frontiers | Frontiers in Environmental Science

TYPE Original Research
PUBLISHED 06 November 2025
DOI 10.3389/fenvs.2025.1707611

OPEN ACCESS

EDITED BY
Peng Liu,
Chinese Academy of Sciences (CAS), China

REVIEWED BY
Meimei Zhang,
Chinese Academy of Sciences (CAS), China
Mingwei Wang,
Hubei University of Technology, China

*CORRESPONDENCE
Shang Zhao,
zhaosh5@chinasatnet.com.cn
Haonan Sun,
sunhaonan666@cug.edu.cn

RECEIVED 17 September 2025
REVISED 16 October 2025
ACCEPTED 24 October 2025
PUBLISHED 06 November 2025

CITATION
Zhang L, Zhao S, An D, Wang P, Guo B and Sun H (2025) Maring ship detection from GF-2 high-resolution remote sensing images with improved YOLOv13 model.
Front. Environ. Sci. 13:1707611.
doi: 10.3389/fenvs.2025.1707611

Maring ship detection from GF-2 high-resolution remote sensing images with improved YOLOv13 model

Liwei Zhang[1], Shang Zhao[1]*, Da An[1], Pengfei Wang[1], Bokai Guo[1] and Haonan Sun[2]*

[1]China Satellite Network Digital Technology Co., Ltd., Xiong'an New Area, Hebei, China, [2]School of Computer Science, China University of Geosciences, Wuhan, China

With the rapid development of global maritime trade and the rising demand for real-time, accurate marine ship monitoring, satellite image-based ship detection has become crucial for marine management and national defense. However, it faces two core challenges: complex backgrounds in high-resolution marine remote sensing images, and great variations in ship sizes–especially difficult small ship extraction. To address these, this study proposes an enhanced method based on improved YOLOv13, using China's Gaofen-2 (GF-2) satellite images. First, GF-2 image data is preprocessed, including radiometric correction to eliminate atmospheric effects, orthorectification to correct image distortion, and fusion of multispectral and panchromatic images to improve spatial resolution and enrich spectral information. Then, three key optimizations are made to the YOLOv13 model: 1) In the backbone network, the A2C2f module is modified by introducing a single-head attention mechanism. By parallelly fusing global and local feature information, it avoids multi-head redundancy and improves the recognition accuracy of small ship targets; 2) In both the backbone and neck networks, the DS_C3K2 module is modified by integrating a lightweight attention mechanism, which enhances the model's feature extraction capability in complex backgrounds while reducing channel and spatial redundancy; 3) In the head network, a path-fused Global Feature Pyramid Network (GFPN) is introduced, which leverages skip-layer and cross-scale connections to strengthen cross-scale feature interaction, refine the representation of small ship features, and effectively address the issues of insufficient deep supervision and feature information loss in multi-scale ship detection. Additionally, the improved YOLOv13 model is pre-trained using the open-source DOTA dataset (rich in non-ship negative samples) to enhance its ability to distinguish between ship foreground and background clutter, and then applied to ship detection in segmented sub-images of GF-2 remote sensing images; finally, the detected sub-images are stitched to restore complete regional images. Experiments show that the accuracy rate reaches 96.9%, the recall rate reaches 91.4%, the mAP50 reaches 95.5%, and the mAP50-95 reaches 75.9%, all of which are higher than the mainstream target detection models. It provides a high-performance solution for complex marine ship detection and has important practical significance for both civilian and military fields.

KEYWORDS

marine ship detection, improved YOLOv13, small target detection, complex marine background, GF-2

1 Introduction

Ship detection in optical remote sensing images (ORSI) stands as a critical and inherently challenging task, with profound implications spanning ecological governance, marine resource management, and military reconnaissance (Kanjir et al., 2018; Wang et al., 2011). As a cornerstone of aerial and satellite image analysis, target detection in optical remote sensing imagery underpins a broad spectrum of practical applications–from environmental monitoring and disaster response to national security operations–by enabling the automated identification and localization of key objects within vast and complex geospatial datasets (Cheng and Han, 2016; Gómez-Chova et al., 2015). Ships, as primary actors in maritime activities, represent pivotal targets for maritime regulation and surveillance. Effective detection and recognition methods are indispensable for executing a range of critical tasks: monitoring and curbing illegal, unreported, and unregulated (IUU) fishing to preserve marine biodiversity; regulating unauthorized resource exploitation, such as unlicensed oil drilling or sand mining, to safeguard coastal ecosystems; investigating smuggling, piracy, and other transnational maritime crimes to maintain maritime order; and tracking the movements of foreign armed vessels to ensure territorial integrity and national security. Satellite remote sensing technology, with its capabilities for large-scale, all-weather, and near-real-time observation, has thus become an irreplaceable tool in safeguarding maritime security, upholding maritime rights and interests, and supporting the seamless operation of both military and civilian maritime transportation networks (Wu et al., 2023; Yue et al., 2021).

Yet, the inherently complex maritime environment poses formidable challenges to accurate ship target detection. Coastal zones, in particular, are rife with confounding factors: dense clusters of port infrastructure (such as docks, cranes, and storage facilities) often exhibit spectral or structural similarities to ships, leading to misclassification; dynamic elements like breaking waves, foam, or tidal fluctuations can obscure ship outlines or create false positives; and the coexistence of small vessels (e.g., fishing boats) with large maritime assets (e.g., cargo ships or warships) exacerbates the difficulty of multi-scale target extraction (Gong et al., 2024; Yan et al., 2025). In addition, changing lighting conditions (such as glare from sunlight on the water, shadowing effects from clouds, and reflections from ice on the water) further degrade image quality, making it more difficult to distinguish ship targets from cluttered backgrounds (Li et al., 2020; Zhang et al., 2026). These complexities collectively hinder the precision and reliability of ship detection systems, underscoring the need for more robust and adaptive methodologies.

Traditional ship detection methods (Zhu et al., 2010; Proia and Pagé, 2009; Shi et al., 2013) predominantly rely on manually designed features–such as edge gradients, texture descriptors, or spectral thresholds–to distinguish ships from backgrounds. However, these approaches suffer from inherent limitations: their performance is highly dependent on expert-defined feature engineering (Li et al., 2025), making them prone to errors in complex scenarios (e.g., overlapping ships or variable lighting); they lack robustness against environmental perturbations like sea fog, wave glint, or coastal clutter; and their labor-intensive feature design processes render them costly and inefficient to adapt to diverse maritime conditions (Guo et al., 2023; Ren et al., 2024).

In contrast, deep learning-based models have revolutionized target detection by enabling end-to-end feature learning, outperforming traditional artificial intelligence systems across numerous domains–from computer vision and natural language processing to speech recognition. In the field of maritime surveillance, recent advancements further highlight their potential: Hu et al. (2024) proposed a laser point cloud-based ship detection method for unconstrained maritime areas, which preprocesses lidar data, converts 3D point clouds into 2D bird's-eye views, and feeds them into a dedicated object detection network, achieving high precision in near-shore scenarios. Wang et al. (2021) integrated edge computing into traditional detection frameworks to mitigate computational bottlenecks, enabling real-time ship monitoring on resource-constrained devices. Ye et al. (2005) developed a visual attention model based on HSI color space feature extraction, converting RGB images to HSI space and generating saliency maps via normalized fusion of multi-scale features, which yielded promising results in detecting ships under varying illumination.

In recent years, transformer-based object detectors have achieved remarkable progress, with RT-DETR (Real-Time Detection Transformer) emerging as a representative architecture that effectively balances detection accuracy and real-time efficiency (Zhao et al., 2024). RT-DETR leverages a hybrid CNN-transformer backbone and an end-to-end query-based detection head, which eliminates the need for post-processing operations such as non-maximum suppression (NMS). This design allows the model to perform global context reasoning across the entire image and enhances detection consistency, making it highly competitive in natural scene detection tasks. Moreover, its dynamic feature aggregation and cross-scale self-attention improve general object localization performance while maintaining real-time inference speeds. However, when applied to high-resolution optical remote sensing imagery for maritime ship detection, RT-DETR still encounters notable limitations. The transformer-based architecture is computationally intensive and memory-demanding, which restricts its deployment in large-scale satellite images or resource-limited environments. Additionally, the global self-attention mechanism tends to dilute fine-grained local features that are crucial for identifying small-scale ships under cluttered backgrounds such as port facilities, sea waves, and cloud shadows. Furthermore, the lack of explicit multi-scale feature fusion in RT-DETR reduces its sensitivity to small object variations compared with CNN-based detectors optimized for dense targets.

Deep learning-based marine ship detection still faces unique challenges, particularly in optical remote sensing scenarios. First, ships exhibit extreme scale diversity–ranging from small fishing boats (a few meters in length) to large cargo vessels (over 300 m)–making it difficult for models to balance sensitivity to tiny targets and precision for large ones. Multi-scale small ships, in particular, are prone to being missed or misclassified due to their low pixel coverage and similarity to background noise (e.g., floating debris). Second, the bounding boxes of small targets often struggle to converge to their true positions during training, as their sparse pixel information provides insufficient gradient signals for model optimization (Shen et al., 2025). Third, mainstream object detection

algorithms like YOLOv13, while lauded for their lightweight architecture and efficiency, have limited multi-scale feature extraction capabilities. In complex marine backgrounds–such as coastal regions with overlapping port infrastructure, dynamic waves, or varying cloud cover–they often fail to distinguish ships from visually similar interference, hindering detection accuracy. Extracting ships from high-resolution remote sensing imagery thus remains a critical bottleneck, as existing methods cannot fully capture fine-grained, multi-scale details necessary for robust identification.

To address these challenges, this study proposes a method for detecting small objects in complex backgrounds based on an improved YOLOv13. First, a single-head attention mechanism is introduced to construct the A2C2f_SHSA (Single-Head Self-Attention) module, replacing the A2C2f module in the backbone. This prevents head redundancy and improves the accuracy of small object ship recognition by combining global and local information in parallel. Then, a lightweight attention mechanism is introduced to construct the C3K2_EFAtt module, replacing the DS_C3K2 module in the backbone and neck components. This enhances the model's feature extraction capabilities in complex backgrounds while reducing channel and spatial redundancy. Finally, a path-fused global feature pyramid network (GFPN) is introduced in the head, including skip-layer and cross-scale connections, to enhance cross-scale feature interactions and refine small object representations. This improves the model's multi-scale feature fusion capabilities and enables better handling of scale diversity and background clutter. These improvements enable our model to achieve a better trade-off between accuracy, robustness, and efficiency than transformer-based approaches like RT-DETR in complex maritime environments. We validate its performance using high-resolution imagery from China's Gaofen-2 (GF-2) satellite, focusing on complex marine environments with dense coastal infrastructure and dynamically changing sea conditions. This research not only provides a reliable technical solution for overcoming the bottleneck of multi-scale ship detection in high-resolution remote sensing imagery, but also promotes the practical application of deep learning in maritime monitoring. By improving the accuracy and reliability of ship detection under complex conditions, this research provides valuable support for strengthening maritime traffic management, enhancing maritime safety law enforcement, and optimizing emergency response capabilities. Ultimately, it will help to more effectively safeguard maritime rights and interests, protect the ecological environment, and ensure national security.

2 Data description and preprocessing

2.1 Data description

2.1.1 Gaofen-2 remote sensing imagery

The Gaofen-2 (GF-2) satellite is one of the key satellites under China's High-Resolution Earth Observation System (CHEOS) program. It was independently developed by China and successfully launched on August 19, 2014. GF-2 primarily serves applications in land resource surveys, crop yield estimation, environmental protection, disaster prevention and mitigation, urban planning, and water resource management, providing high-resolution remote sensing imagery and data support to relevant sectors (Ren et al., 2022).

TABLE 1 GF-2 sensor parameters.

Parameter	Panchromatic/Multispectral camera	
Spectral range	Panchromatic	450~900 nm
	Multispectral	450~520 nm
		520~590 nm
		630~690 nm
		770~890 nm
Spatial resolution	Panchromatic	0.8 m
	Multispectral	3.2 m
Swath width	45 km (Combination of Two Cameras)	
Revisit cycle	5 days	
Coverage cycle	69 days	

GF-2 boasts several technical advantages, including high resolution, wide coverage, flexibility, and versatility. Ship detection based on GF-2 imagery is not only critical for maritime target monitoring and maritime security but also provides essential data support for monitoring illegal fishing, addressing marine pollution, and conducting emergency rescue operations. The technical specifications of GF-2's sensors are shown in Table 1.

GF-2 remote sensing imagery primarily consists of multispectral and panchromatic image data. Multispectral data generally has a lower spatial resolution, as the sensor needs to capture data from multiple bands simultaneously. Given a fixed data volume, each band receives fewer pixels. However, multispectral data has a higher spectral resolution, allowing for the extraction of more spectral characteristics of surface features by combining data from different bands. In contrast, panchromatic data typically has higher spatial resolution. Since it is a single band that integrates information across the visible spectrum, without the need to divide the data into multiple bands as in multispectral imagery, more spatial details can be captured within the same data volume.

2.1.2 DOTA dataset

DOTA is a large-scale benchmark dataset dedicated to object detection in aerial imagery, serving as a critical resource for developing and evaluating object detection algorithms in aerial scenarios. Its image data are sourced from diverse sensors and platforms, with resolutions ranging from 800 × 800 to 20,000 × 20,000 pixels. A key characteristic of DOTA lies in the significant variations exhibited by its annotated objects, including extensive diversity in scale, orientation, and shape. All instances within the dataset are meticulously annotated by experts in aerial image interpretation using arbitrary 8-degree-of-freedom quadrilaterals, ensuring precise localization of targets with complex geometric attributes.

The dataset is continuously updated to expand its scale and enrich its scope, thereby aligning with the dynamic nature of real-world application scenarios. We selected the DOTA dataset for model training and validation, primarily due to its inclusion of

diverse terrain types and the comprehensive coverage of targets with varying scales, orientations, and shapes–features that make it highly representative for aerial object detection tasks.

2.2 Data preprocesses

In the study of ship detection based on GF-2 remote sensing imagery, the processing of image data is crucial for both accuracy and efficiency. GF-2 data primarily includes multispectral and panchromatic images. Multispectral images cover multiple bands (such as visible light and near-infrared), providing high spectral resolution, which is suitable for extracting spectral characteristics of surface features. However, the spatial resolution is relatively low (3.2 m). In contrast, panchromatic images are single-band images that record combined reflectance information across the visible spectrum, offering higher spatial resolution (0.8 m), which allows for finer spatial details.

To obtain imagery that combines both spectral and spatial resolution, data preprocessing was conducted in this study, including radiometric correction (Du et al., 2002; Chen et al., 2004), atmospheric correction (Vermote and Vermeulen, 1999), data orthorectificatio (Aguilar et al., 2013), projection processing and multispectral and panchromatic image fusion (Zhu et al., 2024). The specific process is shown in Figure 1.

3 Methodology

To address the shortcomings of YOLOv13 in detecting small target ships at sea in complex backgrounds, this study proposes a comprehensively improved scheme based on multi-scale feature fusion and lightweight attention mechanisms. Building on the Generalized-FPN (GFPN) (Jiang et al., 2022), the improvements are three-fold: first, a single-head attention mechanism is introduced to construct the A2C2f_SHSA module, replacing the A2C2f module in the backbone. This prevents head redundancy and enhances the accuracy of small ship recognition by parallelly combining global and local information. Second, a lightweight attention mechanism is employed to develop the C3K2_EFAtt module, which replaces the DS_C3K2 module in both the backbone and neck, strengthening feature extraction in complex backgrounds while reducing channel and spatial redundancy. Third, the path-fused GFPN is integrated into the head, incorporating skip-layer and cross-scale connections to enhance cross-scale feature interactions and refine small object representations, thereby boosting multi-scale fusion capabilities to handle scale diversity and background clutter. The DOTA dataset provides abundant negative samples, aiding the model in distinguishing foreground and background–particularly effective in eliminating interference from nearshore non-ship objects. Data augmentation techniques further enhance the model's generalization, ensuring high accuracy across diverse marine meteorological conditions. These combined improvements significantly improved YOLOv13's performance in complex maritime environments, reducing nearshore interference and improving small target detection accuracy. The overall model flow chart is shown in Figure 2.

3.1 Multi-class sample dataset

This study leverages the DOTA (Dataset for Object Detection in Aerial Images) dataset for model training, a comprehensive benchmark widely recognized for its suitability in aerial and remote sensing target detection tasks (Ding et al., 2021). A key advantage of the DOTA dataset lies in its rich and diverse composition: beyond containing a large quantity of ship target samples with varying scales, orientations, and maritime contexts (ranging from small fishing vessels in coastal waters to large cargo ships in open seas), it also incorporates an extensive array of non-ship negative samples. These negative samples encompass a wide spectrum of land-based and coastal features, including buildings (such as port warehouses, coastal residences, and industrial facilities), ground vehicles (trucks, cars, and construction machinery), road networks, bridges, and even natural features like vegetation clusters and rocky outcrops–many of which are specifically distributed in near-shore areas, where ship detection is most prone to interference.

Such a diverse set of negative samples plays a critical role in enhancing the model's discriminative capabilities. By exposing the model to near-shore non-ship objects that often share visual similarities with ships (e.g., rectangular port cranes resembling ship hulls, or large floating structures mimicking vessel outlines), the training process enables the model to learn subtle distinguishing features–such as spectral signatures, contour textures, and contextual associations–that differentiate foreground ship targets from cluttered background information. This is particularly valuable for mitigating confusion between ships and coastal infrastructure, a common source of false detections in complex maritime scenes. Through intensive training on this balanced dataset, the enhanced YOLOv13 model develops a robust ability to filter out irrelevant background interference while accurately identifying true ship targets. This not only reduces the occurrence of false positives caused by near-shore clutter but also strengthens the model's generalization performance across diverse maritime environments–whether in busy port areas with dense infrastructure, turbid coastal waters with dynamic wave patterns, or open seas with sparse but small-scale targets. Ultimately, this training strategy lays a solid foundation for the model to handle real-world complex background ship detection tasks with high precision and reliability.

3.2 Data augmentation

To enhance training data diversity and boost the model's generalization ability, this study applies a range of data augmentation techniques to the DOTA dataset, including random cropping, rotation, flipping, and color jittering. These methods are tailored to simulate real-world variations in maritime remote sensing scenarios: Random cropping exposes the model to ships in diverse spatial contexts (e.g., partially occluded by coastal structures or surrounded by waves), improving its adaptability to target positions. Rotation (0°–360°)

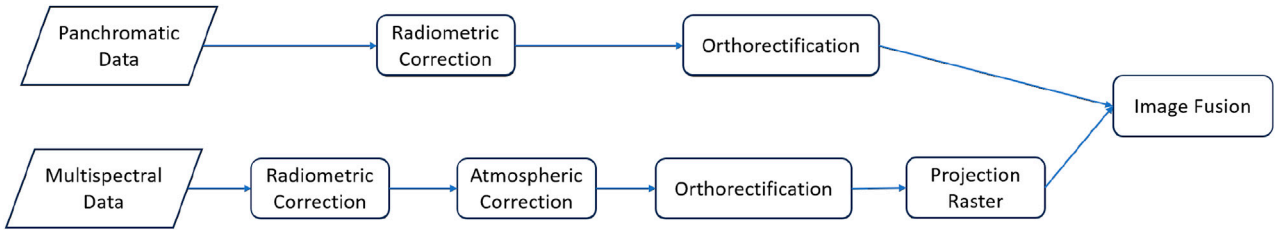

FIGURE 1
Data preprocessing workflow.

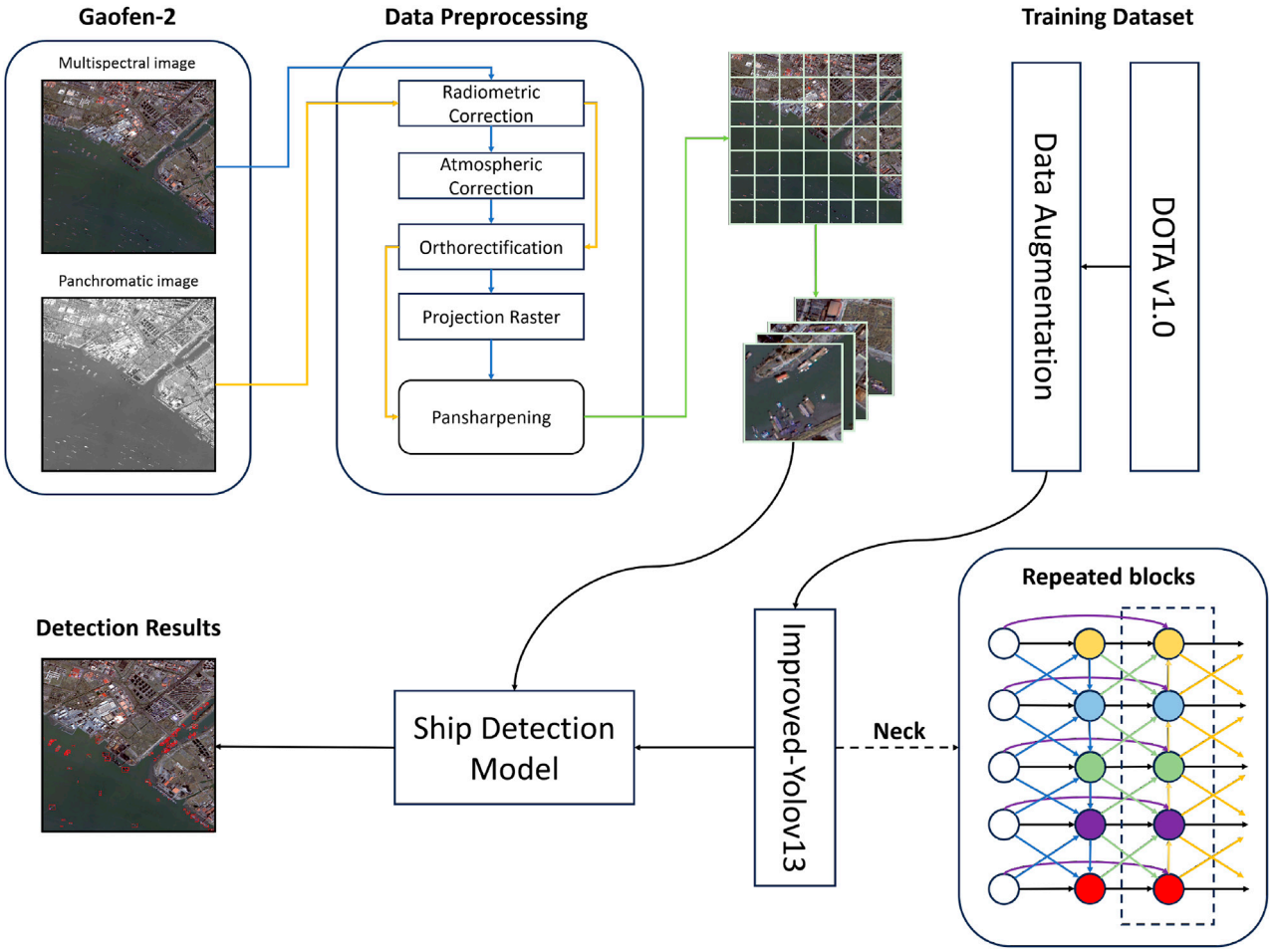

FIGURE 2
Flow chart of ship detection model.

and flipping address the arbitrary orientations of ships–shaped by wind, tides, or navigation routes–preventing overfitting to specific directions. Color jittering adjusts brightness, contrast, and hue to mimic variable lighting at sea (e.g., sunlight glare, overcast conditions), helping the model focus on intrinsic ship features rather than transient spectral changes.

By training on this augmented data, the model learns robust representations that persist across ship sizes, orientations, and lighting conditions. This ensures high detection accuracy even in complex, dynamic maritime environments, where weather and target appearances can shift unpredictably.

3.3 YOLOv13 algorithm

Previous YOLO series follow the computational paradigm of "backbone network - neck network - detection head", which essentially limits the full transmission of information flow. In

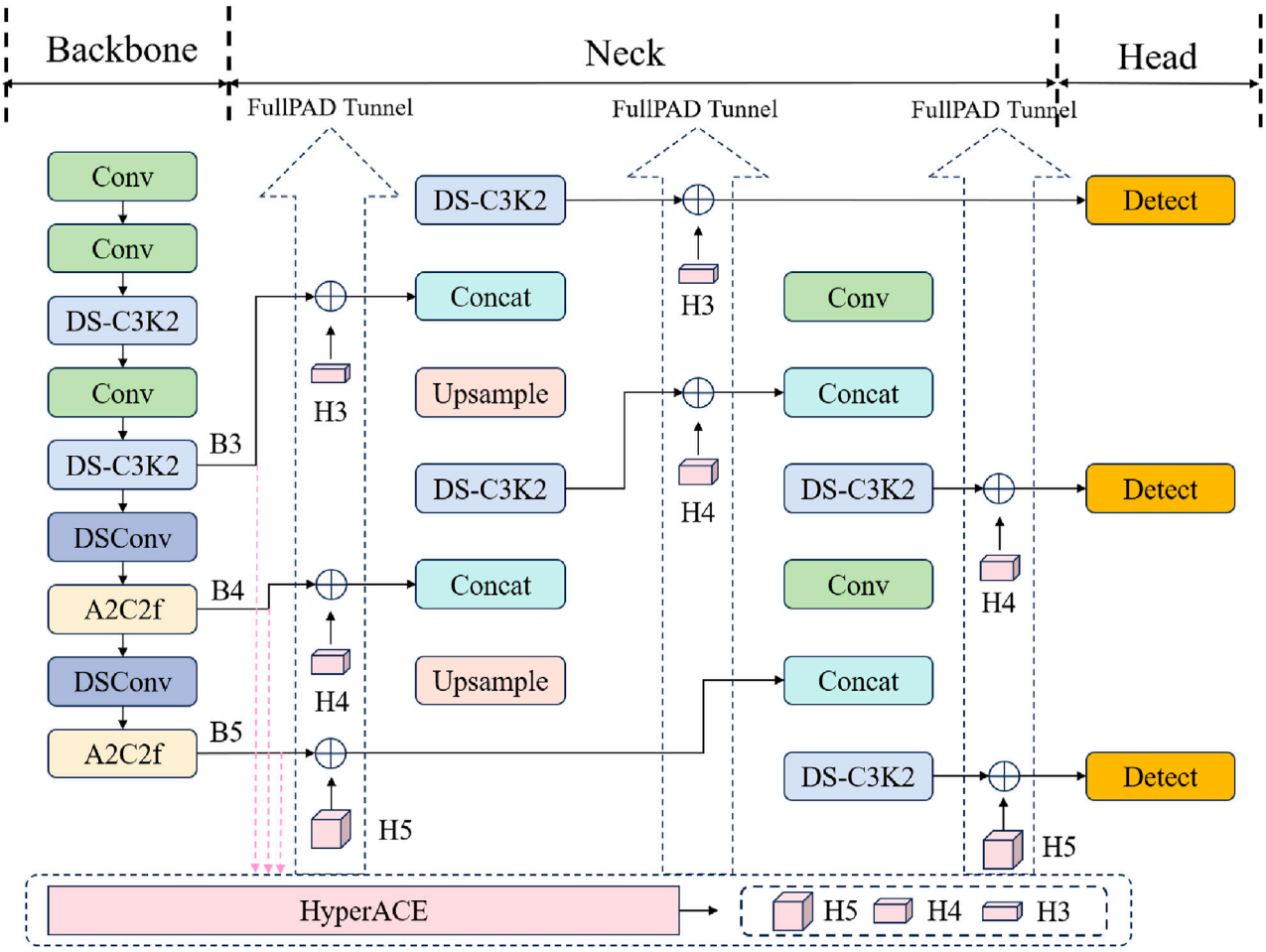

FIGURE 3
YOLOv13 network structure.

contrast, YOLOv13 enhances the traditional YOLO architecture by implementing full-link feature aggregation and allocation (FuIPAD) through the Hypergraph Adaptive Correlation Enhancement (HyperACE) mechanism. Therefore, YOLOv13 achieves fine-grained information flow and representation coordination throughout the network, which can improve gradient propagation and significantly enhance detection performance.

Specifically, as shown in Figure 3, YOLOv13 first uses a backbone network similar to previous work to extract multi-scale feature maps B1, B2, B3, B4, B5, but the large kernel convolution is replaced by the proposed lightweight DS-C3k2 module. Then, unlike the traditional YOLO method that directly inputs B3, B4 and B5 into the neck network, YOLOv13 collects and passes these features to the proposed HyperACE module to achieve high-order correlation adaptive modeling and feature enhancement of cross-scale and cross-position features. Subsequently, its FUPAD paradigm uses three independent channels to distribute the correlation-enhanced features to the connection between the backbone network and the neck network, the internal layers of the neck network, and the connection between the neck network and the detection head to optimize the information flow. Finally, the output feature map of the neck network is passed to the detection head to achieve multi-scale object detection.

3.4 Backbone network optimization

To address the task of detecting small ship targets against complex backgrounds in remote sensing imagery, we focused on improving the A2C2f module within the YOLOv13 backbone network. We incorporated the single-head self-attention (SHSA) mechanism proposed in SHViT (Yun and Ro, 2024), achieving efficient fusion of global and local information with a lightweight design, improving small ship detection accuracy.

Specifically, the improved A2C2f_SHSA module retains the dimensionality transformation and residual connection architecture of the original module. Its core focus is replacing the internal feature mixing unit with computational logic based on single-head attention. Drawing on SHViT's analysis of multi-head attention redundancy, the single-head attention module only computes attention on a subset of channel features (rather than all). By eliminating unnecessary multi-head parallel operations, this significantly reduces computational complexity and memory access costs, adapting to the high-resolution and large-data-volume characteristics of remote sensing imagery and achieving its lightweight design goal. This module also fuses two key pieces of information through parallel paths: First, it uses deep convolution to

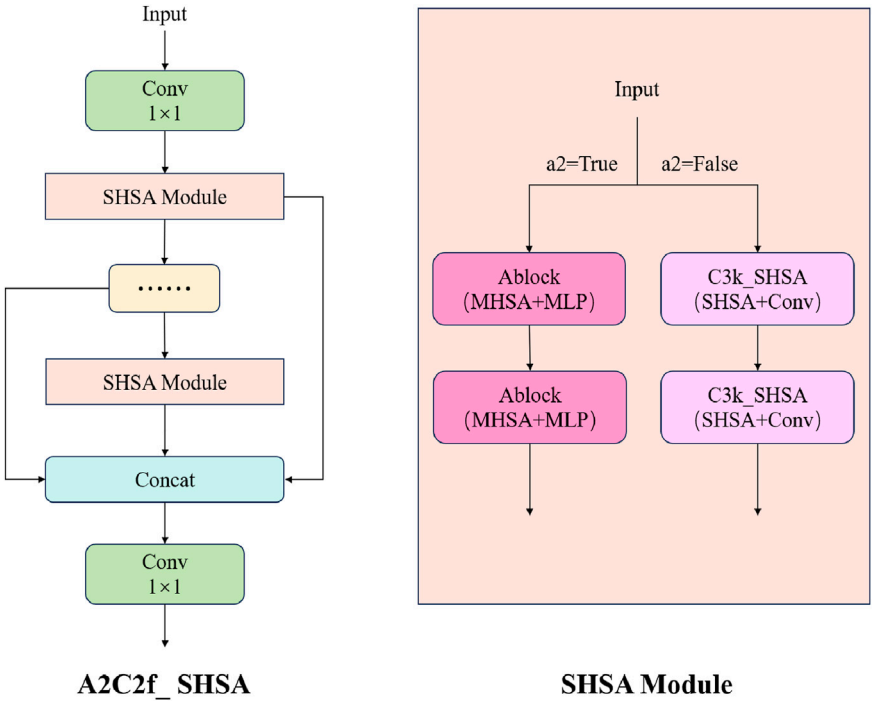

FIGURE 4
A2C2f_SHSA module.

extract fine local features of small ship targets, leveraging convolution's strength in capturing local spatial correlations. Second, it uses single-head attention to model global contextual relationships (such as the spatial distribution of ships and complex backgrounds like waves and islands), addressing the inadequacy of traditional convolution in modeling long-range dependencies. This parallel fusion mechanism enables the model to accurately identify subtle features of small ship targets while effectively distinguishing between targets and complex background interference. This significantly improves the robustness of feature representation, particularly in remote sensing scenarios with small ships and high background noise.

The A2C2f_SHSA module is an improvement based on the A2C2f module in YOLOv13. The core enhances feature expression capabilities by introducing the single-head self-attention (SHSA) mechanism. The process is as follows: the input feature is first compressed from c1 to the hidden channel c_by 1 × 1 convolution (cv1), achieving dimensionality reduction to reduce computational overhead; then enters the multi-branch feature processing stage, and selects different branch structures according to the parameter "a2" - when a2 is True, a branch consisting of two ABlock stacks is used (retaining the original A2C2f multi-head attention mechanism), and each ABlock captures global dependencies through multi-head self-attention (MHSA) and MLP layers; when a2 is False, the branch is replaced by the C3k_SHSA module, which contains two SHSABlock stacks, each SHSABlock first passes through a 3 × 3 Deep convolution extracts local features, then single-head attention (SHSA) computes global correlations for some channels (pdim) while retaining features from the remaining channels. After enhancing channel interactions through a feedforward network (FFN), input and output features are fused via residual connections. The output features of all branches are concatenated by channel and compressed by a 1 × 1 convolution (cv2) to reduce the number of channels from (1 + n) × c_to c2. If the residual mechanism is enabled, the output features are weighted by a learnable parameter gamma and then added to the output of cv1 to form the final module output. By dynamically switching attention modes, this module reduces multi-head redundancy while fusing global and local information simultaneously, making it particularly suitable for feature extraction of small ships against complex backgrounds. As shown in Figure 4.

This improvement, which only involves adjustments to the A2C2f module in the backbone network, leverages SHViT's lightweight attention design to enhance the feature capture of small ship targets while maintaining overall model efficiency, providing a more robust feature foundation for detection tasks in complex backgrounds.

3.5 Lightweight EFAttention optimization

To detect small ships against complex backgrounds in remote sensing imagery, we leveraged the lightweight attention mechanism from LAEDNet to improve the backbone and neck of YOLOv13 (Zhou et al., 2022). We replaced the original DS_c3k2 module with the C3k2_EFAttention module, which incorporates a highly efficient attention mechanism. This aims to enhance the model's feature extraction capabilities for small ships and reduce redundancy in channel and spatial dimensions. The core of this improved module lies in the embedded EFAttention (Efficient Fusion Attention)

mechanism, which achieves refined feature screening and enhancement through a dual-branch structure: the channel branch uses global average pooling to compress spatial information, and then combines it with 1D convolution to model the channel dimension to generate a dynamic channel weight vector. It can effectively highlight channel information related to ship features (such as hull edges and masts) and suppress the interference of background noise channels; the spatial branch compresses features into a single channel through 1×1 convolution, and generates a spatial attention map through sigmoid activation, accurately focusing on the area where the ship target is located and weakening the spatial redundant information of complex background (such as waves, islands, clouds, etc.). The output features of the two are adaptively fused through element-by-element addition to achieve channel and spatial information. At the same time, the AdaptiveFeatureFusionBlock introduced in the module is combined with SELayer (squeeze-excitation module) to further optimize feature flow and fusion efficiency through multi-path convolution and attention reweighting, so that the model can more accurately capture the subtle features of small target ships (such as the blurred outlines of distant ships and the edge information of low-contrast targets) while maintaining its lightweight characteristics, effectively alleviating the problem of feature confusion in complex backgrounds, and providing a more robust feature foundation for subsequent detection tasks, thereby improving the accuracy and stability of the model in small target ship detection in remote sensing images.

The process is as follows: the input feature is first compressed from c1 to the hidden channel c_(c_ = int (c2 × e)) by a 1×1 convolution to achieve dimensionality reduction to reduce computational overhead; then it is split into the main branch and the residual branch, the main branch enters the multi-branch feature processing stage, and different branch structures are selected according to the parameter "c3k" - when c3k is False, a branch processed in sequence by n EFAttention modules is used, and each EFAttention enhances features in parallel through two branches: the channel branch compresses the feature into a $1 \times 1 \times$ c_vector through global average pooling, and then generates a channel attention map through 1D convolution and Sigmoid, which is multiplied with the original feature; the spatial branch compresses the feature to 1 channel through 1×1 convolution, and then generates a spatial attention map through Sigmoid, which is multiplied with the original feature, and the outputs of the two branches are added together to achieve fusion; when c3k is True, the branch is replaced by a C3k module, which contains 2 Bottleneck stacks, and is multiplied by 3×3 Convolution extracts local features; the features processed by the main branch are concatenated with the residual branch features by channel, and then the number of channels is compressed from $2 \times$ c_to c2 through 1×1 convolution. Finally, they are fused with the input features (after dimension adjustment) through residual connection to obtain the module output. As shown in Figure 5.

3.6 Multi-scale feature fusion

GFPN is a feature fusion method designed to enhance multi-scale feature interactions in object detection (Jiang et al., 2022). Its core advantage lies in its flexible and comprehensive cross-scale and cross-layer connection mechanism, which goes beyond the traditional bidirectional flow and achieves more efficient information exchange between features at different levels. The difference between GFPN and several common feature fusion methods such as FPN (Lin et al., 2017a), PANet (Liu et al., 2018), and BiFPN (Tan et al., 2020) is shown in Figure 6. GFPN introduces a "queen fusion" strategy that allows feature information to flow freely across scales and levels like the queen in chess. This includes $\log_2 n$ connections (skip connections), which efficiently pass information from early nodes to later nodes while minimizing redundancy; and dense connections, which enable each feature in the kth layer to receive inputs from all previous layers, ensuring that historical feature information is fully utilized. This design promotes a more thorough fusion of high-level semantic features with low-level spatial details, significantly improving the model's ability to detect objects of different sizes, especially small and medium-sized objects. In GFPN, feature fusion is optimized through an adaptive connection mechanism rather than explicit learnable weights. By dynamically adjusting the information flow according to the importance of different features, GFPN can emphasize key feature components while suppressing less useful feature components, thereby improving the quality of fused features without increasing computational overhead. The structure of GFPN is based on PANet, but it removes unnecessary restrictions on the feature propagation path. It adds cross-scale connections between adjacent layers and cross-layer connections within the same scale, thereby building a more interconnected feature fusion network. This structure can be easily extended to deeper layers, allowing for more complex feature refinement while maintaining computational efficiency. By repeating this generalized fusion structure multiple times, GFPN continuously enhances the interaction between multi-scale features, ultimately improving detection accuracy and robustness in complex scenarios. The structure of the improved YOLOv13 model we proposed is shown in the Figure 7.

4 Experiments and discussion

4.1 Experimental dataset

In this study, we used the DOTAv1.0 dataset for model training and validation, and finally used the model to test the application of the model on GF-2 remote sensing data.

Specifically, the GF-2 test imagery used in this study was acquired on October 11, 2022, covering the Nantong Port area in Nantong City, Jiangsu Province, China (approximately E120.8°, N32.1°). This region is one of the busiest coastal transportation hubs along the Yangtze River Delta, characterized by dense maritime traffic, complex near-shore environments, and frequent vessel activities. The scene contains a diverse range of ship types–including cargo ships, fishing vessels, and port service boats–distributed across both open-water and dockside zones. In addition, the coastal area exhibits high background complexity caused by port infrastructure (e.g., docks, cranes, and warehouses), variable illumination conditions, and wave clutter, all of which pose significant challenges for precise ship detection.

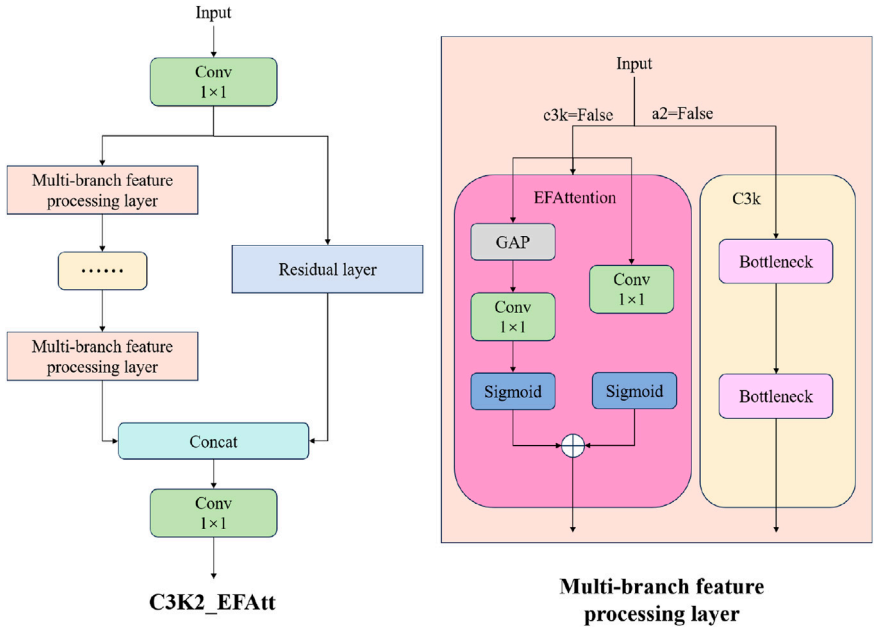

FIGURE 5
C3K2_EFAtt module.

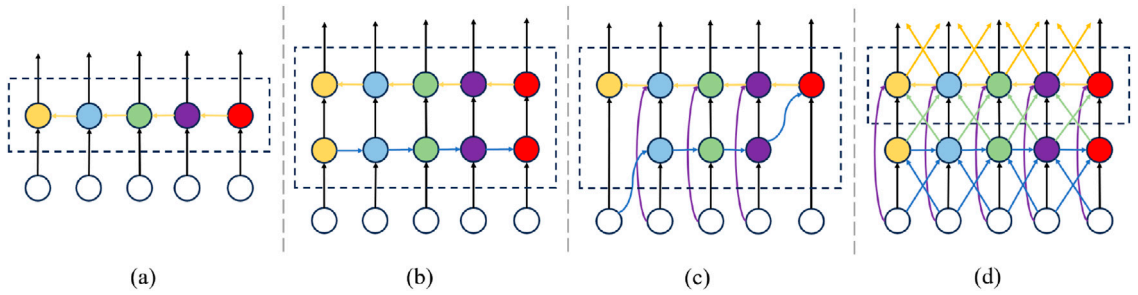

FIGURE 6
Different FPN design. **(a)** FPN. **(b)** PANet. **(c)** BiFPN. **(d)** GFPN.

4.1.1 Data preprocessing

To obtain imagery that combines both spectral and spatial resolution, data preprocessing was conducted in this study, including radiometric calibration, atmospheric correction, orthorectification, projection processing, and data fusion.

After completing the data preprocessing steps, in order to fuse the low-resolution multispectral image with the high-resolution panchromatic image to obtain an image with both high spectral resolution and spatial resolution, we use the Gram-Schmidt transform to pansharpen the image, and finally generate an image with both high spatial resolution and rich spectral information, as shown in Figure 8.

4.1.2 Image enhancement

To further improve the contrast and visual effect in ship recognition, image enhancement is applied to the fused images. This experiment uses the percentage stretch method to adjust the brightness and contrast of the images, making the boundaries between the ship's body and the background water area clearer. The enhancement process allows for better identification of the ship's contours and provides higher visual clarity for target extraction. The low percentile is set to 0.3, and the high percentile is set to 99.7, normalizing the image pixel values to the range of 0~255.

4.1.3 Image cropping

After completing the image enhancement process, the image is cropped twice to more effectively identify ships.

- First Crop: The primary cropping step involves extracting the core research area from the original remote sensing image, with the goal of eliminating most land-based regions (such as coastal buildings, vegetation, and infrastructure) while retaining only the water areas that are critical for ship detection. This targeted cropping significantly increases the proportion of water bodies in the image, thereby concentrating

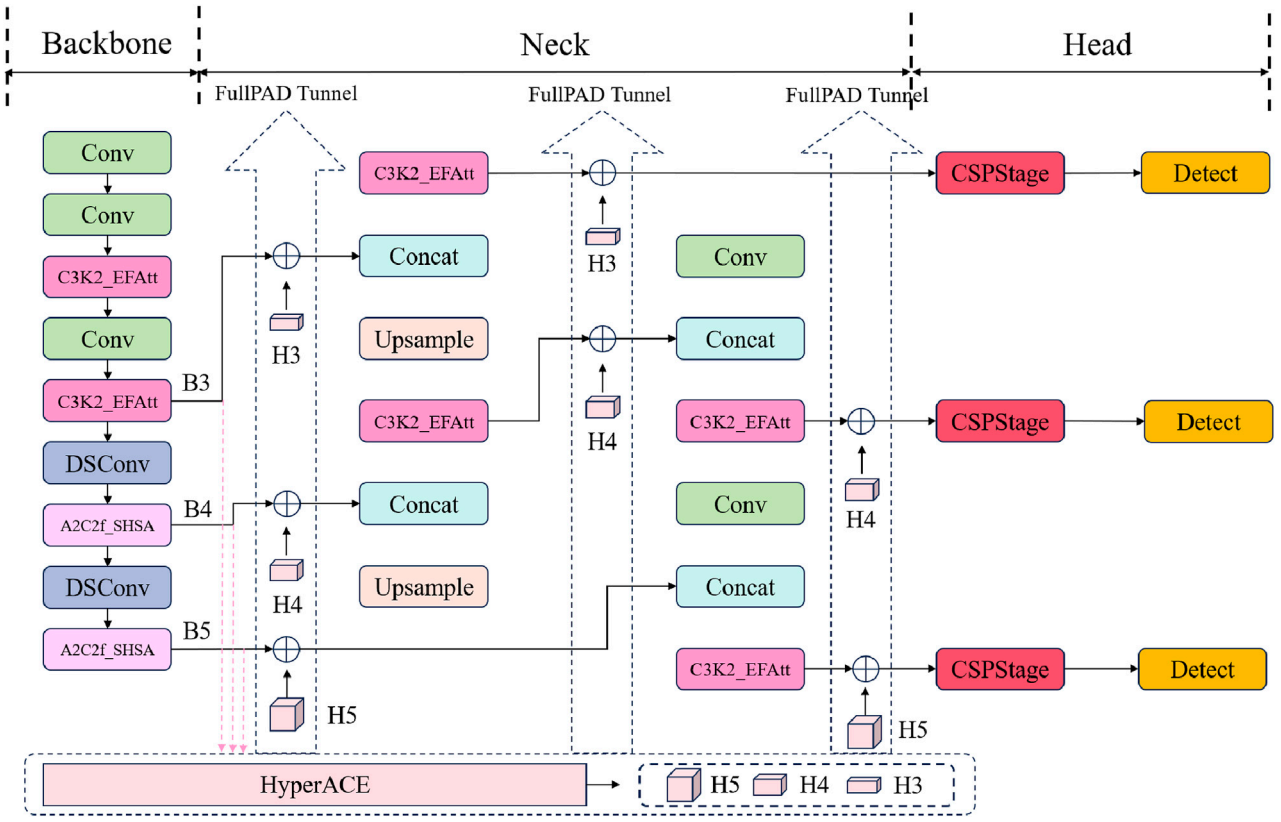

FIGURE 7
Improved-YOLOv13 network structure.

FIGURE 8
Comparison of pre- and post-fusion images.

the model's attention on the relevant maritime domain. By filtering out extensive land interference upfront, this step not only reduces the computational burden of processing irrelevant background information but also minimizes false detections caused by land features that may visually resemble ships (e.g., port cranes or rectangular structures). Consequently, it lays a more efficient foundation for subsequent recognition tasks, enhancing both the accuracy and speed of ship detection. The output of this first cropping step is illustrated in Figure 9.

- Second Crop: Despite the first crop, the resulting image remains relatively large in size, which could exceed the computational limits of the detection model and hinder efficient processing. To address this, the image is further divided into multiple smaller sub-images. Each sub-image is standardized to a resolution of 640 × 640, matching the

FIGURE 9
Image after the first crop.

resolution of the training data to ensure consistency during model inference. During the ship recognition phase, a sliding window approach is employed to process each sub-image individually–this method systematically scans the entire cropped area, ensuring no potential ship targets are overlooked. By breaking down the large image into manageable sub-images, this step not only aligns with the model's input requirements but also reduces the impact of complex background clutter within each processing unit, allowing the model to focus more precisely on local ship features. Ultimately, this two-stage cropping strategy balances efficiency and accuracy, optimizing the overall performance of the ship detection pipeline.

4.2 Model training

This paper uses the DOTA-v1.0 dataset, retaining all samples from nearshore waters and all samples containing ships from high seas. The dataset is randomly divided into training, validation, and test sets in an 8:1:1 ratio. The training set contains 1,818 images, and after data augmentation, 9,090 training samples are obtained.

The experiments in this paper are all based on the Linux operating system, NVIDIA A100 accelerated processor and Pytorch to build a deep learning system. The pre-trained model used in training is YOLOv13-s, and some hyperparameters are as follows: epoch = 100, batch-size = 16, imgsz = 640, initial-lr = 1e-3. The effect of the model on the validation set is shown in Figure 10. It can be seen from the figure that most ships can be accurately identified in complex backgrounds, and only a small number of ships are misinspection or missed inspection.

In the detection results, the model demonstrates excellent detection performance. Specifically, the precision reaches 96.9%, indicating that 96.9% of the targets predicted as ships are correct, with a low false positive rate. The recall is 91.4%, showing that the model successfully identifies most ship targets, with only about 8.6% missed.

Regarding average precision, the mAP50 achieves 95.5%, reflecting the model's high accuracy in detecting ships under a relatively lenient IoU threshold (0.5). However, the mAP50-95 is 75.9%, indicating a decline in accuracy under stricter IoU thresholds, possibly due to complex backgrounds and challenges in precise localization.

Overall, the detection performance for the ship class is excellent, and the model is able to reliably detect ship objects and achieve high precision and recall in most cases.

4.3 Ablation experiment

To verify the effectiveness of each core optimization, we conducted ablation experiments on the DOTA dataset with original YOLOv13 as the baseline, focusing on three components: SHViT single-head attention (A2C2f_SHSA), LANet lightweight attention (C3K2_EFAtt), and GFPN. Results are shown in Table 2.

The baseline YOLOv13 achieves 0.961 precision (P), 0.911 recall (R), and 0.732 mAP. Adding SHViT alone (YOLOv13-SHViT) improves performance to 0.965 P, 0.916 R, and 0.760 mAP, confirming its effectiveness in avoiding multi-head redundancy and enhancing small-ship recognition through efficient global–local feature fusion. However, LANet alone (YOLOv13-LANet) causes a performance drop (0.947 P, 0.880 R, 0.675 mAP) because its lightweight attention excessively compresses feature representations, leading to the loss of discriminative spatial details and weakening the model's sensitivity to small or low-contrast ships. When GFPN is applied independently (YOLOv13-GFPN), modest gains are observed (0.963 P, 0.910 R, 0.753 mAP), indicating that multi-scale feature fusion alone cannot fully compensate for insufficient attention-driven feature refinement. In contrast, two-component combinations exhibit synergy: SHViT + GFPN (0.964 P, 0.916 R, 0.759 mAP) benefits from the integration of global contextual modeling with enhanced feature aggregation, while LANet + GFPN (0.962 P, 0.912 R, 0.753 mAP) partially offsets LANet's loss of detail through improved hierarchical fusion. Notably, the SHViT + LANet combination achieves a balanced global–local feature representation, as SHViT captures long-range semantic dependencies and LANet suppresses background redundancy, resulting in a higher recall (0.920) with stable precision. The full integration of SHViT, LANet, and GFPN yields the best overall performance (0.969 P, 0.914 R, 0.759 mAP), demonstrating that the three modules complement each other–SHViT enhances global perception, LANet refines fine-grained attention, and GFPN strengthens multi-scale interaction–forming a synergistic architecture optimized for accurate and efficient ship detection in complex maritime environments.

4.4 Comparative experiment

In order to verify the effectiveness of the proposed method, the method proposed in this paper was compared with algorithms such as SSD (Liu et al., 2016), Retinanet (Lin et al., 2017b), Centernet (Zhou et al., 2019) and RT-DETR (Zhao et al., 2024), as shown in Table 3.

As observed, traditional detectors perform poorly in complex maritime scenarios. SSD variants (SSD-mobilenetv2, SSD-vgg) exhibit low recall (0.393 and 0.553) and mAP (0.546 and 0.637), while RetinaNet also suffers from limited robustness (recall = 0.360, mAP = 0.610) due to its difficulty in distinguishing ships from

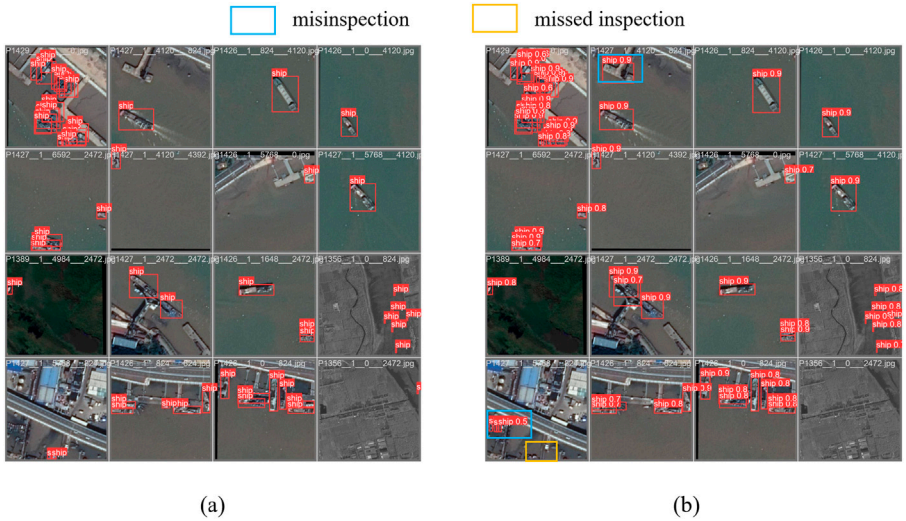

FIGURE 10
Qualitative ship detection results on the validation dataset. **(a)** Ground Truth. **(b)** Yolov9-GFPN.

TABLE 2 The ablation experiment results.

Model	P	R	mAP (0.50:0.95)
YOLOv13	0.961	0.911	0.732
YOLOv13-SHViT	0.965	0.916	0.760
YOLOv13-LANet	0.947	0.880	0.675
YOLOv13-GFPN	0.963	0.910	0.753
YOLOv13-SHViT + LANet	0.906	0.920	0.758
YOLOv13-SHViT + GFPN	0.964	0.916	0.759
YOLOv13-LANet + GFPN	0.962	0.912	0.753
YOLOv13-SHViT + LANet + GFPN	**0.969**	**0.914**	**0.759**

Bold values indicate the optimal results in the experiment.

TABLE 3 Comparison with other algorithms.

Model	P	R	mAP (0.50:0.95)
SSD-mobilenetv2	0.884	0.393	0.546
SSD-vgg	0.930	0.553	0.637
Retinanet	0.912	0.360	0.610
Centernet	0.957	0.854	0.691
RT-DETR	0.933	0.890	0.656
YOLOv13	0.961	0.911	0.732
Improved-YOLOv13	**0.969**	**0.914**	**0.759**

Bold values indicate the optimal results in the experiment.

cluttered backgrounds and detecting small-scale targets. CenterNet shows moderate improvement (0.957 P, 0.854 R, 0.691 mAP) by incorporating keypoint-based localization but still lacks strong multi-scale feature integration.

The transformer-based RT-DETR achieves relatively balanced performance (0.933 P, 0.890 R, 0.656 mAP), benefiting from its end-to-end query-based detection architecture and global attention mechanism. However, its global self-attention introduces heavy computational overhead and often weakens local feature representation, leading to reduced precision in dense or small-object scenarios typical of high-resolution maritime imagery.

The YOLO series performs more effectively under such conditions. The baseline YOLOv13 achieves 0.961 P, 0.911 R, and 0.732 mAP, demonstrating strong efficiency and spatial sensitivity. Our Improved-YOLOv13 further enhances all metrics to 0.969 P, 0.914 R, and 0.759 mAP–surpassing all compared models, including RT-DETR. These results confirm that the proposed optimizations (SHViT, LANet, and GFPN) effectively strengthen multi-scale feature representation and cross-scale fusion while maintaining computational efficiency, thereby improving both the accuracy and robustness of ship detection in complex maritime environments.

4.5 Ship detection

Ship detection is performed on the GF-2 remote sensing image, using the trained YOLOv13 object detection model. The sub-images detection results are shown in Figure 11.

After completing ship recognition on all sub-images, the results are stitched to reconstruct the full regional image, enabling a unified and complete visual output for further spatial analysis. The experimental results, as shown in Figures 12, 13, demonstrate the effectiveness of the proposed method.

Through a complete workflow including data preprocessing, pansharpening, data augmentation, and cropping, this study successfully constructed high-resolution multispectral images based on GF-2 data and applied them to ship detection in complex water environments. The recognition results indicate

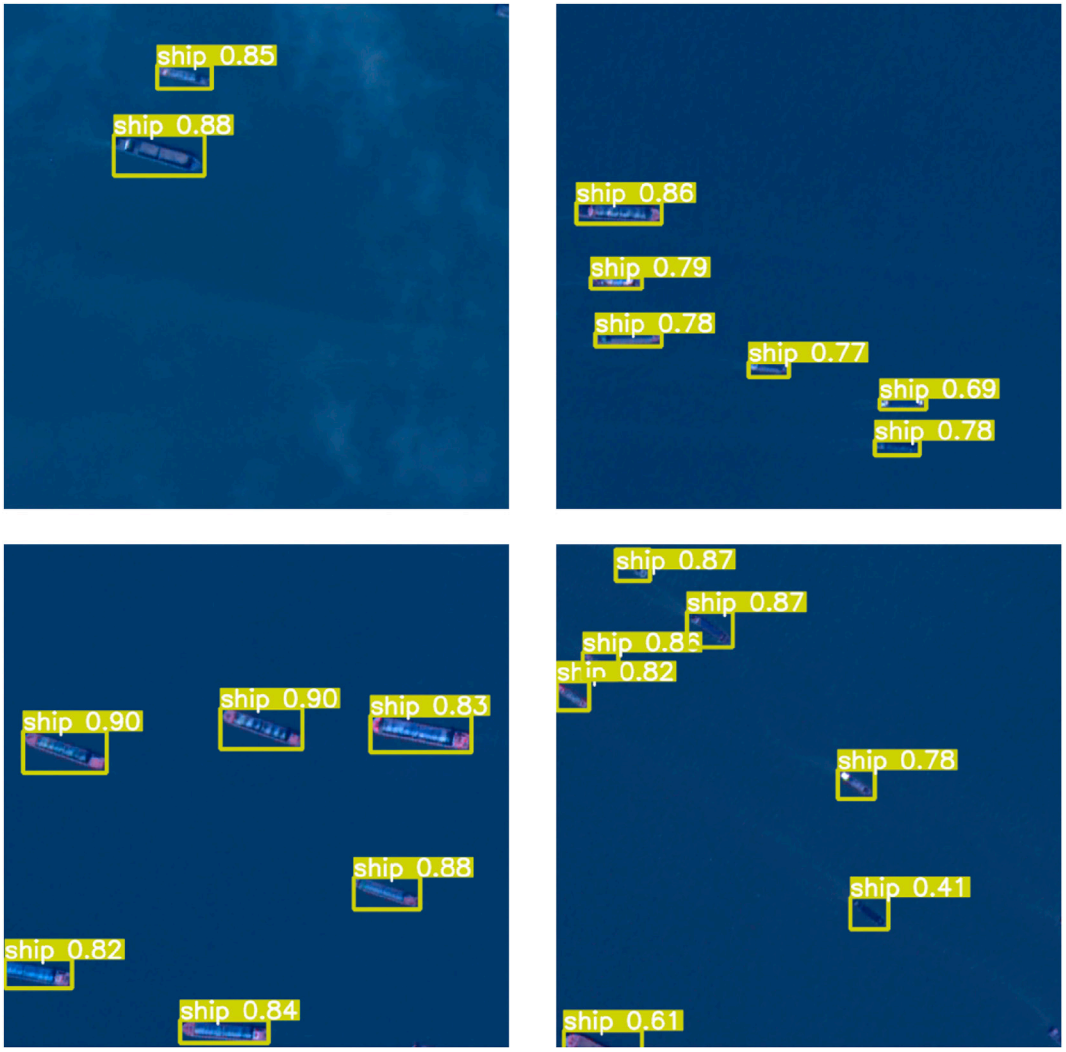

FIGURE 11
Sub-images detection results.

FIGURE 12
Overall detection results.

that the model performs well even under complicated backgrounds, with most ships correctly detected and classified. The bounding boxes are accurately placed, and confidence scores are consistently high (mostly above 0.80), showing strong robustness and precision.

Overall, the proposed approach provides reliable detection performance in challenging scenes with mixed land-water features, dense ship distribution, and varied vessel types, indicating its potential for practical applications in maritime monitoring and remote sensing-based ship detection.

5 Conclusion

In this study, an improved YOLOv13-based ship detection framework was developed and validated using high-resolution GF-2 optical satellite imagery. The proposed method significantly enhances detection accuracy and robustness in complex maritime environments characterized by dense coastal infrastructure, dynamic sea states, and high background interference. By optimizing multi-scale feature representation and introducing attention-guided mechanisms, the framework achieves a superior balance between precision, recall, and computational efficiency. Experimental results demonstrate that the proposed model performs exceptionally well in ship detection, achieving precision and

FIGURE 13
Detection details.

recall rates of 96.9% and 91.4%, respectively. The improved YOLOv13 model provides a reliable and scalable solution for high-precision maritime monitoring, supporting applications in vessel traffic management, coastal surveillance, and marine environmental governance.

Despite its excellent performance, it still has some limitations. Under adverse atmospheric conditions, such as heavy sea fog, haze, and low visibility, optical signals from GF-2 imagery suffer severe attenuation and reduced feature contrast, resulting in diminished detection confidence. Additionally, complex illumination effects–including sun-glint reflections, shadow occlusions, and variable sea surface brightness–may distort spatial features and lead to false or missed detections, particularly for small and low-contrast vessels. Moreover, the current framework primarily relies on single-source optical data and lacks multi-modal fusion mechanisms capable of integrating complementary information from synthetic aperture radar (SAR), multispectral, or infrared imagery to enable all-weather and all-time detection capabilities. Future research will focus on addressing these limitations by developing multi-source data fusion strategies, adaptive illumination and atmospheric correction modules, and temporal feature modeling from sequential or multi-temporal imagery to enhance the system's stability under dynamic marine environments.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://captain-whu.github.io/DOTA/index.html.

Author contributions

LZ: Data curation, Methodology, Validation, Visualization, Writing – original draft, Writing – review and editing. SZ: Data curation, Funding acquisition, Methodology, Supervision, Writing – review and editing. DA: Data curation, Methodology, Supervision, Writing – review and editing. PW: Data curation, Resources, Writing – review and editing. BG: Data curation, Visualization, Writing – review and editing. HS: Data curation, Validation, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the National Natural Science Foundation of China under grant 42471505 and the Guangxi Key Research and Development Program (GuikeAB25069111).

Conflict of interest

Authors LZ, SZ, DA, PW, and BG were employed by China Satellite Network Digital Technology Co., Ltd.

The remaining author 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. Use AI to Translation.

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

References

Aguilar, M. A., del Mar Saldaña, M., and Aguilar, F. J. (2013). Assessing geometric accuracy of the orthorectification process from geoeye-1 and worldview-2 panchromatic images. *Int. J. Appl. Earth Observation Geoinformation* 21, 427–435. doi:10.1016/j.jag.2012.06.004

Chen, X., Vierling, L., Rowell, E., and DeFelice, T. (2004). Using lidar and effective lai data to evaluate ikonos and landsat 7 etm+ vegetation cover estimates in a ponderosa pine forest. *Remote Sens. Environ.* 91, 14–26. doi:10.1016/j.rse.2003.11.003

Cheng, G., and Han, J. (2016). A survey on object detection in optical remote sensing images. *ISPRS J. Photogrammetry Remote Sens.* 117, 11–28. doi:10.1016/j.isprsjprs.2016.03.014

Ding, J., Xue, N., Xia, G.-S., Bai, X., Yang, W., Yang, M. Y., et al. (2021). Object detection in aerial images: a large-scale benchmark and challenges. *IEEE Trans. Pattern Analysis Mach. Intell.* 44, 7778–7796. doi:10.1109/tpami.2021.3117983

Du, Y., Teillet, P. M., and Cihlar, J. (2002). Radiometric normalization of multitemporal high-resolution satellite images with quality control for land cover change detection. *Remote Sens. Environ.* 82, 123–134. doi:10.1016/s0034-4257(02)00029-9

Gómez-Chova, L., Tuia, D., Moser, G., and Camps-Valls, G. (2015). Multimodal classification of remote sensing images: a review and future directions. *Proc. IEEE* 103, 1560–1584. doi:10.1109/jproc.2015.2449668

Gong, Y., Chen, Z., Deng, W., Tan, J., and Li, Y. (2024). Real-time long-distance ship detection architecture based on yolov8. *IEEE Access* 12, 116086–116104. doi:10.1109/ACCESS.2024.3445154

Guo, Q., Wang, Z., Sun, Y., and Liu, N. (2023). Maritime ship target detection based on the yolov7 model. In: 2023 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML); 2023 November 3–5: IEEE. p. 1031–1034.

Hu, H., Li, P., and Qin, Y. (2024). Ship target detection in the restricted parking area of ship lock chamber based on laser point cloud bird's-eye view. In: 2024 6th International Conference on Internet of Things, Automation and Artificial Intelligence (IoTAAI); 2024 July 26–28: IEEE. p. 651–655. doi:10.1109/iotaai62601.2024.10692534

Jiang, Y., Tan, Z., Wang, J., Sun, X., Lin, M., and Li, H. (2022). *Giraffedet: a heavy-neck paradigm for object detection*. arXiv preprint arXiv:2202.04256.

Kanjir, U., Greidanus, H., and Oštir, K. (2018). Vessel detection and classification from spaceborne optical images: a literature survey. *Remote Sens. Environ.* 207, 1–26. doi:10.1016/j.rse.2017.12.033

Li, Z., You, Y., and Liu, F. (2020). Analysis on saliency estimation methods in high-resolution optical remote sensing imagery for multi-scale ship detection. *IEEE Access* 8, 194485–194496. doi:10.1109/ACCESS.2020.3033469

Li, Y., Yan, J., Zhong, L., Bao, D., Sun, L., and Li, G. (2025). Full-coverage mapping of daily high-resolution xco 2 across china from 2015 to 2020 by deep learning-based spatio-temporal fusion. *IEEE Trans. Geoscience Remote Sens.*, 1. doi:10.1109/tgrs.2025.3540289

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 conference on computer vision and pattern recognition; 2024 June 16–22; Seattle, WA: IEEE. p. 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 international conference on computer vision; 2023 October 1–6: IEEE. p. 2980–2988.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., et al. (2016). Ssd: single shot multibox detector. In: Computer Vision–ECCV 2016: 14th European Conference; 2016 October 11–14; Amsterdam, The Netherlands: Springer. p. 21–37.

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; 2024 June 16–22; Seattle, WA: IEEE. p. 8759–8768.

Proia, N., and Pagé, V. (2009). Characterization of a bayesian ship detection method in optical satellite images. *IEEE geoscience remote Sens. Lett.* 7, 226–230. doi:10.1109/LGRS.2009.2031826

Ren, B., Ma, S., Hou, B., Hong, D., Chanussot, J., Wang, J., et al. (2022). A dual-stream high resolution network: deep fusion of gf-2 and gf-3 data for land cover classification. *Int. J. Appl. Earth Observation Geoinformation* 112, 102896. doi:10.1016/j.jag.2022.102896

Ren, Z., Tang, Y., Yang, Y., and Zhang, W. (2024). Sasod: saliency-aware ship object detection in high-resolution optical images. *IEEE Trans. Geoscience Remote Sens.* 62, 1–15. doi:10.1109/tgrs.2024.3367959

Shen, L., Gao, T., and Yin, Q. (2025). Yolo-lpss: a lightweight and precise detection model for small sea ships. *J. Mar. Sci. Eng.* 13, 925. doi:10.3390/jmse13050925

Shi, Z., Yu, X., Jiang, Z., and Li, B. (2013). Ship detection in high-resolution optical imagery based on anomaly detector and local shape feature. *IEEE Trans. Geoscience Remote Sens.* 52, 4511–4523. doi:10.1109/TGRS.2013.2282355

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; 2024 June 16–22; Seattle, WA: IEEE. p. 10781–10790.

Vermote, E., and Vermeulen, A. (1999). *Atmospheric correction algorithm: spectral reflectances (mod09)*. Greenbelt, Maryland: National Aeronautics and Space Administration (NASA) Goddard Space Flight Center. p. 1–107. ATBD version 4.

Wang, N., and Ma, F. (2021). Single ship target detection based on the concept of edge computing. In: 2021 6th International Conference on Transportation Information and Safety (ICTIS); 2021 October 22–24: IEEE. p. 559–564.

Wang, Y., Ma, L., and Tian, Y. (2011). State-of-the-art of ship detection and recognition in optical remotely sensed imagery. *Acta Autom. Sin.* 37, 1029–1039.

Wu, S., Zhang, Y., Tian, T., and Tian, J. (2023). Dfri: detection and fine-grained recognition integrated network for inshore ship. In: IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium (IEEE); 2021 October 22–24: IEEE. p. 5535–5538. doi:10.1109/igarss52108.2023.10283133

Yan, J., Wang, S., Feng, J., He, H., Wang, L., Sun, Z., et al. (2025). New 30-m resolution dataset reveals declining soil erosion with regional increases across Chinese mainland (1990–2022). *Remote Sens. Environ.* 323, 114681. doi:10.1016/j.rse.2025.114681

Ye, C., and Li, C. (2005). Application of hsi based on visual attention model in ship detection. *J. Xiamen Univ.* 44, 484–488.

Yue, T., Yang, Y., and Niu, J.-M. (2021). A light-weight ship detection and recognition method based on yolov4. In: 2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE) (IEEE); 2021 March 26–28: IEEE. p. 661–670.

Yun, S., and Ro, Y. (2024). Shvit: single-head vision transformer with memory efficient macro design. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2024 June 16–22; Seattle, WA: IEEE. p. 5756–5767. doi:10.1109/cvpr52733.2024.00550

Zhang, M., Chen, F., Guan, W., and Zhao, H. (2026). Rapid thinning of lake ice for himalayan glacial lakes since 2010. *Remote Sens. Environ.* 332, 115062. doi:10.1016/j.rse.2025.115062

Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., et al. (2024). Detrs beat yolos on real-time object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2024 June 16–22; Seattle, WA: IEEE. p. 16965–16974. doi:10.1109/cvpr52733.2024.01605

Zhou, X., Wang, D., and Krähenbühl, P. (2019). *Objects as points*. arXiv preprint arXiv:1904.07850.

Zhou, Q., Wang, Q., Bao, Y., Kong, L., Jin, X., and Ou, W. (2022). Laednet: a lightweight attention encoder–decoder network for ultrasound medical image segmentation. *Comput. Electr. Eng.* 99, 107777. doi:10.1016/j.compeleceng.2022.107777

Zhu, C., Zhou, H., Wang, R., and Guo, J. (2010). A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features. *IEEE Trans. Geoscience Remote Sens.* 48, 3446–3456. doi:10.1109/tgrs.2010.2046330

Zhu, Y., Zhang, X., Jiang, B., Su, C., Wang, M., and Zhang, K. (2024). Comparison of performance of fusion methods for sentinel-2 and gf-2 satellite images and the potential in land use classification: a case study from reservoirs in hainan island, China. In: .2024 12th International Conference on Agro-Geoinformatics (Agro-Geoinformatics); 2024 June 15–18; Novi Sad, Serbia: IEEE. p. 1–1.