Tahaoglu, G., and Ulutas, G. (2023). Detection of audio copy-move-forgery with novel feature matching on mel spectrogram. *Expert Syst. Appl.* 213, 118963. doi: 10.1016/j.eswa.2022.118963

Wei, G., Mu, W., Song, Y., and Dou, J. (2022). An improved and random synthetic minority oversampling technique for imbalanced data. *Knowledge-based Syst.* 248, 108839. doi: 10.1016/j.knosys.2022.108839

Wu, Z., Li, J., Wang, Y., Hu, Z., and Molinier, M. (2020). Self-attentive generative adversarial network for cloud detection in high resolution remote sensing images. *IEEE Geosci. Remote Sens. Lett.* 17, 1792–1796. doi: 10.1109/LGRS.2019.2955071

Xu, X., Xie, Z., Wu, M., and Yu, K. (2024). Beyond the status quo: A contemporary survey of advances and challenges in audio captioning. *IEEE/ACM Trans. Audio Speech Lang. Process.* 32, 95–112. doi: 10.1109/TASLP.2023.3321968

Yan, N. (2024). Generating rhythm game music with jukebox. *Front. Artif. Intell.* 7. doi: 10.3389/frai.2024.1296034

Zhu, G., Zhou, K., Lu, L., Fu, Y., Liu, Z., and Yang, X. (2023). Partial discharge data augmentation based on improved wasserstein generative adversarial network with gradient penalty. *IEEE Trans. Industrial Inf.* 19, 6565–6575. doi: 10.1109/TII.2022.3197839

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 18 June 2025
DOI 10.3389/fmars.2025.1534900

OPEN ACCESS

EDITED BY
Weimin Huang,
Memorial University of Newfoundland, Canada

REVIEWED BY
Zheqi Shen,
Hohai University, China
Zhenhua Zhang,
Ministry of Natural Resources, China

*CORRESPONDENCE
Shuai Guo
guoshuai@qut.edu.cn
Meijuan Jia
jiameijuan@dqnu.edu.cn

RECEIVED 26 November 2024
ACCEPTED 19 May 2025
PUBLISHED 18 June 2025

CITATION
Jia M, Mao X, Guo S and Li X (2025) Retrieval algorithm based on locally sensitive hash for ocean observation data. *Front. Mar. Sci.* 12:1534900. doi: 10.3389/fmars.2025.1534900

Retrieval algorithm based on locally sensitive hash for ocean observation data

Meijuan Jia[1]*, Xiaodong Mao[1], Shuai Guo[2]* and Xin Li[1]

[1]College of Computer Science and Information Technology, Daqing Normal University, Daqing, China, [2]College of Science, Qingdao University of Technology, Qingdao, China

As an important technology for eliminating redundant data, data deduplication significantly impacts today's era of explosive data growth. In recent years, due to the rapid development of a series of related industries, such as ocean observation, ocean observation data has also shown a speedy growth trend, leading to the continuous increase in storage costs of ocean observation stations. Faced with the constant increase in data scale, our first consideration is to use data deduplication technology to reduce storage costs. While using duplicate data deletion technology to achieve our goals, we also need to pay attention to some of the actual situations of ocean observation stations. The fingerprint retrieval process in duplicate data deletion technology plays a key role in the entire process. Therefore, this paper proposes a fast retrieval strategy based on locally sensitive hashing. The fast retrieval algorithm based on locally sensitive hashing can enable us to quickly complete the retrieval process in duplicate data deletion technology and achieve the goal of saving computing resources. At the same time, we proposed a bucket optimization strategy for retrieval algorithms based on locally sensitive hashing. We utilized visual information to address the bottleneck problem in duplicate data deletion technology. At the end of the article, we conducted careful experiments to compare hash retrieval algorithms and concluded the strategy's feasibility.

KEYWORDS

local sensitive hashing, ocean observation data, duplicate data deletion technology, fast retrieval algorithm, storage location

1 Introduction

As is well known, today's society is filled with a large amount of data and has entered the era of big data. At the same time, the scale of data generated by various industries has also exploded. According to current research reports, IDC estimates that the storage capacity of the global market will grow exponentially from 33ZB to 173ZB from 2018 to 2025 (Reinsel et al., 2017) As various industries enter the era of big data, the scale of data generated by marine-related industries is unprecedentedly large. The existing marine data includes marine surveying, island monitoring, underwater exploration marine fishery operations, marine fishery operations, marine buoy monitoring, marine scientific research, oil and gas

platform environmental monitoring, satellite remote sensing monitoring, etc., forming a wide range of marine observation and monitoring systems, accumulating a large amount of marine natural science data, including on-site observation and monitoring data, marine remote sensing data, numerical model data, etc. With the rapid advancement of ocean informatization and the increasing sophistication of sensing technologies, ocean data volumes have grown exponentially. For instance, since the launch of the Argo program, over 10,000 profiling floats have been deployed, with approximately 3,800 currently operational in global oceans (Riser et al., 2016) By 2016, Argo-generated data had already surpassed the cumulative ocean observation dataset of the entire 20th century, and both its sampling density and vertical coverage continue to expand. Similarly, as of 2012, the U.S. National Oceanic and Atmospheric Administration (NOAA) hosted annual data archives exceeding 30 petabytes, aggregating over 3.5 billion daily observations from a diverse array of sensor systems (Huang et al., 2015) In recent years, revolutionary changes have occurred in the observation equipment used for observing ocean data. The scale of ocean data represented by satellite remote sensing data is exploding, and the growth rate of ocean observation data is also much faster than most industries. At present, when ordinary people face data growth, they tend to think of increasing storage capacity to solve the problem. However, when we face huge amounts of data, it is unrealistic to solve the problem by increasing storage capacity. Therefore, people usually choose to improve storage efficiency so that more data can be stored in limited storage space. When faced with such problems, people usually think of compression technology first. However, compression technology retrieves the same data block through string matching, mainly using string matching algorithms and their various variants, which achieve precise matching. Implementing precise matching is more complex but more accurate and effective for eliminating fine-grained redundancy.

Data deduplication (Nisha et al., 2016) technology uses the data fingerprint of data blocks to find identical data blocks, and the fingerprint of data blocks is calculated using a fuzzy matching hash function. Fuzzy matching is relatively simple and more suitable for large granularity data blocks, but its accuracy is lower. If we want to save storage space on datasets obtained through ocean observation, we should prioritize duplicate data deletion technology. Data deduplication technology eliminates redundant data in a dataset by removing duplicate data and retaining only one copy. Therefore, data deduplication technology can bring huge practical benefits when facing such problems, such as effectively controlling the rapidly growing data scale, saving sufficient storage space, improving storage efficiency, saving total storage and management costs, and meeting ROI, TCO, etc (Nisha et al., 2016).

The entire process of data deduplication technology is to cut the input file into data blocks and determine whether the data block is a duplicate by querying the fingerprint table in memory. Data deduplication technology can be divided into five stages, including data block segmentation, fingerprint calculation of data blocks, indexing of hash tables, compression techniques, and data management in various storage systems. The compression stage is an optional operation, as it is only applicable to some more traditional compression methods. Data deduplication plays a crucial role in the final stage of storage management. The above explanation shows us that block segmentation and retrieval are the two core stages in data deduplication technology. How to segment data blocks reasonably will seriously affect the final data deduplication rate. However, the focus of this article is on another aspect - retrieval. How to quickly retrieve whether there are data blocks in the fingerprint table will greatly affect the efficiency of the entire data deduplication system. We will save much time if we can achieve fast retrieval. At the same time, reducing the number of comparisons within the fingerprint table will directly affect the computational resource consumption of the entire data deduplication system when facing large-scale data.

At present, there are many research studies on retrieval in duplicate data deletion technology, including Bloom filters, which are used to address challenges in the retrieval process (Lu et al., 2012) HT Indexing accelerates the process by selecting champions, or Sparse Indexing solves real-world problems (Lillibridge et al., 2009).

In this study, we aim to save more resource consumption in the retrieval process of data deduplication technology. Therefore, to address the existing challenges, we propose a fast retrieval algorithm based on locally sensitive hashing (Bucket index), which can reduce the number of comparisons while saving computational resources. B-index is a fast retrieval algorithm based on locally sensitive hashing, which puts similar data blocks into the same bucket. When a data block is passed in, it only needs to be retrieved from the bucket to which the data block belongs without the need to retrieve the entire fingerprint table, thus reducing resource consumption during the retrieval process. The contributions of this article are as follows:

- We propose a fast retrieval algorithm based on locally sensitive hashing, which achieves fast retrieval by splitting and storing many data blocks during the retrieval process.
- We propose a bucket optimization strategy under locally sensitive hashing, which continuously optimizes retrieval efficiency by adjusting the number of buckets when facing different problems.
- Finally, we proposed a strategy for selecting fingerprint tables when faced with ocean observation data.

The content of the remaining chapters of this article is as follows: In Chapter 2, we will provide a detailed introduction to the background of duplicate data deletion technology and the motivation behind this paper. In Chapter 3, we will elaborate on various research related to this paper. Chapter 4 will focus on the fast retrieval algorithm based on locally sensitive hashing. In Chapter 5, we will verify our hypothesis through detailed experiments. In the final chapter, we will make plans for future research.

2 Background and motivation

In the second part, we will briefly introduce the process of data deduplication technology, focus on the importance of retrieval, and briefly introduce other retrieval algorithms. At the end of this section, we will introduce the motivation behind our work.

2.1 The dilemma of duplicate data deletion technology in the retrieval process

When analyzing a problem, the first thing we need to do is to understand where the problem lies. When a data stream is fed into a data deduplication system, we first need to perform a chunk operation on the data stream, cutting it into data blocks of different sizes. How to chunk is based on the content of the data stream, so we can understand that the same content will produce the same data blocks. The first definition of this part was mentioned in the sliding window-based chunk algorithm 1. After the data blocks are cut, we assign fingerprints to each. Each different data block has a different fingerprint. After that, the duplicate data deletion system will compare the fingerprints of each data block with the existing fingerprints in the memory table. In a duplicate data removal system, querying whether a data block is duplicate is done by storing the fingerprint of the data block in a fingerprint table in memory. After cutting out a new data block, the fingerprint of the new data block is searched in the fingerprint table. When the fingerprint of the new data block exists in the fingerprint table, it will be judged as a duplicate data block. Conversely, if the data block does not appear in the fingerprint table, the fingerprint of the data block will be stored in the fingerprint table, and the data block will be saved as shown in Figure 1. While we understand the basic process, we must also be aware of the disk bottleneck issue in data deduplication technology.

Assuming the average size of data blocks is 8KB, the generated fingerprints are approximately 20GB. For 8TB of data, nearly 20GB of fingerprint storage will be required. If all these fingerprints are stored in memory, it will bring a very serious memory burden. At the same time, in a system with an average throughput of 100MB/s, each retrieval will bring a huge burden and increase the system overhead. Even if you use cache memory to accelerate index access, there will not be much change. This is because fingerprint generation is random, and traditional cache memory has a low hit rate and work efficiency. Therefore, in response to the above issues, some people store the fingerprint table in external storage. However, this approach will lead to frequent access to external storage, thereby reducing efficiency. Some people also choose to put some fingerprint tables in memory and some in external storage, but choosing which ones to put in memory and which to put in external storage is not appropriate. Therefore, to improve efficiency, it is better to accelerate the indexing speed directly. Because no matter which method is chosen to avoid the disk bottleneck, it cannot escape the need to retrieve the fingerprint table.

2.2 The particularity of ocean observation data

At this point, we can foresee the problem we are facing. If the fingerprint table becomes larger, we will face great difficulties retrieving it. As a result, if the fingerprint table continues to grow, it will also greatly burden the memory if we keep it in memory. Considering the actual situation we will face, that is, the storage method of ocean observation stations, ocean observation data differs from ordinary data, and most ocean observation data is time series data. Some characteristics need to be understood.

One of them is the existence of non-renewable primitiveness; as the ocean constantly changes, the elemental data of ocean surveys has distinct characteristics of non-renewable primitiveness. Ocean measurement data is a first-hand source of original information obtained from on-site measurements, organization, and calibration by ocean survey ships. The data of ocean remote sensing, whether it is infrared or visible light observations of scanning imaging or microwave measurements, the measured data (including element data inverted according to a certain pattern) is specific in time and

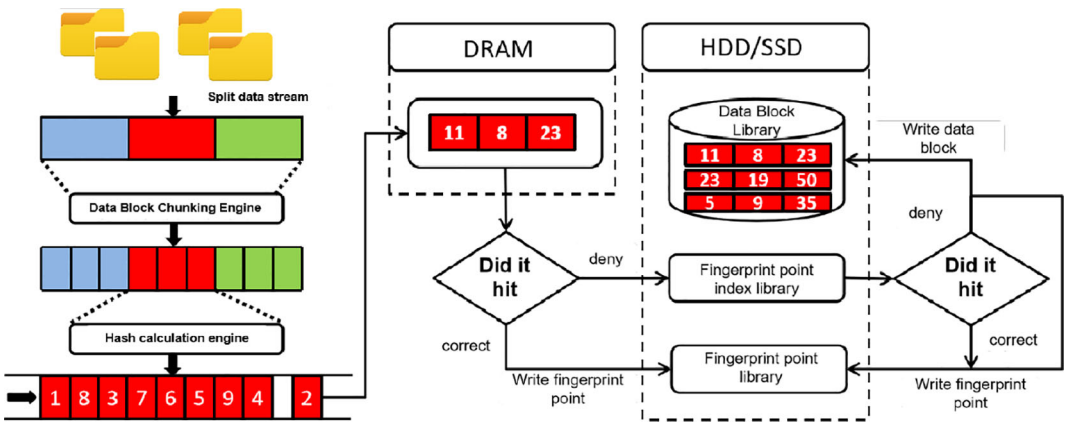


FIGURE 1
Process diagram of duplicate data deletion system.

space, reflecting the characteristics of ocean elements under specific spatiotemporal conditions; Other data such as ship reports also have similar characteristics; Although numerical simulation product data can be obtained repeatedly under certain conditions, a certain type of product data can still be considered as special original obtained data, and therefore also considered as having originality. Moreover, certainty, this characteristic is easy to understand. Certainty refers to the very accurate observation of ocean element data, such as the measurement accuracy of water temperature and depth and the measurement time and space, which are all very precise. There are also many categories included in ocean observation data, such as the inferential, fuzzy, and multi-level nature of ocean element data. We mentioned the characteristics of ocean observation data, and ultimately, the most important point is that ocean observation data may contain non-renewable data after observation. Therefore, storing each observation is crucial, and the disk space challenge caused by storing a large amount of data must be addressed. So, when facing practical problems, we need to consider the various bottlenecks of duplicate data deletion technology and improve duplicate data deletion technology according to the characteristics of ocean observation data.

2.3 Motivation

On this occasion, we have learned about the principle of data deduplication technology and the particularity of ocean observation data. Therefore, we consider applying data deduplication technology to ocean observation stations. Of course, we have also done this. Before this research, we optimized the segmentation module of data deduplication technology and finally applied it to the data deduplication system of ocean observation stations. However, the research at that time mainly aimed to improve the data deduplication rate and neglected some retrieval efficiency. Therefore, we will make up for this overlooked efficiency in this article. The data deduplication system is coherent, so we hope to recover the efficiency lost when we segment it in the subsequent retrieval process. At the same time, we learned about the conflicting issues in the retrieval process, such as how to choose between fingerprint tables in memory and whether to store them in memory or external storage. Of course, no matter how we choose, improving the efficiency of retrieval is crucial because, no matter where it is placed, improving the efficiency of retrieval will accelerate the operation efficiency of the entire system. Therefore, this article chooses algorithms that can accelerate indexing efficiency, and regardless of which method is chosen, the ultimate goal is to improve efficiency. At the same time, we consider that a portion of the fingerprint tables can be stored in memory and another portion in external storage, and how to make a decision is also the main research direction of this article. This article will divide the ocean observation data based on certain characteristics to ensure that the fingerprint tables in memory can receive more access times to improve the efficiency of the entire system. In summary, to address the various problems in the retrieval process of existing duplicate data removal systems, this paper proposes a fast indexing method based on locally sensitive hashing to solve the problem, which can accelerate the efficiency of the entire duplicate data removal system through fast indexing. At the same time, in-depth research has been conducted on the storage of fingerprint tables to ensure the improvement of the speed of duplicate data deletion technology in the retrieval process.

3 Related work

When we learn about data deduplication technology, we first need to understand that the original purpose of CDC was to reduce network traffic consumption when transferring files. Spring and Wetherall (2000) designed the first block-based algorithm using the Border method (Broder, 1997) with the aim of better identifying redundant network traffic and reducing consumption. Muthitacharoen et al. (Spring and Wetherall, 2000) proposed a CDC-based file system called LBFS, which enriches the CDC chunk algorithm to reduce and eliminate duplicate data in lowbandwidth network file systems. You et al. (2005) used the CDC algorithm to reduce data redundancy in archive storage systems. However, due to the time-consuming calculation of Rabin fingerprints in the CDC algorithm, which results in a waste of computing resources, many methods have been proposed to replace Rabin to accelerate the speed of CDC (Xia et al., 2014; Agarwal et al., 2010; Zhang et al., 2015) The encryption function required in the fingerprint recognition process (such as Rabin) can be accelerated through parallel strategies (Xia et al., 2019) Moreover, using the modified version of AE (Zhang et al., 2016) to accelerate the time required for calculating fingerprints.

The retrieval problem in the face of duplicate data deletion technology can be roughly divided into global and partial indexing. The global index maintains the metadata of all stored data blocks. Searching for the fingerprint of each new data block in the index can identify all duplicates and achieve the best data de-duplication rate. Due to the requirement for high search throughput, many studies have focused on improving the read performance of full indexes. With the help of Bloom filters and index segment caching, DDFS (Zhu et al., 2008) reduces the large amount of storage reads required for data block fingerprint lookup. SkimpyStash (Debnath et al., 2011) stores the metadata of data blocks in a flash and indexes them in a memory hash table. Bloom filters are used to improve reading performance. Considering the location of data deletion in the duplicate data removal system, ChunkStash (Debnath et al., 2010) buffers index metadata in memory until it reaches the size of a flash page. Index lookup can benefit from page-based IO, which preserves the location of de-duplicated data blocks. BloomStore (Lu et al., 2012) focuses on improving memory efficiency by using bloom filters to eliminate unnecessary flash reads. Due to the read-intensive search workload in the index of data de-duplication, BloomStore can avoid the flash reading of non-existent data block fingerprints by caching Bloom filters and parallel checking Bloom filters.

Although this technology uses different optimizations to reduce storage reads of global indexes, the efficiency of storage reads

increases with the size of stored data. To address this issue, partial indexing is proposed, effectively reducing storage reads by searching only a small portion of the storage block. Another direction for global indexing is partial indexing, whose basic idea is to search for new data block fingerprints on a selected subset of stored data blocks, thereby reducing the number of storage reads and increasing throughput. According to observations, backup data from the same source are usually highly similar (Wallace et al., 2012; Park and Lilja, 2010) The new data block is deduplicated using a batch processing method called Data Deduplication Window (DW). If we store the metadata of stored data blocks in groups (called tuples) in a "log" manner to maintain locality, we can find tuples that share a certain number of blocks with tuples in DW. These shared blocks are duplicated; the remaining data blocks are considered' unique'. The main goal of partial indexing is to index tuples in memory and quickly select tuples that may overlap highly with tuples in DW. Studies indicate that backup data from the same source generally have highly similar characteristics (Wallace et al., 2012; Park and Lilja, 2010) Therefore, partial indexing techniques are proposed. In order to index all tuples using pure memory structures, partial indexing selects a small portion of data block fingerprints from each tuple as a representative (hook).

The memory's fingerprint table (hook index) maintains the mapping from hooks to their corresponding tuple addresses. After accumulating a new batch of data blocks in DW, check the fingerprint of the new data blocks in the hook index. If it matches one or more hooks (hook hits), there is a high possibility that some data blocks from the same tuple may also appear in the DW due to the excellent positional location of the backup data. Sparse indexing (Lillibridge et al., 2009) extreme binary (Bhagwat et al., 2009) SiLo (Xia et al., 2011) and LIPA (Xu et al., 2019) all use data segments as tuples. The duplicate data removal system generates a recipe based on the order in which data blocks are generated in the input data stream, strictly preserving the order of data blocks during duplicate data removal, regardless of whether the data blocks are duplicates. Sparse indexing calculates the hook hit rate for each tuple and selects the tuple with the highest hook hit rate based on the calculation. Extreme Binning (Bhagwat et al., 2009) is designed for backup based on a single file. It uses the overall recipe of each file as a tuple. When performing duplicate data deletion on a new file, Extreme Binning selects recipe segments from the most similar files and performs duplicate data deletion on the data blocks of the new file based on the data blocks in the selected similar files. SiLo (Xia et al., 2011) further extends extreme boxing by simultaneously considering the similarity of files and the locality of blocks.

SiLo concatenates similar small files together as one data block and divides large files into several data blocks. To perform duplicate data deletion on a new data block, SiLo identifies the most similar data block among existing data blocks. It performs duplicate data deletion based on the data blocks in the block. LIPA (Xu et al., 2019) uses reinforcement learning-based algorithms to determine the similarity between recipe segments and data blocks in DW, thereby achieving higher data deduplication rates. Meanwhile, in recent years, countless technologies have combined distributed systems with data deduplication. Among them, cluster-based sharding methods have achieved considerable data deduplication efficiency on a single system while supporting high throughput (Zhou et al., 2022) Moreover, a system proposed to simultaneously perform client and server duplicate data deletion when faced with forced duplicate data deletion of many concurrent backup streams during peak backup loads (Ammons et al., 2022) In recent years, there has also been a problem of pushing duplicate data removal to the network edge. A new distributed edge-assisted duplicate data removal (EF dedup) framework has been proposed. Maintain a duplicate data removal index structure between them using distributed key-value storage and perform duplicate data removal within these clusters (Li et al., 2022) These frameworks can effectively solve the contradictions of current data deduplication technology. However, this project aims to shift the focus back to the retrieval problem in data deduplication technology, using machine learning-assisted fingerprint table retrieval in combination with distributed operating systems and data deduplication technology. To lay the foundation for subsequent ocean observations in data storage.

Meanwhile, with the vigorous development of various industries in recent years, the application of duplicate data deletion technology is becoming increasingly widespread. The most notable among them is the data deduplication technology in cloud storage (Mahesh et al., 2020) However, there are also more security issues in cloud computing, as PraJapanese et al. (Prajapati and Shah, 2022) made a stunning statement about the security issues in data deduplication technology. Even Yuan et al. (2020) proposed blockchain-based duplicate data removal technology in the popular field of blockchain. In addition to the challenges proposed by Azad et al. At the same time, PG et al. (Shynu et al., 2020) proposed a solution to the network edge problem (Al Azad and Mastorakis, 2022).

4 Fast retrieval algorithm based on locally sensitive hash

This chapter will explore the retrieval part of the duplicate data removal system. The retrieval part is the second most important focus of the entire duplicate data removal system, and the retrieval speed will directly determine the entire system's efficiency. Therefore, this article introduces a fast retrieval method aimed at improving the entire system's efficiency in terms of retrieval. In this chapter, we will provide a detailed introduction to implementing a retrieval algorithm based on locally sensitive hashing and the optimization strategy for buckets. Finally, we will discuss how to choose the storage of fingerprint tables based on the characteristics of ocean observation data.

4.1 Fast retrieval algorithm based on locally sensitive hash

In order to address the existing problems in the retrieval process of the duplicate data removal system, this section proposes a fast retrieval technique based on locally sensitive hashing. By extracting

the similarity of data blocks and constructing multiple data buckets, when similar data blocks appear, the data bucket can be quickly selected and retrieved within the bucket, achieving a fast retrieval function. In this section, we will first introduce the application of locally sensitive hashing, then propose solutions based on existing situations, and finally explain the entire idea of retrieval based on locally sensitive hashing.

4.1.1 Local sensitive hashing strategy

Firstly, local sensitive hashing is an approximate nearest neighbour fast search technique applied in the face of massive high-dimensional data. In many different application fields, we often face an astonishing amount of data that needs processing and generally has high dimensions. Quickly finding the data or a set closest to certain data from a massive high-dimensional data set has become a challenging problem. If the data we face is a small, low-dimensional dataset, we can solve this problem using linear search. However, for the current situation, most of them are high-dimensional and large datasets that need to be processed. If we still use linear search, it will waste much time. Therefore, to solve the problem of dealing with massive high-dimensional data, we need to adopt some indexing techniques to accelerate the search process and speed. This technique is usually called nearest neighbour search, and local sensitive hashing is precisely this technique as shown in Figure 2.

Traditional hashing maps initial data to corresponding buckets, while locally sensitive hashing, compared to traditional hashing, maps or projects two adjacent data points in the initial data space through the same transformation. These two adjacent points in the original space still have a high probability of being close to the new data space. The probability of two non-adjacent data points in the original space being mapped or projected to the same bucket is very low. In summary, if we perform some hash mapping on the initial data, locally sensitive hashing can help us map two adjacent data points to the same bucket with a high probability of having the same bucket number. In ocean observation stations, the daily amount of data generated is astonishing. In duplicate data removal systems, the data blocks cut by data streams are also massive amounts of data, making them very suitable for the application scenario of locally sensitive hashing. We hope to achieve fast retrieval when fingerprints are used in the data deduplication system. We hope that the searched data block can be mapped through local sensitive hashing to find the same data block in its bucket, thus achieving the goal of fast retrieval and saving computing resources. However, to determine whether two data blocks are similar, we have to mention a concept, the Jaccard coefficient. It is expressed as Formula 1, where the larger the Jaccard coefficient, the greater the similarity, and vice versa.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

As shown in Figure 3. The method of local sensitive hashing is to perform a hash mapping on all the data in the initial dataset, and then we can obtain a hash table. These initial datasets will be scattered and shuffled into buckets in the hash table, and each bucket will load some initial data. However, there is a high probability that data belonging to the same bucket will be adjacent, although this is not absolute, and there may also be situations where non-adjacent data is mapped to the same bucket. Therefore, if we can find some hash functions that enable data to fall into the same bucket after being hashed and transformed by these hash functions in the original space, it becomes much simpler for us to perform the nearest neighbour search in the data set. We only need to hash map the data to be retrieved to obtain its mapped bucket number, then extract all the data inside the bucket corresponding to that bucket number, and perform a linear search on these data to find the data adjacent to the query data. As shown in the figure below, after a position-sensitive hash

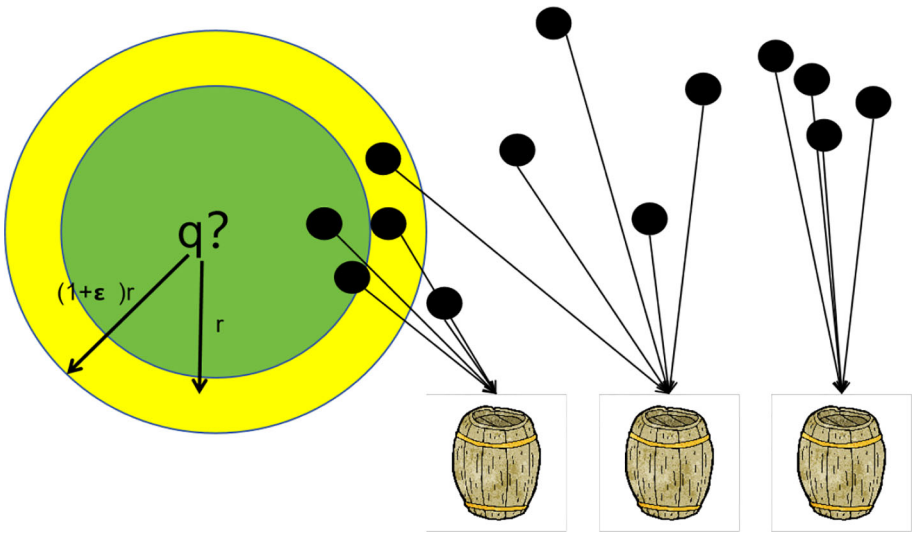


FIGURE 2
Schematic diagram of locally sensitive hash.

function hashed it for q, its *rNN* may be hashed to the same bucket (such as the first bucket). The probability of hashing to the first bucket is relatively high, which will be greater than a certain probability threshold p1. However, objects outside of its $(1 + \varepsilon)$ *rNN* are unlikely to be hashed to the first bucket, meaning the probability of hashing to the first bucket is small and will be less than a certain threshold p2. It is expressed as Equations 2 and 3.

$$p_1 = Pr[I(p) = I(q)](is \quad \text{“}high\text{”} \quad if \quad p \quad is \quad \text{“}close\text{”} \quad to \quad q.) \tag{2}$$

$$p_2 = Pr[I(p) = I(q)](is \quad \text{“}low\text{”} \quad if \quad p \quad is \quad \text{“}far\text{”} \quad from \quad q.) \tag{3}$$

In other words, after the mapping transformation operation of the hash function, we divide the initial data set into many sub-datasets. The data in each sub-data set are close to each other, and the number of elements in the sub-data set is relatively small. Therefore, the problem of finding neighbouring elements in a large set is transformed into the problem of finding neighbouring elements in a relatively small data set, which reduces the computational cost. Alternatively, it can be understood as converting high-dimensional data into low-dimensional data while maintaining the similarity characteristics of the original data within a certain range. However, locally sensitive hashing cannot guarantee determinism. It is probabilistic, or it is possible to map two originally similar data into two completely different hash values or to map originally dissimilar data into the same hash value. High-dimensional data is inevitable in dimensionality reduction, as there will inevitably be some degree of data loss during the operation. However, fortunately, the design of locally sensitive hashing can adjust the corresponding parameters to control the probability of such errors as much as possible. This is also an important reason why locally sensitive hashing is widely used in various fields. The logic of locally sensitive hashing in this article is shown in the following figure. All similar data blocks in the fingerprint table will be divided into the same bucket. When retrieving a new data block, only the bucket where the data block should be stored must be searched. There is no need to traverse the entire fingerprint table for searching, which greatly reduces the time and computational consumption in the data block retrieval process and can accelerate the retrieval efficiency.

As shown in Figure 4, when a data block needs to be retrieved during the retrieval process, we can see that the local sensitive hash will calculate the bucket number that the data block should be placed in and perform the retrieval within that bucket. Regardless of whether the data block is previously stored, it can achieve the goal of fast retrieval.

4.1.2 Local sensitive hash implementation

In this article, the first step is to abstract the actual problem to achieve fast retrieval based on locally sensitive hashing. In practical problems, this article aims to achieve that when a data block's fingerprint is passed in, it can be linearly searched within the range of its similar fingerprints by searching for similar fingerprints rather than retrieving the entire fingerprint table. Therefore, the corresponding local sensitive hash directly searches for the bucket corresponding to a data block fingerprint after inputting it. At the same time, we need to understand several concepts: Euclidean distance, Jaccard distance, Hamming distance, and. The Euclidean distance in locally sensitive hashing refers to Equation 4:

$$H(V) = \frac{V \star R + b}{a} \tag{4}$$

R is a random vector, a is the bucket width, and b is a random variable uniformly distributed between [0, a]. It can also be understood that all vectors are mapped to a straight line through a hash function, and the mapped line is composed of many line segments of length a. Each vector V will be randomly mapped to a different line segment. Jaccard distance is a formula used to calculate the similarity between two data blocks. Hamming distance refers to the number of times the values at the corresponding positions in two vectors of the same length differ. We have completed the integration of practical problems and

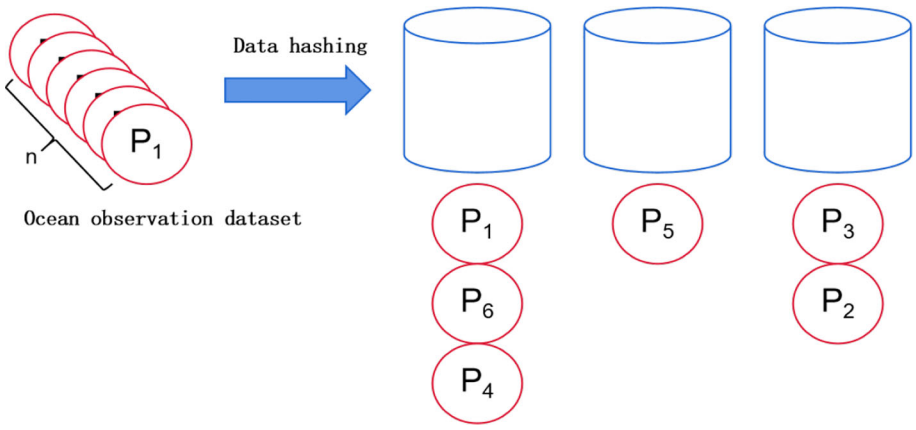


FIGURE 3
Diagram 1 of fast retrieval technology based on local sensitive hash.

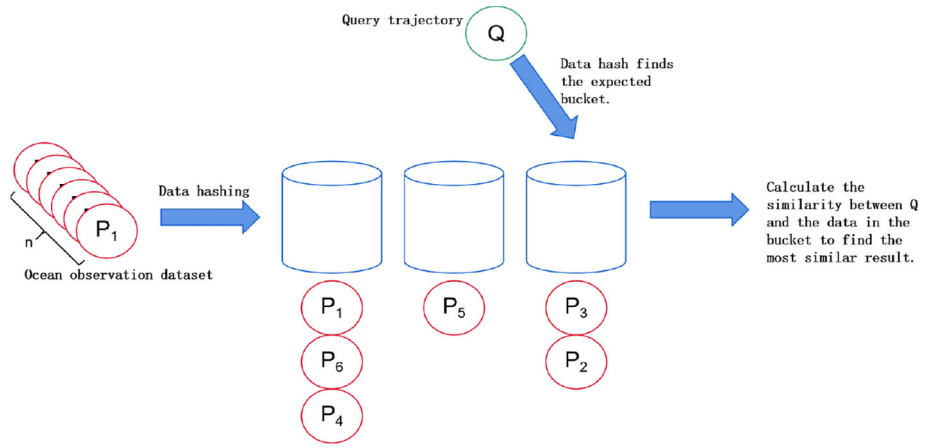


FIGURE 4
Diagram 2 of fast retrieval technology based on local sensitive hash.

locally sensitive hashing. Next, we will provide a detailed introduction to the specific implementation process:

Step 1: Data Preprocessing. Before completing feature extraction, we need to perform a preprocessing step on the data, which may include data cleaning, supplementing missing data, and standardizing the data. However, we only need to supplement the missing data in this article. In addition, the dataset used in this article is the chunking technique explained in the previous chapter, which uses the chunking technique to chunk the data and obtain the hash value of the data block. Finally, the data block size is applied as the second feature.

Step 2: Feature Extraction. In this section, we need to convert various data items in the dataset into feature vectors. This step is based on buckets, where each bucket contains multiple data blocks. The feature λ of each data block is composed of multiple parameters, including the data block identifier D_{ID}, data block size $Chunk_{size}$, and feature λ as shown in Formula 5:

$$\lambda = (D_{ID}, Chunk_{size}) \tag{5}$$

- Data Block Identification: The numerical value obtained by hashing the content of a data block (such as SHA-1) is used as the unique identifier for that data block.
- Data block size: Different sizes of data blocks are obtained based on different data block segmentation methods.

Step 3: Create a locally sensitive hash model. First, in offline mode, map all the vectors completed in the previous step to their respective index positions using the determined hash function. Then, input a vector to be searched and calculate the hash value using the same function as in the previous step. Find all the vectors in that vector's corresponding hash value positions, and calculate the Euclidean distance using the corresponding Euclidean distance calculation method. Finally, select the n vectors with the smallest Euclidean distance as the n results that are closest or most similar to the input vector.

Step 4: Optimize the number of hash buckets. When facing different practical problems, if the data volume is small, we can choose to optimize the number of hash buckets. By increasing or decreasing the number of hash tables for locally sensitive hashes, we can reduce the number of buckets to cope with different situations and practical problems. If the data volume is too large and the features are obvious, we can appropriately increase the number of hash buckets. Conversely, if the features are not obvious and the data volume is small, we can reduce the number of hash buckets to speed up the retrieval process.

Below we will provide pseudocode for local sensitive hashing as shown in Algorithm 1.

We can obtain a set of data block fingerprints through the above code, similar to the input data block fingerprint. If we can search for the input data block fingerprint within this set of data block fingerprints, we can save the need to search for the fingerprint of the data block to be retrieved from the entire fingerprint table. It can be simply finding the bucket number to which the data block to be retrieved belongs, making the number of data blocks in the entire bucket much simpler and more convenient than the entire hash table. It can be understood as simplifying large problems into small ones, achieving global optimization through local optimization. At the same time, it is emphasized that the fast retrieval based on locally sensitive hashing proposed in this section is aimed at saving computational resources when dealing with large-scale data. The purpose is to save the time wasted by linear retrieval, but it does not mean it can achieve fast retrieval in any scenario. The rough flowchart of fast retrieval and computation based on locally sensitive hashing is shown in Figure 5.

```
Input: Fingerprint of the data block to be retrieved;
Output: Fingerprint of data blocks that are similar to the fingerprint of the data block to be retrieved;

def_init_(self, tables_num:int, a:int, depth:int):

self.tables_num = tables_num

self.a = a

self.R = np.random.random([depth, tables_num])

self.b = np.random.uniform(0, a, [1, tables_num])

self.hash_tables = [dict() for i in range(tables_num) do]

def_hash(self, inputs: Union[List[List], np.ndarray]):

hash_val = np.floor(np.abs(np.matmul(inputs, self.R) + self.b)/self.a)

return hash_val

def insert(self, inputs):

inputs = np.array(inputs)

IF len(inputs.shape) == 1 then inputs = inputs.reshape([1, -1])

hash_index = self. hash(inputs)

for inputs_one, indexs in zip(inputs, hash_index) do

for i, key inenumerate(indesx) do self.hash_tables[i].setdefault(ley, []).append(tuple(inputs_one))

end for

end for

def query(self, inputs, nums=20):

hash_val = self._hash(inputs).ravel()

candidates = set()

for i, key in enumerate(hash_val) do candidates.update(self.hash_tables[i][key])

end for

candidates = sorted(candidates, key=lambda x: self.euclidean_dis(x, inputs))

return candidates[:nums]

def euclidean dis(x, y):

x = np.array(x)
```

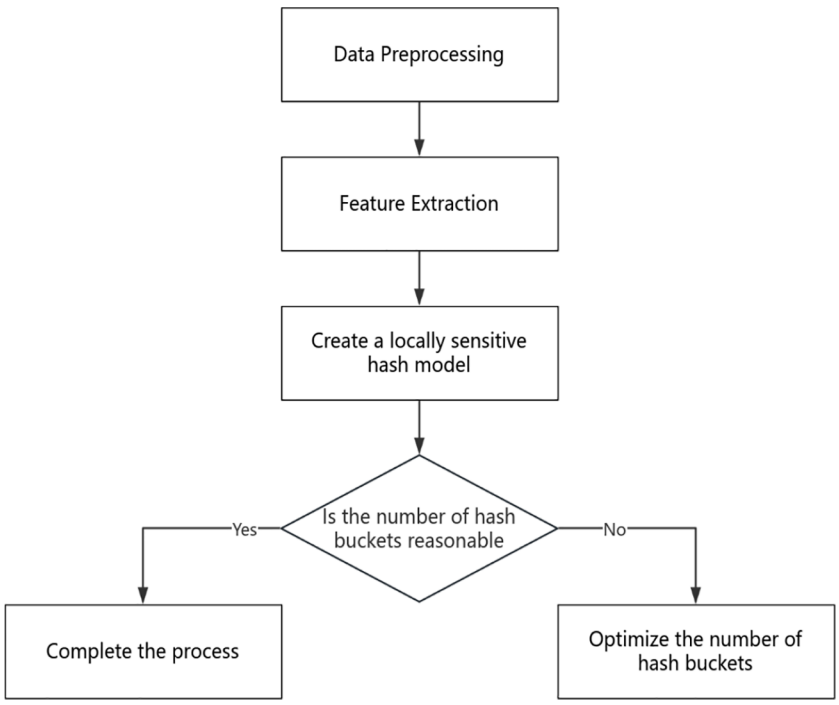


FIGURE 5
Local sensitive hash flowchart.

```
27: y = np.array(y)
28: return np.sqrt(np.sum(np.power(x - y, 2)))
29: IF_name _== '_main_' then
30: data = np.random.random([10000, 100])
31: query = np.random.random([100])
32: lsh = EuclideanLSH(10, 1, 100)
33: lsh.insert(data)
34: res = lsh.query(query, 20)
35: res = np.array(res)
36: print(np.sum(np.power(res - query, 2), axis=-1))
37: sort = np.argsort(np.sum(np.power(data - query, 2), axis=-1))
38: print(np.sum(np.power(data[sort[:20]] - query, 2), axis=-1))
39: print(np.sum(np.power(data[sort[-20]]: - query, 2), axis=-1)) =0
```

Algorithm 1. Locality-sensitive hashing.

4.2 Bucket optimization strategy

Before discussing this issue, we need to think about why we need to optimize the number of buckets. In practical applications, if we initially designed 5 buckets, as the amount of data that needs to be stored continues to increase, if we still scatter the data in five buckets, our retrieval efficiency will become lower and lower. Suppose we can continuously optimize the number of buckets according to the changes in the amount of data that needs to be stored. In that case, the entire duplicate data removal system will have a reasonable usage method. At the same time, we also need to consider another situation. Our initial design still had 5 buckets, but the storage device has just been replaced, and the amount of data we store is small. Therefore, we need to consider whether it is still necessary to use 5 buckets. In these two real-life situations, we need to make changes according to our different needs to achieve a satisfactory state of our duplicate data deletion system.

Implementing this is not difficult. We only need to visualize the number of buckets in various states to intuitively understand whether the number of buckets we are currently using is reasonable. The specific implementation algorithm is as follows as shown in Algorithm 2:

```
Input: The number of hash functions in LSH and the number of buckets for each hash function;
Output: Visualization results;
1: spark=SparkSession.builder.getOrCreate()
2: data=spark.read.csv("",header=True,inferSchema=True)
3: data=data.dropna()
4: assembler=VetorAssembler(inputCols=["featuer1","featuer2"],outputCol="featuer")
5: data=assembler.transform(data)
6: lsh=MinHashLSH(inputCol="featuer",outputCol="hashes",numHashTables=5)
7: model=lsh.fit(data)
8: hashedData=model.transform(data)
9: model=lsh.setNumHashTables(10).fit(data)
10: hashedData=model.transform(data)
11: hashedData.groupBy("hashes").count().show() =0
```

Algorithm 2. Bucket optimization algorithm.

We can solve existing problems intuitively through visual results. At the same time, we can make other optimizations based on the situation inside the bucket, such as the fingerprint table selection strategy under the ocean observation dataset mentioned in our next section. Through intuitive data, we can change the number of buckets for locally sensitive hashes based on storage requirements and analyze the dataset's characteristics through result graphs. However, in this article, we focus more on applying it to optimizing the number of buckets. At the same time, with continuous optimization, we can even analyze within which range the amount of data and how many buckets are more reasonable, laying a solid foundation for future work.

4.3 Fingerprint table selection strategy in ocean observation datasets

Before facing this problem, we need to understand why we need to make a decision strategy for fingerprint tables. Let us imagine that in the storage system of an ocean observation station, we use a duplicate data deletion system to achieve the goal of storing more data. As the amount of data increases, the fingerprint table in our memory will continue to grow. Just like the simple example we gave in our article, assuming the average size of a data block is 8KB, the generated fingerprints will be about 20GB. If we store 8TB of data,

we will generate nearly 20GB of fingerprints, which means we need to store nearly 20GB of fingerprint tables in our memory. The continuous increase of data will undoubtedly bring a huge memory burden.

The strategy of this article is to store a portion of the fingerprint tables in memory and another in external storage. The obvious purpose of this is to reduce the burden on memory. There are many advantages to doing this: 1. Save memory: Storing a portion of the hash table in external storage can effectively save memory resources, allowing the system to process larger datasets without being limited by memory size. 2. Improve performance: By storing hotspot data in memory, the search process for common data blocks can be accelerated without loading from disk every time. 3. Higher scalability: When processing very large amounts of data, the storage capacity of memory is limited, while external storage can provide almost unlimited expansion space, ensuring that the system can handle larger-scale deduplication tasks.

The benefits of doing so are self-evident, but the more important issue is deciding which part of the data to store in memory and which part to store in external storage. Since we only focus on ocean observation data in this article to solve the problem of storage devices for ocean observation stations, can we understand it this way? When facing time series datasets such as ocean observation, as long as there are more similar data blocks, we can understand that the probability of them appearing in the future observation process is also greater, which is what we understand as hot data. In other words, if there are many similar data blocks in some buckets generated by locally sensitive hashing, these data blocks can be defined as hotspot data. So we can store the fingerprint table of this bucket in memory and the rest in external storage if we divide the data into 5 buckets through local sensitive hashing, namely bucket 1, bucket 2, bucket 3, bucket 4, and bucket 5. Briefly introduce the meanings of a few characters: assuming that the access frequency of each data block is the same, the access frequency of each bucket is a_i, the number of data blocks in each bucket is n_i, and the average size of each data block is k_b,S represents the saved memory space size,m is the number of buckets stored in external storage, A_h represents the total required space size, and A_t represents the external storage space size. As shown in Equation 6:

$$S = A_h - A_t = k_l \cdot n - \sum_{i=1}^{m} n_i \cdot k_l \tag{6}$$

So, the consumption of external storage access mainly depends on each bucket's data volume and the pain's access frequency. At this point, we assume that the delay of external storage is the constant time T_{ext}, and each external storage access consumes a fixed time. The consumption of accessing external storage is proportional to the bucket's data volume and access frequency. For bucket i, the external storage access consumption Q_i is Equation 7:

$$Q_i = a_i \cdot n_i \cdot k_l \cdot T_{ext} \tag{7}$$

Therefore, the total access consumption Q is Equation 8:

$$Q = \sum_{i=1}^{m} a_i \cdot n_i \cdot k_l \cdot T_{ext} \tag{8}$$

Usually, we can choose buckets that are painful to put into memory and have high access frequency based on the following criteria: buckets with higher access frequency a_i are usually chosen to put into memory because they will bring higher performance improvement. The bucket in memory should be the largest bucket of a_i. Memory capacity limitation: Due to limited memory, storing some high-frequency access buckets in memory may only be possible. Usually, the storage capacity of memory Mmemis limited, so only buckets with high occupancy and access frequency can be selected until the memory capacity is filled.

5 Experimental results and discussion

Next, we will introduce the experimental results based on locally sensitive hashing. In this section, we will present the experiments based on local sensitive hashing in three directions: the impact of hash tables on the number of buckets, whether the goal of retrieving data blocks can be achieved, and retrieval efficiency. The specific details are as follows.

5.1 Experimental environment and data set source

The computers used in this experiment are shown in Table 1, and the data set used in this experiment is shown in Table 2. The datasets 1–4 used in this article are all from ocean observation datasets, which are a set of time series data sets generated by time changes, while the data set 5–8 is a data set generated by public daily network life. The more important reason for listing different data sets is to observe whether DSW is more suitable for deleting data generated by time series.While the proposed data partitioning framework effectively leverages temporal correlations in ocean observation datasets, its current implementation is tailored to the spatially constrained nature of the target private datasets, which originate from fixed-location sensors. These proprietary datasets exhibit dense temporal sampling but limited spatial coverage, spanning no more than 500 km^2 in targeted zones—contrasting with global-scale datasets like Argo or satellite remote sensing products. As a result, the framework prioritizes temporal partitioning to exploit intra-site time-series dependencies, which are critical for applications such as localized anomaly detection or short-term environmental forecasting in these confined environments as shown in Table 3.

TABLE 1 Specific operating environment of the experiment.

Device name	DELL XPS 8950
Processor	12*th* Gen Intel(R)Core(TM)i7-12700 2.10 GHz
RAM	64.0GB(63.7GB Available)
System type	Windows11/Ubnutu 22.04
Display adapter	NVIDIA GeForce RTX 3060 12GB

TABLE 2 Details of all data sets used in this experiment.

	Name	Source	Size	Specific content of the dataset
1	Ocean observation dataset 1(OD1)	Ocean observation station collection	1.94GB	Voyage data
2	Ocean observation dataset 2(OD2)	Ocean observation station collection	1.83GB	Buoy data
3	Ocean observation dataset 3(OD3)	Ocean observation station collection	1.01GB	Hidden target data
4	Ocean observation dataset 4(OD4)	Ocean observation station collection	1.92GB	Remote sensing data
5	General Dataset 1(GD1)	networkrepository.com	57.3MB	Web document data
6	General Dataset 2(GD2)	networkrepository.com	661MB	Web document data
7	General Dataset 3(GD3)	networkrepository.com	852MB	Web document data
8	General Dataset 4(GD4)	networkrepository.com	878MB	Web document data

5.2 The relationship between hash table and bucket

Firstly, we examined the impact of mapping the hash table on the number of buckets. As shown in the Figure 6, we can indirectly optimize the number of buckets by changing the hash table. This operation can be applied to different environments, as shown in Figure 6. We can see that when we change the number of hash tables, the number of buckets will decrease as the number of hash tables decreases. Through experiments, we can see the relationship between the hash table and the number of buckets, so we can control the number by changing the number of hash tables. When faced with large-scale data, such as the storage environment of ocean observation stations, we can reduce the number of comparisons and thus reduce the computational cost of retrieval by increasing the number of buckets and dispersing the data into more buckets.

5.3 Retrieve test results

On the other hand, we check whether the local sensitive hash model can provide us with a set of fingerprints similar to the fingerprint being queried by inputting the fingerprint to be queried. This section of the experiment mainly tests whether we can obtain the hash value we want through locally sensitive hashing. Therefore,

TABLE 3 Data features algorithm adaptation comparison table.

	Data Attribute	Marine Observation Context	Algorithm Adjustments
1	Time-series dependency	Continuous high-frequency sampling requires retention of time dependency relationships	Sliding Time Window
2	Non-renewable nature	*In situ* sensor failure leads to data loss that cannot be recovered, and data integrity verification is required	Real time CRC verification mechanism during data acquisition
3	Large volume	Single site generates over 10GB of time-series data per day, requiring compatibility with distributed storage	Indexing with Space-Time Grid

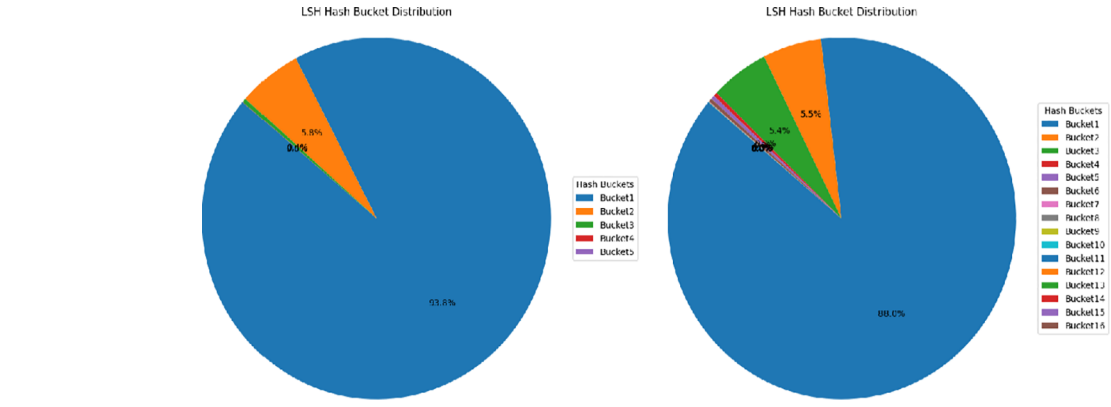


FIGURE 6
Visualization of the impact of bucket count.

TABLE 4 Algorithm provides a schematic table of results.

Similar results provided by LSH	Input data block fingerprint
ea13550d354f178211a33 772f1c46619ffa81114	ea13550d354f178211a33772f1c46619ffa81114, 960053af900262d8647867224b7099dd7b9e61ea, …
d77a30f6e3349b06fc10ae 541698ea1c43927fe0	d77a30f6e3349b06fc10ae541698ea1c43927fe0, 3f6223a1e77363fb10ede586fdfe2f7810d18a23, …
30bcb804a9aaa4e6e4dc7 e990bc7d15115ac856b	30bcb804a9aaa4e6e4dc7e990bc7d15115ac856b, 00d48d219fcd64b392175c4882c6017c9b758e5e, …
30a9318a3cc9fb13700da 0e350ef0a9dbc47ca2f	30a9318a3cc9fb13700da0e350ef0a9dbc47ca2f, 06e7cbd6cb45751cbeefbc2633a9e8989e1ae0db, …
c84d21e904cca69bc4532 c4ec06c1ec981d3fa9e	c84d21e904cca69bc4532c4ec06c1ec981d3fa9e, 7da99e56853c55368528cb793dff6cc54a7a1ccb, …
…	…

in this experiment section, we added a test data block from the training set to test whether this algorithm can accurately find the data block when it reappears. We continuously input 1000 randomly selected sets of data blocks for testing. These 1000 data blocks are all from the data block groups in the

As shown in Table 4, when we input the fingerprint to be queried, the local sensitive hash model can provide us with a set of fingerprints similar to the queried fingerprint. The table shows that after each input, there is an accurate data block in memory with an Euclidean distance of 0 from the input data block fingerprint, which is the backup of the data block in memory. This also indicates that the model can accurately identify whether the data block exists in memory and that the retrieval function is intact and can be applied. After this round of experiments, we can proceed to the next section of the experiment to further verify how much computational consumption can be reduced by the retrieval technology based on locally sensitive hashing in practical applications and to improve the subsequent work.

5.4 Mixed test results

Next, we will mix the data blocks in the training set with those that do not exist in the training set. In this experiment, we will prepare four sets of data, with a total of 1000 data blocks present in the training set, accounting for 20%, 40%, 60%, and 80%, respectively, as inputs to test the optimization ratio of the local sensitive hash based retrieval technique compared to traditional linear search in reducing the number of comparisons. As shown in Figure 7, as the proportion of the training set continues to increase, the retrieval technique based on local sensitive hashing also becomes increasingly effective in reducing the number of comparisons. This proves that in the practical application scenario of ocean observation stations, the retrieval technique based on local sensitive hashing will improve more over time compared to traditional linear search. This also greatly saves computational costs.

5.5 Comparison of differences between internal and external storage fingerprint tables

In the Figure 8, for the convenience of comparison and viewing, we have subtracted the memory consumption from LSH's memory consumption of LSH, aiming to make the comparison clearer. From the figure, we can see that under the same dataset type, the memory consumption of LSH is significantly lower than that of ordinary

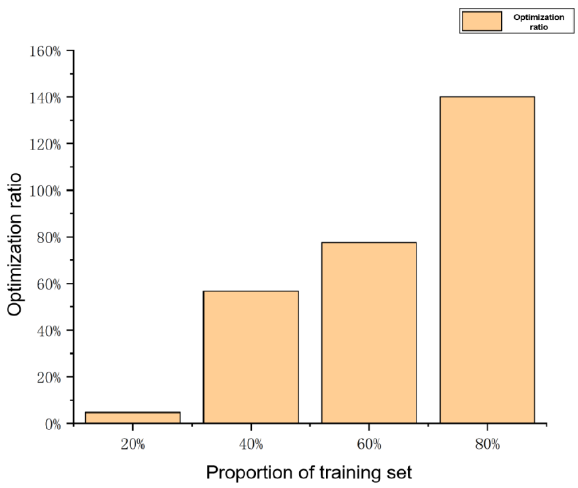


FIGURE 7
Optimization ratio result chart.

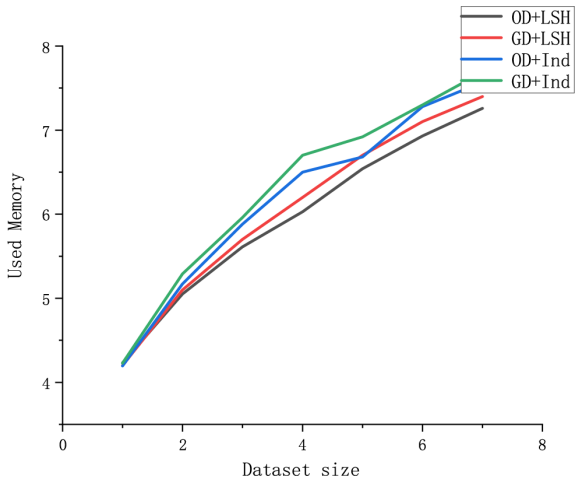


FIGURE 8
Memory usage comparison chart.

retrieval algorithms. This is because we store a part of the fingerprint table externally. Under the same LSH algorithm, memory consumption is lower due to the particularity of ocean observation data. Compared to the normal retrieval algorithm LSH, storing a portion of the fingerprint table externally reduces memory consumption.

5.6 Data duplication removal ratio

As shown in Figure 9, the comparison between the double sliding window segmentation algorithm and the content-based segmentation algorithm in the figure shows the disadvantages of LSH. Due to its occasional errors, the proportion of duplicate data deletion may be slightly reduced. However, reducing the duplicate data deletion ratio is within our acceptable range as it can accelerate retrieval speed. This is an abandonment problem, and we can tolerate abandoning a small portion of the duplicate data deletion ratio to improve the overall system efficiency.

6 Conclusions and future prospects

This article proposes a fast retrieval strategy for ocean observation data based on locally sensitive hashing, aiming to reduce the computational consumption of the duplicate data deletion system during the retrieval process. In order to achieve fast retrieval, similar data blocks are placed in similar buckets. In this way, when searching for the data block, only the corresponding bucket needs to be searched for the data block, without the need to search for all the data blocks. This can achieve the goal of saving computing resources and accelerate the retrieval speed. Finally, this

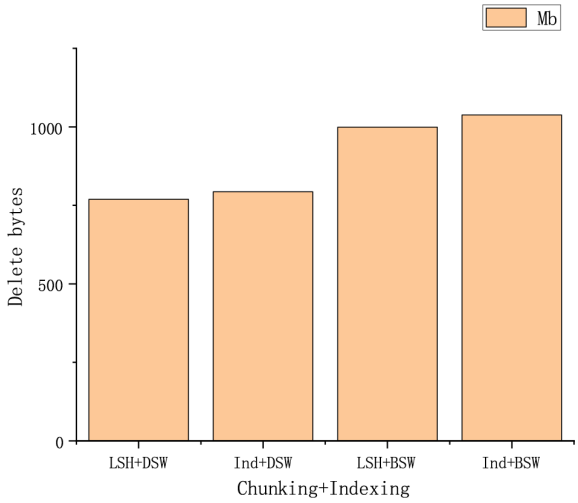


FIGURE 9
Duplicate data deletion ratio result chart.

article demonstrates through reasonable and rigorous experiments that as the amount of data in the storage device increases, the efficiency of fast retrieval algorithms based on local sensitive hashing also increases compared to other retrieval algorithms.

In future work, we will strive to apply fast retrieval algorithms based on locally sensitive hashing to other data, making them more widely applicable.

Data availability statement

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

Author contributions

MJ: Writing – original draft, Writing – review & editing. XM: Conceptualization, Writing – original draft. SG: Conceptualization, Writing – review & editing. XL: Software, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was funded by the Basic Research Business Fund for Undergraduate Universities in Heilongjiang Province. Authorization number is 2024-KYYWF-1248.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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

Agarwal, B., Akella, A., Anand, A., Balachandran, A., Chitnis, P.V., Muthukrishnan, C., et al. (2010). Endre: An end-system redundancy elimination service for enterprises. *NSDI* 10, 419–432.

Al Azad, Md W., and Mastorakis, S. (2022). The promise and challenges of computation deduplication and reuse at the network edge. *IEEE Wireless Commun.* 29.6, 112–118. doi: 10.1109/MWC.010.2100575

Ammons, J., Fenner, T., and Weston, D. (2022). "SCAIL: encrypted deduplication with segment chunks and index locality," in *2022 IEEE International Conference on Networking, Architecture and Storage (NAS)*, (2022 IEEE International Conference on Networking, Architecture and Storage (NAS)) Vol. pp. 1–9 (IEEE).

Bhagwat, D., Eshghi, K., Long, D. D., and Lillibridge, M. (2009). "Extreme binning: Scalable, parallel deduplication for chunk-based file backup," in *2009 IEEE International Symposium on Modeling, Analysis & Simulation of Computer and Telecommunication Systems*. 1–9 (IEEE).

Broder, A. Z. (1997). "On the resemblance and containment of documents," in *Proceedings Compression and Complexity of SEQUENCES 1997 (Cat. No. 97TB100171)*. 21–29 (Proceedings. Compression and Complexity of SEQUENCES 1997).

Debnath, B., Sengupta, S., and Li, J. (2010). "ChunkStash: speeding up inline storage deduplication using flash memory," in *2010 USENIX Annual Technical Conference (USENIX ATC 10)*. (Proceedings of the 2011 ACM SIGMOD International Conference on Management of data).

Debnath, B., Sengupta, S., and Li, J. (2011). "SkimpyStash: RAM space skimpy key-value store on flash-based storage," in *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*. (2010 USENIX Annual Technical Conference (USENIX ATC 10)). 25–36.

Huang, D., Zhao, D., Wei, L., Wang, Z., and Du, Y. (2015). Modeling and analysis in marine big data: Advances and challenges. *Math. Prob. Eng.* 2015, 384742. doi: 10.1155/2015/384742

Li, S., Lan, T., Balasubramanian, B., Lee, H. W., Ra, M.-R., Panta, R. K., et al. (2022). Pushing collaborative data deduplication to the network edge: An optimization framework and system design. *IEEE Trans. Netw. Sci. Eng.* 9, 2110–2122. doi: 10.1109/TNSE.2022.3155357

Lillibridge, M., Eshghi, K., Bhagwat, D., Deolalikar, V., Trezis, G., Camble, P., et al. (2009). Sparse indexing: Large scale, inline deduplication using sampling and locality. *Fast* 9, 111–123. doi: 10.14722/fast.2009.20100

Lu, G., Nam, Y. J., and Du, D. H. C. (2012). "BloomStore: Bloom-filter based memory-efficient key-value store for indexing of data deduplication on flash," in *2012 IEEE 28th Symposium on Mass Storage Systems and Technologies (MSST)*. 1–11 (2012 IEEE 28th Symposium on Mass Storage Systems and Technologies (MSST)).

Mahesh, B., Pavan Kumar, K., Ramasubbareddy, S., and Swetha, E. (2020). "A review on data deduplication techniques in cloud," in *Embedded Systems and Artificial Intelligence: Proceedings of ESAI 2019*, Fez, Morocco. (Springer) 825–833.

Nisha, T. R., Abirami, S., and Manohar, E. (2016). "Experimental study on chunking algorithms of data deduplication system on large scale data," in *Proceedings of the International Conference on Soft Computing Systems: ICSCS 2015*. (Proceedings of the International Conference on Soft Computing Systems: ICSCS 2015) 91–98 (Springer).

Park, N., and Lilja, D. J. (2010). "Characterizing datasets for data deduplication in backup applications," in *IEEE International Symposium on Workload Characterization (IISWC'10)*. 1–10 (IEEE International Symposium on Workload Characterization (IISWC'10)).

Prajapati, P., and Shah, P. (2022). A review on secure data deduplication: Cloud storage security issue. *J. King Saud University-Computer Inf. Sci.* 34, 3996–4007. doi: 10.1016/j.jksuci.2020.10.021

Reinsel, D., Gantz, J., and Rydning, J. (2017). Data age 2025: The evolution of data to life-critical. Don't focus on big data; focus on the data that's big. *Int. Data Corp. (IDC) White Paper*.

Riser, S. C., Freeland, H. J., Roemmich, D., Wijffels, S., Troisi, A., Belbeoch, M., et al. (2016). Fifteen years of ocean observations with the global Argo array. *Nat. Climate Change* 6, 145–153. doi: 10.1038/nclimate2872

Shynu, P.G., Nadesh, R.K., Menon, V. G., Venu, P., Abbasi, M., Khosravi, M. R., et al. (2020). A secure data deduplication system for integrated cloud-edge networks. *J. Cloud Comput.* 9, 61. doi: 10.1186/S13677-020-00214-6

Spring, N. T., and Wetherall, D. (2000). "A protocol-independent technique for eliminating redundant network traffic," in *Proceedings of the conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*. (Proceedings of

the conference on Applications, Technologies, Architectures, and Protocols for Computer Communication) 87–95.

Wallace, G., Douglis, F., Qian, H., Shilane, P., Smaldone, S., Chamness, M., Hsu, W., et al. (2012). Characteristics of backup workloads in production systems. *FAST* 12, 4–4. doi: 10.5555/2208461.2208465

Wen, X., Hong, J., Dan, F., and Yu, H. (2011). "SiLo: A Similarity-Locality based Near-Exact Deduplication Scheme with Low RAM Overhead and High Throughput," in *2011 USENIX Annual Technical Conference (USENIX ATC 11)*, Portland, OR (USENIX Association). Available at: https://www.usenix.org/conference/usenixatc11/silo-similarity-locality-based-near-exact-deduplication-scheme-low-ram.

Xia, W., Feng, D., Jiang, H., Zhang, Y., Chang, V., Zou, X., et al. (2019). Accelerating content-defined-chunking based data deduplication by exploiting parallelism. *Future Gen. Comput. Syst.* 98, 406–418. doi: 10.1016/j.future.2019.02.008

Xia, W., Jiang, H., Feng, D., Tian, L., Fu, M., Zhou, Y., et al. (2014). Ddelta: A deduplication-inspired fast delta compression approach. *Perform. Eval.* 79, 258–272. doi: 10.1016/j.peva.2014.07.016

Xu, G., Tang, B., Lu, H., Yu, Q., and Sung, C. W. (2019). "Lipa: A learning-based indexing and prefetching approach for data deduplication," in *2019 35th Symposium on mass storage systems and technologies (MSST)*. 299–310 (2019 35th Symposium on mass storage systems and technologies (MSST)).

You, L. L., Pollack, K. T., and Long, D. D. E. (2005). "Deep Store: An archival storage system architecture," in *21st International Conference on Data Engineering (ICDE'05)*. 804–815 (21st International Conference on Data Engineering (ICDE'05)).

Yuan, H., Chen, X., Wang, J., Yuan, J., Yan, H., Susilo, W., et al. (2020). Blockchain-based public auditing and secure deduplication with fair arbitration. *Inf. Sci.* 541, 409–425. doi: 10.1016/j.ins.2020.07.005

Zhang, Y., Jiang, H., Feng, D., Xia, W., Fu, M., Huang, F., Zhou, Y., et al. (2015). "AE: An asymmetric extremum content defined chunking algorithm for fast and bandwidth-efficient data deduplication," in *2015 IEEE Conference on Computer Communications (INFOCOM)*. 1337–1345 (2015 IEEE Conference on Computer Communications (INFOCOM)).

Zhang, Y., et al. (2016). A fast asymmetric extremum content defined chunking algorithm for data deduplication in backup storage systems. *IEEE Trans. Comput.* 66, 199–211. doi: 10.1109/TC.2016.2595565

Zhou, P., Zou, X., and Xia, W. (2022). "Dynamic clustering-based sharding in distributed deduplication systems," in *2022 IEEE/ACM 8th International Workshop on Data Analysis and Reduc*. (2022 IEEE/ACM 8th International Workshop on Data Analysis and Reduction for Big Scientific Data (DRBSD)) 54–55. doi: 10.1109/DRBSD56682.2022.00012

Zhu, B., Li, K., and Patterson, R.H. (2008). Avoiding the disk bottleneck in the data domain 644 deduplication file system. *Fast* 8, 1–14. doi: 10.5555/2208461.2208465

frontiers | Frontiers in Marine Science

TYPE Data Report
PUBLISHED 09 July 2025
DOI 10.3389/fmars.2025.1620021

OPEN ACCESS

EDITED BY
Chao Chen,
Suzhou University of Science and Technology, China

REVIEWED BY
Inam Ullah,
Gachon University, Republic of Korea
Mengmeng Yang,
Taishan University, China

*CORRESPONDENCE
Shuping Pan
zhang057xu@163.com

RECEIVED 29 April 2025
ACCEPTED 17 June 2025
PUBLISHED 09 July 2025

CITATION
Zhang Z, Liu J, Deng R, Hu Z and Pan S (2025) Satellite remote sensing of algal blooms in seagoing river in Eastern China. *Front. Mar. Sci.* 12:1620021. doi: 10.3389/fmars.2025.1620021

Satellite remote sensing of algal blooms in seagoing river in Eastern China

Zili Zhang[1], Jinsong Liu[1], Ruru Deng[2], Zunying Hu[1] and Shuping Pan[1*]

[1]Zhejiang Province Ecological Environment Monitoring Centre, Hangzhou, China, [2]Sun Yat-sen University, Guangzhou, China

KEYWORDS

chlorophyll-a (Chl-a), algal bloom, remote sensing, Chl-a remote sensing physical model, seagoing river

1 Introduction

Rivers and lakes serve as critically important water bodies on Earth's surface. Currently, freshwater algal blooms have become a global ecological phenomenon, occurring in diverse water bodies such as lakes, rivers, and reservoirs across temperate and tropical regions, and have increased significantly over the past several decades. Since the 1980s, around 68% of the world's lakes have undergone a persistent rise in algal bloom intensity (Ho et al., 2019; Sukharevich and Polyak, 2020). The degradation of water quality and eutrophication in global freshwater systems have emerged as significant environmental challenges, predominantly driven by anthropogenic activities and accelerating climate change (Suresh et al., 2023; Van Vliet et al., 2023; Liu et al., 2020; Wang et al., 2020). The occurrence of algal blooms in inland waters poses a serious threat to aquatic ecosystems and public health and safety (Brooks et al., 2016). Waters affected by algal blooms often exhibit high levels of eutrophication, and the subsequent death of blooms can deplete dissolved oxygen, resulting in black bloom events. Beyond degrading water aesthetics and severely damaging aquatic ecosystems, algal blooms also pose health risks to humans and animals through their associated toxins.

Remote sensing demonstrates distinct advantages in aquatic environmental monitoring through its comprehensive spatial information acquisition, operational efficiency, and cost-effectiveness. It enables the timely detection of marine environmental risks such as algal blooms and oil spills, and supports the rapid identification of water quality anomalies in coastal bays and estuarine rivers, as well as pollution source tracking. These capabilities have established novel research pathways for forecasting river-to-ocean pollution events through enhanced spatiotemporal monitoring frameworks. For example, Chen et al. (2023) used 30 years (1990–2019) of data to analyze the spatial and temporal characteristics of the Harmful algal blooms (HABs) along the Chinese coasts. To assess the feasibility of remote sensing for detecting HABs in small to medium-sized waterbodies, Liu et al. (2022) applied data from three satellites—Planetscope, Sentinel-2 and Landsat-8—to analyze the impacts of spatial resolution, spectral band availability, and waterbody size on detection accuracy. Similarly, Binding et al. (2018) investigated algal bloom dynamics in Lake Winnipeg using

satellite-derived chlorophyll-a (Chl-a) and metrics of bloom intensity, spatial extent, severity, and duration over the MERIS mission period. Current remote sensing monitoring of algal blooms, however, predominantly targets large water bodies (e.g., oceans, lakes, and reservoirs), while riverine algal blooms have received disproportionately less attention. This disparity is primarily attributed to the lower bloom frequency in river systems compared to lentic ecosystems, coupled with the heightened requirements for retrieval accuracy and stability imposed by complex river hydrodynamics (Rolim et al., 2023).

Algal bloom retrieval models with remote sensing are primarily classified into three categories: empirical, semi-empirical, and physical models (Li et al., 2025; Yang et al., 2025; Vasilakos et al., 2020; Lu et al., 2020; Chen et al., 2022; Yang et al., 2022; Chang et al., 2015). Empirical and semi-empirical models primarily rely on statistical analysis to establish relationships between remote sensing signals and Chl-a concentrations through extensive field-measured datasets (EI-Rawy et al., 2020; Yang et al., 2023; Xiao et al., 2022). These models demonstrate operational simplicity and computational efficiency, but exhibit limited generalizability beyond specific study areas. In contrast, physical models are grounded in rigorous radiative transfer theory, which quantitatively describes the relationship between aquatic components and satellite-derived irradiance through specific absorption coefficients and scattering coefficients of water constituents (Li et al., 2025; Guo et al., 2022). These models simulate light propagation processes in both atmospheric and aquatic environments using radiative transfer equations, enabling quantitative inversion of water quality parameter from remote sensing data. In recent years, the rapid advancement of artificial intelligence (AI) has facilitated the successful application of machine learning algorithms, particularly Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and Random Forest (RF), in remote sensing-based monitoring of organic pollutants in aquatic environments (Vinothkumar and Karunamurthy, 2023; Zhang et al., 2022; Ruescas et al., 2018; Deng et al., 2019). Furthermore, emerging deep learning architectures have demonstrated enhanced capabilities in water color remote sensing for retrieving concentrations of critical water quality parameters (Arshad et al., 2024, 2023; Khan et al., 2023; Ullah et al., 2024). The integrated application of multi-model approaches enables complementary advantages among different methodologies, significantly enhancing the accuracy and stability of water environment remote sensing monitoring.

This study proposes an integrated approach combining satellite remote sensing with *in situ* measurements to establish a multi-technique collaborative monitoring framework for algal blooms in sea-reaching rivers. The framework addresses the challenges of cross-system ecological transitions and salinity gradients in river-to-sea transitional zones. A seasonal algal bloom event in the Qiantang River Basin from July to September 2016 was documented, with its complete phenological cycle (initiation, evolution, and senescence) through multi-temporal remote sensing data at a 30m resolution. The results demonstrate that the Chl-a remote sensing physical mechanism model offers advantages of robust adaptability and algorithm stability, enabling large-scale synchronous monitoring of river algal blooms. Furthermore, the integration of extensive synchronous remote sensing observations with conventional *in situ* point monitoring significantly enhances the spatiotemporal resolution and efficiency of river bloom surveillance.

2 Study area

The Qiantang River Basin is situated between latitude 28°N and 30.5°N and longitude 117.5°E and 120.5°E. Its main stream spans 668 kilometers, encompassing tributaries such as the Xin'an River and Fuchun River, with an average annual runoff of 43.458 m^3 and a drainage area of approximately 60,000 km^2 (Sun et al., 2016). The construction of multistage hydrojunction projects along the mainstem and tributaries of the Qiantang River has induced flow attenuation in certain reaches. Compounded by rapid regional socioeconomic development in recent decades, this anthropogenic modification of hydrological regimes has facilitated the persistent accumulation of nutrients (e.g., nitrogen and phosphorus) within aquatic systems. Such conditions create a critical threshold whereby algal blooms can be rapidly triggered when meteorological and hydrological parameters reach conducive levels. Seasonal algal blooms in the Qiantang River Basin were initially documented as early as the late 20th century, with particularly extensive outbreaks occurring in 2004 and 2010 that triggered severe deterioration of water quality across the watershed (Gao et al., 2025; Reinl et al., 2020; Zhou et al., 2022).

3 Data and methods

3.1 Satellite data sources and ground-based observational data

The HJ-1A/B satellites, China's first domestically developed civilian satellites dedicated to environmental monitoring and disaster mitigation/emergency response, were launched on September 6, 2008. The HJ-1A/B satellites were each equipped with two wide-coverage multispectral CCD cameras, covering four broad spectral bands in the visible and infrared ranges. When operated jointly, these dual-camera systems achieve a 4-day revisit cycle, enabling push-broom imaging with a swath width of 720 km, a spatial resolution of 30 m, and four spectral bands. The HJ-1A/B satellites adopt a band configuration modeled after the U.S. Landsat series, featuring medium spatial resolution and broad spectral coverage, while achieving a wider swath width and shorter revisit cycle compared to their counterparts. One single image of the HJ-1A/B satellites can achieve complete coverage of the Qiantang River Basin, fulfilling the temporal resolution requirements of the study. The key parameters of the HJ-1A/B satellites' CCD sensors are presented in Table 1.

10.3389/fmars.2025.1620021

TABLE 1 The CCD data parameters of HJ-1A/B satellites.

ID	Band (μm)	Spatial Resolution (m)	Swath Width (km)	Revisit Period (day)
1	0.43–0.52	30	360 (single imager) 700 (two imagers)	4
2	0.52–0.60	30		
3	0.63–0.69	30		
4	0.76–0.90	30		

Satellite imagery acquired over the Qiantang River system from July to September 2016 was screened according to the cloud cover threshold of<20%, with the final selected images presented in Table 2.

In addition, synchronized field investigations of algal blooms and emergency water quality monitoring were conducted from July to August 2016. Ground-based investigations employed a dual-mode approach: (1) Fixed-point instrumentation for regular monitoring of water quality parameters and algal bloom occurrence; (2) Event-driven manual sampling with laboratory quantification during bloom episodes in affected areas. Field sampling and analysis were conducted in strict compliance with applicable water quality monitoring standards. The spatial distribution of *in situ* fixed sampling locations is illustrated in Figure 1. The monitored parameters included conventional physicochemical indicators as well as biological indicators such as Chl-a, algal density, and algal species composition and so on.

3.2 Development of remote sensing retrieval model for algal blooms

The water-leaving radiance captured by satellite remote sensing sensors results from the combined effects of: (1) air-water interface reflectance and refraction, (2) water column absorption and scattering processes, and (3) benthic substrate absorption and reflectance. The spectral characteristics of water bodies are predominantly characterized by volume scattering, resulting from the combined effects of water molecules and suspended impurities within the water. The primary pollutants in the water body are suspended sediments, oxygen-consuming organic matter, and Chl-a. The inherent optical parameters (IOPs), such as absorption and scattering coefficients, exhibit wavelength-independent characteristics across the light spectrum.

TABLE 2 Satellite data employed in this study.

Satellite data	Acquisition time (2016)			
	July	Aug	Sep	Nov
HJ-1A/B	22	15	9	3
	25	20	26	6
	26	24		11

The radiative transfer process in water bodies can be expressed as follows (Li et al., 2022; Lu et al., 2020): The water-leaving radiance (Lw) is composed of upwelling scattered radiance from the entire water column (L_s) and the bottom substrate-reflected radiance (L_b), as mathematically represented by Equation 1:

$$L_w = L_s + L_b \tag{1}$$

Based on the radiative scattering characteristics and radiative transfer processes in aquatic environments, the determination of the total radiative transfer model for water bodies requires two essential components: 1) solution of upwelling scattering throughout the entire water column, and 2) calculation of substrate reflectance at the water bottom.

For upwelling scattering, the incident light intensity differs across varying water depths. When considering the scattering contribution from a thin water layer of thickness *dh* at depth *h* (Equation 2).

$$dL_s = \frac{1}{4\pi} E\beta p(\Theta) dh \tag{2}$$

E is the downwelling irradiance at water depth *h*, which can be expressed as by Equation 3

$$E = E_0 \cos\theta' e^{-\frac{(\alpha+\beta)h}{\cos\theta'}} \tag{3}$$

In the equation, α denotes the total absorption coefficient of the water body, while βx4E3A;. represents the total scattering coefficient (Equations 4, 5).

$$\alpha = \alpha_w + D_s\alpha_s + D_u\alpha_u + D_v\alpha_v \tag{4}$$

$$\beta = \beta_w + D_s\beta_s + D_u\beta_u + D_v\beta_v \tag{5}$$

The symbols *w, s, u* and *v* represent pure water, suspended sediment, oxygen-consuming organic matter, and Chl-a, respectively, while *d* denotes the unknown concentrations of each constituent to be solved.

The incident light is first attenuated by the thin water layer. The upward scattered light then undergoes secondary attenuation through the upper water layer before emerging from the water surface. The expression is given as follows by Equation 6:

$$dL_s = \frac{1}{4\pi} E\beta p(\Theta) e^{-\frac{(\alpha+\beta)h}{\cos\varphi'}} dh \tag{6}$$

The upwelling scattered radiance *Ls* of the entire water column can be derived by integrating Equation 6, as represented by Equation 7.

$$L_s = \frac{E_0 \cos\theta' \beta p(\Theta)}{4\pi\mu k}(1 - e^{-\mu k h}) \tag{7}$$

Then $\mu = \frac{1}{\cos\theta'} + \frac{1}{\cos\varphi'}$, the extinction coefficient k = (α + β).

Assuming the water bottom substrate is a Lambertian surface, the radiance exiting the water body is represented by Equation 8.

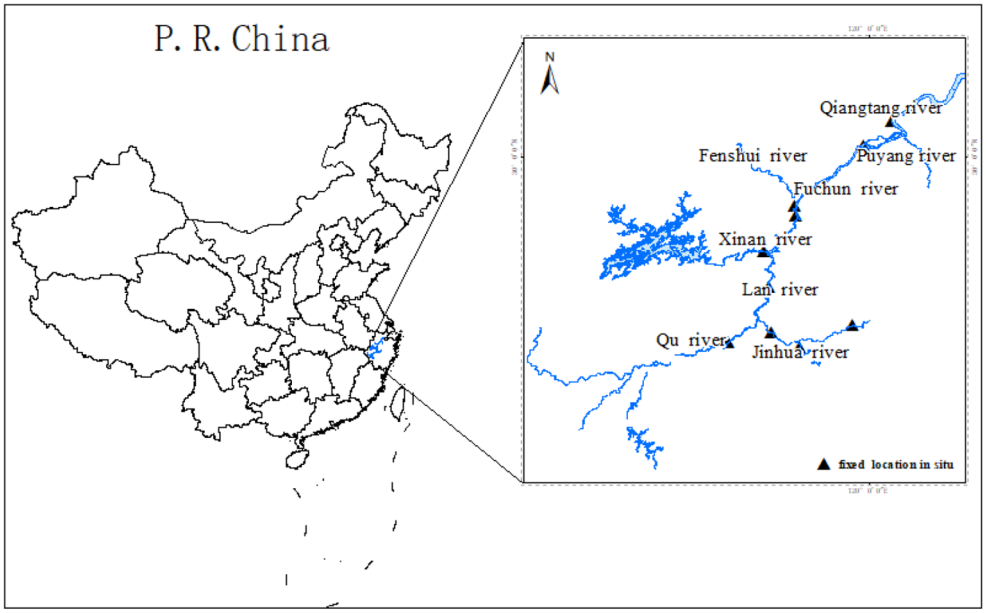


FIGURE 1
Schematic diagram of river reach distribution in the Qiantang River.

$$L_b = \frac{E_0 cos\theta' R_b}{\pi} e^{-\mu kh} \quad (8)$$

The conversion from radiance to surface reflectance (Rw) is expressed as Equation 9:

$$R = \frac{\pi L}{E} \quad (9)$$

Substituting (7) and (8) into (1) and transforming to reflectance via (9) yields the general radiative transfer equation for water bodies represented by Equation 10.

$$R = \frac{\beta p(\Theta)}{4\mu k}(1 - e^{-\mu kh}) + R_b e^{-\mu kh} \quad (10)$$

The scattering and absorption coefficients of suspended sediment, colored dissolved organic matter (CDOM), and Chl-a can be experimentally determined. The unknowns to be resolved by the model include D_s, D_v, D_u and water depth H and benthic substrate reflectance R_b (Equations 4, 5).

For the underdetermined system of equations where the number of unknowns exceeds the number of spectral bands, the Chl-a concentration in water bodies can be solved by analyzing the optical properties of the study area and selecting sensitive bands to construct the equations (Li et al., 2025; Yang et al., 2025; Liang et al., 2024). In this study, since HJ-1A/B employed consist of four spectral bands, the Chl-a concentration in water bodies was derived by solving a system of four simultaneous equations.

After a series of data processing steps, including geometric correction, radiometric calibration, radiometric correction, atmospheric correction, land-water separation, and remote sensing inversion of water quality parameters—followed by calibration using sampled laboratory measurements—the remote sensing monitoring results of Chl-a concentration and algal bloom status in the Qiantang river were finally obtained.

3.3 Geometric correction and radiometric calibration

Radiometric calibration is the process of converting the digital number (DN) values in raw imagery into radiance using calibration parameters provided in satellite data files. Radiometric correction is performed to derive planetary reflectance for all bands at each pixel from radiance using solar spectral irradiance per band, with corrections applied for solar incidence angle and viewing geometry effects.

3.4 Atmospheric correction and land-water separation

Atmospheric correction aims to remove/minimize the influence of atmospheric scattering and absorption in remote sensing data. Atmospheric correction analyzes atmospheric factors, constructs and solves models to maximally mitigate atmospheric interference. The dark target method was employed to derive the relevant atmospheric parameters in this study. To mitigate water vapor effects, clear/deep water pixels were employed as dark targets. After measuring the reflectance of clear water, the atmospheric optical thickness and transmittance were calculated to achieve atmospheric correction.

Following atmospheric correction, water-land separation was performed using a simple threshold approach with the modified Normalized Difference Water Index (NDWI) (Liang et al., 2023; Rad et al., 2021; Mcfeeters, 1996).

3.5 Retrieval of chlorophyll-a concentration in surface water

Given that phytoplankton predominantly accumulate at the air-water interface during extensive bloom events, the sensor-reaching radiance in red and near-infrared bands is mainly representative of surface water characteristics. The reflectance and absorption characteristics of Chl-a lead to a significant enhancement of spectral features in both the red and near-infrared bands, thereby providing more favorable conditions for the retrieval of Chl-a concentration. The study adopts the Chl-a water quality remote sensing physical model described in Section 2 to retrieve water Chl-a concentration through simultaneous inversion using both red and shortwave near-infrared band data (Equations 4, 5). Since the penetration depth of shortwave near-infrared radiation is approximately 25 cm, the retrieved Chl-a concentration primarily represents the surface water Chl-a level, with minimal influence from bottom reflectance.

3.6 Remote sensing index-based classification of algal blooms

Chl-a concentration thresholds serve as critical indicators for classifying algal bloom severity in aquatic ecosystem monitoring standards. Adhering to established *in situ* monitoring protocols and accounting for regional hydrological characteristics, this study implemented the following threshold-based classification: a mild algal bloom corresponds to Chl-a concentrations of 15–25 μg/L, a moderate algal bloom to 25–50 μg/L, and a severe algal bloom to concentrations exceeding 50 μg/L.

4 Results and discussion

4.1 Validation of accuracy

To validate the accuracy of remote sensing inversion for Chl-a concentration in water bodies, ground-based synchronous observations were collected, yielding a total of 49 paired synchronous ground observation points during satellite overpasses. Figure 2 shows a comparison between the remote sensing inversion results of the algal bloom index and the *in situ* measured Chl-a concentrations on August 15 and August 20, 2016, during the algal bloom event in the Qiantang River. Figure 3 shows the accuracy comparison between the remote sensing inversion results of Chl-a and the measured values.

As shown in Figure 3, the remotely sensed Chl-a concentration exhibits good linear agreement with *in situ* measurements (r=0.7984), demonstrating the capability of remote sensing to capture relative Chl-a concentration trends. The field-measured data exhibit significant fluctuations, whereas the remote sensing results demonstrate more gradual variations. The possible reason is that the water sampling collects data from a single point, while the remote sensing result represents the average concentration over a 30×30 meter pixel area of the water surface. This leads to a relatively smoother variation in the remote sensing data, analogous to a low-pass filtering effect.

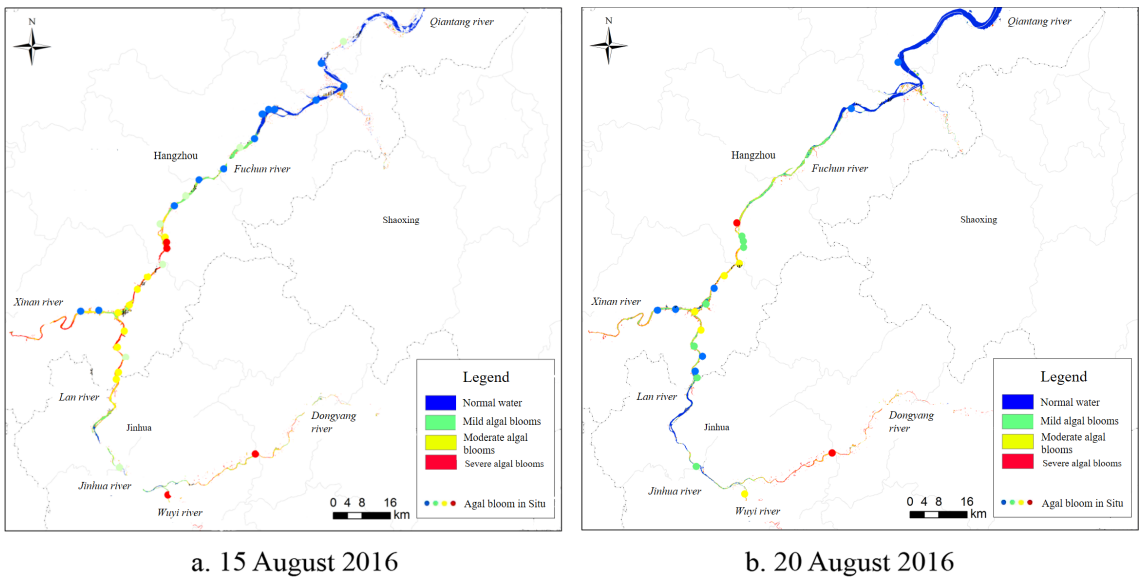


FIGURE 2
Validation of Qiantang river algal bloom remote sensing inversion against *in situ* measurements with **(a)** (August 15, 2016) and **(b)** (August 20, 2016).

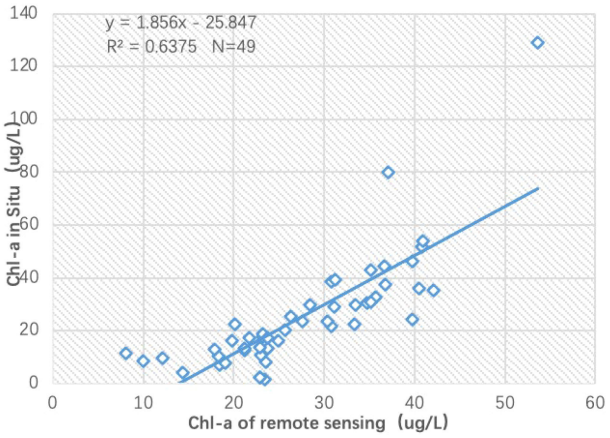


FIGURE 3
Accuracy validation of remote sensing retrieval for chlorophyll-a (Chl-a) in the Qiantang River.

4.2 Spatial distribution characteristics of algal blooms

Through the aforementioned remote sensing methodology, the spatial distribution of algal blooms in the Qiantang River during satellite overpass can be accurately obtained. Unlike ground-based monitoring which can only obtain data at discrete sampling points, satellite-derived results provide the spatial distribution of Chl-a in water bodies during satellite overpasses, enabling rapid identification of the most severe algal bloom areas. Through statistical analysis methods, the occurrence area and proportion of algal blooms in each river section can be obtained. Taking the remote sensing inversion results on August 20, 2016, as an example, the algal bloom area in the main reaches of the Qiantang River within the study region was 138.93 km². Among this, the areas of mild, moderate, and severe algal blooms were 41.72 km², 51.79 km², and 45.42 km², respectively, with moderate algal bloom exhibiting the highest proportion. The results are presented in Figure 4.

Calculate the proportional area coverage of algal blooms at different severity levels for each river segment separately. On August 20, 2016, the Qiantang River mainstream section had the lowest proportion of algal bloom coverage. The Wuyi River and Xin'an River segments accounted for the largest proportion of severe algal bloom coverage, with approximately half of their water areas affected by intense blooms, as shown in Figure 5.

4.3 Remote sensing monitoring of algal bloom evolution processes

Statistics and comparative analysis of multi-temporal remote sensing inversion results can elucidate the temporal evolution of algal blooms. Remote sensing analysis reveals that the algal bloom dynamics in the Qiantang River system, occurring from July to September 2016, exhibited four characteristic phases: incipient stage, proliferation stage, climax stage, and regression stage, as shown in Figure 6.

A. In the initial stage of algal bloom occurrence, as shown in the satellite remote sensing results of the Qiantang River water system on July 22, 2016, the bloom intensity was predominantly mild. A large area of moderate to severe algal blooms occurred in the river section between Lan river and Xin'an river and the confluence of the Fuchun river and Fenshui river. Meanwhile, the algal bloom area in major tributaries such as the Xin'an river further expanded, with increased bloom intensity.
B. During the algal bloom development phase, as shown by the satellite remote sensing results of the Qiantang River system on July 25, the spatial distribution of blooms in the main channel had expanded from the lower reaches of the Lan river to the Fuyang city section in the upper reaches of the Fuchun River, though the bloom intensity remained predominantly mild. A large area of moderate to severe algal blooms occurred in the river section between at the confluence of

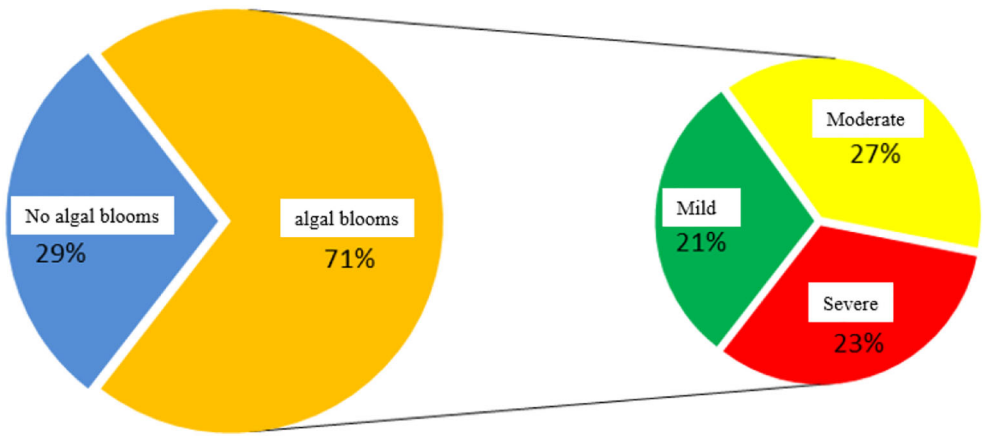


FIGURE 4
Areal percentage of algal blooms in key reaches of the Qiantang River.

the Lan river and Xin'an river and the confluence of the Fuchun river and Fenshui river. Meanwhile, the algal bloom area in major tributaries such as the Xin'an River further expanded, with increased bloom intensity.

C. The peak bloom period, as demonstrated by the satellite remote sensing results of the Qiantang river water system on July 29, 2016. The distribution range of algal blooms in the main channel had covered the river section from the lower reaches of Lan river to the upper reaches of Fuchun river (Fuyang City segment). Meanwhile, large-scale severe algal blooms in other river sections further intensified with continued expansion in coverage area. After August 15th, the intensity of algal blooms in the Qiantang river water system began to intensify again. The satellite remote sensing results of algal blooms in the Qiantang river system on August 20 show that the distribution of algal blooms has spread throughout the main river channel upstream of the Qiantang river, as well as major tributaries such as the Xin'an river and Jinhua river. Moreover, severe algal blooms in the main channel have covered the Fuchun river and Lan river. The algal blooms in major tributaries such as the Xin'an river were also predominantly severe, reaching the peak of both distribution area and intensity during this bloom event. The algal bloom in the Qiantang river on August 24 maintained the outbreak status observed on August 20, with similar distribution patterns and coverage extent. The bloom had spread across the main channel upstream of Qiantang river, as well as major tributaries including the Xin'an river and Jinhua river.

D. The algal bloom decline phase, as demonstrated by the satellite remote sensing results of the Qiantang River water system on September 26, both the intensity and spatial distribution of the bloom gradually diminished after early September. By mid-to-late September, specifically around September 25, the river conditions had essentially returned to their original state.

5 Conclusion and prospects

Grounding in radiative transfer mechanisms, this research established a physics-driven model for Chl-a concentration retrieval from pixel reflectance of remote sensing. The model explicitly accounts for Chl-a, delivering transparent inversion mechanisms and physically interpretable parameters. By exclusively retrieving the surface Chl-a concentration, this approach can effectively reflect the actual conditions of algal blooms. Validation shows a good linear relationship (R=0.7984) between the remote sensing retrieval results and *in situ* measurements, confirming the reliability of satellite data in monitoring algal bloom dynamics in the Qiantang River.

Given the vast area of the Qiantang River Basin (approximately 50,000 km^2), high-resolution satellites with limited swath width were inadequate for complete coverage. The study employed HJ-1A/B satellite data (Environment Satellite) which provides single-scene coverage of the entire study area. However, data availability was constrained by adverse weather

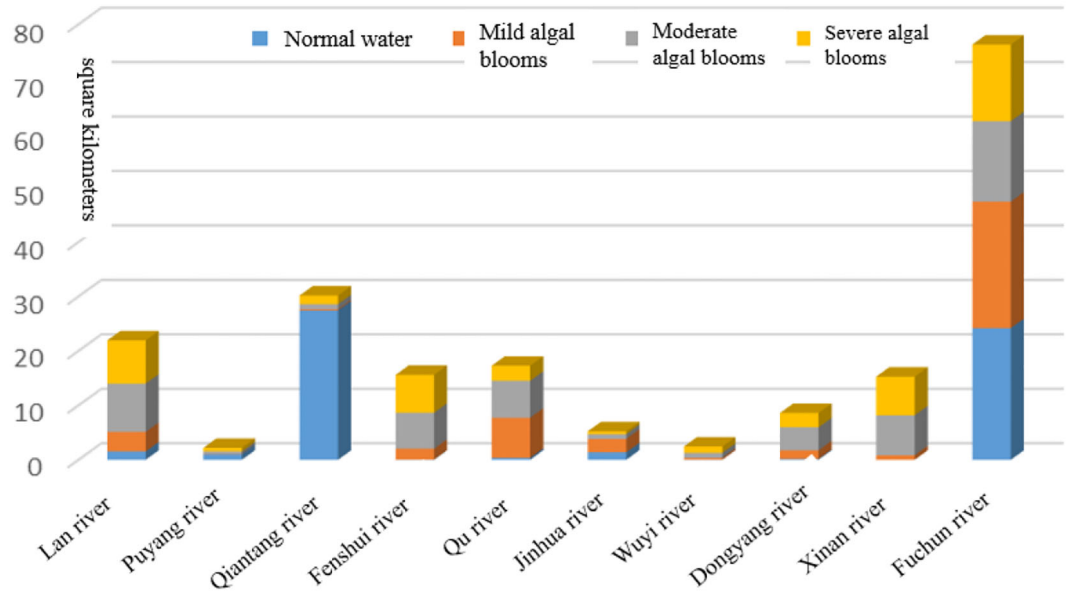


FIGURE 5
Distribution of algal bloom area in the main reaches of the Qiantang River water system.

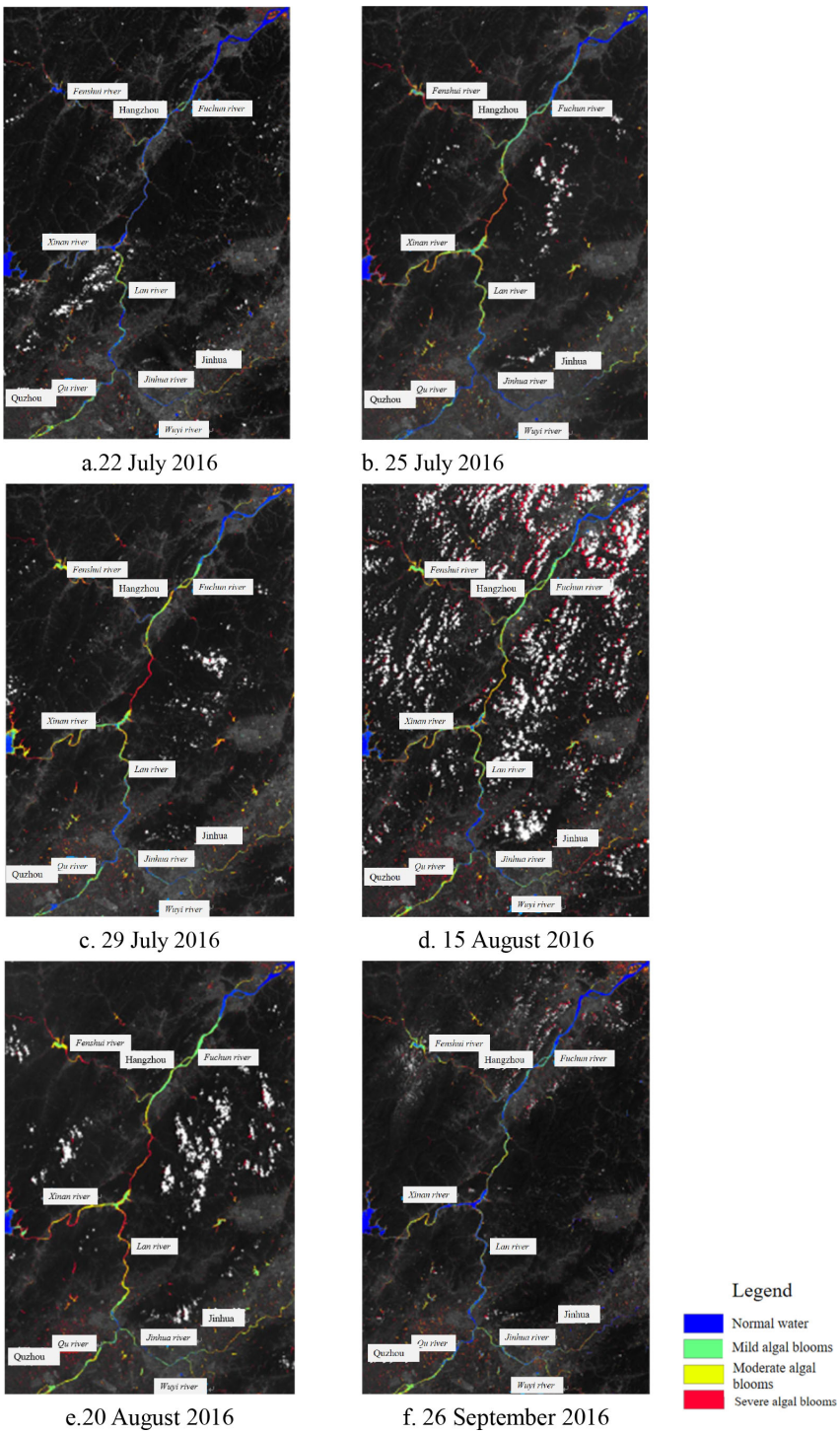


FIGURE 6
Satellite remote sensing of algal blooms of 2016 in the Qiantang river: **(a)** July 22, **(b)** July 25, **(c)** July 29, **(d)** August 15, **(e)** August 20, **(f)** September 26.

conditions and satellite imaging schedules, with particularly scarce acquisitions during early-mid September and October when no usable data were available. Therefore, for emergency water quality monitoring of inland rivers and lakes, multi-source satellites such as Gaofen-1 (GF-1) or Sentinel can be employed to compensate for the temporal coverage limitations of single satellite. With the advancement of deep learning theory, integrating deep learning into quantitative remote sensing physical models is expected to improve the accuracy of remote sensing inversion as well as enhance the model's dynamic predictive capabilities. This will achieve a transition from status monitoring to early warning of algal blooms for river systems.

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

ZZ: Conceptualization, Software, Investigation, Writing – review & editing, Funding acquisition, Resources, Writing – original draft, Project administration, Validation, Methodology, Visualization, Formal analysis, Supervision, Data curation. JL: Writing – original draft, Writing – review & editing. RD: Writing – original draft, Methodology, Investigation. ZH: Writing – review & editing, Validation, Writing – original draft. SP: Resources, Validation, Project administration, Conceptualization, Formal analysis, Data curation, Methodology, Visualization, Supervision, Writing – review & editing, Funding acquisition, Investigation, Software, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by Zhejiang Province "Pioneering Soldier" and "Leading Goose" R&D Project (Grant numbers: 2023C03011) and the National Natural Science Foundation of China (No. 42471445).

Acknowledgments

Author would like to thank potential supervisor, editor and reviewer for their advice and comment.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

Arshad, T., Zhang, J. P., and Ullah, I. (2024). A hybrid convolution transformer for hyperspectral image classification. *Eur. J. Remote Sens.* 57, 2330979. doi: 10.1080/22797254.2024.2330979

Arshad, T., Zhang, J., Ullah, I., Ghadi, Y. Y., Alfarraj, O., and Gafar, A. (2023). Multiscale feature-learning with a unified model for hyperspectral image classification. *Sensors* 23, 7628. doi: 10.3390/s23177628

Binding, C., Greenberg, T., McCullough, G., Watson, S., and Page, E. (2018). An analysis of satellite-derived chlorophyll and algal bloom indices on Lake Winnipeg. *J. Great Lakes Res.* 44, 436–446. doi: 10.1016/j.jglr.2018.04.001

Brooks, B. W., Lazorchak, J. M., Howard, M. D., Johnson, M. V., Morton, S., Perkins, D. A., et al. (2016). Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems? *Environ. Toxicol. Chem.* 35, 6–13. doi: 10.1002/etc.3220

Chang, N.-B., Imen, S., and Vannah, B. (2015). Remote sensing for monitoring surface water quality status and ecosystem state in relation to the nutrient cycle: a 40-year perspective. *Crit. Rev. Environ. Sci. Technol.* 45, 101–166. doi: 10.1080/10643389.2013.829981

Chen, J. Y., Chen, S. S., Fu, R., Li, D., Jiang, H., Wang, C. Y., et al. (2022). Remote sensing big data for water environment monitoring: current status, challenges, and future prospects. *Earth's Future* 10. doi: 10.1029/2021EF002289

Chen, C., Liang, J., Yang, G., and Sun, W. (2023). Spatio-temporal distribution of harmful algal blooms and their correlations with marine hydrological elements in offshore areas. China. *Ocean Coast. Management* 238, 106554. doi: 10.1016/j.ocecoaman.2023.106554

Deng, C. B., Zhang, L. F., and Cen, Y. (2019). Retrieval of chemical oxygen demand through modified capsule network based on hyperspectral data. *Appl. Sci.* 9, 4620. doi: 10.3390/app9214620

EI-Rawy, M., Fathi, H., and Abdalla, F. (2020). Integration of remote sensing data and in *situ* measurements to monitor the water quality of the Ismailia Canal, Nile Delta, Egypt. *Environ. Geochem. Health* 42, 2101–2120. doi: 10.1007/s10653-019-00466-5

Gao, L., Shangguan, Y., Sun, Z., She, Q., and Zhou, L. (2025). A novel algal bloom risk assessment framework by integrating environmental factors based on explainable machine learning. *Ecol. Inform.* 87, 103098. doi: 10.1016/j.ecoinf.2025.103098

Guo, Y., Deng, R., Li, J., Hua, Z., Wang, J., Zhang, R., et al. (2022). Remote sensing retrieval of total nitrogen in the pearl river delta based on landsat8. *Water* 14, 3710. doi: 10.3390/w14223710

Ho, J. C., Michalak, A. M., and Pahlevan, N. (2019). Widespread global increase in intense lake phytoplankton blooms since the 1980s. *Nature* 574, 667–670. doi: 10.1038/s41586-019-1648-7

Khan, S., Ullah, I., Ali, F., Shafiq, M., Ghadi, Y.Y., 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. doi: 10.3389/fmars.2022.1094915

Li, J., Deng, R., Guo, Y., Lei, C., Hua, Z., and Yang, J. (2025). Quantitative estimation of organic pollution in inland water using sentinel-2 multispectral imager. *Sensors* 25, 2737. doi: 10.3390/s25092737

Li, J., Zhang, W., Deng, R., Lu, Z., Liang, Y., Shen, X., et al. (2022). The study of spatial-temporal characteristics for COD_{mn} in shenzhen reservoir based on GF-1 WFV. *J. Remote Sens.* 26, 1562–1574. doi: 10.11834/jrs.20219380

Liang, J., Chen, C., Song, Y., Sun, W., and Yang, G. (2023). Long-term mapping of land use and cover changes using Landsat images on the Google Earth Engine Cloud Platform in bay area-A case study of Hangzhou Bay, China. *Sustain. Horizons* 7, 100061. doi: 10.1016/j.horiz.2023.100061

Liang, Y. J., Deng, R. R., Liang, Y. H., Tang, Y. M, Xiong, L. H, and Li, J. Y. (2024). high-resolution remote sensing of chlorophyll-a across large basins. *Remote Sens. Technol. Application* 39, 1490–1499. doi: 10.11873/j.issn.1004•0323.2024.6.1490

Liu, S., Glamore, W., Tamburic, B., Morrow, A., and Johnson, F. (2022). Remote sensing to detect harmful algal blooms in inland waterbodies. *Sci. Total Environment* 851, 158096. doi: 10.1016/j.scitotenv.2022.158096

Liu, S., Xie, Z., Liu, B., Wang, Y., Gao, J., Zeng, Y. J., et al. (2020). Global river water warming due to climate change and anthropogenic heat emission. *Global Planetary Change* 193, 103289. doi: 10.1016/j.gloplacha.2020.103289

Lu, S., Deng, R., Liang, Y., Xiong, L., Ai, X., and Qin, Y. (2020). Remote sensing retrieval of total phosphorus in the pearl river channels based on the GF-1 remote sensing data. *Remote Sens.* 12, 1420. doi: 10.3390/rs12091420

Mcfeeters, S. K. (1996). The use of the normalized difference water index (NDWI) in the delineation of open water features. *Int. J. Remote Sensing* 17, 1425–1432. doi: 10.1080/01431169608948714

Rad, A. M., Kreitler, J., and Sadegh, M. (2021). Augmented normalized difference water index for improved surface water monitoring. *Environ. Model. Software* 140, 10503. doi: 10.1016/j.envsoft.2021.105030

Reinl, K. L., Sterner, R. W., Lafrancois, B. M., and Brovold, S. (2020). Fluvial seeding of cyanobacterial blooms in oligotrophic Lake Superior. *Harmful Algae* 100, 101941. doi: 10.1016/j.hal.2020.101941

Rolim, S. B. A., Veettil, B. K., Vieiro, A. P., and Kessler, A. B. (2023). Remote sensing for mapping algal blooms in freshwater lakes: a review. *Environ. Sci. Pollut. Res.* 30, 19602–19616. doi: 10.1007/s11356-023-25230-2

Ruescas, A. B., Hieronymi, M., Mateo-Garcia, G., Koponen, S., Kallio, K., and Camps-Valls, G. (2018). Machine learning regression approaches for colored dissolved organic matter (CDOM) retrieval with S2-MSI and S3-OLCI simulated data. *Remote Sens.* 10, 786. doi: 10.3390/rs10050786

Sukharevich, V. I., and Polyak, Y. M. (2020). Global occurrence of cyanobacteria: causes and effects (review). *Inland Water Biol.* 13, 566–575. doi: 10.1134/S1995082920060140

Sun, Q., Zhang, C., Liu, M., and Zhang, Y. (2016). Land use and land cover change based on historical space–time model. *Solid Earth* 7, 1395–1403. doi: 10.5194/se-7-1395-2016

Suresh, K., Tang, T., Michelle, T. H., Marc, F. P., Maryna, S., Florian, S. D., et al. (2023). Recent advancement in water quality indicators for eutrophication in global freshwater lakes. *Environ. Res. Lett.* 18, 063004. doi: 10.1088/1748-9326/acd071

Ullah, I., Ali, F., Ali, A., Naeem, H. M. Y., and Bai, X. (2024). Optimizing underwater connectivity through multi-attribute decision making for underwater IoT deployments using remote sensing technologies. *Front. Mar. Sci.* 11. doi: 10.3389/fmars.2024.1468481

Van Vliet, M. T. H., Thorslund, J., Strokal, M., Hofstra, N., Florke, M., Macedo, H. E., et al. (2023). Global river water quality under climate change and hydroclimatic extremes. *Nat. Rev. Earth Environ.* 4, 687–702. doi: 10.1038/s43017-023-00472-3

Vasilakos, C., Kavroudakis, D., and Georganta, A. (2020). Machine learning classification ensemble of multitemporal sentinel-2 images: the case of a mixed mediterranean ecosystem. *Remote Sens.* 12, 2005. doi: 10.3390/rs12122005

Vinothkumar, J., and Karunamurthy, A. (2023). Recent advancements in artificial intelligence technology: trends and implications. *Quing Int. J. Multidiscip. Sci. Res. Dev.* 2 (1), 1-11. doi: 10.54368/qijmsrd.2.1.0003

Wang, Y. K., Zhang, N., Wang, D., and Wu, J. C. (2020). Impacts of cascade reservoirs on Yangtze River water temperature: Assessment and ecological implications. *J. Hydrol.* 590, 125240. doi: 10.1016/j.jhydrol.2020.125240

Xiao, Y., Guo, Y. H., Yin, G. D., Zhang, X., Shi, Y., Hao, F. H., et al. (2022). UAV multispectral image-based urban river water quality monitoring using stacked ensemble machine learning algorithms-a case study of the zhanghe river, China. *Remote Sens.* 14, 3272. doi: 10.3390/rs14143272

Yang, J., Deng, R., Ma, Y., Li, J., Guo, Y., and Lei, C. (2025). Satellite retrieval and spatiotemporal variability in chlorophyll-a for marine ranching: an example from daya bay, guangdong province, China. *Water* 17, 780. doi: 10.3390/w17060780

Yang, H., Kong, J., Hu, H., Du, Y., Gao, M., and Chen, F. (2022). A review of remote sensing for water quality retrieval: progress and challenges. *Remote Sens.* 14, 1770. doi: 10.3390/rs14081770

Yang, Y. C., Zhang, D. H., Li, X. S., Wang, D. M., Yang, C. H., and Wang, J. H. (2023). Winter water quality modeling in xiong'an new area supported by hyperspectral observation. *Sensors* 23, 4089. doi: 10.3390/s23084089

Zhang, D. H., Zhang, L. F., Sun, X. J., Gao, Y., Lan, Z. Y., Wang, Y. N., et al. (2022). A new method for calculating water quality parameters by integrating space-ground hyperspectral data and spectral-*in situ* assay data. *Remote Sens.* 14, 3652. doi: 10.3390/rs14153652

Zhou, J., Han, X., Brookes, J. D., and Qin, B. (2022). High probability of nitrogen and phosphorus co-limitation occurring in eutrophic lakes. *Environ. pollut.* 292, 118276. doi: 10.1016/j.envpol.2021.118276

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 29 July 2025
DOI 10.3389/fmars.2025.1574991

OPEN ACCESS

EDITED BY
Muhammad Yasir,
China University of Petroleum (East China), China

REVIEWED BY
Yonggang Ji,
China University of Petroleum (East China), China
Xiaoting Sun,
Tongji University, China
Fulong Yao,
Newcastle University, United Kingdom

*CORRESPONDENCE
Yang Li
dreyang@163.com

RECEIVED 11 February 2025
ACCEPTED 27 June 2025
PUBLISHED 29 July 2025

CITATION
Fan J, Guo M, Zhang L, Liu J and Li Y (2025) A marine ship detection method for super-resolution SAR images based on hierarchical multi-scale Mask R-CNN. *Front. Mar. Sci.* 12:1574991. doi: 10.3389/fmars.2025.1574991

A marine ship detection method for super-resolution SAR images based on hierarchical multi-scale Mask R-CNN

Jiancong Fan[1], Miaoxin Guo[1], Lei Zhang[1], Jianjun Liu[1,2] and Yang Li[1,2]*

[1]College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, China, [2]Provincial Key Laboratory for Information Technology of Wisdom Mining of Shandong Province, Shandong University of Science and Technology, Qingdao, China

Synthetic aperture radar (SAR) images have all-weather observation capabilities and are crucial in ocean surveillance and maritime ship detection. However, their inherent low resolution, scattered noise, and complex background interference severely limit the accuracy of target detection. This paper proposes an innovative framework that integrates super-resolution reconstruction and multi-scale maritime ship detection to improve the accuracy of marine ship detection. Firstly, a TaylorGAN super-resolution network is designed, and the TaylorShift attention mechanism is introduced to enhance the generator's ability to restore the edge and texture details of the ship. The Taylor series approximation is combined to optimize the attention calculation, and a multi-scale discriminator module is designed to improve global consistency. Secondly, a hierarchical multi-scale Mask R-CNN (HMS-MRCNN) detection method is proposed, which significantly improves the multi-scale maritime ship detection problem through the cross-layer fusion of shallow features (small targets) and deep features (large targets). Experiments on SAR datasets show that TaylorGAN has achieved significant improvements in both peak signal-to-noise ratio and structural similarity indicators, outperforming the baseline model. After adding super-resolution reconstruction, the average precision and recall of HMS-MRCNN are also greatly improved.

KEYWORDS

synthetic aperture radar (SAR), super-resolution reconstruction, marine ship detection, multiscale feature fusion, Mask R-CNN, TaylorShift attention mechanism

1 Introduction

Synthetic aperture radar (SAR) as an active microwave remote sensing imaging technology, with its all-weather, all-day capability, and low dependence on weather and lighting conditions, has important applications in the fields of marine surveillance, ship detection, etc (Gao et al., 2024; Meng et al., 2024; Wu et al., 2024). SAR imagery is able to

provide high-resolution data under complex weather conditions, which makes it an ideal tool for monitoring ship activities at sea. However, low-resolution SAR images can have an adverse effect on the identification of marine vessels. Due to the low resolution of the equipment and the complex imaging environment, low-resolution images often lack sufficient details, especially in complex backgrounds, and are often affected by scattering noise, background clutter, etc (Li et al., 2023; Cao et al., 2024). Using super-resolution reconstruction technology, more image details can be restored without increasing the hardware costs. Therefore, the development of a method that combines image super-resolution reconstruction with target detection can not only improve the utilization value of SAR images but also provide more efficient technical support for ensuring maritime safety and monitoring maritime traffic (Tang et al., 2024).

Super-resolution (SR) technology, as an effective means to improve image quality, has received widespread attention in SAR image processing (Jiang et al., 2024). SR algorithms for images are mainly categorized into two types: traditional methods and methods of deep learning. In traditional methods, interpolation methods predict unknown pixel information based on known pixel points to improve image resolution. Common interpolation methods include nearest neighbor interpolation (Blu et al., 2004), bilinear interpolation (Tong and Leung, 2007), and bicubic interpolation (Chang et al., 2004). Although the interpolation method is faster in reconstruction, it does not utilize *a priori* knowledge in the low-resolution image, so the reconstructed high-resolution image lacks the main texture information, whereas in reconstruction-based methods *a priori* information is introduced as constraints to reconstruct the image. The main reconstruction-based methods are the convex set projection method (Tom and Katsaggelos, 1996), the iterative inverse projection method (Irani and Peleg, 1991), and the maximum *a posteriori* probability estimation method (Liu and Sun, 2013). Reconstruction-based methods have limited utilization of prior knowledge, and learning-based methods, in order to improve this problem, introduce external datasets for training in order to learn more information about the image so that the reconstruction results contain more high-frequency details. Learning-based methods can be categorized into shallow learning methods and deep learning methods. Shallow learning methods mainly include based sample learning (Freeman et al., 2002), based neighborhood embedding (Chang et al., 2004), and based sparse representation methods (Xu et al., 2019). Shallow learning methods can achieve better results when trained on small-scale datasets, but the learning ability of the model needs to be improved. In recent years, deep learning-based methods have made great breakthroughs in the work of super-resolution reconstruction of images, and the deep learning methods are mainly based on three types of baseline networks: convolutional neural networks, generative adversarial networks, and attention mechanism networks. In 2014, Dong et al. proposed a super-resolution convolutional neural network (SRCNN), which is firstly applied to SR reconstruction, and the network convolves the input image through three layers (feature extraction and representation layer, nonlinear mapping layer, and reconstruction layer), it realizes the mapping from low resolution to high resolution, and the reconstruction effect on image resolution is better than the traditional reconstruction methods (Dong et al., 2014). In 2017, Legid et al. proposed the super-resolution generative adversarial network (SRGAN), which is the first time that generative adversarial networks (GANs) have been applied to the field of SR reconstruction. The network makes good use of the generative-adversarial properties of GAN networks, the generator and discriminator are trained alternately until convergence, the output shows more realistic texture details compared to traditional reconstruction methods, and the resolution is significantly improved visually (Ledig et al., 2017). In 2018, Zhang et al. proposed the residual channel attention network (RCAN), introduced the channel attention mechanism into the SR reconstruction task, and designed a deep residual channel convolutional network (Zhang et al., 2018). The network can learn the information of different channels of the feature map, set *different* weights for each channel, and finally reconstruct a high-resolution image. In recent years, with the excellent performance of Transformer in other image processing fields, scholars have begun to pay attention to the combination of Transformer and SR tasks. In 2020, Yang et al. proposed a texture transformation network (TTSR) for image super-resolution, which can combine low-frequency and high-frequency information to learn the deep correspondence of images, thereby stacking texture details in high-resolution images across scales and enhancing the reconstruction results (Yang et al., 2020). Due to the excellent performance of deep learning in optical image super-resolution, deep learning-based methods have been applied to SAR image super-resolution reconstruction in recent years. In 2018, Wang et al. directly applied the SRGAN network to the Terra-SAR dataset and achieved excellent results in reconstruction accuracy and computational efficiency (Wang et al., 2018). In 2019, Gu et al. proposed a DGAN network for the super-resolution reconstruction of pseudo-high-resolution SAR images, which effectively removed noise from SAR images and improved the resolution of SAR images (Gu et al., 2019). In 2020, Shen et al. used residual convolutional neural networks to improve the spatial resolution of polarimetric SAR images, which was superior to traditional methods in terms of image detail preservation (Shen et al., 2020). In 2022, Smith et al. proposed a SAR image super-resolution reconstruction method based on residual convolutional neural networks, which was superior to traditional methods in terms of reconstruction accuracy and computational efficiency. This method combines ViT with CNN for the super-resolution reconstruction of near-field SAR images, enhancing the details of the generated images (Smith et al., 2022). In 2023, Zhang et al. proposed a learnable probabilistic degradation model, which introduces SAR noise before the cycle-GAN framework, learns the relationship between low-resolution and high-resolution SAR images, and improves the resolution of SAR images (Zhang et al., 2023a). In 2024, Jiang et al. proposed a lightweight super-resolution generative adversarial network (LSRGAN), which improved the resolution of SAR images by introducing deep separable convolution (DSConv) and SeLU activation function, and constructed a lightweight residual module

(LRM) to optimize the GAN network for SAR images (Jiang et al., 2024). In addition, the feature learning capability of the model is significantly improved by combining the optimized coordinated attention (CA) module.

The biggest feature of the traditional SAR image ship detection algorithm is manual extraction. The manual extraction process first preprocesses the image to reduce the image noise; secondly, sea and land segmentation is performed to prevent the near-coastal land area from interfering with the ship detection; finally, the ship is detected. The constant false alarm rate (CFAR) algorithm (Baldygo et al., 1993) is one of the most classical methods in traditional SAR target detection. The algorithm models the ocean background clutter and distinguishes between target ships and background noise. CFAR algorithm does not apply to complex ocean backgrounds or ship targets with different directions, lengths, and widths, and its generalization performance is poor. With the development of artificial intelligence technology, target detection methods based on deep learning are applied by researchers in the field of SAR ship detection, which can be divided into one-stage and two-stage methods. One-stage methods treat all regions of the image as potential target regions and use only one deep convolutional network to recognize the target, which is faster, such as the YOLO series (Redmon, 2016; Ge, 2021). Two-stage methods use region suggestion module or selective search method to localize and recognize targets with higher accuracy (Su et al., 2022), such as R-CNN (Girshick et al., 2014), Faster R-CNN (Ren, 2015), Cascade R-CNN (Cai and Vasconcelos, 2018), Grid R-CNN (Lu et al., 2019), etc. Girshick et al. applied a convolutional neural network (CNN) for the first time to the target detection task and built an R-CNN network, thus achieving good results. Faster R-CNN extracts candidate frames by regional recommendation networks (RPN) and introduces a multi-task loss function, which shows good performance in target detection. In addition, researchers have proposed a large number of methods for the problem of target detection in SAR images. In the same year, Sun et al. (2021) proposed an anchor-free ship detection framework named CP-FCOS, which employs a category-position module to improve localization accuracy by guiding the position regression branch using semantic classification features. Zhang et al. (2021) proposed a novel quadruple pyramid network consisting of four FPNs and conducted experiments on five common SAR datasets, achieving good results. The authors also verified that Quad-FPN has good transferability. In 2022, Tang et al. proposed an algorithm based on Faster R-CNN for target detection in SAR images by using the Bhattacharyya distance (BD) instead of intersection over union (IoU) to avoid the limitations of the commonly used intersection over union ratio in target detection networks for small target recognition, which was evaluated on the LS-SSDD-v1.0 dataset and achieved significant detection results (Tang et al., 2022). In 2023, Zhang et al. proposed the SCSA-Net to address the effects of complex noise and land background interference on target detection in SAR images and also proposed the global average precision loss (GAP loss) to solve the "fractional bias" problem (Zhang et al., 2023b). In 2024, Yasir et al. (2024a) proposed the lightweight YOLOShipTracker model, which was optimized for YOLOv8n via the HGNetv2 reconciliation header and combined with a novel multi-target tracking technique (C-BIoU) to enable efficient, real-time tracking of ships in short-duration SAR image sequences. In the same year, Yasir et al. (2024b) also developed SwinYOLOv7, which combines YOLOv7 with the Swin Transformer and the CBAM Attention Module to demonstrate excellent performance in a variety of SAR datasets, especially in cluttered and near-shore environments. In addition, MGSFA-Net, a multi-scale global scattering feature association network for SAR ship identification, is introduced by Zhang et al. (2024). Their method can effectively capture the intrinsic physical scattering features and significantly improve the identification performance even with limited training data. However, the SAR image itself has limited resolution, which makes it difficult to present key details such as ship contours and deck structures, which will affect the detection algorithm's complete identification of targets. At the same time, the existing target detection methods still have the problem of lack of balance when facing multi-scale targets, which makes it difficult to take into account the small and large targets, resulting in some scale targets being missed.

In order to solve the abovementioned problems, this paper proposes a hierarchical multi-scale marine ship detection method based on Mask R-CNN to accurately detect ships and combines the TaylorGAN super-resolution reconstruction algorithm to enhance the resolution of SAR images.

The main contributions of this paper are as follows:

1. The TaylorGAN super-resolution reconstruction algorithm is proposed by introducing the TaylorShift attention mechanism in the GAN network to improve the resolution of ship image details, especially to enhance the sharpness of ship edges;
2. A hierarchical multi-scale marine ship detection method based on Mask R-CNN is proposed. Different convolutional layers are used to extract the large and small target features of SAR images, respectively. The extracted features are introduced into the RoI Align layer. The multi-scale features are balanced through L2 normalization to improve the detection accuracy.
3. The problem of insufficient detection of small targets is solved by fusing multi-scale feature information to avoid the degradation of detection accuracy due to low resolution.

The subsequent sections of the paper are organized as follows: Section 2 presents a detailed description of the proposed framework, including the TaylorGAN-based super-resolution reconstruction method and the HMS-MRCNN multi-scale ship detection architecture. Section 3 introduces the SAR datasets used in this study, elaborates on the experimental setup, outlines the evaluation metrics, analyzes the detection performance across various models, and reports results from comprehensive ablation studies. Section 4 concludes the paper by summarizing the major findings and highlighting potential directions for future research.

2 Methodology

This section first describes the TaylorGAN super-resolution reconstruction network. Secondly, the hierarchical multi-scale Mask R-CNN architecture proposed in this study is described in detail.

2.1 Super-resolution reconstruction network architecture

Existing super-resolution reconstruction algorithms are often faced with the problems of blurred edges and degraded structures when directly applied to SAR images, which make it difficult to meet the needs of fine reconstruction. For this reason, this paper proposes a network structure called TaylorGAN to improve the super-resolution quality of SAR images, and the overall architecture is shown in Figure 1, including a generator and a discriminator.

The generator takes a low-resolution SAR image as inputs and gradually restores the resolution of the image through the multilayer TaylorShift attentional module and the step-by-step upsampling structure while enhancing the ability to model the global structure and local details of the image. Although the TaylorShift mechanism itself does not directly enhance the image details, the model can effectively capture the high-frequency regions such as hull contours and edges with the help of the hierarchical upsampling and feature fusion structure so as to realize high-quality detail reproduction.

The discriminator adopts a multi-scale structure design, combined with the TaylorShift attention module, to extract features at multiple spatial resolutions, perceive the differences between local details and the global structure of the image, and ultimately output the true/false prediction results through the classification header and to effectively optimize the training direction of the generator.

2.1.1 Generator

The upper part of Figure 1 is the generator of TaylorGAN, which can be divided into three main modules: input module, feature extraction module, and image reconstruction module. The low-resolution SAR image is used as input, denoted as $I_{LR} \in R^{C\times H\times W}$, where C represents the number of image channels, and H and W represent the height and width of the image, respectively. First, the initial features are extracted by the embedding module composed of convolutional layers and ReLU activation function, and it is represented as $x_0 \in R^{C\times H\times W}$. Subsequently, the embedded features are processed by the layer Grid TaylorShift Block, and the attention mechanism is used to capture the long-distance dependencies in the image and model the local semantic information. The TaylorShift attention mechanism replaces the Softmax function by Taylor series expansion, greatly reducing the computational complexity until the resolution is increased to $I_{HR} \in R^{C\times Kh\times Kw}$. In order to gradually improve the image resolution, the generator designs multiple upsampling modules to improve the reconstruction accuracy by gradually expanding the spatial scale. After each level of upsampling, the TaylorShift attention module is stacked to further enhance the feature representation ability, especially the modeling ability of high-frequency details such as edges and contours, thereby improving the clarity and structural consistency of the generated image. Finally, through a set of convolutional layers and Tanh activation functions, the feature map is mapped to the output image at the target resolution $I_{SR} \in R^{C\times Rh\times Rw}$, and r is the magnification factor.

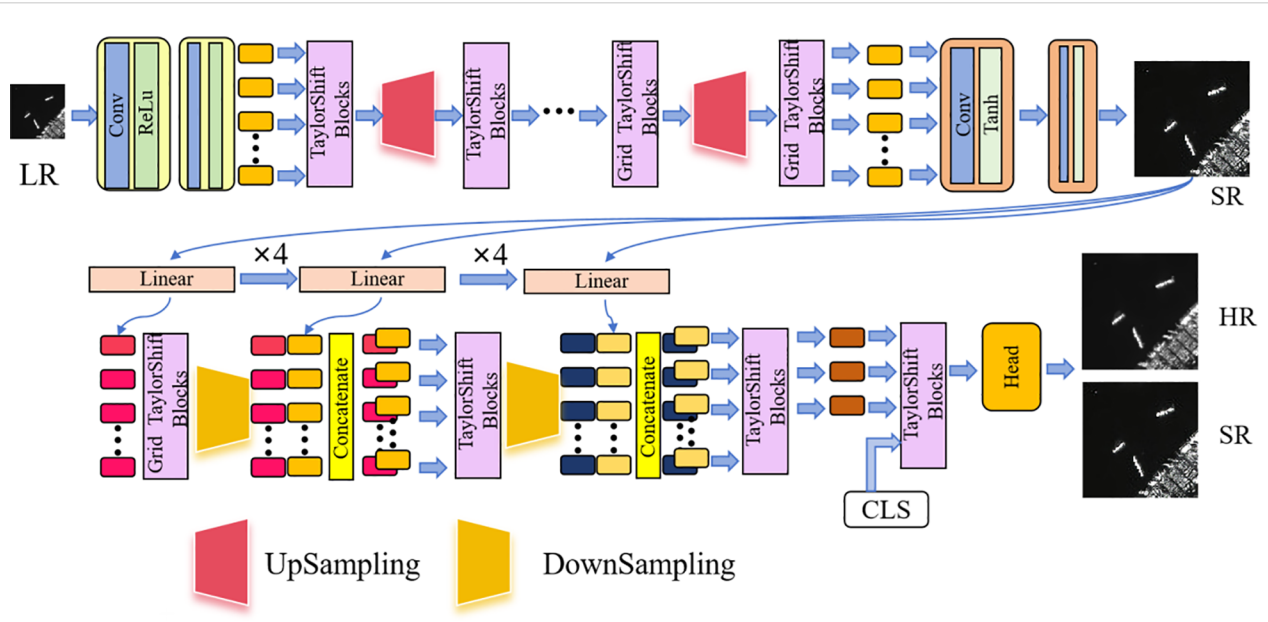


FIGURE 1
TaylorGAN network architecture.

2.1.2 Discriminator

The lower part of Figure 1 shows the discriminator of TaylorGAN, whose input is the super-resolution SAR image generated by the generator. To achieve multi-scale discrimination, the image is divided into three blocks of different scales (P, 2P, and 4P), corresponding to the feature sequences of $y_0 \in R^{C\times\frac{H}{P}\times\frac{W}{P}}$, $y_1 \in R^{C\times\frac{H}{2P}\times\frac{W}{2P}}$, and $y_2 \in R^{C\times\frac{H}{4P}\times\frac{W}{4P}}$, respectively. Each set of sequences is sent to the corresponding TaylorShift block through linear mapping to extract semantic features at different scales. Finally, the discriminator uses a downsampling module to reduce the resolution of the feature map, and the connection block fuses features of different scales so that the model can perceive the global structure and local details of the image at the same time. To evaluate the overall authenticity of the image, a [CLS] tag is added at the end of the discriminator. This tag interacts with all image tokens through a multi-layer attention mechanism, and only the output features of this tag are used as the classifier input so that the discriminator can comprehensively judge the global consistency and detail rationality of the image. Finally, the real/generated discrimination result is output through the classification head to assist the generator in optimizing the image quality.

2.1.3 TaylorShift attention mechanism

The traditional self-attention mechanism has a computational bottleneck, and its time and space complexity are both $O(N^2)$, where N is the length of the token sequence (that is, the number of patches in the image). When processing high-resolution images (such as 256 × 256), the memory usage and inference time increase dramatically, which seriously restricts the scalability of the model. TaylorShift (Nauen et al., 2025) Attention Mechanisms is a variant of Transformer that approximates the exponential operations in a Softmax function by Taylor series expansion. The TaylorShift attention mechanisms are categorized into direct TaylorShift and efficient TaylorShift.

1. Direct-TaylorShift

The Taylor approximation is applied to Softmax in Taylor-Softmax to avoid the computation of the exponential function, and the k-order (kth) Taylor expansion formula is Equation 1:

$$\exp(x) \approx \sum_{n=0}^{k} \frac{x^n}{n!} \tag{1}$$

The Taylor-Softmax formula is Equation 2:

$$T-SM^{(K)}(QK^T) = normalize\left(\sum_{n=0}^{k} \frac{(QK^T)^n}{n!}\right) \tag{2}$$

In Direct-TaylorShift, Taylor-Softmax is used directly instead of Softmax to compute the attention weights and multiply the computed result with the value matrix V. The formula is Equation 3:

$$Y = \frac{\left(\sum_{n=0}^{k} \frac{(QK^T)^n}{n!}\right) V}{\sum_i \left(\sum_{n=0}^{k} \frac{(QK^T)_i^n}{n!}\right)} \tag{3}$$

Y is the outputs, and $\sum_i \left(\sum_{n=0}^{k} \frac{(QK^T)_i^n}{n!}\right)$ is the same Taylor expansion operation performed on each line of QK^T, and normalization is performed on each line. The denominator ensures that the output is normalized across tokens, making the weights valid for each token. This expression is suitable for scenarios with a small number of tokens (such as 32 × 32 and below). It significantly reduces the reliance on exponential functions and is faster to calculate, but the computational complexity is $O(N^2 d)$.

2. Efficient-TaylorShift

For further optimization, TaylorShift introduced an efficient implementation form Efficient-TaylorShift. If the length of the feature sequence exceeds a certain threshold, it is more appropriate to use Efficient-TaylorShift. It is performed by assigning Taylor-Softmax values to the matrices Q and K and moving the normalization operation after multiplying it with the value matrix V. The formula for normalization is Equations 4–6:

$$Y_{nom} = \left(1 + QK^T + \frac{1}{2}(QK^T) \odot 2\right) V \tag{4}$$

$$Y_{denom} = \left(1 + QK^T + \frac{1}{2}(QK^T) \odot 2\right) 1_N \tag{5}$$

$$Y = \frac{Y_{nom}}{Y_{denom}} \tag{6}$$

ȯ denotes Hadamard multiplication (element-level multiplication), 1_N denotes a vector of length N with all ones. Y_{nom} denotes the weighted attention score, and Y_{denom} denotes the value used for normalization.

By changing the calculation order, Efficient-TaylorShift reduces the computational complexity of traditional attention from $O(N^2)$ to $O(Nd^3)$, which is suitable for processing tens of thousands of tokens in high-resolution images. The reduction in computational complexity enables the model to better capture global dependencies and improve the integrity and consistency of image structure.

In this paper, the TaylorShift attention mechanism is integrated into two different module structures: TaylorShift Block and Grid TaylorShift Block, which correspond to two application scenarios of sequence modeling and spatial modeling, respectively. TaylorShift Block is suitable for processing flattened image patch token sequences. The input is a one-dimensional token sequence. The module calculates the long-distance dependencies between different tokens through TaylorShift attention to model the overall semantic information of the image. In contrast, Grid TaylorShift Block is designed for feature map input that retains the spatial structure of two-dimensional images. The module calculates self-attention along the row and column directions of the image, respectively, to more efficiently capture local spatial relationships in the image, such as edge and texture information.

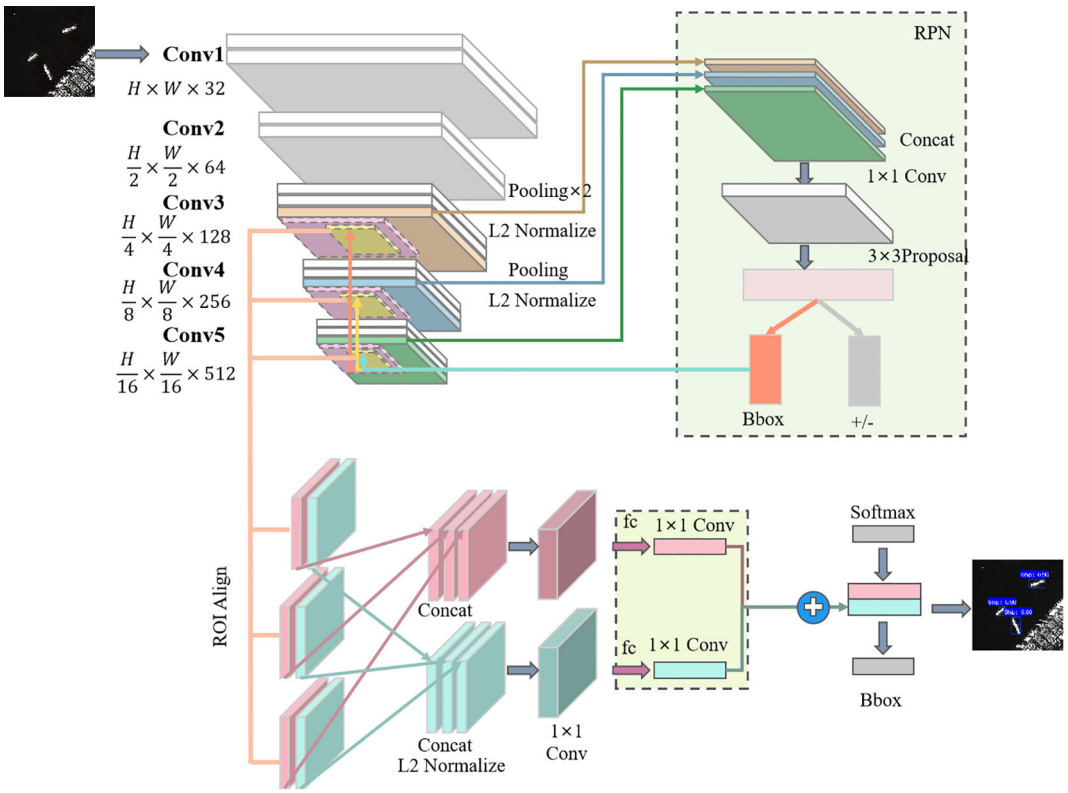


FIGURE 2
HMS-MRCNN network architecture.

2.2 HMS-MRCNN

The paper proposed a hierarchical multi-scale SAR image marine ship detection method based on Mask R-CNN as shown in Figure 2, which is mainly divided into three parts, i.e., the feature extraction module, the region suggestion network, and the prediction module. The feature extraction and fusion module is used to extract the multiscale features of the ship in the SAR image and fuse the different features. The region suggestion network is used to identify potential regions of interest. The prediction module classifies and regresses the candidate boxes and outputs the final bounding box.

2.2.1 Feature extraction module

In the feature extraction module, the SAR images are fed into the network, and the features are first extracted by a backbone network consisting of five convolutional layers that capture the multi-scale features of the ship. Conv1 and Conv2 are feature preprocessing modules in the initial stage of the model. They are mainly used for preliminary feature encoding and spatial downsampling of the input SAR images, helping the model to extract clearer local structural information from the original images. The two middle layers are shallow convolution layers (Conv3–Conv4), which mainly capture the local structural information of small-scale ships. The last layer is a deep convolution layer (Conv5) used to obtain high-level semantic features and contextual relationships of the image. The output feature maps of Conv3 and Conv4 are subjected to 2 × 2 maximum pooling operations to reduce their spatial size so that they maintain the same spatial resolution as the large-scale ship feature maps (such as the feature maps of Conv5). In order to eliminate the numerical differences between feature maps of different layers, the spliced feature maps need to be L2-normalized to ensure that the numerical range of each feature is consistent.

The pooled small-scale feature map, together with the deep feature map (Conv5 output), is input into RoI Align for further processing.

2.2.2 Regional recommended networks—RPNs

The feature maps extracted by the convolutional layers are fed into the RPN, where the small ship feature maps are sequentially passed through cascading convolutional layers of sizes 1 and 3 to ensure that the feature maps can be matched with the output features of the backbone network. The RPN recognizes the ship features of the SAR image for bounding box regression and generates a set of RoI that are considered as possible ship locations, which include the ship regions in the SAR image of the SAR image for the bounding box regression values. In addition, the RPN needs to determine whether each RoI contains a ship and the precise location of the ship.

RPN uses a 3 × 3 convolutional filter to scan the entire feature map. At each ship location in the feature map, RPN generates

multiple anchor boxes with different aspect ratios, which are used to capture ship targets of different sizes and shapes. After generating the anchor boxes, RPN performs two steps: target discrimination and bounding box regression. In the target discrimination task, RPN determines whether the anchor box contains a ship target or not and applies a binary classification method to marine ship detection, i.e., whether the anchor box contains a ship or not, and scores it. RPN performs an accurate bounding box regression task (Bbox) on the anchor boxes that are judged to be ships, adjusting the sizes and shapes of the boxes to better fit the ship targets.

2.2.3 Forecasting module

The low-level feature map and the output results of the five-layer convolution are fused through RoI Align. RoI Align first divides each RoI into a fixed number of sub-regions. In each sub-region, RoI Align uses bilinear interpolation to extract image features. These feature blocks are spliced together to form a unified feature map. On this basis, L2 normalization is performed to ensure that the features between different RoI are numerically consistent. The spliced and normalized feature maps are further processed through a 1×1 convolution layer. The processed feature maps will be used for target classification (Softmax) and bounding box regression (Bbox) tasks. The classification task is responsible for determining whether each RoI contains a target, and the bounding box regression further accurately adjusts the position of the candidate box to ensure that the final output bounding box is more accurate.

2.3 Loss function

2.3.1 Super-resolution reconstruction loss function

The generated network loss function L^{SR} can be divided into three parts: the traditional pixel-by-pixel difference MSE-based loss L^{SR}_{pix}, the content-aware loss L^{SR}_{vgg}, and the adversarial loss L^{SR}_{adv} based on the VGG (Mateen et al., 2018) network.

Define the low-resolution image as L^{HR}, the corresponding high-resolution image as L^{HR}, and the super-resolution reconstructed image as L^{HR}; the super-resolution magnification is r, and $W \times H$ and $rW \times rH$ are used to denote the size of the L^{LR} and L^{HR} images, respectively, while G denotes the super-resolution reconstruction process of the generator, and D denotes the authenticity process of the discriminator.

The formula for the MSE pixel loss L^{SR}_{pix} is Equation 7:

$$L^{SR}_{pix} = \frac{1}{r^2 WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I^{HR}_{x,y} - I^{SR}_{x,y})^2 \tag{7}$$

The formula for perceived loss L^{SR}_{vgg} is Equation 8:

$$L_{vgg} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\varnothing_{ij}(I^{HR})_{x,y} - \varnothing_{ij}(I^{SR})_{x,y})^2 \tag{8}$$

where $\varnothing_{ij}$ denotes the feature mapping map between the i-th largest pooling layer and the j-th convolutional layer in the VGG model, and (i,j) is the corresponding feature map dimension. $W_{i,j}$ and $H_{i,j}$ denote the dimensions of the current layers of the VGG19 network, respectively.

Against loss L^{SR}_{adv} The formula is Equation 9:

$$L^{SR}_{adv} = -E[log(D(I^{SR}))] \tag{9}$$

Because GAN needs to play a game between the generative network and the adversarial network in training, the reconstruction results are prone to the "artifacts" phenomenon because of the poor stability of its network training and the difficulty of convergence of the model. To address the abovementioned problems, this paper combines the reconstruction quality evaluation index Structural Similarity Index (SSIM) to introduce the structural loss function L^{SR}_{SSIM}, whose formula is Equation 10:

$$L^{SR}_{SSIM} = 1 - E\left[\sum_i SSIM_i\right] \tag{10}$$

where $SSIM_i$ is the structural similarity between the i-th batch of reconstructed super-resolution image I^{SR} of the generative network and the reference high-resolution image I^{HR}.

Therefore, the loss function of TaylorGAN is Equation 11:

$$L_{SR} = L^{SR}_{vgg} + \lambda L^{SR}_{adv} + \eta L^{SR}_{pix} + \xi L^{SR}_{SSIM} \tag{11}$$

λ, η, and ξ represent the weights of adversarial loss, pixel-level loss, and structural similarity loss, respectively.

2.3.2 Marine ship detection loss function

The marine ship detection loss function is Equation 12:

$$L_c = L_{cls} + \lambda L_{reg} \tag{12}$$

where L_{cls} and L_{reg} denote the classifier loss and the bounding box regression loss, respectively, and λ is the weight parameter. The focal loss function is used for the classification loss, and the formula is Equation 13:

$$L_{cls} = -\sum_i \alpha_i (1 - p_i)^{\gamma} log(p_i) \tag{13}$$

γ is used to control the weights of easily categorized samples, and α_i is used to solve the problem of category imbalance.

For the bounding box regression loss function, we use the CIoU (complete intersection over union) loss function. CIoU is an extension of IOU, which takes the center offset of the bounding box as well as the aspect ratio into account, and it is suitable for high-precision marine ship detection with Equation 14:

$$L_{reg} = 1 - IoU + \alpha \frac{\rho^2(b, b^{gt})}{c^2} + \beta v \tag{14}$$

where IoU is used to compute the intersection and concurrency ratio between the prediction frame b and the real frame b^{gt}, $\rho^2(b, b^{gt})$ is the Euclidean distance between the prediction frame and the center point of the real frame, c^2 is the length of the diagonal of the smallest outer rectangle, v denotes the consistency of the aspect ratio, and α, and β are used to regulate the hyperparameters of the loss.

3 Experimental results

This section first introduces the SAR image dataset used in this study, then describes the evaluation indicators used in the experiment, and finally gives a comprehensive analysis of the experimental results.

3.1 Datasets

The SSDD dataset (SAR Ship Detection Dataset) was originally proposed by Li et al. (2017) and contains 1,160 SAR image slices, each with a resolution of 500 × 500 pixels. The dataset uses data from multiple satellite sources such as Sentinel-1, TerraSAR-X, and RadarSat-2. In order to improve the computational efficiency, these image slices are resized to 256 × 256 pixels. The selected data is divided into three subsets: training set (70%), validation set (10%), and test set (20%).

The SAR-Ship dataset (Wang et al., 2019) contains 102 images from China's Gaofen-3 satellite and 108 images from Sentinel-1. The dataset contains 43,819 ship slices, each with a resolution of 256 × 256 pixels. This paper selects 3,000 data slices for super-resolution and target detection experiments and divides these slices into three subsets: training set (70%), validation set (10%), and test set (20%).

3.2 Evaluation metrics

In the super-resolution experiment, this paper uses peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and MSE to evaluate the experimental effect of super-resolution reconstruction. This paper takes the original image of each dataset as a high-resolution image and obtains the corresponding low-resolution image through bicubic interpolation.

PSNR is defined by MSE, which is calculated as shown in Equation 15:

$$MSE(I_{SR}, I_{HR}) = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I_{HR}(i,j) - I_{SR}(i,j)]^2 \tag{15}$$

I_{HR} and I_{SR} are high-resolution SAR images and super-resolution SAR images, respectively, both of which have the dimensions $m \times n$.

The formula for PSNR is Equation 16:

$$PSNR(I_{SR}, I_{HR}) = 10 \cdot log_{10}\left(\frac{MAX^2}{MSE(I_{SR}, I_{HR})}\right) \tag{16}$$

SSIM is based on three evaluation metrics: brightness, contrast, and structure, with Equations 17–20:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{17}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{18}$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{19}$$

$$SSIM(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma} \tag{20}$$

α, β, γ are used to adjust the brightness, contrast, and structure of the weight look; when it is 1, SSIM can be simplified as shown in Equation 21:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{21}$$

where μ_x and μ_y denote the means, σ_x^2 and σ_y^2 denote the variances. σ_{xy} denotes the covariance between x and y, and $C_1 = (K_1L)^2$ and $C_2 = (K_2L)^2$ are used to ensure the stability of SSIM.

In marine ship detection experiments, this paper uses accuracy, recall, and mean average precision mean (mAP) as evaluation metrics. Recall is defined as shown in Equation 22:

$$Recall = \frac{TP}{TP + FN} \tag{22}$$

TP is true positives, which denotes the number of correct positive samples, and FN is false negatives, which denotes the number of incorrect negative samples. Recall is used to measure the detection model's rate of checking completeness. Precision is defined as shown in Equation 23:

$$Precision = \frac{TP}{TP + FP} \tag{23}$$

FP stands for false positives and denotes the number of false positive samples. Precision is used to measure the model's checking accuracy, which is related to the false alarm probability Pf. mAP is defined as shown in Equation 24:

$$mAP = \int_0^1 p(r)dr \tag{24}$$

r denotes recall, P denotes precision, and $p(r)$ denotes the precision–recall curve (P–R curve). The computational process of mAP is essentially to find the area under the PRC curve. Because mAP considers both recall and precision, it has been chosen as the sole core measure of detection accuracy.

3.3 Experimental details

In this paper, NVIDIA GTX 4090 GPU is used to train the network model. The training process parameters for the super-resolution reconstruction experiments are set as follows: the initial learning rate is 2e-4, and the learning rate decays by half after 50 iterations. The optimizer is Adam, the batch size is 8, and the total number of epochs is 100. The training process parameters for the marine ship detection experiment are set as follows: the initial learning rate is 0.01, and the final learning rate is reduced to 1e-3. The input image size is 256×256, the optimizer is Adam, and the batch size is 8, with 150 iterations. The software applications used included Pytorch version 1.12.0 with CUDA 12.4 and Python 3.9.

TABLE 1 Comparison of the metrics of different methods at an amplification factor of 4.

Method	SSDD			Ship-SAR		
	PSNR	SSIM	MSE	PSNR	SSIM	MSE
Bicubic	19.51	0.3713	0.2594	20.43	0.4065	0.2619
SRCNN	22.36	0.6176	0.2710	22.56	0.6730	0.2284
SRGAN	23.31	0.7146	0.2519	21.68	0.5211	0.2405
LSRGAN	24.12	0.7508	**0.2373**	23.04	0.6970	0.2151
Cycle-GAN	21.34	0.5175	0.2413	21.80	0.5382	0.2498
TaylorGAN	**25.43**	**0.7931**	0.2481	**24.55**	**0.7721**	**0.2030**

The best results are indicated in bold.

3.4 Experimental results of super-resolution reconstruction of SAR images

In order to evaluate the excellent performance of TaylorGAN in the super-resolution reconstruction of SAR images, we compare it with other super-resolution reconstruction models, and the results of the comparison are analyzed by evaluating the metrics and visual effects. The comparison methods include bicubic, SRCNN, SRGAN, LSRGAN, and cycle-GAN.

3.4.1 Quantitative results

As shown in Table 1, the performance of six super-resolution methods is evaluated across two SAR datasets, SSDD and Ship-SAR, under an amplification factor of 4. The results indicate that TaylorGAN achieves consistent improvements across all evaluation metrics, outperforming both GAN-based and non-GAN-based baselines.

On the SSDD dataset, TaylorGAN attains the highest PSNR (25.43 dB) and SSIM (0.7931), alongside the lowest MSE (0.2481). Among GAN-based models, it surpasses LSRGAN—the second best performer—by 1.31 dB in PSNR, 0.0423 in SSIM, and a 0.0102 reduction in MSE. Compared to Cycle-GAN, TaylorGAN shows more pronounced enhancements, with a 4.09-dB gain in PSNR, 0.2766 in SSIM, and 0.0072 lower MSE. Notably, when benchmarked against non-GAN approaches such as SRCNN, TaylorGAN yields an increase of 3.07 dB in PSNR, 0.1765 in SSIM, and 0.0229 decrease in MSE, reflecting its superior capability in structure preservation and noise suppression.

On the Ship-SAR dataset, similar trends are observed. TaylorGAN has a PSNR of 24.55 dB, a SSIM of 0.7721, and an MSE of 0.2030, outperforming other methods in all indicators. Compared with GAN-based models, TaylorGAN surpasses LSRGAN by 1.51 dB, 0.0754, and 0.0121 in PSNR, SSIM, and MSE, respectively. In addition, compared with Cycle-GAN, TaylorGAN improves by 2.75 dB, 0.2339, and 0.0468 in the three indicators, respectively. Compared with the non-GAN baseline SRCNN, its improvement is also very significant, with a PSNR increase of 2.00 dB, a SSIM increase of 0.0991, and an MSE reduction of 0.0254.

3.4.2 Qualitative results

Figures 3 and 4 qualitatively compare the super-resolution reconstruction results on the SSDD and SAR-Ship datasets, respectively. These figures show the visual effects of different models on improving the resolution of SAR images. As shown in the figure, TaylorGAN is able to consistently generate images with clearer textures and higher visual fidelity than other methods. In particular, TaylorGAN is able to effectively recover the structural details of the ship and suppress background noise, showing its advantage in recovering fine-grained features. In contrast, non-GAN-based models such as bicubic interpolation and SRCNN produce significantly blurred results. Although SRCNN was originally proposed for the super-resolution reconstruction of natural images, it does not generalize well on SAR data due to its simple structure and limited ability to model high-frequency components. GAN-based models, such as SRGAN, LSRGAN, and cycle-GAN, provide better performance than non-GAN baselines by generating clearer contours and richer textures. However, these methods often suffer from artifacts or excessive noise. Overall, the visual results in Figures 3 and 4 demonstrate the superior perceptual quality of TaylorGAN across different SAR image scenarios, further confirming its effectiveness in high-fidelity SAR image reconstruction tasks.

3.5 Experimental results of marine ship detection for SAR images

To verify the effectiveness of the proposed HMS-MRCNN method, this paper compares it with several representative object detection algorithms, including YOLO v8, Quad-FPN (Zhang et al., 2021), Faster R-CNN, Cascade R-CNN, and Grid R-CNN. In addition, this paper also tests high-resolution images without super-resolution reconstruction methods to evaluate the contribution of SR methods.

3.5.1 Quantitative results

Table 2 presents the quantitative comparison of the proposed HMS-MRCNN framework against several object detection models on the SSDD and Ship-SAR datasets. The evaluation metrics include precision, recall, and mAP50, which comprehensively reflect the accuracy and robustness of each method.

On the SSDD dataset, the proposed HMS-MRCNN (SR) achieves the highest performance in all metrics, with accuracy of 93.0%, recall of 90.3%, and mAP50 of 93.1%. These values exceed those of the high-resolution input version (HMS-MRCNN (HR)) as well as other traditional detectors. Notably, the mAP50 of HMS-MRCNN (SR) is improved by 1.9% compared to Quad-FPN, demonstrating the effectiveness of integrated super-resolution reconstruction in enhancing detection results.

On the Ship-SAR dataset, the proposed method maintains its leading position, achieving an accuracy level of 91.9%, recall of 93.0%, and mAP50 of 92.6%. This performance exceeds that of Quad-FPN and other classic detectors.

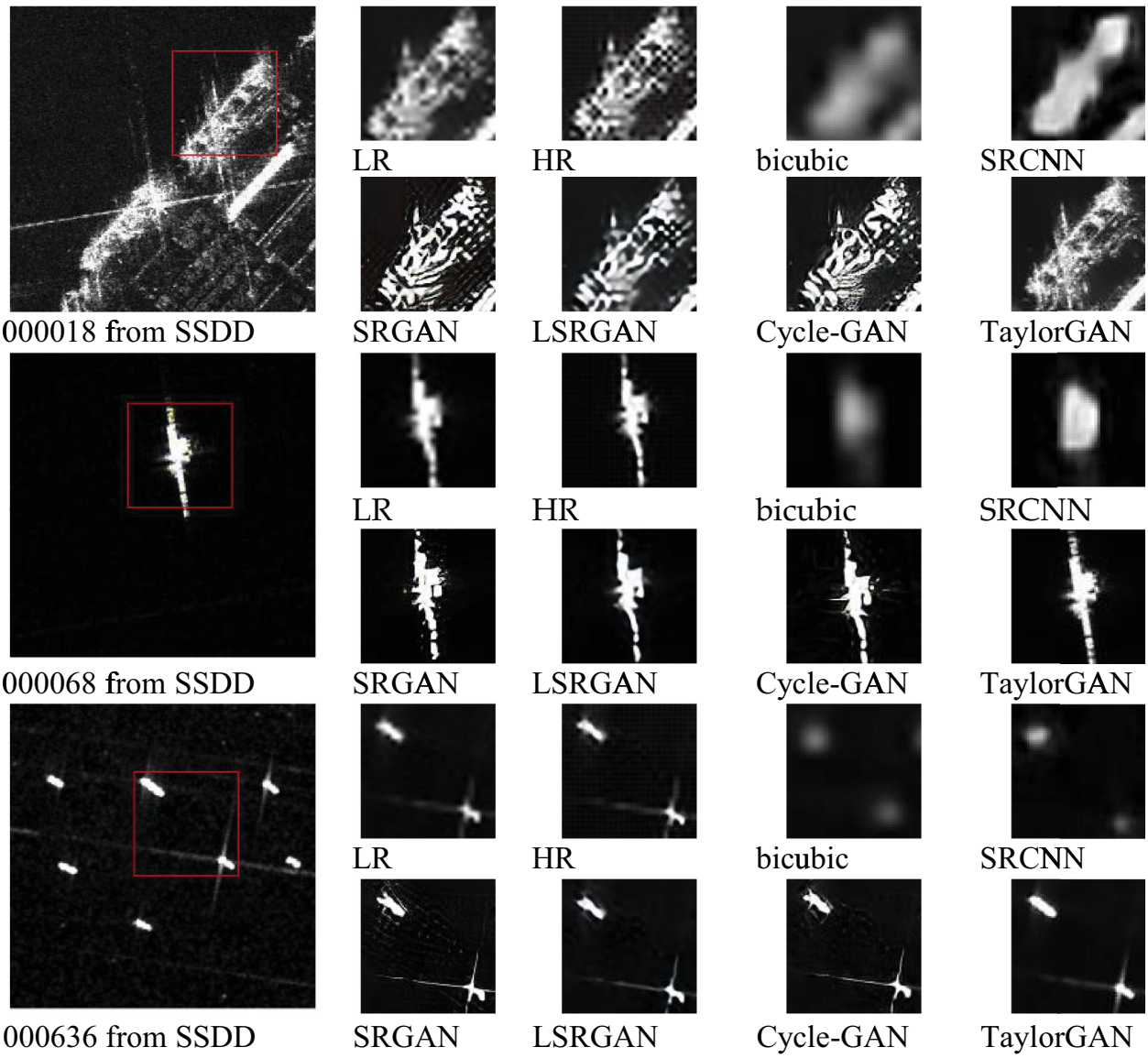


FIGURE 3
Comparison results of the super-resolution reconstruction of the SSDD dataset.

Overall, the experimental results verify that the proposed HMS-MRCNN (SR) not only improves the average detection accuracy but also enhances its stability at different scales and scene complexity, making it very suitable for practical SAR-based ship detection tasks.

3.5.2 Qualitative results

Figures 5 and 6 qualitatively compare the detection results of different target detection algorithms on the SSDD and SAR-Ship datasets. The methods include YOLO v8, Quad-FPN (Zhang et al., 2021), Faster R-CNN, Cascade R-CNN, Grid R-CNN, and HMS-MRCNN. Red circles indicate missed detections, and yellow circles indicate incorrectly detected target objects.

As can be seen from the figure, YOLO v8 and Faster R-CNN are prone to more false positives or missed detections, especially when detecting small or low-contrast ships. Quad-FPN shows higher positioning accuracy and recall rate than traditional models, but it occasionally produces false detections in complex near-shore scenes or cluttered wave backgrounds. Cascade R-CNN and Grid R-CNN also have more missed detections and false detections.

In contrast, HMS-MRCNN, proposed in this paper, shows obvious advantages in detection results. In particular, after using TaylorGAN to reconstruct the image for super-resolution, HMS-MRCNN can better detect the target ship.

3.6 Ablation experiments

To evaluate the contribution of key structural components in the proposed TaylorGAN, this paper conducts ablation experiments focusing on two core modules: the TaylorShift Attention (TSA) module and the feature fusion (FF) module. The TSA module is

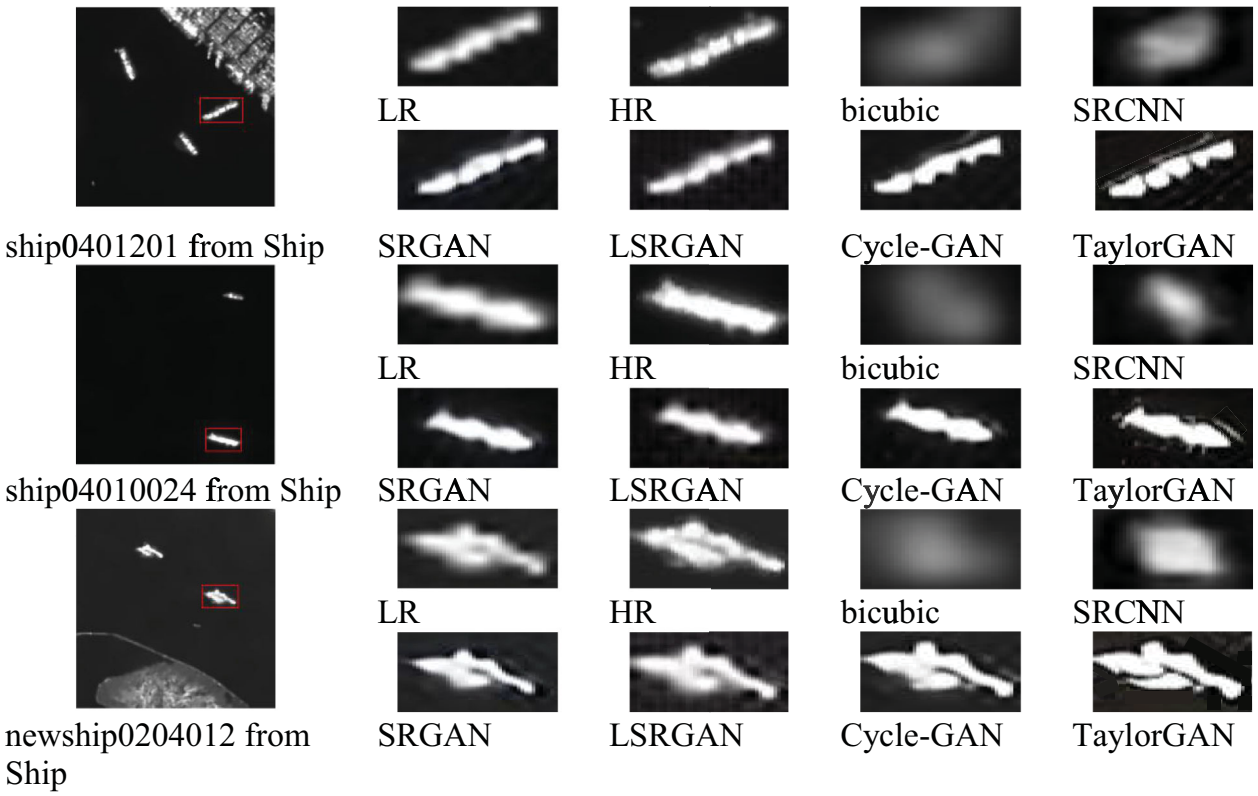


FIGURE 4
Comparison results of the super-resolution reconstruction of the SAR-Ship dataset.

designed to enhance the network's global and local modeling capability through a position-aware attention mechanism, while the FF module facilitates the integration of multi-scale features to recover high-frequency structures such as ship contours and edges.

As shown in Table 3, this paper begins with a baseline configuration that excludes both TSA and FF modules. This version achieves relatively low performance (20.89 dB PSNR and 0.5852 SSIM on SSDD), indicating its limited capability in recovering structural and fine-grained details. Introducing the TSA module alone yields a noticeable improvement, increasing PSNR by 1.51 dB and SSIM by 0.0177 on SSDD. This again demonstrates the effectiveness of TaylorShift attention in enhancing feature representation, even without structural fusion.

When both modules are integrated, the model achieves its highest performance, with 25.43 dB PSNR and 0.7931 SSIM on SSDD and 24.55 dB PSNR and 0.7721 SSIM on Ship-SAR. This final configuration outperforms all ablated variants, confirming that the combination of attention-based modeling and feature fusion significantly improves image quality, especially in restoring high-frequency textures under complex SAR imaging conditions.

TABLE 2 Comparative experimental results.

Method	SSDD			Ship-SAR		
	Precision	Recall	mAP50	Precision	Recall	mAP50
YOLO v8	87.8	81.9	89.9	80.1	85.3	83.6
Quad-FPN	90.6	88.4	91.2	89.3	91.7	90.5
Faster R-CNN	87.4	86.0	87.2	84.5	84.1	83.1
Cascade R-CNN	91.7	86.5	88.3	87.7	83.0	84.8
Grid R-CNN	88.4	87.0	87.9	81.5	82.3	81.9
HMS-MRCNN (HR)	91.9	89.7	92.5	90.8	89.9	91.3
HMS-MRCNN (SR)	**93.0**	**90.3**	**93.1**	**91.9**	**93.0**	**92.6**

The best results are indicated in bold.

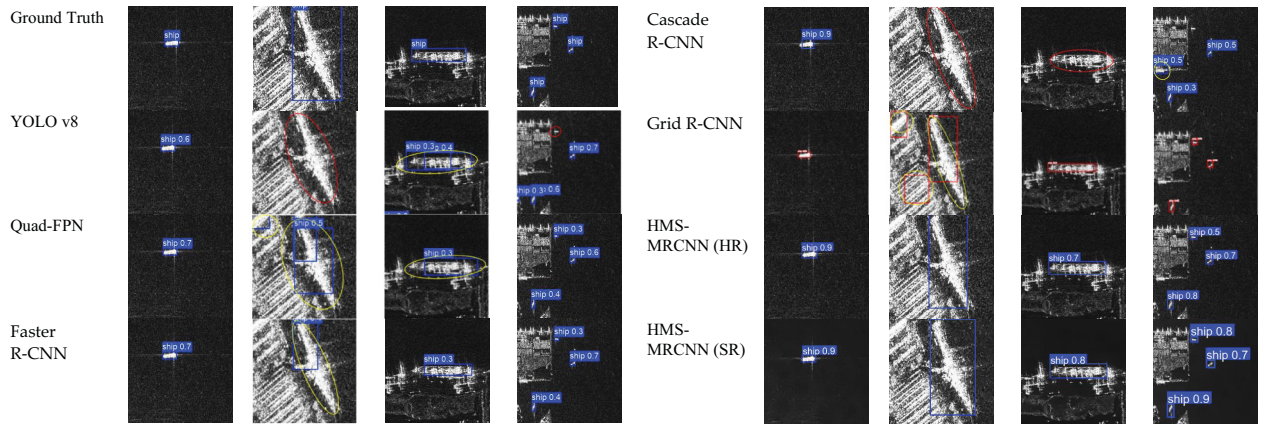


FIGURE 5
Comparison of the detection results of each model for the SSDD dataset.

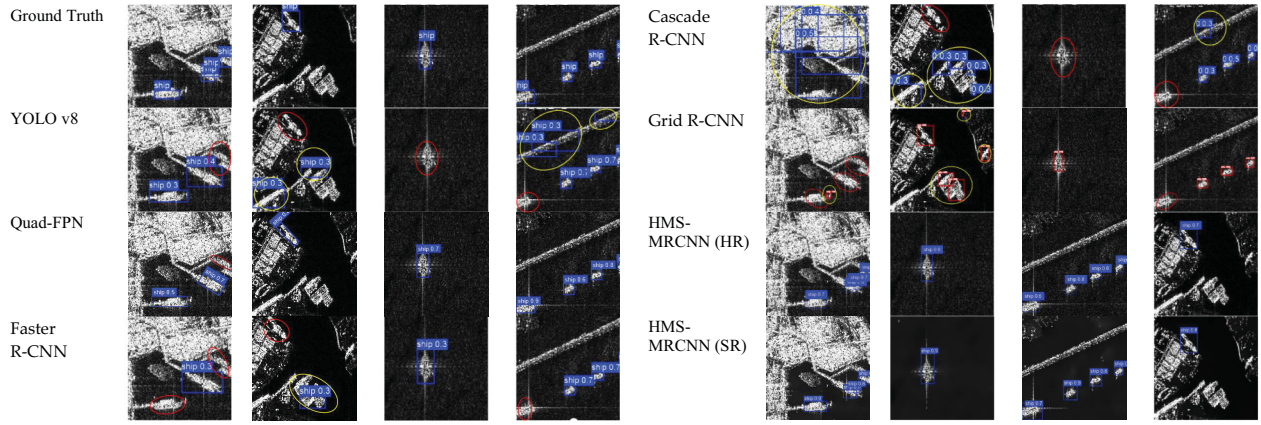


FIGURE 6
Comparison of the detection results of each model for the SAR-Ship dataset.

TABLE 3 Ablation experiment results of different blocks of TaylorGAN.

TSA	FF	SSDD		Ship-SAR	
		PSNR	SSIM	PSNR	SSIM
		20.89	0.5852	21.07	0.5631
✓		22.40	0.6029	22.46	0.6234
✓	✓	**25.43**	**0.7931**	**24.55**	**0.7721**

The best results are indicated in bold.

To evaluate the contribution of each module to detection performance, this paper conducts a controlled ablation study analyzing the impact of the DCR (feature imitation) and DCN (deformable convolution) modules within the HMS-MRCNN framework. The DCR module enhances semantic-level feature representation, while the DCN module improves spatial adaptability. The performance metrics of each module configuration are detailed in Table 4.

As shown in Table 4, the DCR module alone yields notable improvements in recall, while the DCN module contributes more to precision and localization. However, the combination of DCR and DCN achieves higher overall performance than either module individually, demonstrating their complementary strengths. The full model, integrating both modules, significantly enhances detection accuracy on both SSDD and Ship-SAR datasets.

These results indicate that fusing semantic feature imitation with spatially adaptive convolution can effectively enhance network robustness and accuracy under complex SAR imaging conditions.

4 Conclusion

Given the challenges of low resolution of SAR images and the susceptibility of marine ship detection to noise and multi-scale target interference, this paper proposes a "super-resolution reconstruction-multi-scale detection" collaborative optimization solution. The main contributions are as follows:

TABLE 4 Ablation experiment results of different blocks of HMS-MRCNN.

DCR	DCN	SSDD			Ship-SAR		
		Precision	Recall	mAP50	Precision	Recall	mAP50
✓		91.4	89.2	91.8	89.4	90.8	89.9
	✓	92.1	88.6	92.2	90.0	92.1	91.0
✓	✓	**93.0**	**90.3**	**93.1**	**91.9**	**93.0**	**92.6**

The best results are indicated in bold.

TaylorGAN super-resolution network: It aims to recover high-frequency detail information from low-resolution SAR images. The method works by feeding the low-resolution image into the generator taking the corresponding high-resolution image as the target of discriminator learning and continuously optimizing the generator through adversarial training so that its output image is closer to the real high-resolution image in terms of structural clarity and detail restoration. In order to enhance the detail modeling ability, TaylorGAN introduces the TaylorShift attention mechanism, replacing the traditional Softmax operation with Taylor series expansion, which improves the ability to recover high-frequency details (e.g., ship contours, deck structures). Experiments prove that TaylorGAN significantly outperforms mainstream models such as SRGAN and cycle-GAN in terms of PSNR, SSIM, and subjective visual quality.

HMS-MRCNN multi-scale detection framework: HMS-MRCNN is designed for marine ship detection, extracting small target details from shallow layers (Conv3-4) and capturing global semantic context from deep layers (Conv5). Through feature map downsampling and L2 normalization, the model achieves accurate cross-scale feature alignment. Experiments show that HMS-MRCNN (SR) achieves 93.1% mAP50 accuracy on SSDD and 92.6% mAP50 accuracy on Ship-SAR, outperforming traditional detectors such as Faster R-CNN and Grid R-CNN.

End-to-end performance verification: The combination of super-resolution reconstruction and marine ship detection improves the mAP50 of ship image detection by 0.6% and 1.3% on the SSDD and Ship-SAR datasets, indicating that the resolution improvement directly improves the performance of downstream tasks.

Data availability statement

The datasets SSDD and SAR-Ship for this study can be found in the https://github.com/TianwenZhang0825/Official-SSDD/blob/main/README.md and https://radars.ac.cn/web/data/getData?dataType=SDD-SAR.

Author contributions

JF: Funding acquisition, Writing – review & editing. MG: Methodology, Visualization, Writing – original draft. LZ: Data curation, Software, Writing – original draft. JL: Supervision, Validation, Writing – review & editing. YL: Methodology, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by Shandong Provincial Natural Science Foundation of China under Grant ZR2018MF009, the State Key Research Development Program of China under Grant 2017YFC0804406, National Natural Science Foundation of China under Grant 42472324, the Special Funds of Taishan Scholars Construction Project, and the foundation of Key Laboratory of Mining Disaster Prevention and Control (Shandong University of Science and Technology).

Acknowledgments

The authors would like to thank the editors and reviewers for their valuable comments and suggestions.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

Baldygo, W., Brown, R., Wicks, M., Antonik, P., Capraro, G., and Hennington, L. (1993). "Artificial intelligence applications to constant false alarm rate (CFAR) processing," in *The record of the 1993 IEEE national radar conference* (United States: IEEE), 275–280.

Blu, T., Thévenaz, P., and Unser, M. (2004). Linear interpolation revitalized. *IEEE Trans. Image Process.* 13, 710–719. doi: 10.1109/tip.2004.826093

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*, (United States: IEEE Press) 6154–6162.

Cao, Q., Chen, H., Wang, S., Wang, Y., Fu, H., Chen, Z., et al. (2024). LH-YOLO: A lightweight and high-precision SAR ship detection model based on the improved YOLOv8n. *Remote Sensing* 16, 4340. doi: 10.3390/rs16224340

Chang, H., Yeung, D.-Y., and Xiong, Y. (2004). "Super-resolution through neighbor embedding," in *Proceedings of the 2004 IEEE computer society conference on computer vision and pattern recognition 2004. CVPR 2004* (United States: IEEE), I–I.

Dong, C., Loy, C. C., He, K., and Tang, X. (2014). "Learning a deep convolutional network for image super-resolution," in *Computer vision–ECCV 2014: 13th european conference* (Springer, Zurich, Switzerland), 184–199. Proceedings, Part IV 13.

Freeman, W. T., Jones, T. R., and Pasztor, E. C. (2002). Example-based super-resolution. *IEEE Comput. Graphics Appl.* 22, 56–65. doi: 10.1109/38.988747

Gao, G., Chen, Y., Feng, Z., Zhang, C., Duan, D., Li, H., et al. (2024). R-LRBPNet: A lightweight SAR image oriented ship detection and classification method. *Remote Sensing* 16, 1533. doi: 10.3390/rs16091533

Ge, Z. (2021). Yolox: Exceeding yolo series in 2021. doi: 10.48550/arXiv.2107.08430

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*, (United States: IEEE Press) 580–587.

Gu, F., Zhang, H., Wang, C., and Wu, F. (2019). "SAR image super-resolution based on noise-free generative adversarial network," in *IGARSS 2019–2019 IEEE international geoscience and remote sensing symposium* (Yokohama, Japan: IEEE), 2575–2578.

Irani, M., and Peleg, S. (1991). Improving resolution by image registration. *CVGIP: Graphical Models Image Process.* 53, 231–239. doi: 10.1016/1049-9652(91)90045-l

Jiang, N., Zhao, W., Wang, H., Luo, H., Chen, Z., and Zhu, J. J. R. S. (2024). Lightweight super-resolution generative adversarial network for SAR images. *Remote Sensing* 16, 1788. doi: 10.3390/rs16101788

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., et al. (2017). "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (United States: IEEE Press) 4681–4690.

Li, J., Qu, C., and Shao, J. (2017). "Ship detection in SAR images based on an improved faster R-CNN," in *2017 SAR in big data era: models, methods and applications (BIGSARDATA)* (United States: IEEE), 1–6.

Li, X., Chen, P., Yang, J., An, W., Zheng, G., Luo, D., et al. (2023). TKP-net: A three keypoint detection network for ships using SAR imagery. doi: 10.1109/JSTARS.2023.3329252

Liu, C., and Sun, D. (2013). On Bayesian adaptive video super resolution. *14th European Conference Comput. Vision*. 36, 364–376. doi: 10.1109/TPAMI.2013.127

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.

Mateen, M., Wen, J., Nasrullah,, Song, S., and Huang, Z. (2018). Fundus image classification using VGG-19 architecture with PCA and SVD. *Symmetry* 11, 1. doi: 10.3390/sym11010001

Meng, F., Qi, X., and Fan, H. (2024). LSR-det: A lightweight detector for ship detection in SAR images based on oriented bounding box. *Remote Sensing* 16, 3251. doi: 10.3390/rs16173251

Nauen, T. C., Palacio, S., and Dengel, A. (2025). "Taylorshift: Shifting the complexity of self-attention from squared to linear (and back) using taylor-softmax," in *International conference on pattern recognition* (Cham, Switzerland: Springer), 1–16.

Redmon, J. (2016). "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. (United States: IEEE Press)

Ren, S. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 39:1137–1149. doi: 10.1109/TPAMI.2016.2577031

Shen, H., Lin, L., Li, J., Yuan, Q., and Zhao, L. (2020). A residual convolutional neural network for polarimetric SAR image super-resolution. *ISPRS J. Photogrammetry Remote Sens.* 161, 90–108. doi: 10.1016/j.isprsjprs.2020.01.006

Smith, J. W., Alimam, Y., Vedula, G., and Torlak, M. (2022). "A vision transformer approach for efficient near-field SAR super-resolution under array perturbation," in *2022 IEEE texas symposium on wireless and microwave circuits and systems (WMCS)* (United States: IEEE), 1–6.

Su, L., Sun, Y., and Yuan, S. (2022). A survey of instance segmentation research based on deep learning. *CAAI Trans. Intell. Syst.* 17, 16–31. doi: 10.11992/tis.202109043

Sun, Z., Dai, M., Leng, X., Lei, Y., Xiong, B., Ji, K., et al. (2021). An anchor-free detection method for ship targets in high-resolution SAR images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 14, 7799–7816. doi: 10.1109/jstars.2021.3099483

Tang, J., Cheng, J., Xiang, D., and Hu, C. (2022). Large-difference-scale target detection using a revised Bhattacharyya distance in SAR images. *IEEE Geosci. Remote Sens. Lett.* 19, 1–5. doi: 10.1109/lgrs.2022.3161931

Tang, Y., Zhang, Y., Xiao, J., Cao, Y., and Yu, Z. (2024). An enhanced shuffle attention with context decoupling head with wise ioU loss for SAR ship detection. *Remote Sensing* 16, 4128. doi: 10.3390/rs16224128

Tom, B. C., and Katsaggelos, A. K. (1996). "Iterative algorithm for improving the resolution of video sequences," in *Visual communications and image processing'96* (United States: SPIE), 1430–1438.

Tong, C., and Leung, K. (2007). Super-resolution reconstruction based on linear interpolation of wavelet coefficients. *Multidimens. Syst. Signal Process.*18, 153–171. doi: 10.1007/s11045-007-0023-2

Wang, Y., Wang, C., Zhang, H., Dong, Y., and Wei, S. (2019). A SAR dataset of ship detection for deep learning under complex backgrounds. *IGARSS 2019–2019 IEEE Int. Geosci. Remote Sens. Symposium* 11, 765. doi: 10.3390/rs11070765

Wang, L., Zheng, M., Du, W., Wei, M., and Li, L. (2018). "Super-resolution SAR image reconstruction via generative adversarial network," in *2018 12th international symposium on antennas, propagation and EM theory (ISAPE)* (China: IEEE), 1–4.

Wu, F., Hu, T., Xia, Y., Ma, B., Sarwar, S., and Zhang, C. (2024). WDFA-YOLOX: A wavelet-driven and feature-enhanced attention YOLOX network for ship detection in SAR images. *Remote Sensing* 16, 1760. doi: 10.3390/rs16101760

Xu, Y., Wu, Z., Chanussot, J., and Wei, Z. (2019). Nonlocal patch tensor sparse representation for hyperspectral image super-resolution. *IEEE Trans. Image Process.* 28, 3034–3047. doi: 10.1109/tip.2019.2893530

Yang, F., Yang, H., Fu, J., Lu, H., and Guo, B. (2020). "Learning texture transformer network for image super-resolution," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, (United States: IEEE Press) 5791–5800.

Yasir, M., Liu, S., Pirasteh, S., Xu, M., Sheng, H., Wan, J., et al. (2024a). YOLOShipTracker: Tracking ships in SAR images using lightweight YOLOv8. *Int. J. Appl. Earth Obs. Geoinf.* 134, 104137. doi: 10.1016/j.jag.2024.104137

Yasir, M., Shanwei, L., Mingming, X., Jianhua, W., Nazir, S., Islam, Q. U., et al. (2024b). SwinYOLOv7: Robust ship detection in complex synthetic aperture radar images. *Appl. Soft Comput.* 160, 111704. doi: 10.1016/j.asoc.2024.111704

Zhang, X., Feng, S., Zhao, C., Sun, Z., Zhang, S., Ji, K., et al. (2024). MGSFA-Net: Multiscale global scattering feature association network for SAR ship target recognition. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 17, 4611–4625. doi: 10.1109/jstars.2024.3357171

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., and Fu, Y. (2018). "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European conference on computer vision (ECCV)*, (Heidelberg, Germany: Springer-Verlag GmbH) 286–301.

Zhang, L., Liu, Y., Qu, L., Cai, J., and Fang, J. (2023b). A spatial cross-scale attention network and global average accuracy loss for SAR ship detection. *Remote Sensing* 15, 350. doi: 10.3390/rs15020350

Zhang, C., Zhang, Z., Deng, Y., Zhang, Y., Chong, M., Tan, Y., et al. (2023a). Blind super-resolution for SAR images with speckle noise based on deep learning probabilistic degradation model and SAR priors. *Remote Sensing* 15, 330. doi: 10.3390/rs15020330

Zhang, T., Zhang, X., and Ke, X. (2021). Quad-FPN: A novel quad feature pyramid network for SAR ship detection. *Remote Sensing* 13, 2771. doi: 10.3390/rs13142771

frontiers | Frontiers in Marine Science

TYPE Original Research
PUBLISHED 15 August 2025
DOI 10.3389/fmars.2025.1613061

OPEN ACCESS

EDITED BY
Muhammad Yasir,
China University of Petroleum (East China), China

REVIEWED BY
Peter Feldens,
Leibniz Institute for Baltic Sea Research (LG), Germany
Arthur C. Trembanis,
University of Delaware, United States

*CORRESPONDENCE
Tobias Ziolkowski
tziolkowski@geomar.de

RECEIVED 16 April 2025
ACCEPTED 21 July 2025
PUBLISHED 15 August 2025

CITATION
Ziolkowski T, Devey CW and Koschmider A (2025) Detecting small seamounts in multibeam data using convolutional neural networks.
Front. Mar. Sci. 12:1613061.
doi: 10.3389/fmars.2025.1613061

Detecting small seamounts in multibeam data using convolutional neural networks

Tobias Ziolkowski[1]*, Colin W. Devey[1] and Agnes Koschmider[2]

[1]GEOMAR - Helmholtz Centrum for Ocean Research Kiel, Kiel, Germany, [2]Process Analytics Group, University of Bayreuth, Bayreuth, Germany

Seamounts play a crucial role in marine ecosystems, ocean circulation, and plate tectonics, yet most remain unmapped due to limitations in detection methods. While satellite altimetry provides large-scale coverage, its resolution is insufficient for detecting smaller seamounts, necessitating high-resolution multibeam bathymetry. This study introduces a deep-learning-based framework for automated small seamount detection in multibeam bathymetry, combining a CNN-based filtering step with U-Net segmentation to enhance accuracy and efficiency. Using multibeam bathymetric data from the SO305–2 expedition, the proposed approach successfully identified 30 seamounts, many of which were undetectable using satellite altimetry. A hyperparameter optimization study determined the optimal U-Net configuration, achieving a Dice Coefficient of 0.8274 and a Mean IoU of 0.7514. While the model performed well within the training dataset, cross-regional generalization remains challenging, with reduced accuracy observed in areas of highly variable seafloor morphology. The results highlight the limitations of satellite altimetry, as only 14 of the 30 detected seamounts were visible in satellite-derived datasets. This underscores the necessity of high-resolution multibeam surveys for capturing fine-scale seafloor features. In contrast to time-intensive manual annotation—which can require several hours to accurately delineate each individual seamount—the automated U-Net-based segmentation approach analyzed 146,060 km^2 of multibeam data within seconds, offering substantial time savings and scalability for large-scale mapping efforts. Beyond geological mapping, automated seamount detection has broad applications in marine ecology, environmental monitoring, and plate tectonics research. Future work should focus on integrating physical principles and geological constraints, such as typical seamount morphology, size distributions, and tectonic setting, to improve classification accuracy.

KEYWORDS

multibeam, seamount, convolutional neural network, seamount catalog, feature vector, bathymetry, U-net, seafloor mapping

1 Introduction

Seamounts, underwater mountains formed by volcanic activity, are significant features of the ocean floor, providing important information about plate tectonics and influencing, for example, marine ecosystems, ocean circulation and global geochemical cycles. Mapping these structures is essential for advancing oceanographic and geological research. However, most seamounts remain unmapped due to limitations in detection methods.

Satellite altimetry has been widely used to detect large seamounts through gravity anomalies, but its resolution constraints hinder the identification of smaller structures. Kim and Wessel (2011) detected seamounts taller than 1,500 meters, estimating between 25,000 and 140,000 seamounts exceeding 1,000 meters in height while suggesting that up to 25 million seamounts above 100 meters remain uncharted. More recently, Gevorgian et al. (2023) expanded the global seamount catalog by identifying 19,325 new seamounts, increasing the total to 43,454. Despite these advances, the reliability of satellite altimetry in detecting small seamounts remains uncertain, particularly given the influence of data resolution and noise.

Multibeam bathymetry enables direct, high-resolution mapping of the seafloor, offering far greater detail than satellite-based methods. However, while the surveys themselves remain time-intensive and spatially constrained, the subsequent analysis and annotation of collected data present an additional bottleneck. To address this challenge, this study introduces an automated deep-learning-based framework to accelerate the detection and classification of small seamounts in multibeam datasets. The approach combines convolutional neural networks (CNNs) for initial filtering with a U-Net-based segmentation model to delineate potential seamount regions. By replacing manual annotation with a scalable two-step pipeline, the method significantly reduces the time and effort required for post-survey analysis—especially for identifying small seamounts often missed in global databases.

An additional challenge lies in understanding the morphological properties of small seamounts. Smith (1988) proposed a height-to-base radius ratio of 0.21, but it remains unclear whether this relationship holds for smaller seamounts or if geometric variations require adjustments in altimetry-based models. Addressing this question is critical for improving detection methodologies.

To systematically evaluate this approach, the study addresses the following research questions:

1. How does a filtering-based approach improve the identification of small seamounts in multibeam bathymetric data compared to manual identification?
2. What are the optimal hyperparameters for training a U-Net model to achieve the highest segmentation accuracy for small seamount detection?
3. How well does the proposed framework generalize across different geographic regions, and what limitations arise when applying a model trained in one ocean basin to another?
4. What is the effective lower detection limit of satellite altimetry for small seamounts, and how does this compare to detections from high-resolution multibeam bathymetric data?

Beyond geological mapping, automated seamount detection has broad applications in marine science. In submarine topography studies, this methodology can be extended to detect and classify other undersea features, such as ridges, trenches, and hydrothermal vent fields (Huang et al., 2024). In marine ecology, seamounts serve as biodiversity hotspots, providing habitat for deep-sea organisms; automating their detection can support conservation efforts (Clark et al., 2010). Additionally, accurate seamount mapping contributes to research on seafloor geodynamics, volcanic activity, and plate tectonics (Matabos et al., 2022). Automated bathymetric analysis also plays a critical role in environmental monitoring and deep-sea mining, assisting in landslide risk assessment and resource extraction planning (Jones et al., 2021; Usui and S, 2022).

2 Literature review

Seamount classification has been a focal point in marine geosciences, employing a range of methods from satellite altimetry to high-resolution multibeam bathymetry. Early studies, such as Smith (1988) and Mitchell (2001), primarily relied on satellite-derived gravity data to detect and classify seamounts. While effective for large-scale features, these approaches are inherently constrained by resolution limitations, as only larger seamounts generate sufficiently strong gravitational anomalies to be visible in global datasets. Multibeam bathymetry provides a higher-resolution alternative, enabling the detection of smaller features. However, its limited spatial coverage and the manual effort required for classification restrict its scalability for global mapping.

To address these challenges, machine learning techniques have been explored for automated feature extraction in bathymetric datasets. Cracknell and Reading (2014) compared various supervised learning algorithms for lithology classification, identifying Random Forests as a robust choice due to its spatial accuracy, while SVMs and k-NN exhibited computational inefficiencies and sensitivity to noise. Despite their success in broad geological classification, these methods rely on hand-engineered features, making them unsuitable for detecting complex, small-scale seamounts.

Recent advances in deep learning have significantly improved seafloor classification by automatically extracting hierarchical features from raw data. Valentine and Kalnins (2013) introduced an autoencoder-based framework to detect seamount-like features based on reconstruction errors, reducing human bias but requiring extensive training data. Similarly, Liu et al. (2024) employed YOLO V7 Tiny for detecting deepsea features under challenging imaging conditions, achieving high accuracy but struggling to generalize across diverse bathymetric terrains.

Several CNN architectures have been widely explored in geospatial and seafloor classification applications, including VGG16 (Simonyan and Zisserman, 2015), ResNet50 (He et al., 2016), InceptionV3 (Szegedy et al., 2016), and MobileNetV2 (Sandler et al., 2018). These models offer varying trade-offs in feature representation, computational efficiency, and robustness:

- VGG16 is a deep yet simple architecture, utilizing small convolutional filters to extract structured features, making it effective for hierarchical representation. However, its high computational demand limits its efficiency for large-scale datasets.
- ResNet50 introduces residual connections, allowing deeper networks while mitigating vanishing gradient issues, making it well-suited for complex pattern recognition in bathymetric data.
- InceptionV3 employs multi-scale convolutions, enhancing adaptability to seamounts of varying size and morphology.
- MobileNetV2, optimized for computational efficiency, uses depthwise separable convolutions but lacks the necessary depth and architectural components for detailed segmentation.

Given the need for efficient large-scale filtering in seamount detection, we conduct a comparative analysis of these models in Section 4.1 to evaluate their effectiveness in generating feature vectors for clustering and classification.

For detailed bathymetric segmentation, U-Net (Ronneberger et al., 2015) was selected as the core architecture due to its proven ability to combine high segmentation accuracy with computational feasibility. Originally developed for biomedical imaging, U-Net's encoder-decoder design, augmented with skip connections, ensures that both contextual and spatial information is preserved—critical for detecting small, irregularly shaped seamounts in multibeam bathymetric data. Unlike classification models that provide a single output per image or object detectors that require bounding boxes, U-Net performs dense pixel-wise labeling, which is particularly suited for the continuous and ambiguous topography of the seafloor. Its relatively low data requirements and efficient training regime further make it a practical solution for seafloor mapping tasks where labeled data is limited.

Other segmentation architectures, though effective in image processing, exhibit notable limitations:

- DeepLabV3+: Chen et al. (2018), while capturing multi-scale context through atrous convolutions, is computationally expensive.
- Mask R-CNN: He et al. (2017) excels in instance segmentation but relies on predefined object boundaries, making it less suitable for the continuous, often ambiguous topographies of seamounts.
- YOLO-based models: Wang and Bochkovskiy (2022), while optimized for real-time object detection, lack the granularity required for detailed segmentation.

Given these considerations, U-Net provides the best balance between segmentation accuracy and computational efficiency. Its architecture is uniquely suited for small seamount segmentation, enabling robust detection even under conditions of sparse training data and morphologically complex targets.

The increasing availability of high-resolution bathymetric datasets has led to a surge in the application of deep learning across marine geosciences, environmental monitoring, and geospatial data fusion. Chitre et al. (2024) demonstrated machine learning applications in bathymetric data processing, while Cherubini et al. (2024) utilized Copernicus Marine Service and EMODnet data for marine habitat modeling. Similarly, Deng et al. (2024) applied deep learning to analyze the environmental impact of floating offshore wind turbines.

Beyond environmental modeling, deep learning has also been applied in geospatial data fusion and numerical homogenization. Khalil et al. (2024) integrated airborne electromagnetic and borehole data with bathymetric analysis to enhance coastal mapping, while Qin et al. (2024) developed multi-scale satellite-derived bathymetry models to improve spatial resolution.

Despite these advancements, detecting small seamounts remains challenging due to:

- Limited labeled training data for small-scale features.
- High variability in seafloor morphology, making classification difficult.
- Distinguishing true seamounts from noise in multibeam bathymetry.

Many models, including Random Forests, SVMs, and XGBoost, struggle to generalize across diverse regions. Unsupervised clustering techniques, though useful in segmenting bathymetric images, often fail to distinguish small seamounts from background noise.

To address these challenges, this study introduces a two-step deep learning framework combining CNN-based feature filtering with U-Net segmentation:

1. Feature Clustering: The dataset is first filtered to pre-select seamount candidates using CNN-generated feature vectors.
2. Seamount Segmentation: The U-Net model is then applied to refine the classification, ensuring robust detection.

Additionally, this study explicitly tests cross-regional generalization, training on Atlantic Ocean bathymetry and evaluating on Indian Ocean datasets to assess model adaptability.

By integrating these innovations, this study presents a scalable, high-accuracy framework for small seamount detection, addressing the key limitations in machine learning-based seafloor classification. The following sections outline the methodology, experimental setup, and results to demonstrate the effectiveness of this approach.

3 Methodology

3.1 Image filtering

Our framework processes large-scale bathymetric data into manageable subsets, facilitating the efficient detection of potential seamount features. The methodology, illustrated in Figure 1, involves five key steps: image creation, segmentation, feature calculation, clustering, and manual labeling.

Image Creation: The input consists of multibeam bathymetric data that has been preprocessed to correct artifacts and improve overall quality. To ensure clean input data, outlier detection and removal were performed using the optimized filtering method described by Ziolkowski et al. (2024), which enhances data reliability by eliminating spurious depth values from multibeam echo-sounder measurements. High-resolution seafloor maps were subsequently generated using Python's Matplotlib library, applying the viridis color scheme to represent depth variations. This perceptually uniform colormap enhances contrast between flat seafloor and elevated features such as seamounts, facilitating the ability of the U-Net architecture to learn and distinguish relevant morphological patterns during the segmentation process. To ensure that every 256×256 image uses the same absolute depth-to-color mapping (and thus identical contrast), we compute a single pair of "global" depth limits (global min, global max) over the entire input survey before tiling. After interpolating each 24×24 chunk and resizing it to 256×256, we clip every pixel to [global min,global max] and linearly rescale to [0,1]. In this way, no two images from the same survey ever have different contrast ranges—each pixel's color always maps back to the same meter-value. *Segmentation:* To efficiently manage the computational challenges of seafloor mapping, the data is divided into 256 × 256 pixel images with 10% overlap, ensuring that no seamount is truncated at the segment edges. This step enables efficient downstream processing while retaining critical morphological details in each region and ensuring that the image size is large enough to fully visualize entire seamount structures. We chose 256 × 256 as our image size because it is a common power-of-two input for U-Net. We briefly tested 128 × 128 (faster but lost small-feature fidelity) and 512 × 512 (higher fidelity but 4× more memory/time) and found that 256 × 256 provided the best trade-off. Pixel resolution is kept fixed across all surveys: each chunk is first interpolated to a 24 × 24 grid at 0.001° resolution—covering 0.024° × 0.024° in latitude/longitude—and then resized to 256 × 256 pixels. Hence each pixel corresponds to 9.375 × 10° (10 m, depending on latitude), both during training and application. Even if a new dataset has different raw point densities, our pipeline "forces" it onto that same 0.024°footprint per image, so the model always sees a consistent meter-per-pixel scale. In summary, by fixing grid resolution=0.001 and chunk size=24 and always resampling to 256 × 256, we guarantee identical pixel resolution from training to application, regardless of which survey file is used. *Feature Calculation:* CNNs compute feature vectors for each segmented image, capturing key characteristics such as texture and structure. These vectors provide a compact, descriptive representation of the seafloor features, enabling effective analysis.

Clustering: Feature vectors are clustered into 10 groups using unsupervised methods, which ensures that images with similar morphological characteristics are grouped together, significantly reducing dataset complexity and focusing attention on potential seamount regions. The choice of k=10 was not arbitrary but reflects a balance between two competing needs: capturing the major morphological variations in our CNN-derived feature space and keeping the number of clusters low enough for efficient human review. In practice, we found that ten clusters cleanly separated large, flat or gently sloping patches from steeper, seamount-like textures. Increasing k beyond 10 rarely produced qualitatively new seamount candidate groups—most extra clusters simply subdivided empty or flat-area images—while fewer than ten clusters began to merge distinct seamount morphologies with background.

Labeling: A domain expert reviews and labels the clusters. This human oversight ensures accurate identification of potential seamount candidates. Images of flat seafloor and background are excluded from further analysis, while the potential seamount cluster is retained for subsequent steps. On average, the cluster-level review takes under five minutes per survey: the expert scans a handful of thumbnails from each of the 10 clusters (100 images total) in about 2–3 minutes, discards the clearly "background" clusters, and flags only a few as "seamount candidates." If desired, they can then page through those candidate clusters for extra confidence—but the minimal filtering step is complete in under five minutes, since no individual "yes/no" decision is made on all 5–804 images. The result of this methodology is a refined dataset of labeled clusters. Only the

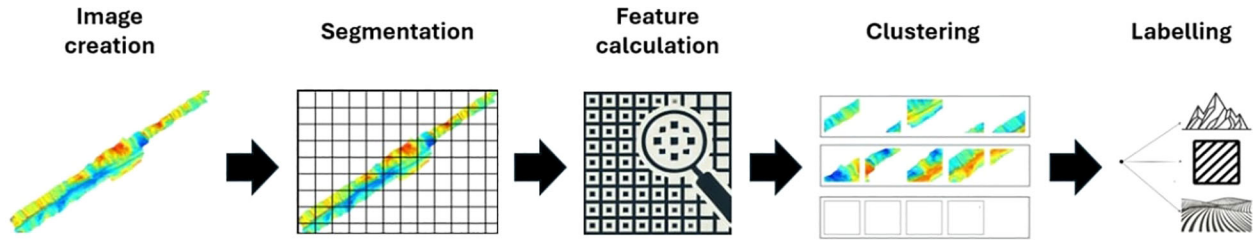


FIGURE 1
Workflow for filtering images containing potential small seamounts from multibeam bathymetric data. The process involves five key steps: (1) creation of a high-resolution seafloor map, (2) segmentation into smaller images for efficient processing, (3) feature extraction using CNNs, (4) filtering of images based on calculated feature vectors, and (5) manual labeling of filtered images to identify potential seamount candidates for further analysis.

clusters containing potential seamounts undergo further analysis to detect the summits and extents of each seamount.

3.2 Workflow for training and evaluating a CNN for seamount detection

The workflow shown in Figure 2 outlines the process for preparing, training, and evaluating a UNet architecture to detect seamounts in multibeam data, beginning after the pre-selection of images likely to contain seamounts (Figure 1). *Data Input:* The workflow starts with the selected images containing regions that most likely include seamounts, as shown in Figure 2. These images serve as the input for the subsequent labeling and model training steps.

Summit Detection and Extent Mapping: Using labelme, seamount features are manually annotated to create masks for training. Black polygons outline the extent of each seamount. This step ensures accurate identification of key morphological features essential for the training process. Although all annotations were created by a single domain expert to maintain consistency, this introduces potential subjectivity and bias into the ground truth masks. Future work should consider inter-annotator agreement studies or collaborative labeling strategies to better quantify annotation reliability and improve robustness of training data.

Mask Generation: The annotated summit and extent data are used to generate binary masks for each seamount, where black areas represent the seamount and white areas indicate the background. These masks serve as the ground truth for training the UNet architecture, establishing the expected output for each input image.

Model Training: The UNet architecture is trained using the input images and their corresponding masks. The model learns to map the input image features to the expected output, enabling it to detect and delineate seamounts in multibeam bathymetric data accurately.

Model Evaluation: The trained model is evaluated using the mean Intersection over Union (mean IoU) metric, which measures the overlap between the predicted and manually labeled masks and ranges from 0 (no overlap) to 1 (perfect overlap). A higher mean IoU indicates better model performance in identifying and segmenting seamounts, providing a reliable assessment of its accuracy. Generally, mean IoU values between 0.75 and 0.85 are considered acceptable for complex medical segmentation tasks, particularly when segment boundaries are difficult to define, such as in tumor segmentation or vessel segmentation (Amri et al., 2025; Peng et al., 2025; Moradmand and R, 2025). In addition to mean IoU, the Dice coefficient is another widely used metric in image segmentation, particularly in medical imaging. It measures the similarity between predicted and ground-truth segmentations and ranges from 0 (no overlap) to 1 (perfect overlap). The Dice coefficient is particularly useful in imbalanced datasets, where positive class pixels (e.g., segmented structures) are much fewer than background pixels (Chamseddine et al., 2025; Yang et al., 2025; Alyahyan, 2025).

To optimize model training, the Dice loss function is employed, which is derived from the Dice coefficient. It is commonly used in medical image segmentation because it mitigates the effect of class imbalance by emphasizing the similarity of foreground structures rather than treating all pixels equally. Dice loss is especially beneficial for detecting small and irregularly shaped structures, making it a suitable choice for seamount segmentation, where feature boundaries are often ambiguous (Zhang et al., 2025; Shen et al., 2025). In future studies, implementing cross-validation labeling rounds with multiple annotators and calculating inter-annotator metrics such as Cohen's kappa (Cohen, 1960) could further strengthen the training dataset quality and reduce the likelihood of label noise.

This workflow represents a comprehensive pipeline for training and evaluating a UNet model tailored for the automatic detection of small seamounts. It combines human expertise in labeling with advanced machine learning techniques, enabling efficient and accurate analysis of multibeam bathymetric data.

4 Results and discussion

4.1 Analysis of model performance in seamount image filtering using feature vectors

Seamount images show complex patterns and structural ambiguity, posing significant challenges for automated feature

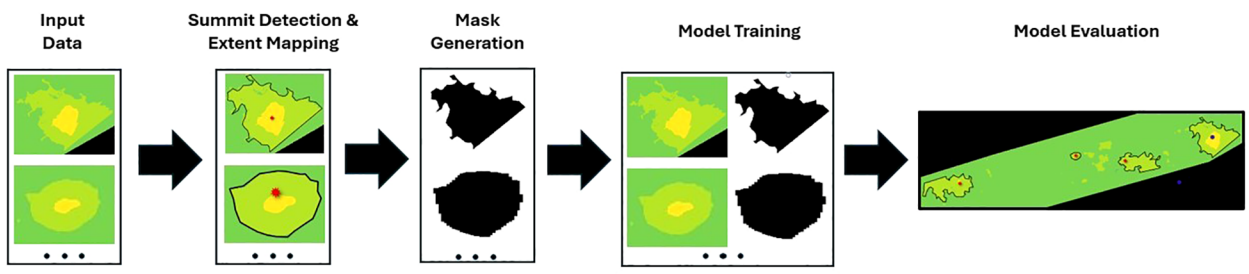


FIGURE 2
Workflow illustrating the processing pipeline for training a UNet architecture to detect small seamounts in multibeam data. The pipeline begins with raw multibeam data input, followed by extent mapping to annotate seamount features. Masks are then generated to prepare training images, which are used to train the UNet model. The workflow concludes with model evaluation to assess performance and accuracy in detecting and delineating seamounts.

extraction. The results indicate that models with stronger feature extraction capabilities, such as VGG16 and ResNet50, produced more precise feature vectors, leading to better clustering performance and higher agreement with manual labeling.

Models that rely on lightweight architectures and reduced feature complexity, such as MobileNetV2, demonstrated lower performance, particularly in separating clusters when faced with highly uneven cluster sizes. InceptionV3, while effective in capturing variations in shape and color, exhibited reduced clustering precision when confronted with uniform textures across different clusters. Below, the performance of each model is discussed in terms of cluster separation, agreement with manual labeling, strengths, and weaknesses, as summarized in Table 1.

- VGG16: VGG16 achieved the highest agreement with manual labeling (97–100%), demonstrating robust and interpretable feature extraction, leading to clear cluster separation. Its architecture is particularly suited to datasets with clear patterns, making it ideal for applications requiring consistent and robust feature extraction. However, its tendency to over represent large clusters limited its effectiveness for highly complex or imbalanced datasets.
- ResNet50: ResNet50 performed well in scenarios requiring the extraction of more complex or abstract patterns, achieving 81–90% agreement with manual labeling. It is a viable alternative for datasets with higher structural variability or subtle morphological differences. However, its performance was less consistent than VGG16, particularly in datasets with limited textural differentiation, where it struggled to maintain stable clustering.
- InceptionV3: InceptionV3 showed strong multi-scale feature extraction but lower consistency, with agreement scores ranging from 66–94%. It is recommended for datasets with significant variability in patterns and colors, but it is less effective for uniform image distributions. Its performance was hindered when dealing with color homogeneity within clusters, leading to occasional misclassification.
- MobileNetV2: MobileNetV2 had the lowest agreement with manual labeling (18–82%), reflecting its difficulty in handling fine-grained textures and separating clusters effectively. This was especially evident in datasets where clusters varied significantly in size, ranging from single instances to over 3000 images. While computationally efficient, MobileNetV2 should be avoided in tasks requiring detailed feature extraction, such as seamount identification, due to its inability to handle complex patterns.

The analysis highlights VGG16 as the optimal model for seamount identification due to its ability to extract robust features and achieve high agreement with manual labeling. ResNet50 is a strong alternative for datasets with complex patterns but suffers from inconsistencies in cluster separation. InceptionV3 is useful for datasets with diverse features but struggles with uniform patterns, while MobileNetV2 is unsuitable for this application due to its limited feature extraction capabilities and poor clustering performance. These insights provide a clear basis for selecting

TABLE 1 Summary of clustering results across different CNN architectures for seamount image classification.

Model	Total	Selected	% sel.	Cluster distribution	Matches	Manual	% match	Clusters
VGG16	16816	11266	66.99	300–2000 per cluster, good separability	124	127	97.64	15
VGG16	16816	5834	34.69	Uniformly distributed	127	127	100.00	10
VGG16	16816	6818	40.54	One large cluster with 10k, others balanced	120	127	94.49	5
ResNet50	16816	11927	70.93	400–1600 per cluster, good differentiation	103	127	81.10	15
ResNet50	16816	8151	48.47	Uniformly distributed	86	127	67.72	10
ResNet50	16816	9430	56.08	One large (8k), others 2500 each	115	127	90.55	5
InceptionV3	16816	11178	66.47	129–1900, not perfectly separable	102	127	80.31	15
InceptionV3	16816	14404	85.66	Slightly uneven distribution	120	127	94.49	10
InceptionV3	16816	15419	91.69	One larger cluster, rest uniform	84	127	66.14	5
MobileNetV2	16816	6687	39.77	Very uneven, 1–3059 per cluster	23	127	18.11	15
MobileNetV2	16816	9430	56.08	Very uneven distribution	105	127	82.68	10
MobileNetV2	16816	9433	56.10	Very uneven distribution	69	127	54.33	5

The table lists the number of total and selected images, the percentage selected, and a qualitative description of cluster distribution. "Cluster Distribution" refers to how images are grouped in terms of size variation, uniformity, and dominance. The "% Match" indicates the percentage of selected images matching the manually curated seamount labels.

appropriate models based on the specific requirements of seamount image clustering tasks.

4.2 Training of the U-Net architecture for seamount detection

The dataset used for training the U-Net architecture consists of high-resolution bathymetric and geological data collected during two research cruises, MSM75 and MSM88, in the Atlantic Ocean. These datasets provide detailed information on seafloor morphology, fault structures, and small seamounts, making them well-suited for an image-based deep learning approach.

The MSM75 cruise, conducted in 2018, focused on four key areas along the Reykjanes Ridge, a slows preading ridge influenced by the Iceland hotspot. This dataset includes 15 m resolution ship-based bathymetry, ROV-based ground-truthing, and geochemical analyses of glass samples, capturing variations in magma composition, fault density, and seamount morphology. These features are strongly influenced by factors such as distance from the hotspot and the magmatic or tectonic accretion state of axial volcanic ridges (AVRs) (Le Saout et al., 2023). Given the distinct geological and morphological variations within the dataset, it provides an excellent basis for training a segmentation model capable of distinguishing complex seafloor structures.

Complementing this, the MSM88 cruise dataset, collected using a Kongsberg EM 122 multibeam system at approximately 100 m horizontal resolution, covers a much larger area—approximately 153,121 square kilometers—spanning from the Cabo Verde Exclusive Economic Zone (EEZ) to the EEZs of Guadeloupe, Dominica, and Martinique. This dataset includes diverse Atlantic seabed morphologies, ranging from flat sedimented plains to seamounts, fracture zones, and the Mid-Atlantic Ridge. The large volume of depth soundings (86 million) ensures high spatial coverage and variability, further enhancing the robustness of the training data.

Table 2 provides an overview of the spatial extent, resolution, and depth ranges of the three datasets used for training and testing. The diversity of these datasets enhances the robustness and applicability of the model across different seafloor morphologies.

These datasets are particularly well-suited for training the U-Net model, as they provide high-resolution seafloor imagery with detailed geological labels. The combination of fine-scale bathymetry from MSM75 and the broader regional coverage of MSM88 ensures that the model learns to generalize across varying seafloor structures, improving its ability to segment and classify geological features effectively.

4.3 Hyperparameter selection and training strategy

In this section, we analyze the impact of different hyperparameter configurations on the performance of the U-Net architecture for seamount detection. The evaluation focuses on validation loss, mean Intersection over Union (IoU), and validation mean IoU, as summarized in Table 3. The goal of this analysis is to identify the optimal parameter constellation for final model training, ensuring high segmentation accuracy and robustness.

Mean Intersection over Union (IoU) is a widely used metric in image segmentation, quantifying the overlap between predicted and ground truth masks. It is calculated as the ratio of the intersection to the union of both masks, ranging from 0 to 1, where higher values indicate better segmentation performance (Dwarakanath and Kuntiyellannagari, 2025).

Several key hyperparameters were varied during the grid search, including the number of filters, kernel size, dropout rate, learning rate, and batch size. One of the primary considerations is the number of filters in the convolutional layers, which defines the depth of feature extraction. A lower filter count, such as 16, may fail to capture sufficient spatial details, whereas a significantly higher count, such as 256 or more, increases computational costs and the risk of overfitting, particularly given the relatively small dataset size. To balance feature richness and computational efficiency, 32 and 128 filters were selected, following insights from prior research in biomedical segmentation tasks (Iqbal et al., 2022; Srinivasan et al., 2024).

Another crucial factor is the kernel size, which determines the receptive field of convolutional layers. Smaller kernels, such as 3×3, are effective for fine-grained detail extraction, while larger kernels, such as 5×5, allow for broader spatial pattern detection in bathymetric structures. The study focused on comparing these two kernel sizes, as excessively large kernels (e.g., 7×7) could introduce computational challenges and potentially over-smooth small-scale features.

To mitigate overfitting and enhance generalization, dropout rate was varied between 0.1 and 0.5. Dropout serves as a regularization technique by randomly deactivating neurons during training, preventing the model from relying too heavily on specific features. This variation allowed for an assessment of the trade-off between preventing overfitting and ensuring sufficient information retention for effective segmentation.

Additionally, the learning rate plays a vital role in determining how quickly the model updates its weights during training. A low learning rate encourages stable convergence, whereas a higher learning rate accelerates training but increases the risk of

TABLE 2 Summary of bathymetric datasets used for training and evaluation.

Dataset	Cruise ID	Area covered (km^2)	Resolution	Depth range (m)
MSM75	MSM75	~10,000	15 m	102–2,044
MSM88	MSM88	~153,000	100 m	1,500–6,000
SO305/2	SO305/2	~12,000	100 m	492–5,664

overshooting optimal weight values. To identify an optimal balance, the study compared learning rates of 0.0001 and 0.001, ensuring that the model could learn effectively without instability or divergence. We use the Adam optimizer (with the learning rate chosen via our hyperparameter search). Adam combines the benefits of momentum and adaptive learning rates, which helps stabilize training on our relatively small U-Net dataset.

Lastly, the batch size was explored to assess its effect on training efficiency and model performance. Smaller batch sizes allow for more frequent weight updates per iteration, while larger batch sizes contribute to more stable gradient estimations. To maintain a balance between computational efficiency and convergence stability, batch sizes of 16 and 64 were evaluated.

Their selection was guided by best practices in deep learning, computational efficiency, and the unique characteristics of bathymetric data. Specifically, the dataset was divided into 80% training and 20% validation sets using stratified sampling to preserve class balance and ensure a robust evaluation of model performance. We combined labeled images from both MSM75 and MSM88 into a single pool, then applied an 80/20 split with random state=42, so the training/validation split is fixed across all runs. For augmentation, we rotated each normalized 256×256 image by 90° and 180°, producing two extra images per original (three total). These practices are widely adopted in the geospatial and marine sciences communities and have been recommended for applications involving multibeam bathymetry and habitat mapping (Summers et al., 2021; Roelfsema et al., 2021). Their influence on model performance is analyzed in the following sections, with a focus on preventing overfitting and supporting generalization across diverse seafloor morphologies.

The following sections discuss the results of these hyperparameter configurations, analyzing their influence on model performance and the trade-offs they introduce in the context of seamount segmentation.

4.3.1 Number of filters

In our implementation, the number of filters doubles at each successive "down" step in the encoder and then halves again in the decoder. The results indicate that models using 32 filters generally outperform those with 128 filters in terms of mean IoU and validation mean IoU. The best-performing configuration (32 filters, kernel size 5, dropout rate 0.1, learning rate 0.0001, batch size 64) achieves a validation mean IoU of 0.722, higher than configurations with 128 filters, which generally yield IoU values below 0.66.

Models with 128 filters and a large kernel size (5) tend to perform poorly, particularly in cases where the dropout rate is high or the learning rate is large. Several configurations with 128 filters, kernel size 5, dropout rate 0.5, and a learning rate of 0.001 resulted in extremely poor performance (mean IoU < 0.21). These results suggest that larger models may overfit or fail to generalize when handling small-scale features in seamount detection.

4.3.2 Kernel size

A kernel size of 5 consistently improves model performance compared to a kernel size of 3. The best-performing models all use a 5 × 5 kernel, which appears to enhance the model's ability to capture seamount structures in multibeam data. Notably, the highest validation mean IoU (0.722) is obtained with a 5 × 5 kernel, 32 filters, dropout rate 0.1, learning rate 0.0001, and batch size 64.

Configurations with a 3 × 3 kernel tend to yield slightly lower performance, with validation mean IoU values ranging from 0.584 to 0.660. While smaller kernels may still be effective, the data suggests that capturing larger contextual information with a 5 × 5 kernel improves segmentation quality. Larger kernels (e.g., 7 × 7) were not tested due to increased computational complexity and potential over-smoothing of small seamount features.

4.3.3 Dropout rate

The best-performing models use a dropout rate of 0.1, while higher dropout rates (0.5) lead to a decline in performance. Configurations with dropout 0.5 frequently result in unstable training, with validation mean IoU values dropping below 0.55 in most cases. This suggests that excessive regularization hinders the network's ability to learn fine-grained features necessary for segmenting small seamounts. Lower dropout values (<0.1) were avoided to prevent potential overfitting, while higher values (>0.5) were not considered due to excessive information loss during training.

4.3.4 Learning rate

A learning rate of 0.0001 is generally more stable and results in higher mean IoU values than 0.001. Many configurations with a learning rate of 0.001 exhibit poor performance, with validation loss values reaching 0.828, indicating divergence or unstable training.

Notably, when a learning rate of 0.0001 is used in combination with a kernel size of 5 and dropout rate of 0.1, the model achieves the best performance. These findings suggest that a lower learning rate prevents the model from overshooting optimal weights, leading to better generalization. Higher learning rates (>0.01) were excluded due to the risk of divergence, while lower rates (<0.0001) were avoided as they could lead to excessively slow training.

4.3.5 Batch size

The best-performing models generally use a batch size of 64. While some configurations with batch size 16 perform well (validation mean IoU around 0.66), they do not outperform batch size 64 when combined with optimal hyperparameters.

Interestingly, several models with batch size 16 and 128 filters perform significantly worse, possibly due to instability in training. A larger batch size appears to contribute to better gradient estimation and stable convergence. Extremely large batch sizes (>128) were not tested due to GPU memory constraints and the risk of poor generalization.

4.3.6 Optimal configuration and conclusions

Based on this analysis, the best-performing configuration is:

32 filters, kernel size 5, dropout rate 0.1, learning rate 0.0001, batch size 64.

This configuration achieves the highest validation mean IoU of 0.722, suggesting that it provides the most reliable segmentation performance for small seamounts. These results emphasize the importance of choosing a balanced architecture that prevents overfitting while ensuring stable learning dynamics. The findings also reinforce that hyperparameter tuning is essential for optimizing deep learning models in seamount segmentation, as poor configurations can severely impact model accuracy and generalization ability.

4.3.7 Balancing generalization and model complexity

In deep learning applications, particularly those involving image segmentation, managing the balance between model complexity and generalization is crucial to avoid overfitting or underfitting. These phenomena directly influence a model's ability to perform accurately on unseen data and are especially critical when working with spatially diverse and sparsely labeled bathymetric datasets.

Overfitting occurs when a model learns the training data too well, including its noise and minor fluctuations, leading to poor generalization on validation or test data. This typically manifests as a low training loss combined with a high validation loss. In contrast, underfitting arises when the model is too simplistic to capture the underlying patterns of the data, resulting in high errors on both training and validation sets.

To ensure that the U-Net model maintains a strong balance between learning capacity and generalization, we monitored validation loss, mean Intersection over Union (IoU), and validation mean IoU across training epochs (see Table 3). These metrics help assess both segmentation accuracy and model robustness. In particular, consistently high validation mean IoU values without significant divergence from training performance indicate strong generalization ability.

Additionally, to avoid overfitting, regularization strategies such as dropout, early stopping, and data augmentation were applied. Stratified sampling was used to divide the dataset into 80% training and 20% validation subsets, preserving class balance and ensuring that all seamount categories are proportionally represented.

The observed performance trends align with best practices established in machine learning literature. For example, Sivakumar et al. (2024) emphasize the trade-off between training and testing ratio and its effect on generalization in image processing. Similarly, Manikandan et al. (2024) highlight the impact of architectural complexity on overfitting and underfitting in segmentation tasks using U-Net, supporting our methodological choices for small seamount detection.

While the model demonstrates strong validation performance, a limitation remains its sensitivity to out-of-domain (OOD) data—bathymetric inputs that differ significantly from the training distribution in terms of seafloor morphology, resolution, or noise characteristics. Such domain shifts frequently occur in real-world deployments and may degrade model reliability. Future work should therefore explore strategies such as domain adaptation, transfer learning, and uncertainty quantification. These approaches can improve robustness by enabling the model to generalize to morphologically diverse regions, reducing the risk of false positives or negatives in unfamiliar tectonic settings. Transfer learning, in particular, has shown promise in segmentation tasks with sparse annotations and heterogeneous data domains, such as in medical imaging (Tajbakhsh et al., 2016).

4.4 Application of workflow to real-world data

The data shown in Figure 3 were acquired during SO305-2, a transit across the Indian Ocean after exiting the territorial waters of Indonesia and Malaysia. Using the EM122 swath mapping system, high-resolution bathymetric data were collected along this tectonically active region, which exhibits significant deformation of the oceanic plate. The dataset reveals detailed seafloor morphology, uncovering previously uncharted geological features in this underexplored area.

As the survey approached the Central Indian Ridge (CIR), it focused on the Argo transform fault and its fracture zones. The EM122 system detected numerous small seamounts, many less than 1000 meters in diameter, which remain undetectable in lower-resolution satellite altimetry. This highlights the limitations of satellite-based mapping for smaller topographic features and underscores the advantages of multibeam systems in resolving fine-scale bathymetric details.

This high-resolution dataset provides a valuable resource for developing and validating automated seamount detection algorithms. It offers detailed bathymetric imagery across varied tectonic settings, making it an ideal testbed for refining detection methods and improving our understanding of small seamount distribution and morphology.

A total of 11,139 images were generated from the SO305–2 expedition data during preprocessing, with 30 seamounts manually labeled. To prepare the dataset for seamount detection, a filtering and clustering process (Section 4.1) reduced the dataset to 6,626 images, effectively eliminating 40% of the original data. As shown in Figure 4, clusters 1, 6, and 7 were selected for further processing, as they most likely contain seamount images, while the remaining clusters primarily represent flat seafloor or other irrelevant features. Clusters 0 and 2, identified as potential artifacts likely caused by noise, were excluded from further analysis. Additionally, 30 seamounts were manually identified within the dataset, and all 30 seamounts from the original data were retained in the selected images, ensuring comprehensive coverage of the target structures for model training. For model training, we used the 256×256 images generated from the MSM75 and MSM88 datasets, in which 138 seamounts had been manually labeled. The U-Net was trained on this combined pool of MSM75/MSM88 images.

The model was trained using the optimal hyperparameters identified in Section 4.3. To prevent overfitting and ensure

optimal performance, early stopping was implemented, monitoring validation loss and halting training once no further improvements were observed. Additionally, model checkpointing was used to save the model whenever a lower validation loss was achieved, ensuring retention of the best-performing version for further evaluation. The progression of validation loss throughout training is shown in Figure 5, exhibiting a steady decline until approximately epoch 37, after which further reductions become minimal.

Throughout training, the model demonstrated a progressive improvement in segmentation performance, as reflected in the increasing Dice Coefficient and Mean IoU, while validation loss steadily decreased. Dice Loss, commonly used in segmentation tasks to mitigate class imbalance, is derived from the Dice Coefficient, a similarity measure evaluating the overlap between two sets. As for class imbalance, most images have no seamount pixels, and even in images that do, seamounts cover only about 10–20% of pixels—hence our use of Dice Loss. By emphasizing misclassified regions, Dice Loss helps capture fine-grained structures, making it particularly effective for segmenting objects with irregular boundaries (Zheng et al., 2025).

During the initial training phase (epochs 1–10), the model exhibited low Dice scores, ranging from approximately 0.14 to 0.21. However, validation loss dropped sharply from 0.85 to 0.26 within this period, with the first major performance improvement occurring around epoch 5, marking the transition to more stable learning. This trend is illustrated in Figure 6, which depicts the evolution of the Dice Coefficient over epochs.

In the mid-training phase (epochs 11–30), the model continued improving, with validation loss reaching its minimum (0.1734) at epoch 37. The Dice Coefficient rose significantly, surpassing 0.82, while the Mean IoU exhibited a steady upward trend, further indicating the model's ability to generalize effectively. The trajectory of the Mean IoU over training epochs, as visualized in Figure 7, reflects this improvement.

During the late training phase (epochs 30–50), signs of overfitting emerged as validation loss plateaued. The Dice Coefficient fluctuated between 0.81 and 0.86, while the Mean IoU remained relatively stable, showing minimal gains beyond epoch 37. These observations suggest that further training did not yield additional benefits, indicating that the model had reached its optimal performance.

The best performance was recorded at epoch 37, where validation loss reached its minimum (0.1734), and the model attained a Dice Coefficient of 0.8274 and a Mean IoU of 0.7514—representing the peak segmentation accuracy observed during training. These results suggest that the model successfully learned meaningful feature representations for image segmentation, with performance stabilizing beyond this epoch. Consequently, epoch 37 was identified as the optimal balance point between learning and generalization.

After training, the U-Net model was applied to the filtered dataset to generate segmentation results for seamount detection. Model predictions were compared to manually labeled seamounts to assess performance. The U-Net successfully identified all 30 seamounts in the dataset, demonstrating high detection accuracy. The predicted outlines closely matched the ground truth, with only minor deviations in shape and boundary precision, suggesting that the model effectively captures key morphological characteristics of seamounts.

Figure 8 displays all 30 manually labeled seamounts alongside their corresponding model predictions, highlighting the robustness of the proposed workflow in accurately detecting and segmenting small seamount structures.

To further illustrate the challenges of detecting small seamounts, Figure 9 presents examples of false predictions made by the U-Net model. The misclassification of certain regions as seamounts can be attributed to the complexity of seafloor morphology and the inherent subjectivity of manual labeling. Seafloor features vary significantly, and even human interpreters may disagree on what qualifies as a seamount. Given this subjectivity, discrepancies between model predictions and reference labels are expected due to human error or differing interpretations of the data.

A common characteristic among false positives is the presence of localized seafloor elevations, which appear as yellow regions in the bathymetric data. Although not actual seamounts, these features share topographic similarities with true seamount structures, making misclassification understandable. However, a key limitation of the U-Net model is its occasional inability to accurately capture the typical circular morphology of small seamounts. Instead, elongated or irregularly shaped elevations are sometimes misclassified as seamounts despite lacking the distinct topographic characteristics that define them.

These observations suggest that while the model effectively identifies seafloor elevations, it could be further improved in distinguishing true seamounts from other raised features. Future refinements could involve integrating morphological constraints during training or applying post-processing techniques to filter out elongated structures that do not conform to the expected circular shape of small seamounts.

In Section 2, this study identified three key challenges in small seamount detection: (1) the scarcity of manually labeled training data, (2) the difficulty of segmenting irregular and morphologically diverse features, and (3) the need for models that generalize across varying seafloor conditions. To address the first challenge, a training set of 138 seamounts was manually labeled using high-resolution multibeam bathymetry, providing a diverse and representative dataset for supervised learning. The second challenge was mitigated through the use of the U-Net architecture, whose encoder-decoder structure and skip connections allow for precise pixel-wise segmentation of irregular and fine-scale seafloor features. Lastly, the model's generalization capability was enhanced by filtering the dataset with a CNN-based clustering approach, reducing noise and guiding the network's attention to relevant regions. Together, these strategies enabled effective training and application of a robust segmentation model capable of detecting small seamounts with high accuracy in real-world data.

Figure 10 highlights the 30 seamounts identified in the SO305–2 dataset, revealing a significant number of previously undetected features. A major limitation of satellite-derived global seamount

TABLE 3 Hyperparameter tuning results for U-Net.

Filters	Kernel size	Dropout rate	Learning rate	Batch size	Val loss	Mean IoU	Val mean IoU
32	3	0.5	0.0001	64	0.403	0.584	0.584
128	3	0.5	0.0001	16	0.315	0.513	0.516
128	3	0.1	0.001	16	0.575	0.541	0.546
32	3	0.1	0.0001	64	0.298	0.616	0.617
32	3	0.1	0.0001	16	0.292	0.606	0.606
32	5	0.1	0.0001	16	0.214	0.637	0.636
32	3	0.1	0.001	16	0.301	0.609	0.615
128	5	0.1	0.001	64	0.828	0.207	0.101
128	5	0.5	0.0001	16	0.828	0.193	0.095
32	5	0.1	0.001	64	0.828	0.299	0.153
128	5	0.1	0.0001	16	0.318	0.606	0.628
32	3	0.5	0.001	16	0.828	0.309	0.161
128	5	0.1	0.001	16	0.828	0.199	0.094
32	5	0.1	0.0001	64	0.225	0.721	**0.722**
32	3	0.1	0.001	64	0.798	0.463	0.460
32	3	0.5	0.0001	16	0.493	0.544	0.548
128	5	0.1	0.0001	64	0.353	0.596	0.609
128	3	0.1	0.0001	64	0.275	0.659	0.660
128	5	0.5	0.001	16	0.505	0.462	0.469
128	5	0.5	0.0001	64	0.369	0.606	0.616
32	5	0.5	0.001	16	0.545	0.427	0.492
128	3	0.5	0.0001	64	0.295	0.650	0.652
128	3	0.5	0.001	64	0.828	0.209	0.100
128	3	0.5	0.001	16	0.828	0.194	0.094
32	3	0.5	0.001	64	0.828	0.318	0.167
32	5	0.5	0.0001	16	0.220	0.696	0.698
128	5	0.5	0.001	64	0.828	0.196	0.093
32	5	0.1	0.001	16	1.000	0.343	0.363
128	3	0.1	0.001	64	0.555	0.459	0.460
128	3	0.1	0.0001	16	0.291	0.649	0.651
32	5	0.5	0.001	64	0.828	0.201	0.095
32	5	0.5	0.0001	64	0.218	0.713	0.715

The configuration with 32 filters, kernel size 5, dropout rate 0.1, learning rate 0.0001, and batch size 64 achieved the highest validation mean IoU of 0.722 (highlighted in bold), indicating optimal performance for seamount segmentation.

datasets, such as those based on vertical gravity gradient (VGG) data, is their inability to resolve smaller seamounts (Yesson et al., 2011). Consequently, only 14 of the 30 identified seamounts were visible in satellite altimetry data, while the remaining 16 were too small to be detected. This underscores the importance of high-resolution.

As shown in Table 4, 16 of the 30 identified seamounts (well-known = 2) were completely absent from global satellite datasets. In their study, Gevorgian et al. (2023) improved upon previous altimetry-based seamount detection methods, identifying seamounts as small as 421 meters in height, with most detections exceeding 700 meters due to the limitations of the VGG method.

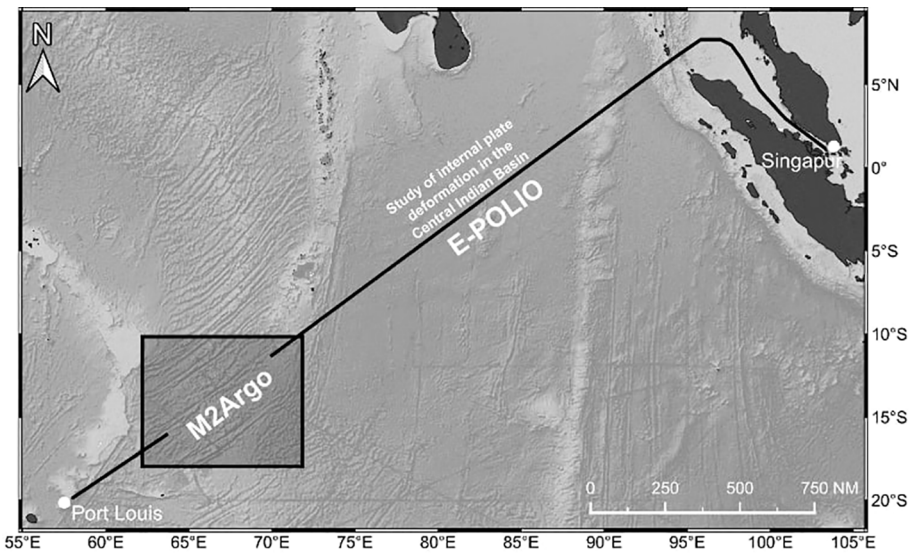


FIGURE 3
Joint working area of the E-POLIO and M2Argo projects during the SO305–2 cruise, shown along the transit route from Singapore to Port Louis. The boxed area indicates the survey region focused on the ARGO fracture zone.

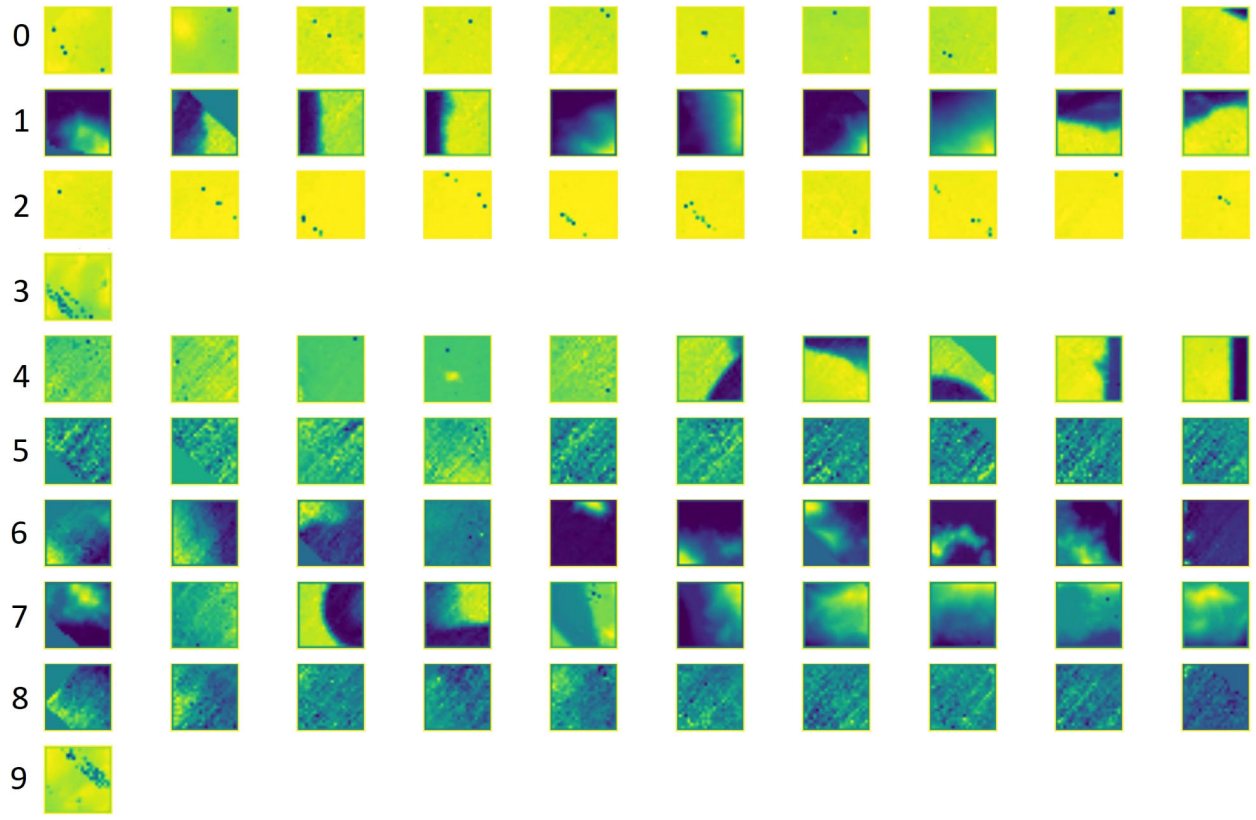


FIGURE 4
Clusters generated after the filtering process, highlighting distinct seafloor morphologies. Clusters 1, 6, and 7 likely contain seamount images, making them suitable for further analysis. All other clusters primarily represent flat seafloor or similar features and can be excluded from the U-Net detection process for small seamount detection.

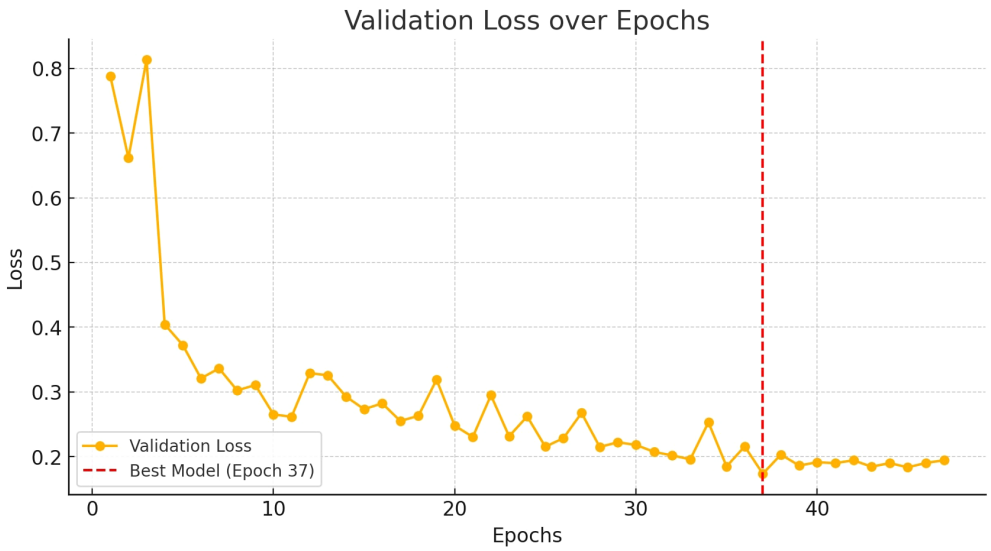


FIGURE 5
Validation loss over epochs during model training. The red dashed line indicates the epoch where the best model was saved based on the lowest validation loss.

This marked a significant improvement over earlier studies, such Kim and Wessel (2011), who noted that traditional altimetry-based methods struggled to detect features smaller than 1,500 meters due to the limited resolution of gravity anomaly data. These findings underscore the limitations of satellite-derived data in detecting smaller seamounts and highlight the necessity of high-resolution multibeam bathymetric surveys for comprehensive seafloor mapping.

Additionally, 10 seamounts (well-known = 1) were previously cataloged by Gevorgian et al. (2023) and Kim and Wessel (2011), but our method provided independent validation of their existence using direct multibeam observations. These seamounts, ranging from 315 to 2,005 meters in height, demonstrate that our approach can both confirm and refine existing seamount inventories through high-precision bathymetric measurements.

Interestingly, one feature predicted in previous seamount catalogs (well-known = 0) was not confirmed in our multibeam dataset. This discrepancy suggests a false positive in the satellite-derived data, potentially caused by noise, interpolation artifacts, or misclassification of other seafloor features as seamounts. Such cases highlight the importance of direct validation using high-resolution mapping to ensure the accuracy of global seamount databases.

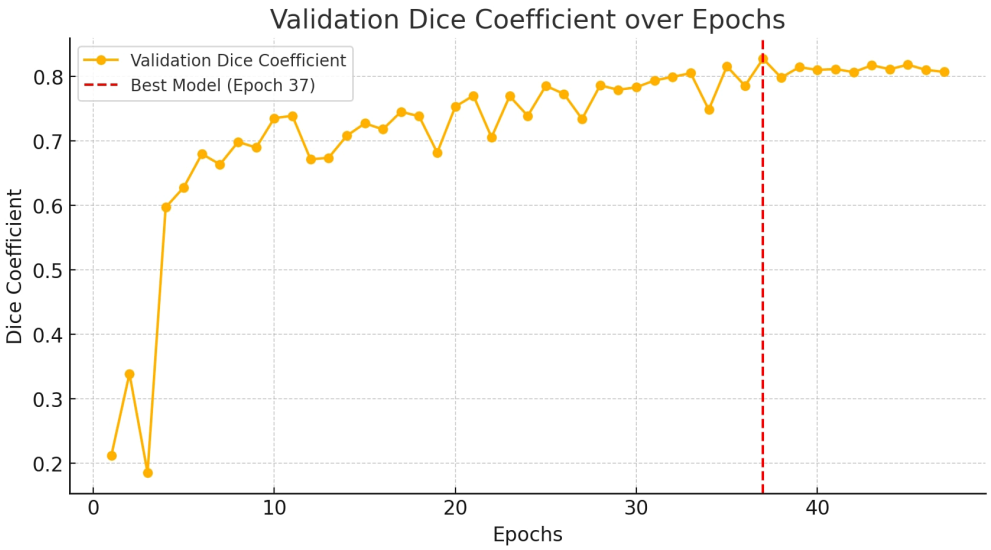


FIGURE 6
Validation Dice coefficient over epochs. A higher Dice coefficient indicates better segmentation performance. The red dashed line highlights the best model.

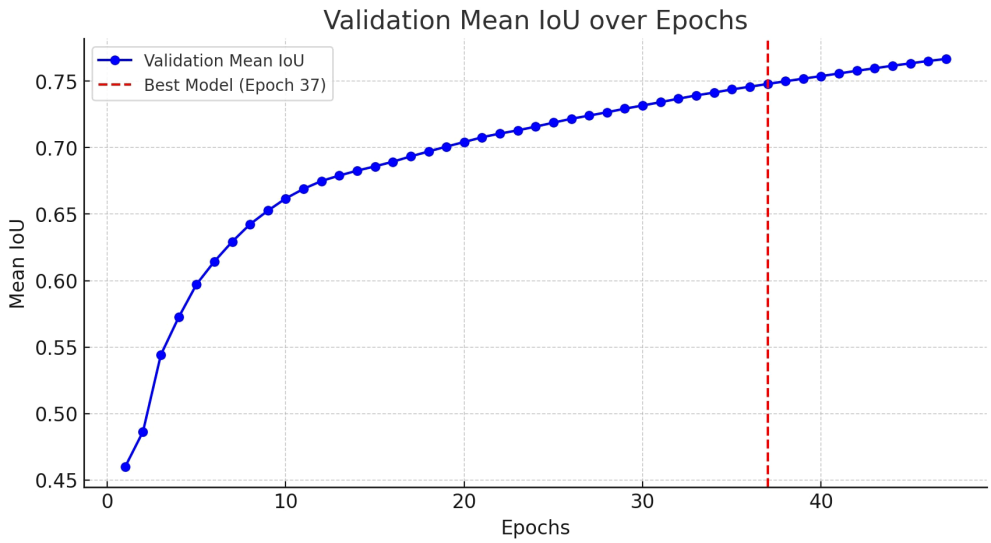


FIGURE 7
Validation mean Intersection over Union (IoU) over epochs. The IoU metric evaluates the overlap between predicted and ground truth segmentation masks, with higher values indicating better performance.

Overall, these findings emphasize the crucial role of multibeam sonar in capturing fine-scale seafloor topography and identifying small seamounts that remain undetected in satellite altimetry data. While altimetry-based methods provide valuable large-scale global coverage, they systematically underestimate the number of small seamounts due to resolution constraints. By applying machine learning-based segmentation on high-resolution bathymetry, our approach bridges the gap between broad-scale satellite surveys and precise, localized mapping techniques.

5 Conclusion

This study introduced a deep-learning-based framework for detecting small seamounts in multibeam bathymetric data, addressing key limitations of traditional classification methods and satellite altimetry. The proposed two-step approach—combining CNN-based filtering with U-Net segmentation—significantly improved detection accuracy and efficiency. The findings provide insights into each of the research questions posed in the introduction:

- How does a filtering-based approach improve the identification of small seamounts in multibeam bathymetric data compared to direct classification methods? The results demonstrated that CNNbased filtering enhances seamount detection by pre-selecting relevant image subsets, reducing noise and improving segmentation accuracy. Unlike direct classification methods, which attempt to classify entire images, the filtering process focuses only on regions likely to contain seamounts, reducing false positives and computational

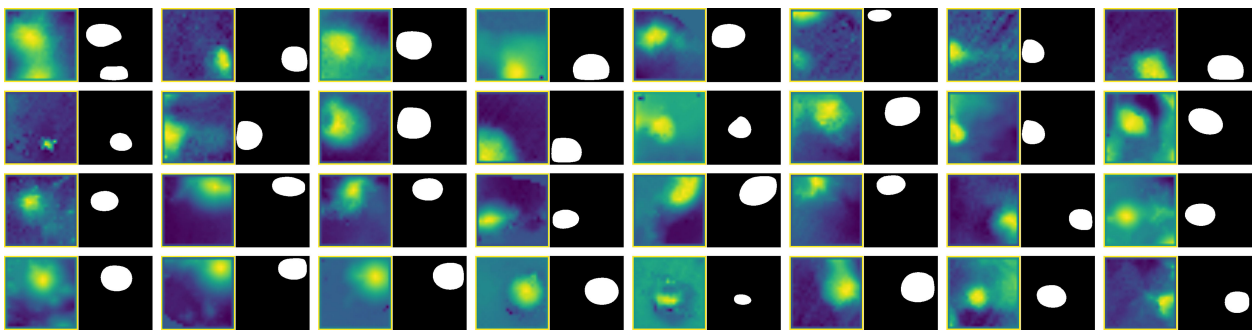


FIGURE 8
Comparison of manually selected images containing seamounts and their corresponding U-Net model predictions. The predicted seamounts closely align with the actual seamount locations, indicating the model's high detection performance.

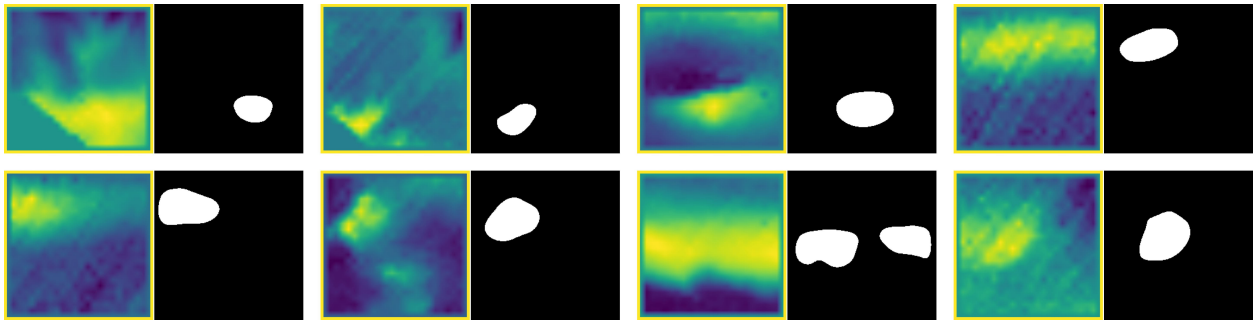


FIGURE 9
Examples of false predictions made by the U-Net model. In these cases, the model incorrectly labeled certain seafloor features as seamounts, likely due to local elevation changes or elongated structures that share some morphological characteristics with true seamounts. These misclassifications highlight the challenges in distinguishing small seamounts from other topographic variations in bathymetric data.

complexity. This two-step strategy outperformed traditional direct classification methods, ensuring that the segmentation model processes only meaningful data.

- What are the optimal hyperparameters for training a U-Net model to achieve the highest segmentation accuracy for small seamount detection? A grid search analysis identified the best-performing hyperparameter configuration: 32 filters, a kernel size of 5 × 5, a dropout rate of 0.1, a learning rate of 0.0001, and a batch size of 64. These settings balanced feature extraction depth, regularization, and training stability, yielding the highest segmentation accuracy, with a Dice Coefficient of 0.8274 and a Mean IoU of 0.7514. Models with excessively high filter counts or dropout rates exhibited overfitting or unstable convergence, highlighting the need for a balanced architecture in seamount segmentation tasks.
- How well does the proposed framework generalize across different geographic regions, and what limitations arise when applying a model trained in one ocean basin to another? While the model performed well on the SO305–2 dataset, cross-regional generalization remains a challenge. When tested on new datasets, the model maintained high accuracy for seamounts with well-defined topographic signatures, but performance declined in regions with

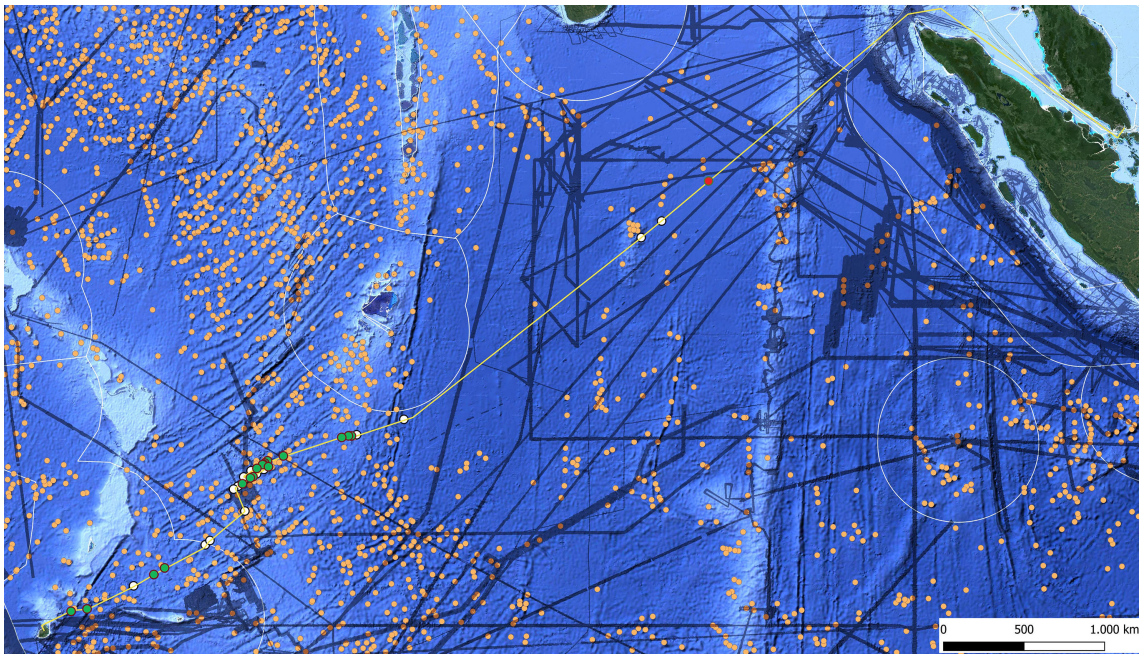


FIGURE 10
Map showing the 30 seamounts identified during the SO305/2 expedition. Seamounts detected by our model and also reported by Gevorgian et al. (2023) and Kim and Wessel (2011) are marked in white. Green dots indicate newly discovered seamounts that were not previously documented, while the red dot represents a location where a seamount was expected based on Gevorgian et al. (2023) and Kim and Wessel (2011), but no actual seamount was found. These findings highlight the effectiveness of multibeam systems in detecting previously unknown seamounts. The expedition started in Singapore (top right) and concluded in Mauritius (bottom left).

TABLE 4 Seamount characteristics including height, altimeter-derived height, well-known status, and coordinates.

Seamount ID	Height (m)	Altimeter (m)	Well-known	Longitude	Latitude
6	111	1078	2	70.788	-11.854
20	124	1260	2	66.558	-13.593
13	148	1430	2	66.968	-13.083
8	151	1578	2	70.373	-11.912
12	176	1755	2	67.013	-13.042
7	180	1888	2	70.663	-11.840
9	190	1833	2	67.860	-12.708
17	193	1969	2	67.224	-13.173
21	243	2410	2	66.444	-13.649
14	248	2532	2	66.724	-13.242
11	249	2429	2	67.063	-13.014
10	252	2584	2	67.177	-12.916
32	308	3046	2	58.736	-19.382
18	315	3164	1	67.022	-13.287
22	345	3434	2	66.102	-13.904
28	357	3641	2	62.755	-17.514
5	402	3953	1	71.005	-11.802
23	454	4567	1	65.905	-14.011
25	512	5097	1	66.207	-15.075
29	548	5555	2	62.293	-17.798
31	571	5704	2	59.426	-19.263
26	625	6308	1	64.703	-16.332
30	632	6274	1	61.429	-18.285
24	662	6667	1	65.720	-14.133
27	789	7880	1	64.504	-16.515
2	824	8317	1	84.125	-2.619
16	998	9927	1	66.325	-13.516
4	1091	10930	1	73.057	-11.147
3	1142	11413	1	83.245	-3.324
15	1402	13975	1	66.474	-13.341
19	2005	20124	1	66.710	-13.476
1	–	–	0		

highly variable seafloor morphology. This suggests that further fine-tuning or domain adaptation strategies may be necessary when applying the model to seamounts formed under different tectonic and geological conditions.

- To what extent can satellite altimetry reliably detect small seamounts, and how do its results compare to high-resolution multibeam bathymetric data? The results confirmed that satellite altimetry systematically underestimates the number of small seamounts due to resolution constraints. Of the 30 seamounts detected in the SO305–2 dataset, only 14 were visible in satellite-derived data, highlighting the importance of high-resolution multibeam bathymetry for capturing fine-scale seafloor features. Additionally, satellite-derived databases contained false positives, underscoring the need for direct validation using multibeam sonar.

Beyond geological applications, the automated detection of seamounts has broader implications for marine ecology, environmental monitoring, and plate tectonics research. Future work should focus on improving cross-regional generalization, integrating morphological priors into deep-learning models, and expanding the dataset to further enhance classification accuracy. In addition, improving the annotation process through inter-annotator validation or collaborative labeling could reduce subjectivity and improve the reliability of training labels, which is particularly important in seafloor datasets where feature boundaries can be ambiguous. In particular, the integration of domain adaptation or transfer learning methods holds promise for improving model performance in morphologically diverse or OOD seafloor regions, enabling broader applicability of the framework without requiring extensive manual relabeling or retraining. To further enhance robustness in real-world applications, future models should also incorporate strategies for handling OOD inputs, including uncertainty estimation and domain-specific priors to reduce prediction errors in unfamiliar seafloor environments. By bridging the gap between machine learning and marine geosciences, this framework contributes to the advancement of automated seafloor mapping and global seamount inventories, improving our understanding of the ocean floor.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://doi.pangaea.de/10.1594/PANGAEA.972385.

Author contributions

TZ: Formal analysis, Visualization, Investigation, Validation, Writing – review & editing, Software, Writing – original draft, Conceptualization, Methodology. CD: Investigation, Writing – review & editing, Validation, Supervision, Conceptualization. AK: Validation, Investigation, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. We acknowledge the financial support provided by MarDATA — Helmholtz School for Marine Data Science, Germany for this research project. We thank Captain Bjorn¨ Maaß and Captain Oliver Meyer and their crew for support at sea while collecting the training data set. I additionally want to thank the GEOMAR Open Access Publikationsfonds for providing support for the publication fee.

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.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. Solely for text correction and language refinement. All scientific content, analyses, and interpretations were developed by the author(s) without the aid of generative AI.

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

Alyahyan, S. (2025). A novel nemonet framework for enhanced rcc detection and staging in ct images. *Discov. Comput.* 28, 4. doi: 10.1007/s10791-025-09499-0

Amri, Y., M, Z., S, R., and Slama, A. B. (2025). Automatic glioma segmentation based on efficient u-net model using mri images. *Artif. Intelligence-Based. Med.* 11, 100216. doi: 10.1016/j.ibmed.2025.100216

Chamseddine, E., Tlig, L., and Sayadi, M. (2025). Gabrain-net: An optimized gabor-integrated u-net for multimodal brain tumor mri segmentation. *Brain Tumor. MRI. Segment. - SSRN*. doi: 10.2139/ssrn.5105664

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 801–818. doi: 10.1007/978-3-030-01234-249

Cherubini, C., Piazzolla, D., and Saccotelli, L. (2024). Habitat suitability modeling of loggerhead sea turtles in the central-eastern mediterranean sea: a machine learning approach using satellite tracking data. *Front. Mar. Sci* 11. doi: 10.3389/fmars.2024.1493598

Chitre, V., Nashipudmath, M. M., and Shinde, S. (2024). Exploring machine learning techniques for predictive analytics in computational mathematics. *Comput. Math. J.* doi: 10.52783/pmj.v34.i2.919

Clark, M. R., S, T., and Rowden, A. A. (2010). The ecology of seamounts: Structure, function, and human impacts. *Annu. Rev. Mar. Sci.* 2, 253–278. doi: 10.1146/annurev-marine-120308-081109

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. psychol. Measure.* 20, 37–46. doi: 10.1177/001316446002000104

Cracknell, M. J., and Reading, A. M. (2014). Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Comput. Geosci.* 63, 22–33. doi: 10.1016/j.cageo.2013.10.008

Deng, S., Ning, D., and Mayon, R. (2024). The motion forecasting study of floating offshore wind turbine using self-attention long short-term memory method. *Ocean. Eng.* 310, 2024. 127899. doi: 10.1016/j.oceaneng

Dwarakanath, B., and Kuntiyellannagari, B. (2025). Glioma segmentation using hybrid filter and modified african vulture optimization. *Bull. Electric. Eng. Inf.* 14, 2. doi: 10.11591/eei.v14i2.8730

Gevorgian, J., Sandwell, D. T., Yu, Y., Kim, S.-S., and Wessel, P. (2023). Global distribution and morphology of small seamounts. *Earth Space. Sci.* 10, e2022EA002331. doi: 10.1029/2022EA002331

He, K., Gkioxari, G., Dollár, P., and Girshick, R. B. (2017). "Mask r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (IEEE), 2961–2969. doi: 10.1109/ICCV.2017.322

He, K., S., R., Zhang, X., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. doi: 10.1109/CVPR.2016.90

Huang, D. X., W., G., W., X., W., W., and Sun, Y. F. (2024). Research on seamount substrate classification method based on machine learning. *Front. Mar. Sci.* 11. doi: 10.3389/fmars.2024.1431688

Iqbal, A., Sharif, M., Khan, M. A., Nisar, W., and Alhaisoni, M. (2022). Ff-unet: a u-shaped deep convolutional neural network for multimodal biomedical image segmentation. *Cogn. Comput.* 14, 1563–1579. doi: 10.1007/s12559-022-10038-y

Jones, D. O. B., D., A. B. B., and Simon-Lledó, E. (2021). Environment, ecology, and potential effectiveness of an area protected from deep-sea mining (clarion clipperton zone, abyssal pacific). *Prog. Oceanogr.* 197, 102653. doi: 10.1016/j.pocean.2021.102558

Khalil, S. M., Haywood, E., and Forrest, B. (2024). *Standard Operating Procedures for Geoscientific Data Management, Louisiana Sand Resources Database (LASARD)* (Louisiana: Tech. rep., Coastal Protection and Restoration Authority of Louisiana).

Kim, S.-S., and Wessel, P. (2011). New global seamount census from altimetry-derived gravity data. *Geophys. J. Int.* 186, 615–631. doi: 10.1111/j.1365-246X.2011.05076.x

Le Saout, M., Pałgan, D., Devey, C. W., Lux, T. S., Petersen, S., Thorhallsson, D., et al. (2023). Variations in volcanism and tectonics along the hotspot-influenced reykjanes ridge. *Geochem. Geophys. Geosyst.* 24, e2022GC010788. doi: 10.1029/2022GC010788

Liu, A., Liu, Y., Xu, K., Zhao, F., and Zhou, Y. (2024). Deepseanet: A bio-detection network enabling species identification in the deep sea imagery. *IEEE Trans. Geosci. Remote Sens.* 62, 1–13. doi: 10.1109/TGRS.2024.10415449

Manikandan, J., Harini, K., and Saranya, M. (2024). "Deep learning-based lung cancer detection and classification with hybrid sampling for imbalanced data," in *2024 International Conference on Smart Technologies (IEEE).*

Matabos, M., J, S., and Barreyre, T. (2022). Integrating multidisciplinary observations in vent environments (imove): Decadal progress in deep-sea observatories at hydrothermal vents. *Front. Mar. Sci.* 9. doi: 10.3389/fmars.2022.866422

Mitchell, N. C. (2001). Transition from circular to stellate forms of submarine volcanoes. *J. Geophys. Res.: Solid. Earth* 106, 1987–2003. doi: 10.1029/2000JB900263

Moradmand, H., and R, L. (2025). Multistage deep learning methods for automating radiographic sharp score prediction in rheumatoid arthritis. *Sci. Rep.* 15, 3391. doi: 10.1038/s41598-025-86073-0

Peng, M., Y., W., Q., M., T., W., and Li, J. (2025). Accurate and robust segmentation of cerebral distal small arteries by dvnet with dual contextual path and vascular attention enhancement. *Quant. Imaging Med. Surg.* 15, 2. doi: 10.21037/qims-24-1514

Qin, X., L, X., S, J., Z, D., Z, J., and Wu, Z. (2024). Musrfm: Multiple scale resolution fusion-based precise and robust satellite derived bathymetry model for island nearshore shallow water regions. *ISPRS. J. Photogramm. Remote Sens.* 128, 150–169. doi: 10.1016/j.isprsjprs.2024.09.007

Roelfsema, C. M., Lyons, M., Murray, N., and Phinn, S. R. (2021). Workflow for the generation of expert-derived training and validation data: A view to global scale habitat mapping. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.643381

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention – MICCAI *2015*. Eds. N. Navab, J. Hornegger, W. M. Wells and A. F. Frangi (Springer International Publishing, Cham), 234–241.

Sandler, M., M. Z., A. Z., Howard, A., and Chen, L.-C. (2018). "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4510–4520. doi: 10.1109/CVPR.2018.00474

Shen, Y., J., L., H., C., C., W., H., D., and Chen, L. (2025). Pads-net: Gan-based radiomics using multi-task network of denoising and segmentation for ultrasonic diagnosis of parkinson disease. *Med. Imaging Anal.* 120, 102490. doi: 10.1016/j.compmedimag.2024.102490

Simonyan, K., and Zisserman, A. (2015). "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations (ICLR).*

Sivakumar, M., Parthasarathy, S., and Padmapriya, T. (2024). Trade-off between training and testing ratio in machine learning for medical image processing. *PeerJ. Comput. Sci.* 10, e2245. doi: 10.7717/peerj-cs.2245

Smith, D. K. (1988). Shape analysis of pacific seamounts. *Earth Planet. Sci. Lett.* 90, 457–466. doi: 10.1016/0012-821X(88)90143-4

Srinivasan, S., Durairaju, K., Deeba, K., Mathivanan, S. K., Vijayakumar, M., and Murugan, P. (2024). Multimodal biomedical image segmentation using multi-dimensional u-convolutional neural network. *BMC Med. Imaging* 24, 1–17. doi: 10.1186/s12880-024-01197-5

Summers, G., Lim, A., and Wheeler, A. J. (2021). A scalable, supervised classification of seabed sediment waves using an object-based image analysis approach. *Remote Sens.* 13, 2317. doi: 10.3390/rs13122317

Szegedy, C., S., I.-J. S., Vanhoucke, V., and Wojna, Z. (2016). "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2818–2826. doi: 10.1109/CVPR.2016.308

Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., et al. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Trans. Med. Imaging* 35, 1299–1312. doi: 10.1109/TMI.2016.2535302

Usui, A., and S, K. (2022). Geological characterization of ferromanganese crust deposits in the nw pacific seamounts for prudent deep-sea mining. *Deep-Sea. Mining.: Sustainabil. Technol. Environ. Issues*. 81–113. doi: 10.1007/978-3-030-87982-24

Valentine, A. P., and Kalnins, L. M. (2013). Discovery and analysis of topographic features using learning algorithms: A seamount case study. *Geophys. Res. Lett.* 40, 3596–3600. doi: 10.1002/grl.50615

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

Yang, J., H., W., Y., L., Y., S., X., F., and Huang, S. (2025). A novel 3d lightweight model for covid-19 lung ct lesion segmentation. *Med. Eng. Phys.* 137, 104297. doi: 10.1016/j.medengphy.2025.104297

Yesson, C., Clark, M. R., Taylor, M., and Rogers, A. D. (2011). The global distribution of seamounts based on 30-second bathymetry data. *Deep. Sea. Res. Part I.: Oceanogr. Res. Papers.* 58, 442–453. doi: 10.1016/j.dsr.2011.02.004

Zhang, L., X, Z., and Zhang, W. (2025). Automatic liver tumor segmentation based on improved yolo-v5 and b-spline level set. *Cluster. Comput.* 28, 214. doi: 10.1007/s10586-024-04918-1

Zheng, Y., Tian, B., Yu, S., Yang, X., Yu, Q., and Zhou, J. (2025). Adaptive boundary-enhanced dice loss for image segmentation. *Biomed. Signal Process. Control.* 106, 107741. doi: 10.1016/j.bspc.2025.102526

Ziolkowski, T., Koschmider, A., and Devey, C. (2024). An optimized outlier detection function for multibeam echo-sounder data. *Comput. Geosci.* 186, 105572. doi: 10.1016/j.cageo.2024.105572

frontiers | Frontiers in Remote Sensing

TYPE Original Research
PUBLISHED 10 September 2025
DOI 10.3389/frsen.2025.1578841

OPEN ACCESS

EDITED BY
Chao Chen,
Suzhou University of Science and Technology, China

REVIEWED BY
Dušan P. Nikezić,
University of Belgrade, Serbia
Xinyue Yang,
Curtin University, Australia

*CORRESPONDENCE
Yuelin Xu,
m220200700@st.shou.edu.cn

RECEIVED 18 February 2025
ACCEPTED 28 July 2025
PUBLISHED 10 September 2025

CITATION
Zhu W, Xu Y, Zhang L, Liu Z, Liu S and Li Y (2025) A deep-learning framework to detect green tide from MODIS images.
Front. Remote Sens. 6:1578841.
doi: 10.3389/frsen.2025.1578841

A deep-learning framework to detect green tide from MODIS images

Weidong Zhu[1], Yuelin Xu[2]*, Lei Zhang[3], Zitao Liu[2], Shuai Liu[2] and Yifei Li[2]

[1]College of Oceanography and Ecological Science, Shanghai Ocean University, Shanghai, China, [2]Shanghai Engineering Research Center of Estuarine and Oceanographic Mapping, Shanghai Ocean University, Shanghai, China, [3]College of Surveying and Geo-Informatics, Tongji University, Shanghai, China

Introduction: Monitoring *Ulva prolifera* blooms over the long term is crucial for maintaining marine ecological balance. MODIS images, with their wide spatial coverage, high temporal resolution, and rich historical data, are commonly used for this purpose. However, their relatively low spatial resolution may lead to inaccuracies in precisely defining the bloom extents, thereby impeding the formulation of effective management strategies.

Methods: To address this issue, our study developed the WaveNet model. This model integrates VGG16 with the Bidirectional Feature Pyramid Network (BiFPN) and is further enhanced with a Convolutional Block Attention Module (CBAM). We applied this framework to MODIS imagery for the detection and monitoring of *U. prolifera*.

Results: WaveNet demonstrated superior performance in long-term sea surface *U. prolifera* monitoring compared to traditional methods, achieving an accuracy of 97.14% and an F1 score of 93.26%. This represents a significant improvement over existing techniques.

Discussion: These results highlight WaveNet's improved capacity for accurate spatial recognition and classification, overcoming the limitations of previous methods. Applying this approach, we analyzed the spatiotemporal distribution of *U. prolifera* blooms in the Yellow Sea of China from 2018 to 2024. Our framework offers valuable insights for early prevention and targeted management of green tides, contributing to the development of more effective mitigation strategies.

KEYWORDS

deep learning model, green tide detection, MODIS, satellite remote sensing, yellow sea

1 Introduction

Green tide blooms, particularly those caused by *U. prolifera* (*Ulva prolifera*) in the Yellow Sea, have become a major environmental issue, causing significant ecological and socio-economic impacts (Ye et al., 2011; Liu et al., 2013). These blooms are fueled by *U. prolifera*'s remarkable tolerance to high temperatures and intense light (Cui et al., 2015), which enables rapid and persistent growth. The decomposition of these algae releases harmful gases like hydrogen sulfide and ammonia, threatening marine ecosystems, human health, and coastal economies (Ye et al., 2011; Smetacek and Zingone, 2013). This study

aims to address this gap by developing a dynamic monitoring framework, which can serve as a foundation for constructing real-time monitoring systems for green tide blooms. Remote sensing, particularly satellite image analysis, offers a promising solution, but challenges remain in achieving a balance between spatial resolution, temporal coverage, and processing efficiency. This study aims to address this gap by developing a dynamic monitoring framework for real-time detection of green tide blooms. By leveraging advanced image-processing techniques, the study seeks to improve the accuracy, scalability, and efficiency of *U. prolifera* monitoring, providing valuable insights for timely intervention and sustainable management of green tides. This approach not only enhances monitoring but also contributes to the development of effective strategies for mitigating the environmental and socio-economic consequences of green tide blooms.

Satellite remote sensing, in contrast to field surveys, provides several advantages including wide coverage, rapid data acquisition, short update cycles, strong timeliness, and cost-effectiveness, rendering it an effective tool for monitoring and management of *U. prolifera* events (Hu et al., 2010; Hu et al., 2017). 10 m resolution Sentinel-2 imagery is suitable for monitoring smaller features like *U. prolifera* (Brisset et al., 2021), but its narrow swath width and coarse temporal resolution make it unsuitable for large areas like the Yellow Sea. Imagery from the Moderate Resolution Imaging Spectroradiometer (MODIS) has significantly advanced the assessment and prediction of algal bloom mechanisms (Lee et al., 2011; Cao et al., 2019; Hu et al., 2019; Xing et al., 2019). From May to June each year, *U. prolifera* blooms rapidly spread across the Yellow Sea. MODIS imagery, with its near-daily updates and 2,330 km2 coverage, effectively monitors the entire lifecycle of these blooms. However, the coarse resolution of MODIS images, with a maximum spatial resolution of only 250 m, introduces a degree of error in the extracted estimates of algal biomass (Hu et al., 2010; Hu et al., 2015). Minimizing this extraction error has emerged as a bottleneck in optical remote sensing for algae detection.

Various remote sensing threshold methods are used to extract *U. prolifera* information, utilizing the unique spectral characteristics of green algae in visible and infrared bands. Common approaches include the Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Algae Index (NDAI) (Shi and Wang, 2009), applicable across multiple satellite sensors. Other methods, such as the Floating Algae Index (FAI) (Hu, 2009), Virtual-baseline Floating Macro Algae Height index (VB-FAH) (Xing and Hu, 2016), and RGB Floating Algae Index (Jiang et al., 2020), are robust to environmental variations, including thin cloud cover (Xu et al., 2016). However, these optical-based methods are hindered by challenges such as cloud interference, variable backgrounds, and the need for meticulous threshold selection, which often requires expert knowledge (Shi and Wang, 2009; Hu et al., 2010).

While deep learning methods show promise in overcoming these limitations (Schmidhuber, 2015; Li et al., 2020), existing studies, such as those utilizing ERISNet for Sargassum algae extraction in the coastal waters of Mexico and approaches employing AlexNet for large algae extraction from UAV images, have made progress in specific environments but still face challenges in achieving large-scale and accurate monitoring of green tides (Arellano-Verdejo et al., 2019; Wang et al., 2019). Recent advancements, like models designed to detect green tide information from both SAR and optical images, highlight the potential of deep learning in this domain, paving the way for more accurate, scalable, and efficient monitoring (Gao et al., 2022). This study aims to advance the application of deep learning for dynamic monitoring of *U. prolifera*, addressing the gap in real-time, large-scale, and precise green tide detection. It also focuses on improving the accuracy of *U. prolifera* extraction from low-resolution satellite imagery and enabling dynamic daily monitoring of green tides on a large scale.

The objectives of this paper includ 1) developing of a deep learning network to more effectively extract information about green tide from coarse-resolution optical imagery; 2) implementing of large-scale dynamic monitoring of green tide; and 3) extracting and analysing of the spatiotemporal distribution changes of green tide outbreaks in the Yellow Sea region from 2018 to 2024 on both interannual and intermonthly scales. The organization of the paper is as follows. Section 2 presents the study area and related datasets, including optical MODIS data, Sentinel-2 data, and the training dataset for the deep learning model. Section 3 introduces the proposed deep learning network model, encompassing physical model optimization and model performance verification methods. Section IV details the training of the model and the research findings. Discussions and conclusions are presented in Sections 4 and 5.

2 Study area and datasets

2.1 Study area

The study area, situated within the Yellow Sea between 32°N and 37°N and 119°E–124°E, is shown in Figure 1. Influenced by the East Asian monsoon, the climatic regime of the region under study is characterized by cold, arid winters and warm, humid summers (Xing and Hu, 2016; Qi et al., 2017; Zhang et al., 2019). The confluence of these climatic conditions with substantial terrestrial influences results in the Yellow Sea exhibiting moderate to high levels of turbidity, which are characteristic of the region (Shi and Wang, 2009; Zhang et al., 2010; Xing et al., 2019). These environmental parameters significantly influence the proliferation of *U. prolifera*, as its growth dynamics are intricately tied to water temperature and nutrient availability. Since the onset of the 21st century, *U. prolifera* has exhibited periodic summer blooms in the region, with each event demonstrating extensive areal coverage, substantial biomass accumulation, and significant long-distance transportation (Wang et al., 2015). These phenomena have had profound negative repercussions on the coastal tourism industry, aquaculture activities, and the integrity of the ecological environment, underscoring the urgency and relevance of our research in real-time, large-scale, and precise green tide detection.

Given the influence of the East Asian monsoon, the climate in this region is marked by cold, dry winters and hot, humid summers (Xing and Hu, 2016; Qi et al., 2017; Zhang et al., 2019). These climatic conditions, coupled with the significant terrestrial impact, contribute to the Yellow Sea's moderate to high turbidity levels, which are typical of the area (Shi and Wang, 2009; Zhang et al., 2010; Xing et al., 2019). Due to these geographical and climatic reasons,

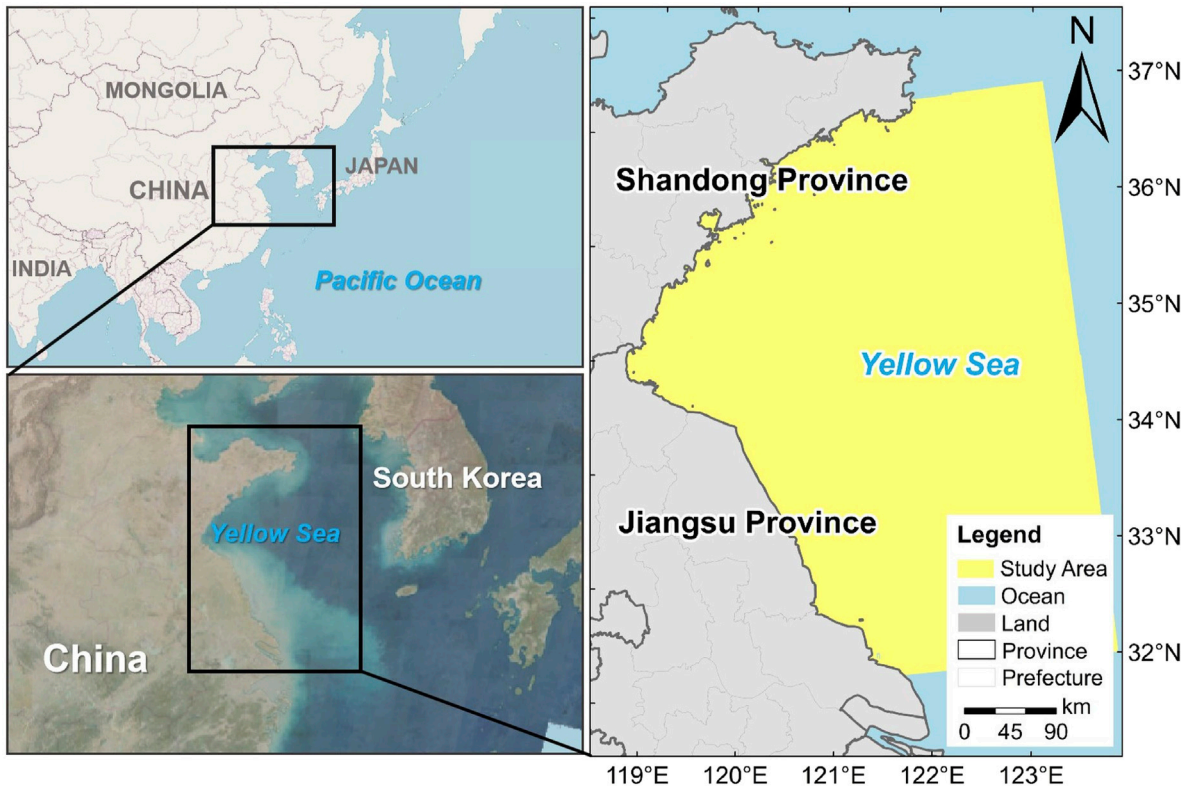


FIGURE 1
Location of the study area.

the proliferation of *U. prolifera*, a green tide-forming macroalga, is influenced, as its growth is closely linked to water temperature and nutrient availability. Understanding these environmental factors is crucial for this study, as they provide insights into the conditions that may favor or inhibit the development of *U. prolifera* blooms, which are the focus of our research. Since 2007, *U. prolifera* has periodically erupted in this region every summer. Its characteristics, such as broad coverage, large biomass, and extensive distance transport (Wang et al., 2015), have severely impacted coastal tourism, aquaculture, and the ecological environment.

2.2 Datasets

The MODIS satellite, initiated by NASA in 1999, is a prominent space remote sensing instrument, providing surface spectral reflectance estimates for 36 bands every 1–2 days. Research indicates that the peak *U. prolifera* period in the South Yellow Sea spans from May to August annually (Zhou et al., 2021). To extend monitoring to March, EOS MODIS 1B (Terra/Aqua) remote sensing dataset from March to August 2018–2024 were chosen from NASA's data repository (https://ladsweb.modaps.eosdis.nasa.gov/search). This selection included MOD02QKM and MOD02HKM products with resolutions of 250 m and 500 m, respectively. Among126 images, those with minimal cloud cover during *U. prolifera* blooms from 2018 to 2024 were selected. Data were processed using SNAP software for reprojection, calibration, and band synthesis, with MOD02HKM resampled to 250 m. Subsequently, sea-land separation was conducted to extract relevant sea areas.

Sentinel-2, part of the European "Copernicus" program, consists of two satellites, Sentinel-2A (launched 23 June 2015) and Sentinel-2B (launched 27 March 2017). These satellites operate on a sun-synchronous orbit with individual revisit periods of 10 days and a collective revisit period of 5 days. Sentinel-2 Level-2A (L2A) dataset, comprising atmospherically corrected bottom-of-atmosphere reflectance dataset, were acquired from the European Space Agency's Copernicus Open Access Hub (https://scihub.copernicus.eu/dhus/#/home). Cloud-free images with 10-m resolution overlapping with the MODIS data dates in the study area were selected. Red (R), Green (G), and Blue (B) bands were utilized to generate true-color composite images for subsequent *U. prolifera* extraction validation.

In this research, MODIS images were selected to create a dataset. Initially, bands 1 (red), 2 (near-infrared), and 4 (green) were chosen, corresponding to R, G, and B channels, respectively, to generate false-color composite images. Subsequently, these images were segmented into 512*512-pixel tiles using a sliding window approach. Within the MODIS images, *U. prolifera* exhibits more prominent green patches compared to seawater. Therefore, Lableme software was employed to label the segmented images with *U. prolifera* samples. Out of 608 sets of MODIS images and corresponding labels, 425 sets were allocated for training, and 183 sets were reserved for testing. During training, the dataset was divided into 70% for training and 30% for validation.

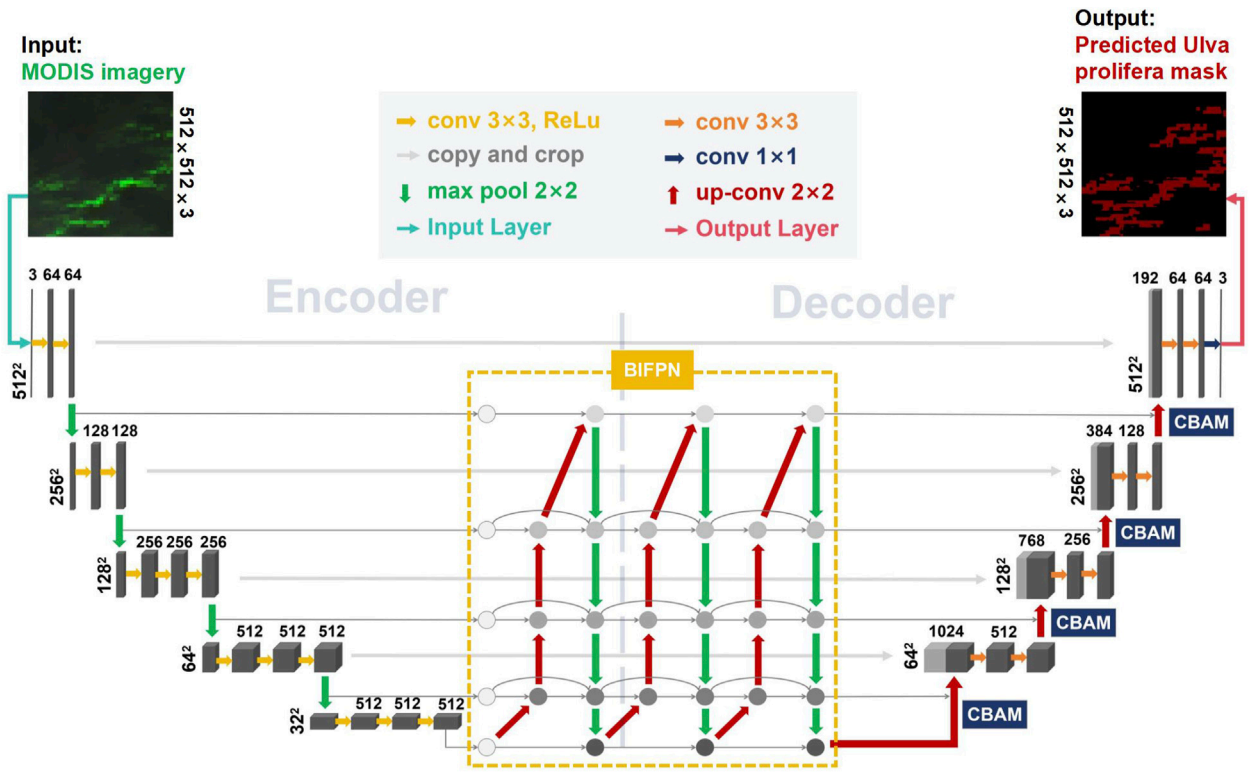


FIGURE 2
The structure of the WaveNet.

3 Methods

3.1 The structure of the WaveNet model

Figure 2 illustrates the architecture of the proposed model, which integrates VGG16, BiFPN, and CBAM in a collaborative hierarchy.

Firstly, the Visual Geometry Group 16-layer model (VGG16) is adopted as the backbone feature extraction network. It transforms the input MODIS image into multi-level feature maps through four convolutional and downsampling stages, effectively capturing both low-level textures and mid-level semantic patterns. Secondly, three layers of Bidirectional Feature Pyramid Network (BiFPN) are incorporated to enhance multi-scale feature fusion. BiFPN enables bidirectional information flow, allowing high-level semantic information from deep layers to guide low-level spatial details, and *vice versa*. This preserves fine-grained localization critical for identifying *U. prolifera* boundaries. Thirdly, four upsampling stages are applied to restore the spatial resolution of feature maps to match the original image dimensions. At each stage, a Convolutional Block Attention Module (CBAM) is introduced to emphasize the most relevant spatial regions and spectral channels. CBAM refines features by applying sequential channel and spatial attention, thereby improving feature saliency and reducing background noise.

This hierarchical design enables VGG16 to focus on core visual patterns, BiFPN to integrate information across scales, and CBAM to selectively enhance discriminative features. Together, they collaboratively improve the model's accuracy in detecting green tide areas under complex oceanographic conditions.

3.2 Visual Geometry Group 16-layer model, VGG16

In our study, VGG16 was chosen as the backbone network due to its proven effectiveness in image feature extraction, particularly in tasks requiring high accuracy and localization precision, such as the ILSVRC-2014 ImageNet challenge. Its architectural design, which includes five sets of convolutional layers followed by max-pooling layers and three fully connected layers, allows for efficient feature representation and nonlinearity enhancement while preserving the perceptual field. The use of 3 × 3 convolutional kernels and 2 × 2 max-pooling layers increases the network depth, enabling the detection of intricate patterns critical for identifying *U. prolifera* in complex marine environments. VGG16 offers a deeper structure with more precise feature extraction, which is crucial for achieving the high accuracy (97.14%) and F1 score (93.26%) demonstrated in our green tide monitoring framework. This integration provides a robust foundation for the dynamic monitoring of *U. prolifera*, outperforming previous methods in large-scale green tide detection and classification.

3.3 Bidirectional feature pyramid network, BiFPN

The traditional VGG16 architecture fails to effectively utilize multiscale information from the backbone network, as it directly connects to fully connected layers after the fifth convolutional layer.

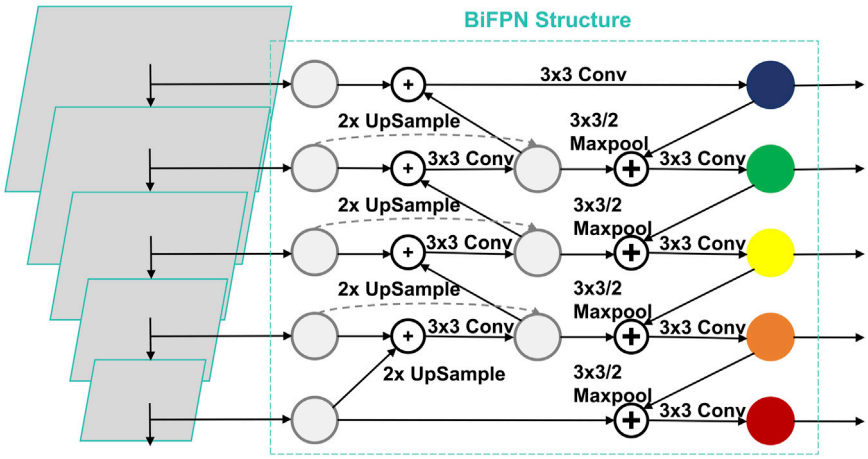


FIGURE 3
The structure of BiFPN.

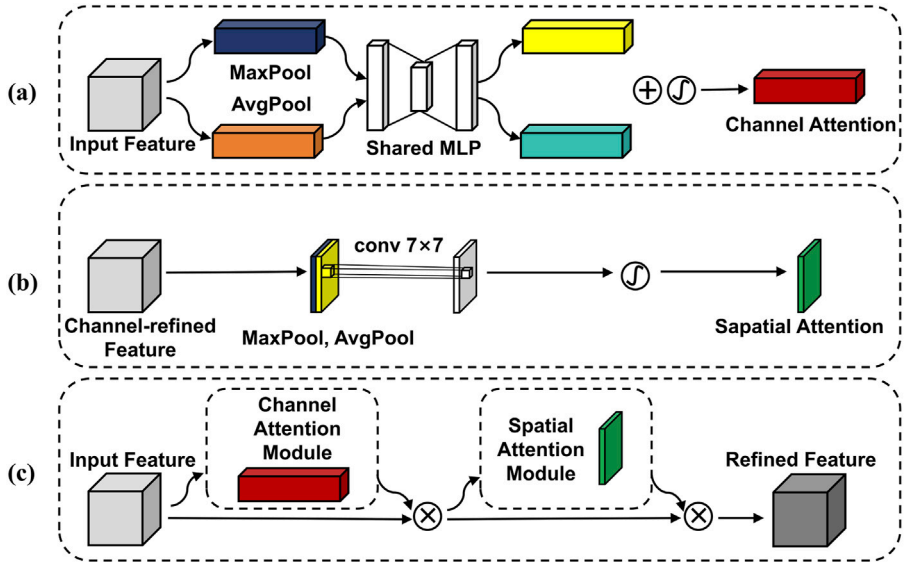


FIGURE 4
Structure of CBAM: **(a)** Channel attention module; **(b)** Spatial attention module; **(c)** CBAM.

To remedy this, we incorporate the BiFPN module, which processes shallow features and integrates multiscale information. BiFPN, introduced by Google in 2020 within the EfficientDet model (Tan et al., 2020), employs a weighted bidirectional feature pyramid network. This network consists of top-down and bottom-up pathways, enabling the propagation of both semantic and positional information, as shown in Figure 3.

To address the challenge of accurately classifying *U. prolifera* at the pixel level, we introduced a novel modification to the VGG16 architecture. Specifically, instead of relying on the traditional fully connected and softmax output layers, we replaced them with the fusion results generated by the BiFPN structure. This approach utilizes attribute maps extracted from multiple stages of the backbone network, capturing spatial and contextual information at different resolutions. By adaptively integrating these multi-scale features, the BiFPN structure enhances the model's ability to preserve fine-grained details and resolve ambiguities in areas with similar spectral characteristics. This modification significantly improves the network's feature representation capabilities, ensuring better accuracy in distinguishing *U. prolifera* from surrounding elements.

Mathematically, the fusion process in BiFPN can be expressed through Equations 1 and 2 as follows:

$$F_{out} = \sum_{i=1}^{n} \omega_i F_i \tag{1}$$

$$\omega_i = \frac{exp(\alpha_i)}{\sum_{j=1}^{n} exp(\alpha_j)} \tag{2}$$

where F_i is the feature map at the i-th scale from VGG16; ω_i is the learned attention weight for scale i; α_i is a trainable scalar associated with each input scale. This formulation ensures that features contributing most to discrimination are emphasized in the final output.

3.4 Convolutional Block Attention Module, CBAM

Attention mechanisms allow the network to focus on the most relevant features of the target and have been extensively capabilities of convolutional neural networks (Lian et al., 2018), and improve feature extraction efficiency and accuracy (Ma et al., 2023). In our work, attention mechanisms are central to enhancing feature deployed in deep learning applications, like natural language processing and visual recognition, to enhance the learning extraction and representation, addressing the challenge of accurately identifying *U. prolifera* in MODIS imagery. We employ the Convolutional Block Attention Module (CBAM), which integrates both channel and spatial attention mechanisms to refine feature representations dynamically, and its structure is illustrated in Figure 4. The channel attention component aggregates information across feature map channels, highlighting the most relevant spectral features for distinguishing *U. prolifera*. Simultaneously, the spatial attention mechanism focuses on critical spatial regions within each channel, enabling the model to capture localized patterns associated with green tides.

To further enrich the model's feature extraction capacity, we integrate CBAM within the last four upsampling layers of the network. By processing input feature maps and applying attention mechanisms, CBAM outputs weighted feature maps, emphasizing both channel and spatial information. This dual-focus strategy ensures the preservation of spectral and spatial nuances, significantly improving classification precision. Such an approach is particularly effective given the moderate spatial resolution and complex spectral characteristics of MODIS imagery, providing a robust framework for green tide detection.

3.5 Accuracy assessment

The *U. prolifera* extraction method underwent evaluation using standard metrics: accuracy, precision, recall, F1 score, mIoU, and mPA, with their calculation formulas detailed in Equation 3. Results were classified into four groups: True Positive (TP) for accurately identified *U. prolifera* pixels, True Negative (TN) for accurately classified background pixels, False Positive (FP) for background pixels erroneously identified as *U. prolifera*, and False Negative (FN) for *U. prolifera* pixels mistakenly classified as background. Manual determination of the true value was based on MODIS false-color images.

$$\begin{cases} Accuracy = (TP+TN)/(TP+TN+FP+FN) \\ Precision = (TP+FP) \\ Recall = TP/(TP+FN) \\ F1 = 2*(Precision*Recall)/(Precision+Recall) \\ \mathrm{mIoU} = 1/(k+1)*\sum_{i=0}^{k}\left[p_{ii}/\left(\sum_{i=0}^{k}p_{ij}+\sum_{i=0}^{k}p_{ji}-p_{ii}\right)\right] \\ \mathrm{mPA} = 1/(k+1)*\sum_{i=0}^{k}\left(p_{ii}/\sum_{i=0}^{k}p_{ij}\right) \end{cases} \quad (3)$$

where denotes predicting i as j, which is a false negative (FN); denotes predicting j as i, which is a false positive (FP); denotes predicting i as i, which is a true positive (TP).

3.6 Estimation of *Ulva prolifera* area

Area of *U. prolifera* depicts the ground area size, which can be computed by the product of spatial resolution and the corresponding number of pixels (Cui et al., 2018), as outlined in Formula 4.

$$Area_{\_GT} = PS * N_{\_GT} \quad (4)$$

where, $\mathrm{Area}_{\_GT}$ represents the area of *U. prolifera* in km^2; PS represents the ground area size corresponding to one pixel of satellite imagery in km^2; $N_{\_GT}$ represents the number of detected *U. prolifera* pixels.

4 Experiments and results

4.1 Training and experimental settings

The deep learning tasks were performed on a Windows 10 system equipped with an NVIDIA GeForce RTX 3060Ti GPU boasting 8 GB of storage. CUDA version 11.6 was utilized, alongside the PyTorch 11.0 deep learning platform for model construction. The software environment was Anaconda (Python 3.8). Throughout the training phase, the Adam optimization algorithm (Ronneberger et al., 2015) dynamically adjusted the network weights and biases. The parameters are set as follows: $\beta_1 = 0$, $\beta_2 = 0.99$. The learning rate (α) of the network is initialized to 0.001, and after every 40 epochs (with a total training epoch limit set to 250), α is multiplied by a decay factor of 0.1 to reduce the parameter search space.

4.2 Evaluation of model performance

This paper evaluates the detection performance of *U. prolifera* using the WaveNet deep learning model in comparison with the Normalized Difference Vegetation Index (NDVI) and the Adjusted Floating Algae Index (AFAI) methods, both widely applied in algae detection tasks.

The NDVI method, proposed by Rouse et al., leverages the characteristic spectral reflectance of vegetation in the near-infrared and red bands. Its adaptability to large floating algae, due to their spectral similarities with vegetation, makes it a commonly used approach for algae extraction. The calculation formula for NDVI is given in Equation 5.

$$NDVI = (R_{rc,NIR} - R_{rc,RED})/(R_{rc,NIR} + R_{rc,RED}) \quad (5)$$

where, and represent the reflectance of the near-infrared band (860 nm) and the red band (660 nm), respectively.

The AFAI method is designed to reduce the impact of atmospheric effects, thin clouds, and moderate solar glint. It employs a linear baseline between adjacent bands to compute near-infrared reflectance (Fang et al., 2018). It employs a linear

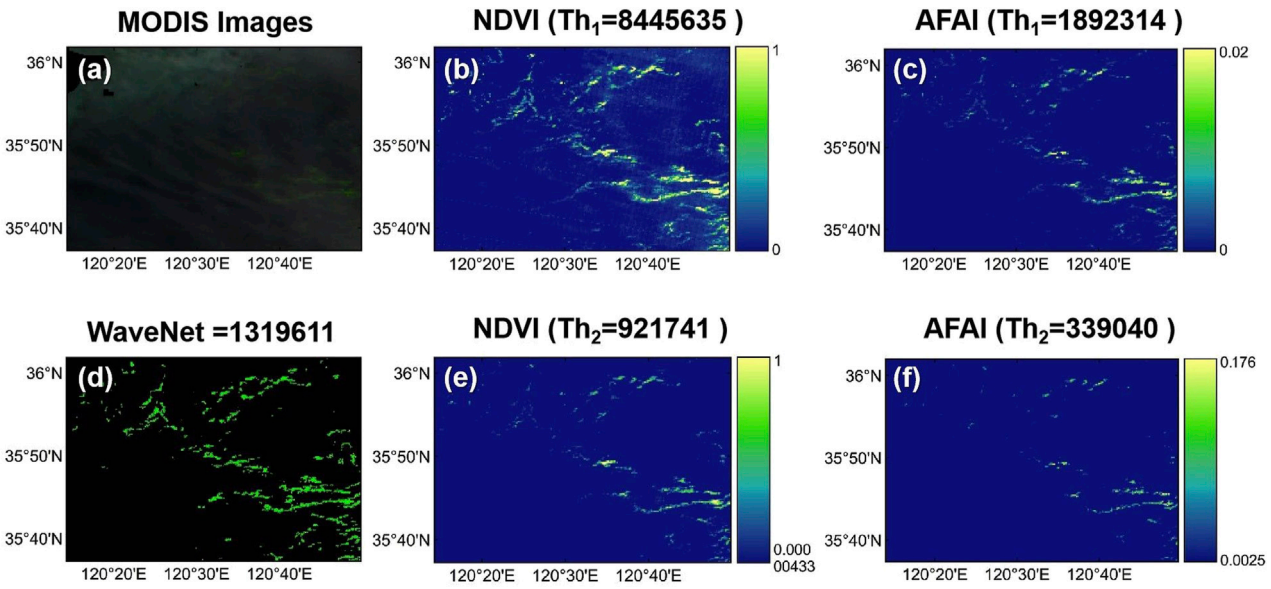


FIGURE 5
Comparison result of the WaveNet model and traditional biological index methods. **(a)** MODIS true color; **(b,e)** NDVI results; **(c,f)** AFAI results; **(d)** WaveNet result.

baseline between adjacent bands to compute near-infrared reflectance. The formulas are given in Equations 6, 7:

$$AFAI = R_{rc,NIR} - R'_{rc,NIR1} \quad (6)$$

$$R'_{rc,NIR1} = R_{rc,RED} + (R_{rc,NIR2} - R_{rc,RED}) * (\lambda_{NIR1} - \lambda_{RED})/(\lambda_{NIR2} - \lambda_{RED}) \quad (7)$$

where, $R_{rc,NIR1}$, $R_{rc,RED}$, $R_{rc,NIR2}$ represent the reflectance in the near-infrared band (748 nm), the red band (667 nm), and the long-wave near-infrared band (869 nm), respectively.

Both NDVI and AFAI have demonstrated effectiveness in detecting floating algae using satellite imagery, such as MODIS and Landsat. However, their accuracy is limited by manual threshold selection. In this study, the WaveNet model, trained on MODIS imagery, demonstrated superior performance in *U. prolifera* detection. Unlike NDVI and AFAI, the fixed threshold in WaveNet (0.5) relies on the model's optimized weights, eliminating manual adjustments (Liu et al., 2009; Qi et al., 2016a; Hu et al., 2019; Zheng et al., 2022). Results show that WaveNet not only reduces threshold dependency but also achieves significantly higher precision and coverage accuracy, highlighting its potential for dynamic green tide monitoring.

Figure 5a shows the MODIS true color image of the Yellow Sea from 5 July 2023, where *U. prolifera* appears in light green. Our method, alongside the two index methods (NDVI and AFAI), confirmed that these colored patches are floating *U. prolifera* Figures 5b-f. Through our deep learning model, 1,319,611 algal pixels were identified. When using thresholds of >0 and >0.00000433, the NDVI index method identified 8,445,635 and 921,741 algal pixels, respectively. With thresholds of (0, 0.02) and (0.0025, 0.0176), the AFAI index method identified 1,892,314 and 339,040 algal pixels, respectively. Due to differences in threshold selection, both NDVI and AFAI methods exhibit considerable uncertainty; in fact, the algal identification results could vary by orders of magnitude (Hu, 2009; Liu et al., 2009; Xu et al., 2014; Qi et al., 2016b; Hu et al., 2019). In contrast, the deep learning-based model mitigates the potential bias introduced by selecting different extraction thresholds for NDVI or AFAI.

TABLE 1 Accuracy evaluation of the extraction effect.

Models	Precision	Recall	F1-score	Accuracy
WaveNet	92.83	93.69	93.26	97.14
NDVI (Th_2)	82.37	88.48	85.32	90.55
AFAI (Th_1)	87.92	90.02	88.96	92.76

The WaveNet model achieved precision and recall metrics, as well as a comprehensive F1 score for *U. prolifera* extraction, all exceeding 90.0%. The accuracy reached 97.14%, which is 7.6% higher on average compared to the NDVI method and 4.3% higher on average compared to the AFAI method. In summary, our method excels at extracting *U. prolifera* from MODIS images, achieving the highest recognition accuracy. The outcomes are provided in Table 1.

Four areas were randomly selected to compare the segmentation results of the WaveNet, NDVI, and AFAI methods (Figure 6). In addition to MODIS false-color images, Sentinel-2 true-color images with a resolution of 10 m were added as references. Since MODIS images have a resolution of 250 m, the *U. prolifera* patches derived using these methods will appear larger than those in the Sentinel-2 images. Although the selected reference images were taken on the same day, slight differences in *U. prolifera* patches may occur due to different transit times.

In Region 1 (R1), the aim was to compare the extraction performance of *U. prolifera* over a large area, while Regions 2

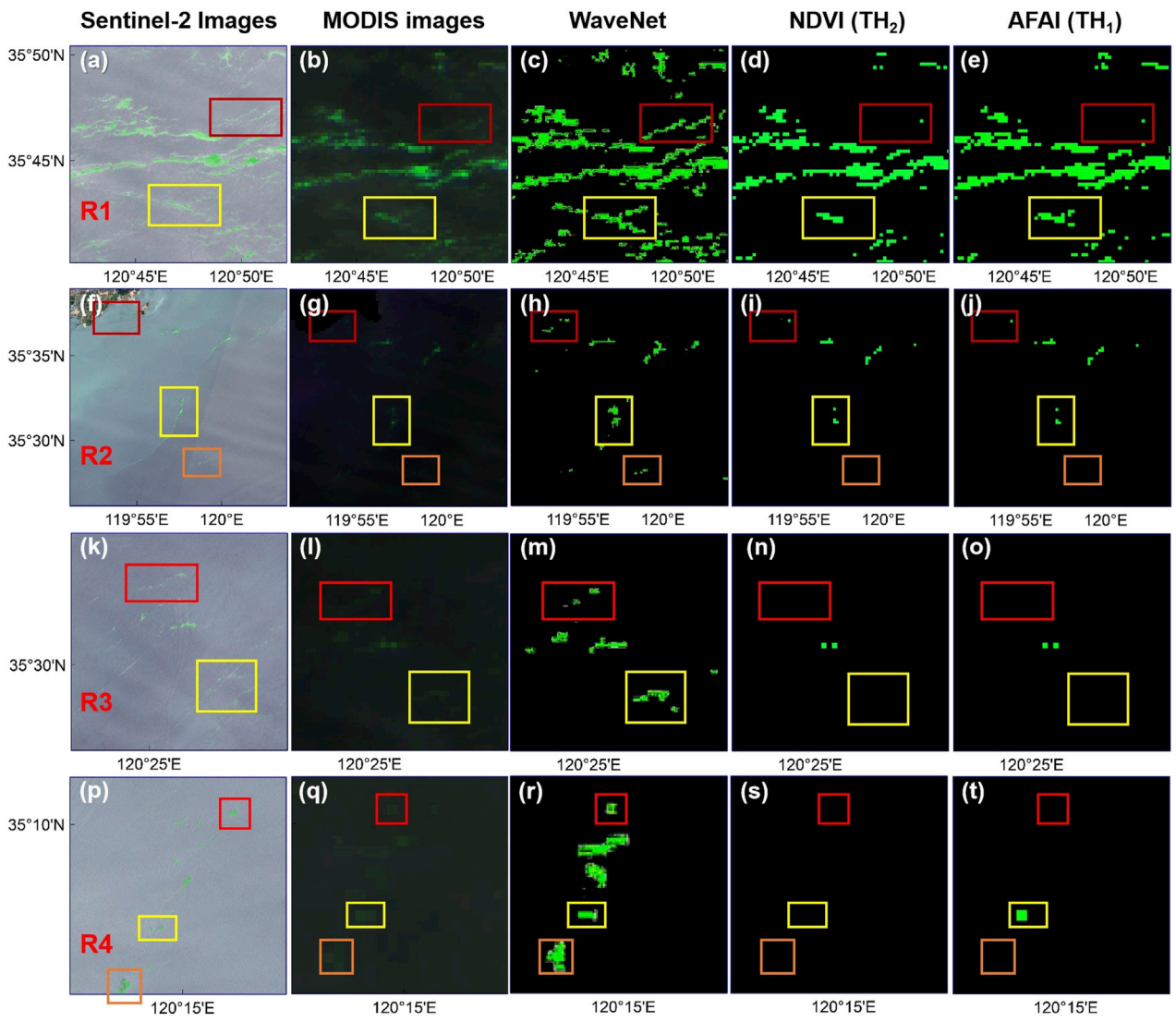


FIGURE 6
The comparison of *U. prolifera* extraction results. R1-R4 regions: **(a,f,k,p)** Sentinel-2; **(b,g,l,q)** MODIS; **(c,h,m,r)** WaveNet; **(d,i,n,s)** NDVI (TH_2); (e,j,o,t) AFAI (TH_1)

(R2), 3 (R3), and 4 (R4) focused on smaller areas. In the upper and lower parts of R1, both the NDVI and AFAI methods exhibited instances of under-segmentation, while the WaveNet method provided the most complete extraction of *U. prolifera*. However, in R2, both the NDVI and AFAI methods had instances of under-segmentation in the nearshore area on the upper left, and the WaveNet method showed misclassification in the central part.

In R3, both the NDVI and AFAI methods exhibited a significant amount of under-segmentation. In R4, the extraction performance of all three methods was relatively poor, with instances of under-segmentation in the NDVI and AFAI methods, and misclassification in the WaveNet model. As shown in Table 2, the pixel count and area of extracted *U. prolifera* using different methods were also compared, with the same conclusions as depicted in Figure 6.

5 Discussion

5.1 Performance evaluation of different composite models

Group 1: VGG16+(BiFPN + SA). This group integrates the Spatial Attention (SA) mechanism into the BiFPN module within the VGG16 framework. The design focuses on enhancing spatial perception, improving feature fusion efficiency, and minimizing information loss. However, as shown in the heatmap comparison (Figure 7), the performance of this configuration remains limited. The mean Intersection over Union (mIoU) reaches only 86.15%, and the F1 score achieves 92.36%, both of which are lower than those of attention-enhanced

TABLE 2 The comparison of pixel count and area of extract.

Method	Pixel number of floating *U. prolifera* blooms				A_{sar}/km^2			
	R1	R2	R3	R4	R1	R2	R3	R4
WaveNet	533448	34977	56010	123146	333.44	21.86	35.01	76.97
NDVI (Th_2)	268537	10665	4,233	0	167.84	6.67	2.66	0
AFAI (Th_1)	315447	11159	4,233	8,283	197.16	6.97	2.66	5.18

Note: A_{sar} represents distribution area of floating *Ulva prolifera* blooms.

dual-dimensional models. This indicates that spatial-only attention mechanisms may be insufficient for robust green tide discrimination and comprehensive feature preservation.

Group 2: VGG16+(BiFPN + CBAM). This model group substitutes the SA mechanism with CBAM while maintaining other aspects unchanged to compare the two attention mechanisms. SA emphasizes spatial information, while CBAM integrates channel and spatial attention, enhancing the model's perception of both types of information. This adjustment aims to enhance overall model performance. As depicted in Figure 7, the replacement leads to slight improvements in mIoU, Precision, F1 score, and Accuracy, ranging from 0.5% to 1%.

Group 3: VGG16+skip (BiFPN + CBAM). This group introduces skip connections to directly transfer low-level features into upper layers, theoretically improving representation comprehensiveness. Nevertheless, the model shows decreased accuracy compared to Group 2. Specifically, the mIoU drops to 83.69%, and the F1 score declines to 90.75%. These results may stem from redundant or noisy features introduced via the skip paths, which interfere with semantic abstraction and reduce final prediction quality.

Group 4: VGG16+[BiFPN + dual (CBAM + SA)]. This model incorporates both CBAM and SA as a dual-attention mechanism. Although it expands the model's attention diversity, the additional complexity leads to performance degradation. As shown in Figure 7, the mIoU is only 83.50%, and the F1 score is 90.61%. This suggests that overly complex attention fusion may introduce conflicts or overfitting, limiting the effectiveness of feature integration.

Group 5: VGG16 + 3*BiFPN + CBAM. As the proposed WaveNet configuration, this group employs a triple BiFPN structure for deep multi-scale fusion and integrates CBAM during the upsampling stages. According to the comparative heatmap (Figure 7), this model achieves the highest overall performance: the mIoU reaches 87.79%, and the F1 score improves to 93.26%, with an accuracy of 97.14%. These results confirm that deeper fusion layers and attention refinement significantly enhance both feature preservation and perceptual discriminability

5.2 Monthly spatial-temporal distribution characteristics of *Ulva prolifera*

Based on MODIS remote sensing satellite imagery, the spatial coverage area and impact scope of *U. prolifera* were extracted for different years from 2018 to 2024 (Figure 8).

In late May and early June 2018, *U. prolifera* was first detected in the shallow waters of northern Jiangsu Province, China. It then drifted northeastward, affecting the coastal waters of the northern Yellow Sea. Initially, its coverage area was only 80 km^2, but within 11 days, it sharply increased to 164 km^2. Subsequently, *U. prolifera* drifted northward, accumulating extensively in the coastal areas of the northern Yellow Sea by the end of June, reaching its maximum coverage area and impact range. On July 14, *U. prolifera* extensively landed in coastal cities in the northern Yellow Sea. Gradually, its coverage area decreased and disappeared, with only sporadic patches remaining in the region by July 18.

In 2019, *U. prolifera* was first spotted in the southeastern Yellow Sea on May 9, covering an area of 14 km^2. Towards the end of May, it appeared in the shallow waters off the coast of northern Jiangsu Province, China, before drifting eastward and merging with the existing *U. prolifera* in the southeastern Yellow Sea, then moving northward. By June 23, it reached its peak coverage area of 2,127 km^2. In early July, *U. prolifera* landed in the northern Yellow Sea, with coverage shrinking to 703 km^2 before gradually fading away.

In 2020, the observation of *U. prolifera* was about 2 weeks later than the previous year. Initially appearing in the southeastern Yellow Sea on April 29, it covered an area of 18 km^2. From late April to late May, it drifted northwestward, steadily expanding its coverage. By May 27, it reached 219 km^2, growing to 302 km^2 the next day. Peaking at 950 km^2 on June 4, it then moved northward, landing in the northern Yellow Sea by the end of June and dissipating approximately 2 weeks earlier than in 2019. Overall, the *U. prolifera* bloom in 2020 was less severe than in 2019.

In 2021, *U. prolifera* was first spotted on April 8, initially appearing in scattered amounts in the southeastern Yellow Sea. By May 21, it had drifted northward, covering 55 km^2 in the shallow waters off northern Jiangsu. Throughout June, *U. prolifera* proliferated extensively in the central Yellow Sea. From June 19 to July 10, a severe *U. prolifera* bloom affected coastal cities in the northern Yellow Sea, peaking at an extent of 3,534 km^2. By mid-July, the bloom gradually dissipated, with coverage shrinking to 31 km^2 by July 19. Compared to previous years, 2021 experienced the largest coverage area, longest duration, and most severe *U. prolifera* disaster, with the widest coverage area four times that of the previous year.

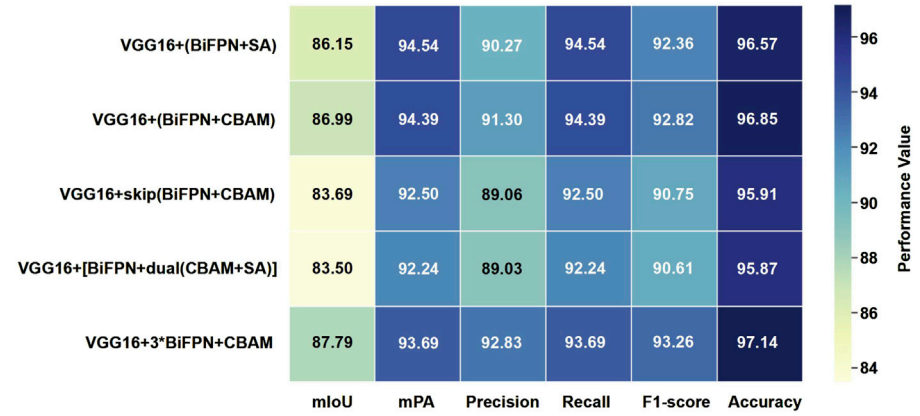


FIGURE 7
Heatmap of accuracy metrics for different model combinations.

In 2022, *U. prolifera* was first spotted on May 27 in the southeastern Yellow Sea, a month and a half later than the previous year. By June 25, it had reached coastal cities along the northern Yellow Sea, peaking at a coverage area of 548 km^2 before gradually dissipating over the following month. Compared to 2021, the *U. prolifera* coverage area significantly decreased in 2022, suggesting a relatively mild *U. prolifera* bloom overall.

On 16 May 2023, scattered *U. prolifera* patches were spotted in the shallow waters of northern Jiangsu, covering just 4 km^2. Within 2 weeks, the area surged to 276 km^2 by June 3, drifting northward thereafter. By the end of June, *U. prolifera* had proliferated massively in the central Yellow Sea, peaking at 2,170 km^2 on June 22, four times larger than in 2021. By July 24, it had dissipated significantly, leaving only 8 km^2 scattered off Qingdao and Yantai. Overall, the *U. prolifera* bloom in 2023 ranked second only to 2021 in severity.

On 18 May 2024, sparse patches of *U. prolifera* were initially detected in the shallow waters off northern Jiangsu and in the southeastern Yellow Sea, spanning an area of merely 17 km^2. Within just 10 days, however, *U. prolifera* proliferated rapidly from the northern Jiangsu shallows to the northern Yellow Sea, with its coverage expanding by a factor of 13. By mid-June, the *U. prolifera* extent had decreased to approximately 128 km^2. On June 26, it experienced a notable resurgence, reaching a peak area of 454 km^2, before gradually dissipating by mid-July. In summary, the *U. prolifera* bloom in 2024 demonstrated significant improvement, with a shorter duration and the smallest maximum coverage observed in the past 7 years.

5.3 Yearly spatial-temporal distribution characteristics of *Ulva prolifera*

Figure 9 illustrates the temporal and spatial dynamics of green tide (*U. prolifera*) coverage area in the Yellow Sea from 2018 to 2024. The trends indicate that peak green tide coverage varies significantly each year. For instance, 2021 shows the highest recorded green tide coverage, with a peak area exceeding 3,500 km^2, observed between June and July. In contrast, 2024 reflects a noticeable improvement, with substantially reduced peak coverage. The annual progression generally follows a similar pattern: a gradual increase in early year, peaking around late June, and decreasing in July.

This temporal pattern, along with area fluctuations, aligns with existing research suggesting that annual environmental conditions, such as temperature, nutrient availability, and ocean currents, significantly influence the extent and duration of green tides (Qi et al., 2016a; Zhang et al., 2019). Our identification results corroborate these findings, demonstrating that years with higher peak coverage often correspond to elevated sea surface temperatures and increased nutrient inputs, potentially driven by anthropogenic activities and seasonal upwelling. Furthermore, the spatial distribution of *U. prolifera* mirrors the prevailing ocean currents, which likely facilitate its dispersal across the Yellow Sea. These insights underscore the interplay between biological processes and physical drivers in shaping green tide dynamics, highlighting the importance of integrating environmental monitoring with algae detection systems.

The comprehensive analysis of *U. prolifera*'s yearly distribution patterns from 2018 to 2024 (Figure 10) reveals significant fluctuations in the intensity and extent of *U. prolifera* blooms over the past 7 years. Notably, 2021 experienced the most severe bloom within the study period, with the coverage area peaking of approximately 3,534 km^{2}on June 23. In contrast, 2022 marked a milder bloom and the lowest peak area of roughly 548 km^2. However, in 2023, the bloom coverage area surged to the second-highest value in nearly 7 years, underscoring the ongoing need for robust and consistent management strategies to mitigate green tide impacts.

This consistency, along with the observed peak times, typically around late June, aligns with previous findings on *U. prolifera* bloom cycles and suggests that these blooms may be influenced by recurring environmental conditions, such as temperature and nutrient availability, during this period. Notably, the unprecedented bloom in 2021, with its record-high coverage, was strongly linked to the impact of typhoons. Typhoons enhance nutrient enrichment in coastal waters by stirring sediments and promoting upwelling, creating ideal conditions for *U. prolifera* growth. This exceptional event underscores the significance of incorporating extreme weather events into bloom analyses. The temporal alignment of peak coverage across years, including the

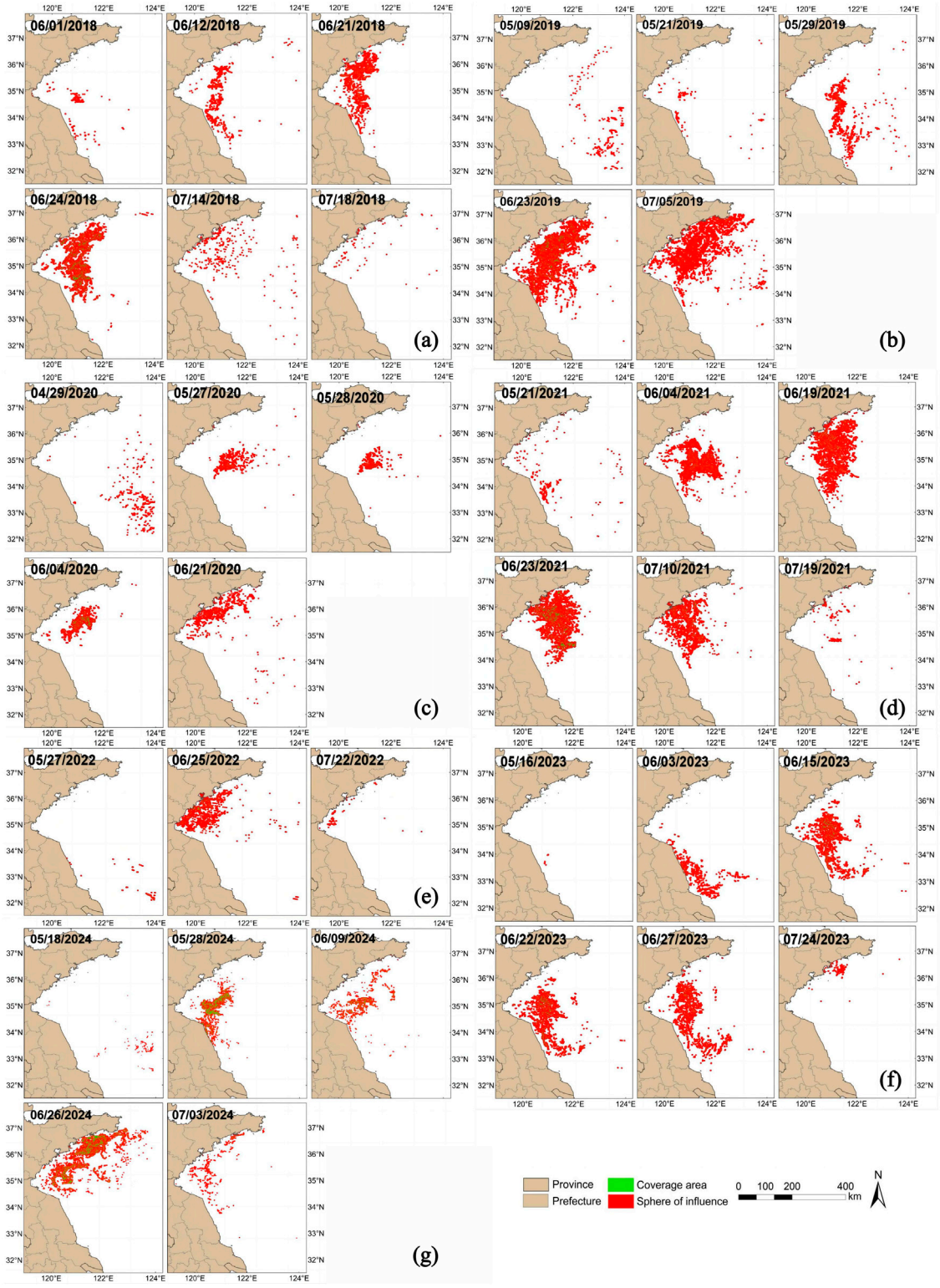


FIGURE 8
Figures **(a–g)** present the spatiotemporal distribution patterns of *U. prolifera* in the Yellow Sea during the period from 2018 to 2024.

typhoon-induced surge in 2021, highlights the importance of targeted monitoring in late June to better anticipate and manage bloom intensity. By integrating factors such as extreme weather events, more effective prediction and management strategies can be developed to mitigate the impact of green tides (Liu et al., 2009; Cui et al., 2018; Fang et al., 2018).

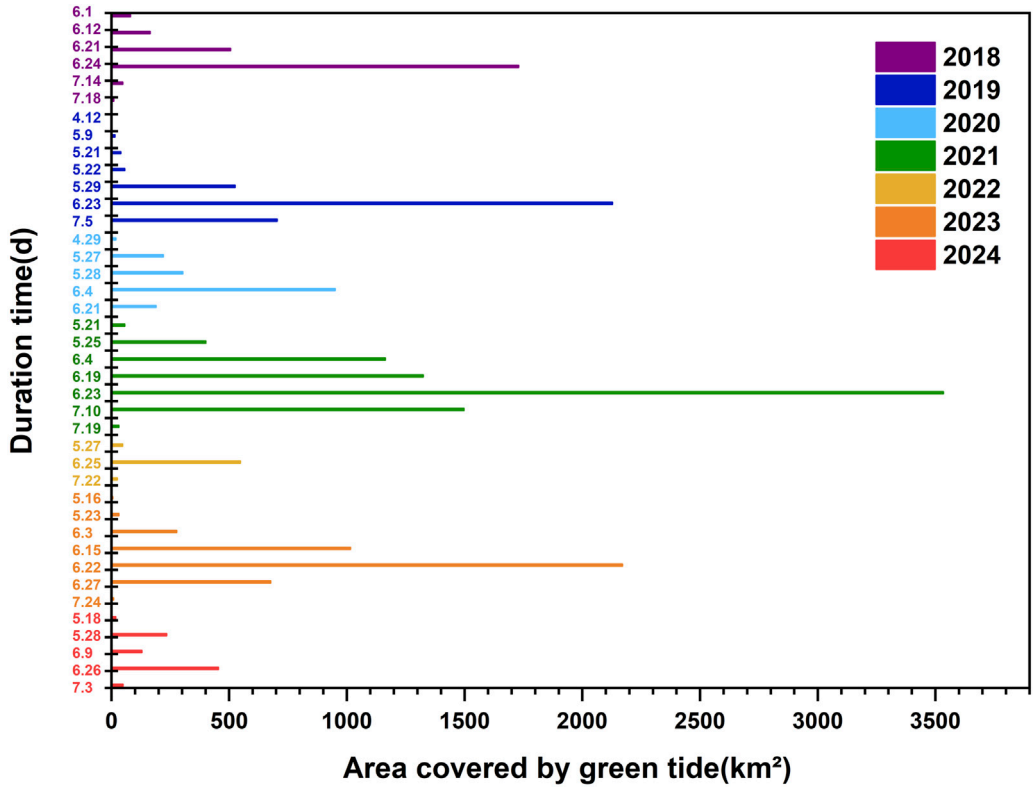


FIGURE 9
The changes in *U. prolifera* area in the Yellow Sea from 2018 to 2024.

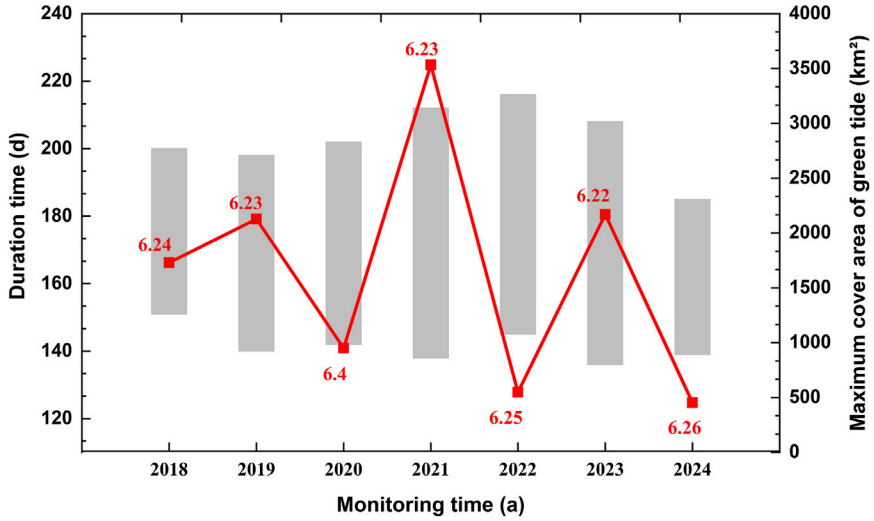


FIGURE 10
Duration and maximum coverage area of *U. prolifera* in the Yellow Sea from 2018 to 2024.

6 Conclusion

U. prolifera, known for forming green tides, poses significant ecological threats in coastal regions. We propose a tailored WaveNet deep learning model for *U. prolifera* detection using MODIS images, taking advantage of their extensive coverage and high data collection frequency. WaveNet employs VGG16 as its backbone feature extraction network and integrates BiFPN feature pyramid network, replacing fully connected layers and softmax outputs, to enhance feature extraction across various resolutions. We also introduce a lightweight CBAM

attention mechanism to filter background noise, ensuring more accurate and efficient feature extraction. With 608 annotated sample pairs, WaveNet achieved a detection accuracy of 97.14%, precision of 92.83%, recall of 93.69%, and an F1 score of 93.26%, significantly outperforming the NDVI and AFAI methods by mitigating uncertainties arising from threshold selection discrepancies. Through analyzing *U. prolifera* bloom dynamics in the Yellow Sea from 2018 to 2024, we confirmed a significant increase in *U. prolifera* area every June. Through our analysis, we observed that the maximum coverage area of *U. prolifera* exhibited an oscillating trend, initially increasing and then decreasing on an interannual basis. Furthermore, our research identified the southeastern Yellow Sea as the source of *U. prolifera* blooms in 2019, 2020, 2021, 2022, and 2024. These findings provide valuable insights into the early detection, prevention, and control of green tide formation, especially in identifying key geographical sources and underlying factors contributing to *U. prolifera*.

Data availability statement

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

Author contributions

WZ: Funding acquisition, Resources, Writing – original draft, Writing – review and editing. YX: Investigation, Methodology, Software, Writing – original draft, Writing – review and editing. LZ: Supervision, Validation, Writing – review and editing. ZL: Conceptualization, Data curation, Writing – review and editing. SL: Methodology, Software, Writing – review and editing. YL: Formal Analysis, Validation, Writing – review and editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

Arellano-Verdejo, J., Lazcano-Hernandez, H. E., and Cabanillas-Teran, N. (2019). ERISNet: deep neural network for Sargassum detection along the coastline of the Mexican Caribbean. *PeerJ* 7, e6842. doi:10.7717/peerj.6842

Brisset, M., Van Wynsberge, S., Andréfouët, S., Payri, C., Soulard, B., Bourassin, E., et al. (2021). Hindcast and near real-time monitoring of green macroalgae blooms in shallow coral reef lagoons using sentinel-2: a new-caledonia case study. *Remote Sens.* 13 (2), 211. doi:10.3390/rs13020211

Cao, Y., Wu, Y., Fang, Z., Cui, X., Liang, J., and Song, X. (2019). Spatiotemporal patterns and morphological characteristics of *Ulva prolifera* distribution in the Yellow Sea, China in 2016–2018. *Remote Sens.* 11 (4), 445. doi:10.3390/rs11040445

Cui, J., Zhang, J., Huo, Y., Zhou, L., Wu, Q., Chen, L., et al. (2015). Adaptability of free-floating green tide algae in the Yellow Sea to variable temperature and light intensity. *Mar. Pollut. Bull.* 101 (2), 660–666. doi:10.1016/j.marpolbul.2015.10.033

Cui, T., Liang, X., Gong, J., Tong, C., Xiao, Y., Liu, R., et al. (2018). Assessing and refining the satellite-derived massive green macro-algal coverage in the Yellow Sea with high resolution images. *ISPRS J. Photogrammetry* 144, 315–324. doi:10.1016/j.isprsjprs.2018.08.001

Fang, C., Song, K., Shang, Y., Ma, J., Wen, Z., and Du, J. (2018). Remote sensing of harmful algal blooms variability for Lake Hulun using adjusted FAI (AFAI) algorithm. *J. Environ. Inf.* 34 (2), 108–122. doi:10.3808/jei.201700385

Gao, L., Li, X., Kong, F., Yu, R., Guo, Y., and Ren, Y. (2022). AlgaeNet: a deep-learning framework to detect floating green algae from optical and SAR imagery. *IEEE J. Sel. Top. Appl. Earth Observations* 15, 2782–2796. doi:10.1109/JSTARS.2022.3162387

Hu, C. (2009). A novel ocean color index to detect floating algae in the global oceans. *Remote Sens. Environ.* 113 (10), 2118–2129. doi:10.1016/j.rse.2009.05.012

Hu, C., Li, D., Chen, C., Ge, J., Muller-Karger, F. E., Liu, J., et al. (2010). On the recurrent *Ulva prolifera* blooms in the Yellow Sea and East China sea. *J. Geophys. Res. Oceans* 115 (C5). doi:10.1029/2009JC005561

Hu, C., Feng, L., Hardy, R. F., and Hochberg, E. J. (2015). Spectral and spatial requirements of remote measurements of pelagic Sargassum macroalgae. *Remote Sens. Environ.* 167, 229–246. doi:10.1016/j.rse.2015.05.022

Hu, L., Hu, C., and Ming-Xia, H. (2017). Remote estimation of biomass of *Ulva prolifera* macroalgae in the Yellow Sea. *Remote Sens. Environ.* 192, 217–227. doi:10.1016/j.rse.2017.01.037

Hu, L., Zeng, K., Hu, C., and He, M.-X. J. R. s.o.e. (2019). On the remote estimation of *Ulva prolifera* areal coverage and biomass. *Remote Sens. Environ.* 223, 194–207. doi:10.1016/j.rse.2019.01.014

Jiang, X., Gao, M., Gao, Z., and Science, S. (2020). A novel index to detect green-tide using UAV-based RGB imagery. *Estuar. Coast.* 245, 106943. doi:10.1016/j.ecss.2020.106943

Lee, J. H., Pang, I. C., Moon, I. J., and Ryu, J. H. (2011). On physical factors that controlled the massive green tide occurrence along the southern coast of the Shandong Peninsula in 2008: a numerical study using a particle-tracking experiment. *J. Geophys. Res. Oceans* 116 (C12), C12036. doi:10.1029/2011JC007512

Li, X., Liu, B., Zheng, G., Ren, Y., Zhang, S., Liu, Y., et al. (2020). Deep-learning-based information mining from ocean remote-sensing imagery. *Natl. Sci. Rev.* 7 (10), 1584–1605. doi:10.1093/nsr/nwaa047

Lian, S., Luo, Z., Zhong, Z., Lin, X., Su, S., and Li, S. (2018). Attention guided U-Net for accurate iris segmentation. *J. Vis. Commun.* 56, 296–304. doi:10.1016/j.jvcir.2018.10.001

Liu, D., Keesing, J. K., Xing, Q., and Shi, P. J. M. p.b. (2009). World's largest macroalgal bloom caused by expansion of seaweed aquaculture in China. *Mar. Pollut. Bull.* 58 (6), 888–895. doi:10.1016/j.marpolbul.2009.01.013

Liu, D., Keesing, J. K., He, P., Wang, Z., Shi, Y., and Wang, Y. (2013). The world's largest macroalgal bloom in the Yellow Sea, China: formation and implications. *Estuar. Coast.* 129, 2–10. doi:10.1016/j.ecss.2013.05.021

Ma, X., Xu, J., Pan, J., Yang, J., Wu, P., and Meng, X. (2023). Detection of marine oil spills from radar satellite images for the coastal ecological risk assessment. *J. Environ. Manag.* 325, 116637. doi:10.1016/j.jenvman.2022.116637

Qi, L., Hu, C., Xing, Q., and Shang, S. (2016a). Long-term trend of *Ulva prolifera* blooms in the western Yellow Sea. *Harmful Algae* 58, 35–44.

Qi, L., Hu, C., Xing, Q., and Shang, S. J. H. A. (2016b). Long-term trend of *Ulva prolifera* blooms in the western Yellow Sea. *Harmful Algae* 58, 35–44. doi:10.1016/j.hal.2016.07.004

Qi, L., Hu, C., Wang, M., Shang, S., and Wilson, C. (2017). Floating algae blooms in the East China sea. *Geophys. Res. Lett.* 44 (22), 501–511. doi:10.1002/2017GL075525

Ronneberger, O., Fischer, P., and Brox, T. (2015). *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer. doi:10.1007/978-3-319-24574-4_28

Schmidhuber, J. (2015). Deep learning in neural networks: an overview. *Neural Netw.* 61, 85–117. doi:10.1016/j.neunet.2014.09.003

Shi, W., and Wang, M. (2009). Green macroalgae blooms in the Yellow Sea during the spring and summer of 2008. *J. Geophys. Res. Oceans* 114 (C12). doi:10.1029/2009JC005513

Smetacek, V., and Zingone, A. (2013). Green and golden seaweed tides on the rise. *Nature* 504 (7478), 84–88. doi:10.1038/nature12860

Tan, M., Pang, R., and Le, Q. V. (2020). Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.

Wang, Z., Xiao, J., Fan, S., Li, Y., Liu, X., and Liu, D. (2015). Who made the world's largest green tide in C hina? an integrated study on the initiation and early development of the green tide in Yellow Sea. *Limnology* 60 (4), 1105–1117. doi:10.1002/lno.10083

Wang, S., Liu, L., Qu, L., Yu, C., Sun, Y., Gao, F., et al. (2019). Accurate *Ulva prolifera* regions extraction of UAV images with superpixel and CNNs for ocean environment monitoring. *Neurocomputing* 348, 158–168. doi:10.1016/j.neucom.2018.06.088

Xing, Q., and Hu, C. (2016). Mapping macroalgal blooms in the Yellow Sea and East China Sea using HJ-1 and Landsat data: application of a virtual baseline reflectance height technique. *Remote Sens. Environ.* 178, 113–126. doi:10.1016/j.rse.2016.02.065

Xing, Q., An, D., Zheng, X., Wei, Z., Wang, X., Li, L., et al. (2019). Monitoring seaweed aquaculture in the Yellow Sea with multiple sensors for managing the disaster of macroalgal blooms. *Remote Sens. Environ.* 231, 111279. doi:10.1016/j.rse.2019.111279

Xu, Q., Zhang, H., Ju, L., and Chen, M. (2014). Interannual variability of *Ulva prolifera* blooms in the Yellow Sea. *Int. J. Remote Sens.* 35 (11-12), 4099–4113. doi:10.1080/01431161.2014.916052

Xu, Q., Zhang, H., Cheng, Y., Zhang, S., and Zhang, W. (2016). Monitoring and tracking the green tide in the Yellow Sea with satellite imagery and trajectory model. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 9 (11), 5172–5181. doi:10.1109/jstars.2016.2580000

Ye, N.-h., Zhang, X.-w., Mao, Y.-z., Liang, C.-w., Xu, D., Zou, J., et al. (2011). 'Green tides' are overwhelming the coastline of our blue planet: taking the world's largest example. *Ecol. Res.* 26, 477–485. doi:10.1007/s11284-011-0821-8

Zhang, M., Tang, J., Dong, Q., Song, Q., and Ding, J. (2010). Retrieval of total suspended matter concentration in the Yellow and East China Seas from MODIS imagery. *Remote Sens. Environ.* 114 (2), 392–403. doi:10.1016/j.rse.2009.09.016

Zhang, Y., He, P., Li, H., Li, G., Liu, J., Jiao, F., et al. (2019). *Ulva prolifera* green-tide outbreaks and their environmental impact in the Yellow Sea, China. *Neurosurgery* 6 (4), 825–838. doi:10.1093/nsr/nwz026

Zheng, L., Wu, M., Cui, Y., Tian, L., Yang, P., Zhao, L., et al. (2022). What causes the great green tide disaster in the South Yellow Sea of China in 2021? *Ecol. Indic.* 140, 108988. doi:10.1016/j.ecolind.2022.108988

Zhou, F., Ge, J., Liu, D., Ding, P., Chen, C., and Wei, X. (2021). The Lagrangian-based floating macroalgal growth and drift model (FMGDM v1. 0): application to the Yellow Sea green tide. *Geosci. Model Dev.* 14 (10), 6049–6070. doi:10.5194/gmd-14-6049-2021

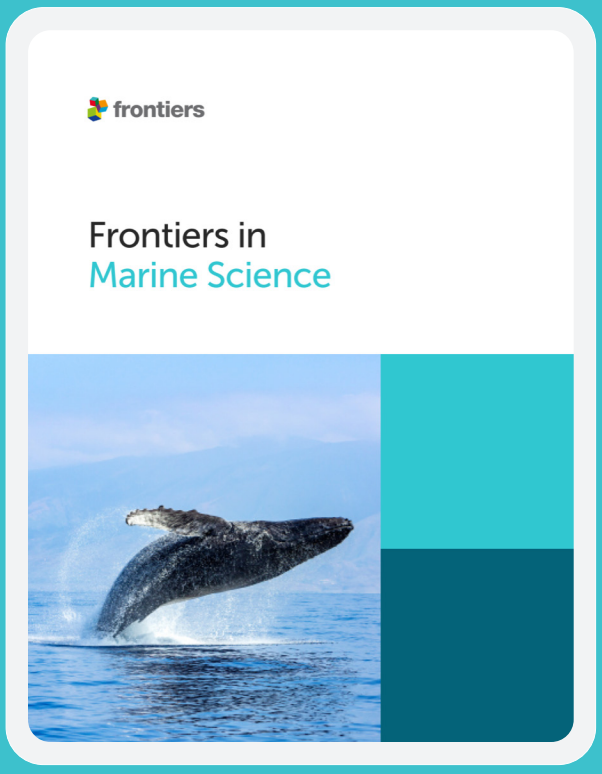
Frontiers in Marine Science

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