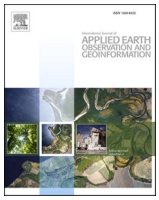
International Journal of Applied Earth Observation and Geoinformation 134 (2024) 104137

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



YOLOShipTracker: Tracking ships in SAR images using lightweight YOLOv8

Muhammad Yasir [a], Shanwei Liu [a,*], Saied Pirasteh [b,c], Mingming Xu [a], Hui Sheng [a], Jianhua Wan [a], Felipe A.P. de Figueiredo [d], Fernando J. Aguilar [e], Jonathan Li [f]

[a] College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China
[b] Institute of Artificial Intelligence, Shaoxing University, Shaoxing, 508 West Huancheng Road, Yuecheng District, Zhejiang Province 312000, China
[c] Department of Geotechnics and Geomatics, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India
[d] National Institute of Telecommunications - INATEL - Santa Rita do Sapucaí, Minas Gerais, Brazil
[e] Department of Engineering and CIAIMBITAL Research Centre, University of Almería, Carretera de Sacramento s/n, 04120 Almería, Spain
[f] Department of Geography and Environmental Management and Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada

ARTICLE INFO

Keywords:
YOLOShipTracker
YOLOv8n
Lightweight model
Ship tracking
Short-time SAR images

ABSTRACT

This paper presents a novel approach to tracking ships in Synthetic Aperture Radar (SAR) images based on an improved lightweight YOLOv8 Nano (YOLOv8n), specially devised to improve efficiency without compromising accuracy. In our method, we replaced the heavy backbone and neck of YOLOv8 with HGNetv2 and slim-neck, respectively. We also implemented a lightweight decoupling head using EMSConvP. Additionally, we integrated a knowledge distillation module to further enhance detection capabilities. Furthermore, we conducted extensive experiments on the short-time sequence SAR dataset to demonstrate superior accuracy metrics compared to the original YOLOv8n model. Regarding tracking ships in SAR images, we developed a multi-object tracking (MOT) technique called Cascaded-Buffered IoU (C-BIoU). This method enlarges the detection and trajectory matching space by increasing the buffer zone, effectively combining detection and trajectory information from short-time sequence SAR images. The findings reveal that our method significantly reduces the computational complexity, parameters, and model size by up to 54.7 %, 68.4 %, and 68.3 %, respectively, with respect to the original model metrics. As a direct consequence of these reductions, our proposed model demonstrates a remarkable 133.1 % improvement in image processing speed expressed as frames per second (FPS). Moreover, Our C-BIoU method shows outstanding performance in tracking accuracy and efficiency, with superior Higher Order Tracking Accuracy (HOTA), Multiple Object Tracking Precision (MOTP), and Identification F1 score (IDF1) scores of 72.8 %, 87.9 %, and 80.7 %, respectively, compared to existing tracking algorithms. The results from testing on multiple datasets highlight our method's excellent performance in ship detection and tracking, offering high-speed processing capabilities with an average image processing speed of 81 FPS. In this sense, this method provides reliable real-time monitoring and management of maritime traffic, enhancing situational awareness for maritime operations.

1. Introduction

Ship tracking is central in maritime operations and covers various activities, from ship navigation and logistics to safety and environmental monitoring. Among the variety of technologies used for vessel tracking, Synthetic Aperture Radar (SAR) imaging has emerged as a cornerstone due to its unique all-weather, day, and night surveillance capabilities (Zhang et al., 2022a,b; Mao et al., 2023; Gao et al., 2023; Ma et al., 2024; Yasir et al., 2024c). Unlike optical imaging systems, SAR is not affected by atmospheric conditions or darkness, making it ideal for maritime surveillance where continuous monitoring is critical (Zhang et al., 2020a; Zhang et al., 2021a; Mao et al., 2022; Zha et al., 2023; Yasir et al., 2024a). In this sense, SAR imagery has revolutionized ship tracking by providing unparalleled coverage and resolution over vast ocean areas in recent years. Indeed, SAR imagery offers clear advantages, including detecting ships regardless of weather conditions, distinguishing between different types of ships based on their radar signature, and tracking ship movements with high precision (Yasir et al., 2024b).

In recent years, deep learning technology has rapidly developed and found wide application in various fields such as computer vision (CV),

* Corresponding author.
E-mail address: 20110052@upc.edu.cn (S. Liu).

https://doi.org/10.1016/j.jag.2024.104137
Received 10 June 2024; Received in revised form 18 August 2024; Accepted 31 August 2024
Available online 11 September 2024

natural language processing (NLP), speech and audio processing, medicine, and biology (O'Shea and Hoydis, 2017; Aceto et al., 2019). Among deep learning architectures, convolutional neural networks (CNNs) stand out, particularly in image data processing, because they can automatically extract local features from images through convolutional layers. This feature extraction capability enables CNNs to effectively capture various visual patterns, from simple edges and textures to complex object shapes and scene structures (Alzubaidi et al., 2021). Some researchers have applied computer vision and image processing, particularly for general deep learning applications. For example, (O'Shea and Hoydis, 2017) discussed the broad application of deep learning in CV, NLP, speech and audio processing, medicine, and biology. Moreover, (Aceto et al., 2019) explored the extensive use of deep learning across various fields. In addition, (Kumar and Renuka, 2023) highlighted the development and applications of deep learning technologies in multiple domains. Other researchers applied specific image processing techniques. For example, Alzubaidi et al. (2021) focused on the capability of CNNs to capture visual patterns through convolutional layers, enhancing their performance in image data processing, while Chen et al. (2023) developed a rotational ship detector based on the widely known object detection deep learning algorithm YOLOv5 for extracting rotating ships from maritime surveillance videos.

Applications in ship detection using SAR imagery are among the recent researches. For example, Wei et al. (2020) introduced HR-SDNet for ship detection in high-resolution SAR images. In the same way, Li et al. (2019) combined features at different levels to improve ship detection at various sizes. Another example is Wang et al. (2020), who incorporated an attention module (SENet) to reduce scattered radiation and highlight targets. Later, Gao et al. (2020) proposed a dense attention feature aggregation network with deeply differentiable convolution to improve efficiency with fewer parameters. In the same context, Zhou et al. (2022) developed MSSDNet, a multi-scale hull detection network capable of balancing model size and detection accuracy. Further, Zhu et al. (2022) applied fully convolutional single-stage target detection, eliminating anchor points through pixel-by-pixel detection. Recently, Wen et al. (2024) fused deep learning with traditional hand-crafted features by integrating Constant False Alarm Rate (CFAR) features into the YOLOv5s model. In addition to this, integration with auxiliary data is one of the techniques that researchers have recently considered. For example, Dechesne et al. (2019) used deep learning for ship identification and characterization, synergizing Sentinel-1 SAR images with Automatic Identification System (AIS) data to demonstrate the value of integrating multiple data sources to improve accuracy. In this same way, Chen et al. (2024) collected data from AIS in New York Harbor in order to analyze the spatiotemporal distribution of pollutant emissions from ships. This classification showcases the broad applications of deep learning technologies, particularly CNNs, in various fields, also emphasizing their specific use in enhancing ship detection and tracking using SAR imagery.

Recently, researchers have focused on developing lightweight models to make ship detection in SAR images more practical (i.e., real-time processing capability). These efforts have led to various innovative approaches to reduce model complexity while maintaining or enhancing detection capabilities. For example, making modifications to existing models such as RetinaNet and YOLOv4. In this sense, (Miao et al., 2022) modified RetinaNet by replacing shallow convolutional layers and reducing deep convolutional layers. This effectively reduced both floating-point operations and model parameter counts while maintaining good ship detection capability. In another study, Liu et al. (2022) introduced YOLOv4-LITE, a lightweight ship detection model utilizing MobileNetv2 as the backbone network and incorporating RFB for enhanced feature extraction, achieving a model size of 49.34 M. Moreover, some researchers enhanced YOLO models like YOLOv5. For example, Xu et al. (2022) devised a lightweight and accuracy-enhancing module integrated into the YOLOv5 model, applying network pruning to shrink the model to only 14.18 % of its original size. Another example is the one developed by Xiong et al. (2022), who revamped the pyramid pooling structure and integrated attention mechanisms into the YOLOv5n model. As for YOLOv7, Tang et al. (2024) enhanced the YOLOv7 model by integrating AMMRF, a multi-scale feeling field convolution block, to capture the relationship between vessels and their background. Accordingly, (Yu et al., 2022) simplified YOLOX-s's complex pyramid structure and introduced the Residual Asymmetric Dilated Convolution (RADC) block to enhance semantic information extraction from SAR images. Furthermore, regarding new lightweight models, (Ren et al., 2023) introduced a lightweight feature-enhanced backbone called LFEBNet to reduce computational costs. Later, optimized ship detectors were proposed through a hybrid approach centered on data and model improvements (Humayun et al., 2024). These classified investigations underscore the progress made in developing lightweight detection and tracking models for identifying ships in SAR images, highlighting the various strategies researchers have employed to reduce complexity and improve efficiency.

Deep neural network-based algorithms for multiple object tracking (MOT) (Chu et al., 2017) are commonly employed to detect and track multiple targets using optical image features, resulting in efficient tracking capabilities. The Simple Online and Real-Time Tracking (SORT) algorithm (Bewley et al., 2016) stands out as a prominent MOT algorithm capable of swiftly tracking multiple targets using optical images. This algorithm leverages the widely known Faster R-CNN object detection model to detect targets, followed by applying Kalman filtering and the Hungarian algorithm to track them effectively. Building upon the foundation laid by SORT, other MOT algorithms have emerged. For instance, the Deep-SORT algorithm (Wojke et al., 2017) integrates visual characteristics into the association module of the SORT algorithm to improve tracking accuracy. Conversely, the MOTDT method (Chen et al., 2018) constructs a scoring mechanism using the region-based fully convolutional network (R-FCN) to assess potential targets, utilizing a cascade association approach to enhance tracking resilience. Nevertheless, conventional MOT techniques that rely on optical imaging of ship targets may experience notable impacts due to complex maritime environments, such as adverse weather conditions encountered at sea. On the contrary, SAR imaging provides continuous surveillance regardless of weather conditions and operates effectively over long distances due to the exceptional penetration capabilities of the radar beam (Zhang et al., 2022a,b; Chen and Yu, 2023; B. Zhang et al., 2023). SAR images not only capture target features but also furnish vital positional data, making them especially well-suited for detecting and tracking targets in intricate maritime environments. However, utilizing MOT techniques on SAR images to detect multiple targets poses various challenges. Unlike their optical counterparts, SAR images lack color information due to grayscale representations, potentially leading to information loss. Moreover, the presence of cluttered backgrounds like sea clutter and land regions in SAR images can trigger false alarms during detection. Additionally, fluctuations in ship movements may cause defocusing in certain image areas, resulting in tracking box deviations that compromise tracking accuracy. Furthermore, variations in target resolution, height, and angle during SAR imaging hinder the acquisition of adequate prior samples for training, negatively affecting the network's overall performance (Zhang et al., 2023).

While deep-learning-based ship identification boasts impressive detection capabilities, applying these methods directly to SAR images poses several challenges, as highlighted by several authors (Zhang et al., 2020b; Zhang et al., 2021b; Zhao et al., 2023; Zhang et al., 2024). SAR imaging introduces unique characteristics such as reduced contrast, increased scattering noise, and interference from sea debris, adding complexity to ship identification. The varying sizes and shapes of ships, particularly smaller vessels vulnerable to speckle noise, further complicate the identification process. In addition, ships may be inadequately represented by only a few pixels in large SAR scenes due to resampling, potentially compromising identification accuracy. Moreover, heavy parameterization and computational intensity hinder the

deployment of deep learning target detection models in SAR image analysis, making them problematic for real-time detection tasks. It is worth noting that previous studies in SAR image ship recognition have prioritized accuracy over detection speed, exacerbating the computational challenges.

Therefore, we introduce YOLOShipTracker to tackle these prevailing issues. In this way, YOLOShipTracker addresses the challenge of ship detection and tracking in SAR short-time sequence images by developing a lightweight object detection model based on YOLOv8n that has been optimized for real-time performance. YOLOv8n was chosen for its compact size, anchor-free detection, and optimized backbone, which are crucial for achieving high-speed and real-time performance in resource-constrained environments. YOLOv8n builds upon the foundation of YOLOv5, introducing a revamped backbone network, an Anchor-Free detection head, and an improved loss function. These improvements provide superior object detection accuracy and reduced model size, making YOLOv8n an ideal choice for the challenges faced in ship identification using SAR imagery. Nevertheless, our model introduces four main changes to the original structure of the YOLOv8n model to improve its efficiency and accuracy in detecting ships from SAR images:

i. We propose the HGNetv2 backbone, designed to improve feature extraction for SAR imagery while minimizing complexity.
ii. We incorporate a lightweight neck architecture utilizing separable convolutions to optimize feature extraction and compress the model structure.
iii. A lightweight decoupled head is implemented to maintain detection accuracy while significantly reducing model parameters.
iv. We integrate a knowledge distillation module, which mitigates accuracy loss from the lightweight design without altering the model structure.

After detection, we develop a multi-object tracking approach based on the method Cascaded-Buffered Intersection over Union (C-BIoU) proposed by Yang et al. (2023). This approach increases the detection and trajectory tracking area by adding a buffer area, effectively merging detection data with trajectory information. As a result, the C-BIoU

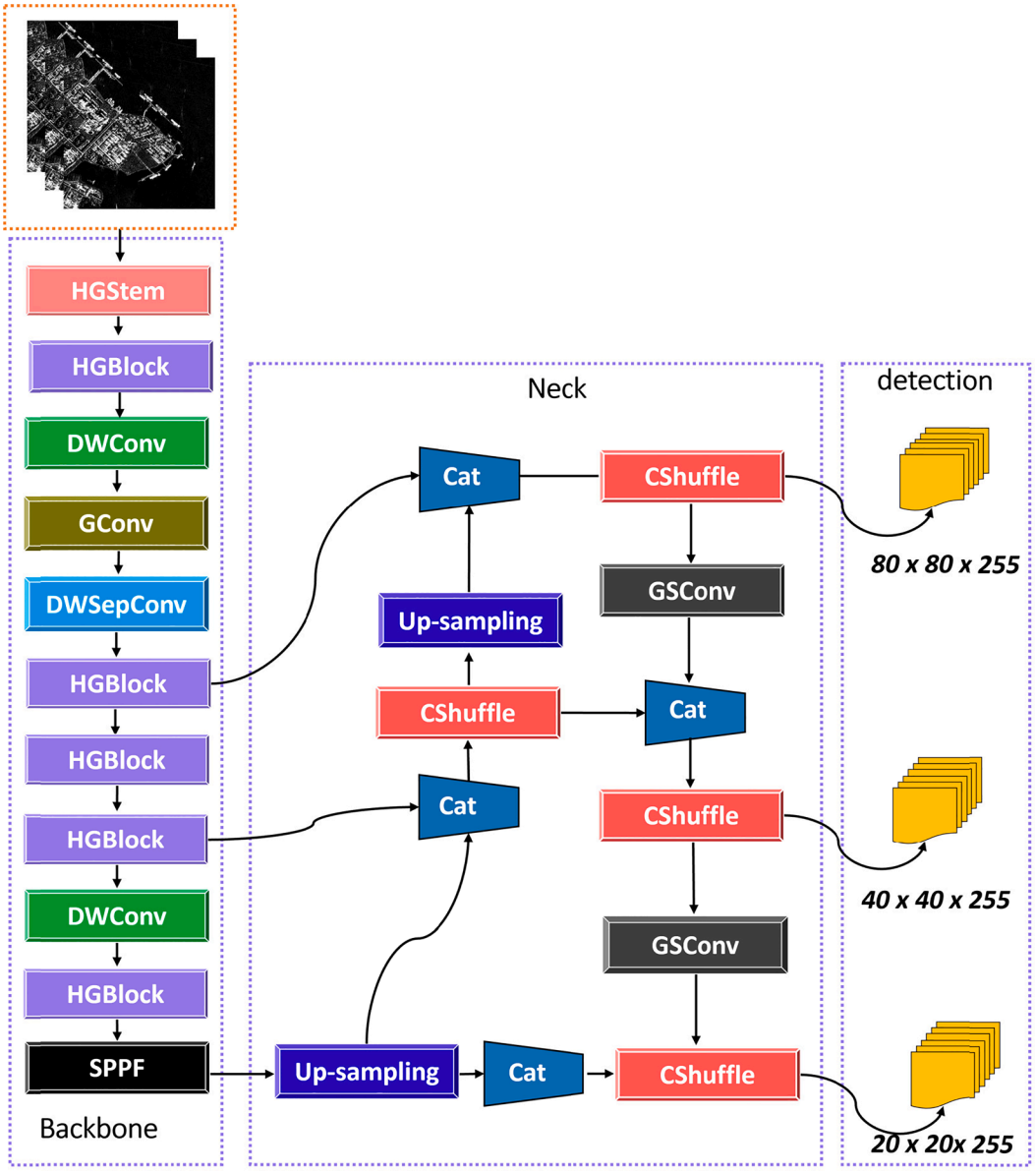


Fig. 1. The overall architecture of the proposed YOLOShipTracker model. CShuffle in red box refers to the module 'ChannelShuffle' in PyTorch, which is used to divide and rearrange channels in a tensor. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

method guarantees precise tracking of multiple ships, providing strong performance and real-time capability, which are crucial for monitoring maritime traffic and improving situational awareness.

The subsequent sections of the paper are organized as follows: Section 2 delves into an intricate examination of the proposed method and its architecture. In Section 3, we provide an overview of the SAR dataset utilized in this study, elucidate the experimental setup, detail the assessment metrics employed, analyze detection results, compare the findings with state-of-the-art (SOTA) models, discuss the conducted ablation study, and evaluate multi-object tracking metrics. Section 4 furnishes a comprehensive analysis of the obtained results and compares them with several SOTA models. Finally, Section 5 concludes the paper, summarizing key findings and presenting recommendations for future research.

2. Method

2.1. Proposed model architecture

The architecture of the proposed lightweight tracking model is presented in Fig. 1. It is composed of four key components: i) a lightweight backbone, ii) a streamlined neck, iii) a distillation module, and iv) a decoupled head designed for single-category ship detection. We began by introducing a lightweight backbone to bolster the network's feature extraction capabilities and enable multi-level feature fusion. This not only lowers the parameter count but also enhances the perception, generalization, and computational efficiency of the network. Subsequently, incorporating depth-wise separable convolution technology into the neck architecture markedly raises computational efficiency while retaining comparable feature extraction quality to standard convolution. As a result, this expedites ship target recognition in SAR images. By preserving the favorable aspects of the original YOLOv8n decoupled head, our design facilitates independent tuning and precise learning of task-specific features, thereby enhancing the effectiveness and performance of the model in single-target identification tasks. By integrating shared parameters and decoupling head structures, our detection head minimizes computational complexity while maintaining its inherent advantages. The subsequent sub-sections delve into the specifics of our proposed approach.

2.2. Lightweight backbone

Drawing inspiration from RT-DETR (Zhao et al., 2023), we integrated a lightweight network module called HGNetV2 as the backbone extraction network. The network structure of HGNetV2 is depicted in Fig. 2. Our model adheres to the single-stage object detection framework of the YOLO series, proficiently extracting features and downsampling at the model's forefront with the incorporation of HGStem blocks. HGNetV2 was selected for several reasons, specifically tailored to address the unique challenges posed by SAR imagery. Firstly, HGStem enhances computational efficiency by reducing the dimensionality of the initial feature map and removing redundant parameters. This reduction in complexity is crucial for SAR applications, where real-time processing is often required. Secondly, the HGBlock module (Fig. 3), essential to the HGNetV2 backbone, plays a crucial role in multi-scale feature extraction. Stacked HGBlocks ensure that features extracted from the network integrate information from multiple scales and depths, which is critical for managing the high variance in ship sizes and shapes found in SAR imagery. This design optimization enhances the network's multi-scale processing capability, enabling it to effectively manage scale variations in images. Such capabilities are invaluable for accurately identifying ships of various sizes in SAR imagery.

Furthermore, incorporating Depthwise Convolution (DWConv) in the backbone architecture over traditional convolutional operations significantly reduces the model parameter count, simplifying the model while maintaining efficiency. DWConv conducts convolutions independently across each channel of the input feature maps, ensuring the preservation of fine-grained details within individual channels, which is crucial for maintaining the fidelity of feature information across the network. It is important to highlight that SAR images are characterized by their high resolution and the presence of speckle noise, which can obscure fine details. The multi-scale and depth-aware feature extraction capabilities of HGNetV2 help mitigate these challenges by ensuring robust feature representation. Moreover, the lightweight nature of HGNetV2 allows for faster processing times, which is essential for real-time SAR applications.

In summary, the integration of HGNetV2 into YOLOShipTracker is not merely for its lightweight nature and multi-scale feature extraction capabilities but also for its ability to handle the specific challenges associated with SAR imagery, such as varying ship sizes, speckle noise, and the need for real-time processing. These tailored optimizations make HGNetV2 an excellent choice for improving the efficiency and accuracy of SAR ship tracking.

2.3. Streamlined neck

A lightweight neck component for YOLOv8, inspired by Slim-neck (Li et al., 2024), was devised. Considering the distinct challenges in ship detection within SAR imagery—like sea clutter, noise, and environmental variations—the YOLOv8n architecture was substantially

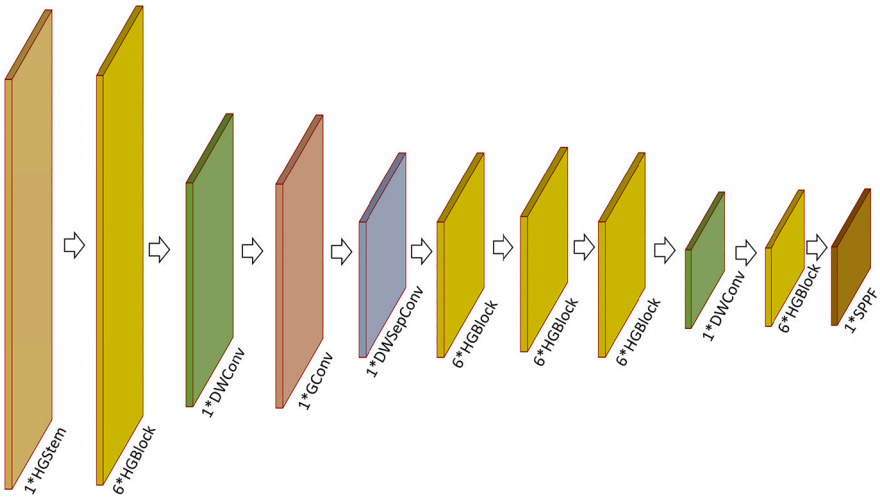


Fig. 2. The backbone structure of the proposed tracking model.

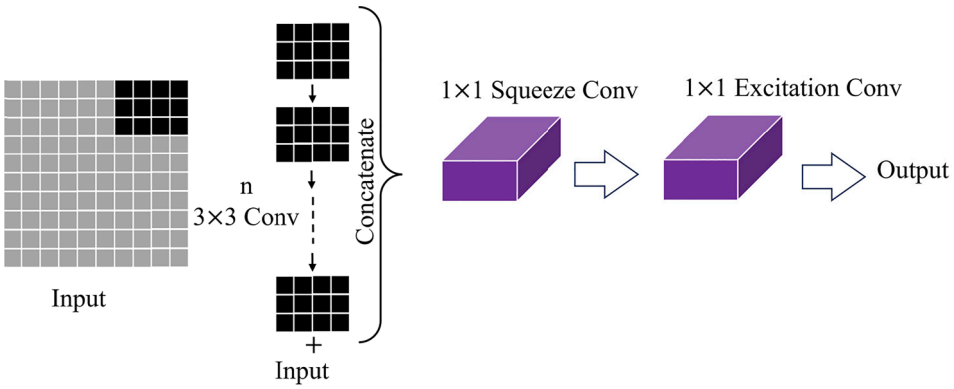


Fig. 3. The overall structure of HGBlock.

modified. These modifications are essential to enhance the model's performance and efficiency in dealing with SAR imagery's unique characteristics.

Firstly, we replaced the conventional C2F module within the YOLOv8n network's neck with the Channel Shuffle module, as depicted in Fig. 4 (a, b). The central element of this module, that is GSBottleneck, harnesses the group separable convolution (GSConv) structure. This structure is particularly suited for SAR imagery as it efficiently extracts subtle ship features while reducing the computational burden. The GSConv technique combines the outputs of a 1×1 convolution with those from a series of 1×1 and 5×5 convolutions, followed by channel shuffling. This design significantly reduces computational complexity and parameter count while maintaining the model's ability to identify ship features accurately.

Additionally, the Channel Shuffle module utilizes 1×1 convolutions to compress and integrate feature information, thus simplifying the feature representation and decreasing the model's overall complexity. This is crucial for SAR imagery, where maintaining a balance between computational efficiency and feature extraction capability is vital due to the presence of sea clutter and noise.

The modifications to the neck network, specifically the use of GSConv, provide dual benefits:

- **Computational Efficiency**: GSConv reduces the computational load and the number of parameters, making the model more suitable for deployment on resource-constrained devices. This efficiency is critical for real-time applications, such as maritime surveillance and operational responsiveness.
- **Enhanced Feature Extraction**: By incorporating GSConv and the Channel Shuffle technique, our model preserves the efficiency in feature extraction and enhances its adaptability to the diverse environmental variations present in SAR data. This ensures that subtle ship features are accurately detected even amidst noise and clutter.

2.4. Decoupled head

The decoupled head is a vital component of the YOLOv8n architecture, accounting for nearly half of the total network parameters. This significant allocation highlights its crucial role in the model's overall functionality. Each of the three decoupled heads employs a dual-branch structure, enabling efficient feature extraction across different scales. Notably, this design separates bounding box prediction from class

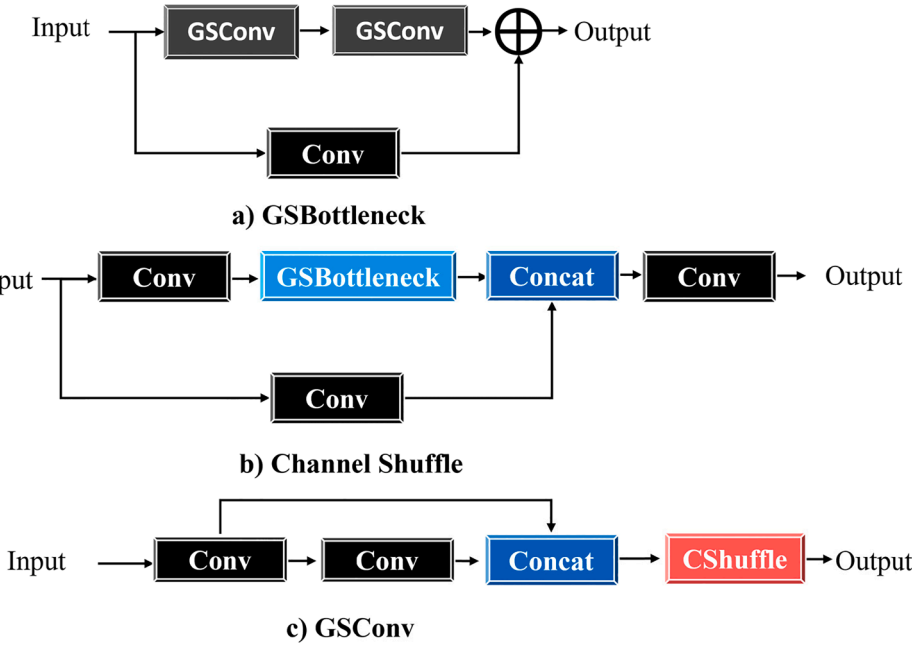


Fig. 4. The overall structure of Channel Shuffle. a) GSBottleneck technique, b) Channel Shuffle technique, and c) GSConv technique. CShuffle in red box refers to the module 'ChannelShuffle' in PyTorch, which is used to divide and rearrange channels in a tensor. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

classification tasks, resulting in more focused and effective object detection. This decoupling mechanism enhances feature extraction accuracy and improves YOLOv8n adaptability and robustness in various object detection challenges. To maintain the high feature extraction accuracy of the original decoupled head while reducing parameter count and computational load, we devised a novel decoupled head. Illustrated in Fig. 5, our design introduces EMSConvP (Efficient Multi-Scale - Convolution), which leverages convolutional filters of various sizes (1×1, 3×3, 5×5, 7×7) to process input feature maps. Afterward, 1×1 convolutions merge features from different groups, and the fused output is processed through a dual-branch structure for recognition and classification. By replacing four standard convolutions with EMSConvP, we significantly reduce computational and parameter loads while maintaining the ability to handle high-resolution feature maps. Additionally, using convolutional kernels of different sizes improves the model's capacity to detect ships of varying sizes in SAR images.

2.5. Distillation module

Knowledge distillation involves transferring knowledge from one network to another using attention mechanisms (Hinton et al., 2015; Wang et al., 2021). A commonly used technique in this process is L2 distillation (Smith et al., 2023; Popp et al., 2024), which is widely prevalent. The L2 distillation process is illustrated in Fig. 6.

In short, the knowledge distillation module transfers knowledge from a larger and well-trained teacher model to our lightweight model. This process is crucial for several reasons:

- **Enhanced Feature Extraction**: The distillation process helps improve the feature extraction capabilities of the lightweight model. By learning from the teacher model, the student model gains the ability to extract more nuanced features from SAR imagery, which is essential for accurate ship detection and tracking.
- **Compensation for Simplified Architecture**: Given the use of DWConv in our model, there might be potential loss in feature extraction capability due to the lack of direct interaction between channels. The distillation module helps mitigate this loss by transferring the detailed feature representations learned by the teacher model to the student model, thus enhancing the overall accuracy and performance.
- **Efficiency without Complexity**: The distillation process allows the lightweight model to achieve high performance without increasing its complexity. This is particularly important for real-time applications where computational resources are limited.
- **Empirical Validation**: Our empirical results (see section 3.5 about the ablation study) demonstrate that the inclusion of the knowledge distillation module significantly enhances the model's performance in SAR ship detection and tracking tasks. The specific weights used in the loss function were determined through rigorous experimentation to ensure optimal results.

The implementation applied in this work utilized a Mean Squared Error (MSE) loss function based on the L2 norm. This loss function is divided into two parts. The first part, denoted as L_{box}, corresponds to the loss associated with bounding boxes, being its Eq. (1) as follows:

$$L_{box} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\left(s_{ij}^{box} - t_{ij}^{box}\right)^{2\times\omega_i} \tag{1}$$

N represents the number of samples in a batch, while M denotes the number of bounding box parameters for each sample or the dimensionality of other structured outputs. s_{ij}^{box} signifies the predicted value of the $i-th$ bounding box parameter for the $i-th$ sample by the student model, while t_{ij}^{box} denotes the predicted value of the same bounding box parameter for the same sample by the teacher model. Additionally, ω_i represents the weight for the $i-th$ sample, which is calculated based on the maximum sigmoid value of the teacher model's classification scores and being applied to each bounding box parameter. Another part of the loss function pertains to the classification loss function, which is given by Eq. (2):

$$L_{cls} = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{C}\left(s_{ik}^{box} - t_{ik}^{box}\right)^{2\times\omega_i} \tag{2}$$

where C represents the total number of categories and s_{ik}^{box} denotes the predicted score by the student model for the $i-th$ sample belonging to the $k-th$ category. t_{ik}^{box} signifies the predicted score by the teacher model for the same sample belonging to the $k-th$ category, while ω_i represents the same sample weight as in the bounding box loss. The final distillation loss is obtained through a weighted summation of the two-loss values. Empirical validation has shown that setting the weight for the bounding box loss to 1.5 and the weight for the classification score loss to 0.5 yields superior results. Therefore, the final distillation loss function is given by Eq. (3) as follows:

$$L_{distill} = 1.5\times L_{box} + 0.5\times L_{cls} \tag{3}$$

2.6. Ship tracking model

The study proposes a MOT algorithm for maritime real-world environments, as illustrated in Fig. 7. First, the proposed lightweight version of YOLOv8n network is trained on short-time sequence SAR images to recognize patterns. This network swiftly and precisely identifies every ship in the images, providing vital data such as bounding boxes and confidence scores. These detection results are then fed into a multi-object tracking algorithm, which uses the bounding box data and

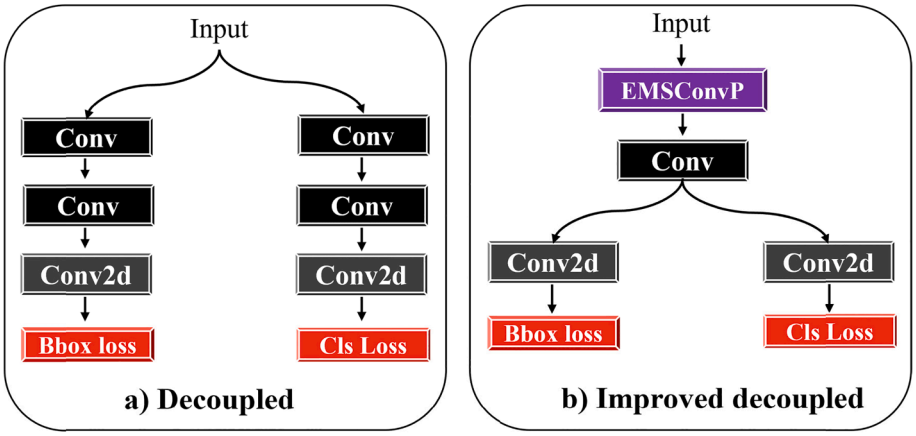


Fig. 5. Differences between the two structures. a) The original decoupled head implemented in YOLOv8n. b) Our improved decoupled head.

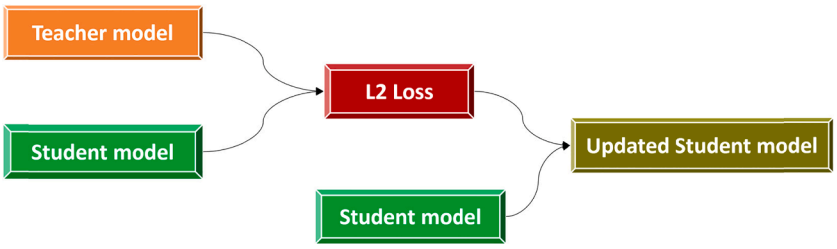


Fig. 6. The process of L2 distillation.

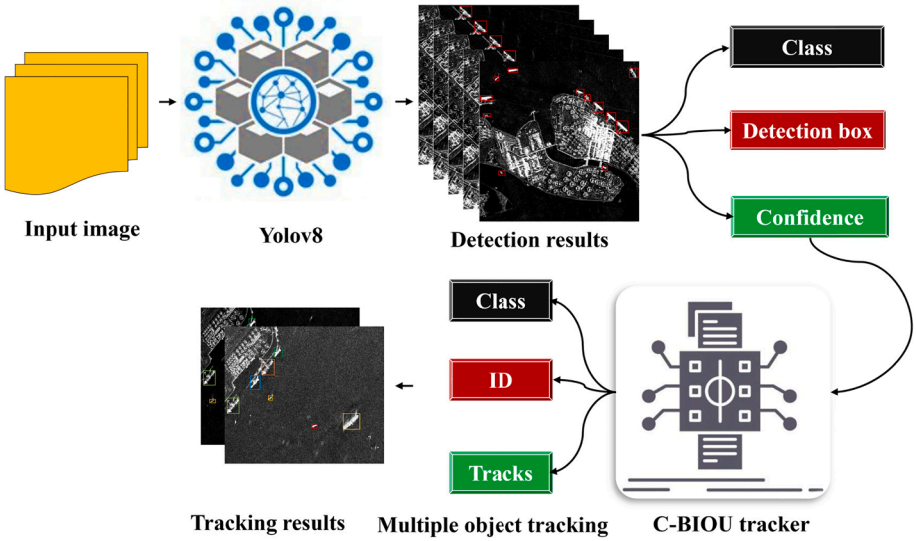


Fig. 7. The ship tracking process.

confidence scores to simultaneously track multiple ships and determine their motion trajectories. The proposed approach enhances the previous tracking algorithm by integrating ship data with target tracking trajectories.

In multi-object tracking studies, appearance consistency and geometric coherence are fundamental assumptions for linking detections between successive frames. Generally, an object should appear consistent in earlier frames if its current appearance closely matches its past appearance. Similarly, its former position and shape should match its current position and shape when taking into account its projected movement. In this way, many multi-object tracking algorithms are now incorporating re-identification modules to utilize appearance features, thereby improving tracking performance. Nevertheless, depending only on appearance can be unreliable for ships with erratic movement patterns and substantial changes in appearance. Additionally, integrating a re-identification module elevates the computational demands, slowing down the model's inference time and making deployment complicated. To tackle these obstacles, this research is based on the MOT method called Cascaded Buffered IoU (C-BIoU) (Yang et al., 2023), which widens the buffer area to increase the matching space between detections and trajectories.

This research introduces an innovative multi-ship tracking approach that expands the buffer area by enhancing trajectory detection and tracking. This expansion increases the search space and consequently enhances the effectiveness of multi-object tracking. Fig. 8 shows how the tracking procedure works. Initially, the BIoU phase uses the primary

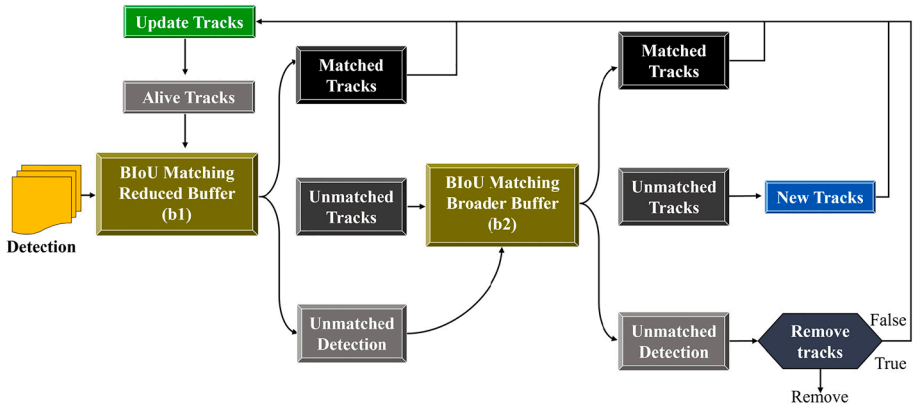


Fig. 8. Visualization of the tracking process in the C-BIoU Tracker, demonstrating the multi-stage matching strategy employing varying buffer sizes for improved trajectory association.

metric to associate active tracks and detections using a reduced buffer (b1). In the subsequent stage, a broader buffer (b2) is employed to reconcile all unmatched tracks and detections, employing the BIoU metric again. In this second stage, there is increased flexibility during the matching process, enabling consideration of objects with slight spatial variations caused by movement or other factors. The tracking method with C-BIoU enhances the robustness and precision of associating tracks with detections by utilizing two matching stages with different buffer dimensions, resulting in more reliable object tracking.

A buffer zone is integrated into the C-BIoU Tracker to improve alignment between trajectory and detection. Building upon the initial Intersection over Union (IoU) metric, Fig. 9 demonstrates the multiplication of the Buffered IoU (BIoU) by a factor. BIoU enlarges the fitting space while preserving the original detection and the center position, scale, and shape of the trajectory. This enlargement increases the chances of successful trajectory mapping, thereby enhancing overall tracking accuracy. Eq. (4) establishes the calculation of the buffered target detection zone. By encompassing a buffer region around the target bounding box in BIoU, we can assess identical but non-overlapping detection boxes and trajectories across consecutive frames. This approach strengthens the ability to maintain consistent tracking even despite small discrepancies between frames.

It is important to highlight that the buffer zone preserves the contextual details of the target, allowing the algorithm to adjust to variations in ship position and shape. To avoid excessive expansion of the matching area, a cascaded matching approach is implemented. Initially, a smaller buffer zone is employed for matching detections and trajectories, succeeded by a larger buffer zone for further matching. According to Eq. (4).), the expansion factor for the small buffer zone in this study was set to b1 = 0.3, while the expansion factor for the large buffer zone was set to b2 = 0.5.

$$ShipBox = [x - bw, y - bh, w + 2bw, h + 2bh] \tag{4}$$

where x and y represent the coordinates of the top-left corner of the original ship detection box, w and h represent the width and height of the original ship detection box, respectively, bw and bh represent the buffer width and buffer height, respectively, and b is the expansion factor.

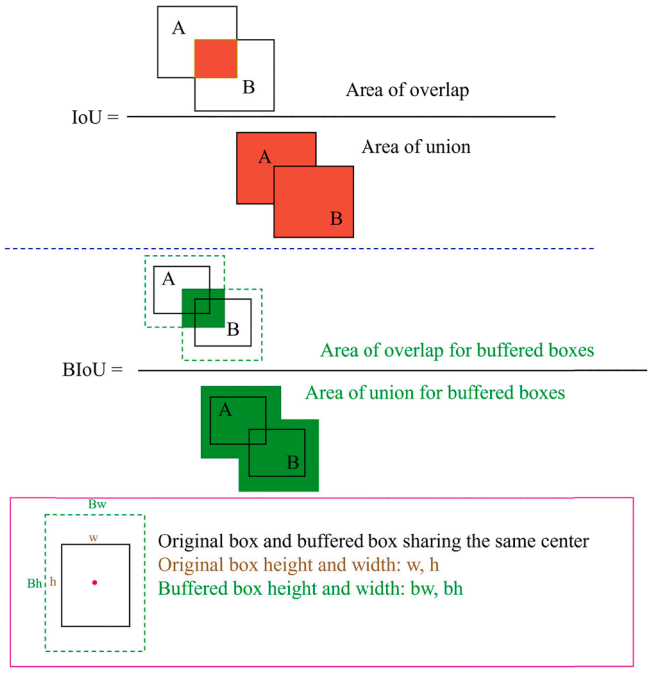


Fig. 9. Illustration of BIoU, which enlarges the matching space around detection boxes to improve trajectory association in the C-BIoU tracking method.

The C-BIoU tracker predicts motion by averaging motion data over recent frames, a deviation from conventional multi-object tracking techniques that depend on Kalman filters for state estimation. This enables it to respond swiftly to unforeseen changes in motion. If a track has matching detections at a frame t for more than n frames, its state S is updated after Δ non-matching frames. Note that $TrackBox = (x, y, w, h)$ represents the result of detection, S represents the estimated state, Δ represents the number of unmatched frames and n represents the hyperparameter used to compute the average velocity in a given time interval.

3. Results

3.1. SSTD dataset

This study utilized SAR Ship Tracking Dataset (SSTD), which includes two collections of satellite-derived SAR images covering the Zhejiang Chinese area (bordering the East China Sea) and the Strait of Malacca (southern part of the Malay Peninsula). These images were captured by Radarsat-2 and TerraSAR-X. The Radarsat-2 data was initially captured across the Strait of Malacca on January 7, 2012, at 22:51:06, employing the VV polarization mode. Simultaneously, TerraSAR-X images using VV polarization were also generated at the same time and area but with a slightly enhanced resolution of 2.00 × 2.00 m. In addition, TerraSAR-X data was collected in the Zhejiang region on December 12, 2018, at 09:19:24 in the HH polarization mode, while Radarsat-2 imagery (HH polarization) was collected in the same area and day at 10:01:00 (Yasir et al., 2024b). The main features of the employed datasets are listed in Table 1.

3.2. Experimental setup

The experimental setup for our study consisted of an Intel(R) Core (TM) i9-13980HX CPU paired with an NVIDIA GeForce RTX 4080 GPU, which had 12 GB of RAM. Python 3.10 served as the programming language of choice. We used PyTorch version 2.1.0 together with CUDA 11.8 for GPU acceleration. The initial learning rate was 0.01. We used the SGD optimizer with a momentum of 0.937 and a weight drop of 0.0005 for the optimization. To improve the model's performance and avoid overfitting, we iterated over 300 epochs. These initial hyperparameters were refined through iterative experimentation.

3.3. Evaluation metrics

Several evaluation metrics to assess our model's performance were employed in this study (Everingham et al., 2010). These metrics encompass Precision (P), Recall (R), F1-score (F1), Mean Average Precision at a 50 % IoU threshold (mAP50), Floating Point Operations per Second (FLOPs) in Gigaflops (G), Parameters (Params), Model Size, and Frames Per Second (FPS). This comprehensive set of metrics was selected to comprehensively evaluate the model's accuracy, efficiency, and real-time detection capabilities from various perspectives. The accuracy metrics used in this work are defined in Eq. (5):

$$\begin{cases} P = \dfrac{N_{TP}}{N_{TP} + N_{FP}} \\ R = \dfrac{N_{TP}}{N_{TP} + N_{FN}} \\ F1 = 2 \times \dfrac{P \times R}{P + R} \\ mAP50 = \displaystyle\int_0^1 P(R)\,dR \end{cases} \tag{5}$$

where N_{TP} represents the number of true positives or correctly detected

Table 1
List of datasets used in this study.

Satellite	Location	Imaging time(local time)	Time difference	Spatial resolution (m)	Polarization mode
RADARSAT-2	Malacca Strait	January 7, 201222:51:06	152 s	2.50*2.50	VV
TerraSAR-X	Malacca Strait	January 7, 201222:53:38		2.00*2.00	VV
TerraSAR-X	Zhejiang	December 12, 201809:19:24	41 min 36 s	2.08*2.08	HH
RADARSAT-2	Zhejiang	December 12, 201810:01:00		1.95*1.95	HH

ships, N_{FP} is the number of false positives or objects incorrectly classified as ships (commission error), and N_{FN} corresponds to the number of false negatives or not detected ships (omission error). Note that while R measures the under-detection error (related to omission error), P is focused on the over-detection error (commission error). Finally, the *F1-score* considers the overall performance by considering both under- and over-detection errors.

The mAP50 threshold evaluates the detection accuracy of the model by computing the mean average precision at an IoU threshold of 50 % across all object categories. Using metrics such as FLOPs, Parameters, Model size, and FPS aims to assess our model's portability concerning computational demand, parameter count, model size, and processing speed.

3.4. Results of ship detection

Table 2 depicts the results of the quantitative performance comparison between the proposed model (YOLOShipTracker) and the baseline model (YOLOv8n). It is notable that the proposed model exhibited enhancements across various metrics. Specifically, the precision metric witnessed a marginal increase of 3.7 %, rising from 89.2 % to 92.9 %. Additionally, there was a notable improvement of 5.1 % in the recall metric, ascending from 85.6 % to 90.7 %, while the F1 metric experienced a significant boost of 4.4 %, elevating from 87.4 % to 91.8 %. Finally, the mAP50 metric registered a slight uptick of 5.3 %, climbing from 90.1 % to 95.4 %. Table 2 also compares specifications between the proposed model and the base model, indicating that the proposed model features a considerably lighter design, resulting in a reduction of approximately 45.7 % in the number of FLOPs or computational costs. Similarly, the enhanced model also has a lower number of parameters (2.05 million versus 3.02 million) and a smaller model size (3.92 MB versus 5.63 MB), but has a capacity of processing up to 339.7 FPS, that is, approximately 25.2 % higher than the base model. In summary, the results show that the proposed model significantly reduces the number of parameters, calculations, and model size without reducing the accuracy. On the contrary, it even increases all metrics used to evaluate accuracy. Therefore, the proposed model can be considered comprehensively better than the baseline model.

Fig. 10 shows the F1 confidence curves of the baseline model and the proposed model. It can be seen that the F1-score of our model is higher than that of the baseline model when the confidence value is greater than 0.4. Because the F1-score is a comprehensive evaluation indicator of Precision and Recall, this not only indicates that the proposed model has better detection performance at high confidence thresholds but also shows a more balanced performance with respect to commission and omission errors. It is important to note that high confidence is generally required to ensure the model's reliability in ship detection from SAR images. In this case, our model is clearly more reliable than the baseline model. To illustrate the difference between our model and the original YOLOv8n in terms of ship detection from SAR images, we deliberately selected two images from the dataset that are known for their complicated backgrounds. Fig. 11 depicts that the baseline model has problems with false detections, especially when facing multiple offshore vessels that cannot detect all vessels near the coast. In contrast, the enhanced model identifies every vessel in these images, albeit with slightly less emphasis on the vessel in the lower right corner of the image. For better understanding and comparison, we highlight the miss detection with the yellow circle and false detection with a light blue color within the figures.

We thoroughly compared the YOLOShipTracker model with several state-of-the-art algorithms (SOTA). The models to be compared were carefully selected based on their relevance, popularity, and applicability to the task at hand. Extensive experimentation on our datasets and rigorous evaluation using performance metrics such as accuracy, precision, and computational efficiency were carried out to analyze the potential of the model proposed in this work compared to several SOTA models (see Table 3). This analysis not only highlights the strengths of our approach but also provides valuable insights into its comparative performance and potential advantages over existing solutions. As shown in Table 3, the proposed model not only yields clearly better accuracy metrics in terms of P, R, F1 and mAP50, but also outperforms other evaluation indicators, namely FLOPs, number of parameters to estimate, model size, and FPS. Remarkably, its computational complexity is only 4.5 GFLOPs, which means a reduction of 81 % compared to the second-ranked YOLOv5 model. The number of parameters to be estimated in the proposed model is 2.05 million, which is only 22.5 % of the second-ranked YOLOv5 model. Similarly, the new model presents a size of 3.92 MB, which is only 22.5 % of the second-placed YOLOv5 model. Finally, our model achieves a frame rate of 339.6 FPS, that is, 87.2 FPS higher than that of the second-placed YOLOv7 model. These results show that our model strikes a balance between accuracy and efficiency. Compared to other models, the proposed model is more suitable for use on edge devices as it ensures accuracy while enabling real-time detection.

3.5. Ablation study

An ablation study is an important aspect of scientific research in which certain system components are systematically removed or "ablated" to evaluate their individual effects on overall performance. By isolating these components, researchers can understand their individual contributions, leading to more informed decisions about system design or optimization. In this case, the ablation experiment aimed at dissecting the impact of the YOLOShipTracker model's different modules on accuracy metrics and computational cost (FLOPs, parameters, and model size). As illustrated in Table 4, each proposed module serves as a lightweight component, effectively reducing the model's operational load, parameter count, and overall size to varying extents. Notably, including the knowledge distillation module does not alter the structural data of the model but directly influences its recognition performance.

Table 2
Performance and specifications comparison between the proposed YOLOShipTracker model and the baseline model (YOLOv8n).

Model	P (%)	R(%)	F1(%)	mAP50 (%)	FLOPs(G)	Params(M)	ModelSize (MB)	FPS
Baseline model(YOLOv8n)	89.2	85.6	87.4	90.1	8.3	3.02	5.63	254.1
YOLOShipTracker	**92.9**	**90.7**	**91.8**	**95.4**	**4.5**	**2.05**	**3.92**	**339.7**

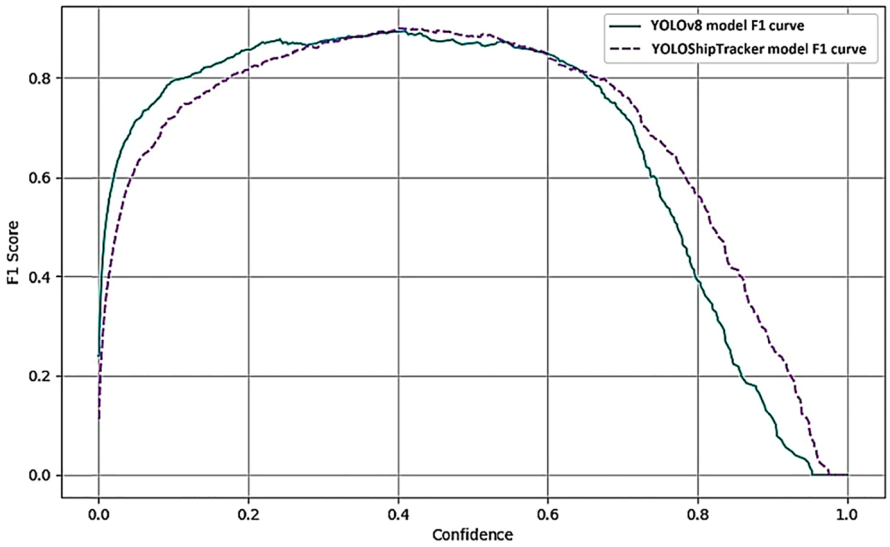


Fig. 10. F1-score confidence curves of the baseline model and the proposed YOLOShipTracker model.

Consequently, upon amalgamating these modules to construct our proposed model, the final model has witnessed a significant reduction in the model's structural parameters. The experimental outcomes underscore the lightweight nature of our holistic model architecture design.

Adding various modules to the baseline model has led to alterations in the accuracy evaluation metrics, albeit without significant changes. Each lightweight module inclusion increased the mAP50 metric. Upon integrating HGNetv2, Slim-neck, and Efficient Head (plus distillation module), our proposed model demonstrated a 3.7 % enhancement in precision (rising from 89.2 % to 92.9 %) and a 5.1 % increase in recall (from 85.6 % to 90.7 %). Although the F1-score increased by 4.4 % (from 87.4 % to 91.8 %), the mAP50 experienced a more noticeable rise of 5.3 % (from 90.1 % to 95.4 %). These findings underline the effectiveness of our model structure in improving the detection accuracy for ship detections in SAR images.

3.6. Multi-object tracking evaluation metrics

The primary evaluation metrics for ship multi-object tracking used in this work were the following: Higher Order Tracking Accuracy (HOTA), Multi-Object Tracking Accuracy (MOTA), Multiple Object Tracking Precision (MOTP), and IDF1 (similar to F1-score in tracking performance). HOTA, MOTA, MOTP and IDF1 are calculated as expressed in Eq. (6).

$$\begin{cases} HOTA = \sqrt{DetA.AssA} = \sqrt{\dfrac{\sum_{C\in TP}A(c)}{TP+FN+FP}} \\ A(C) = \dfrac{TPA(c)}{TPA(c)+FPA(c)+FNA(c)} \\ MOTA = 1 - \dfrac{\sum_t FP+FN+IDS}{\sum_t g_t} \\ MOTP = \dfrac{\sum_{t,i} d_{i,t}}{\sum_t c_t} \times 100\% \\ IDF1 = \dfrac{2IDTP}{2\mathrm{IDTP}+IDFP+IDFN} \times 100\% \end{cases} \tag{6}$$

In Eq. (6), HOTA is calculated as the square root of the product of DetA (detection accuracy score) and AssA (association accuracy score). Being C a point belonging to True Positives (TP), $A(C)$ represents the association accuracy, that is, how well a tracker links detections over time into the same identities (IDs), given the ground-truth set of identity links in the ground-truth tracks. Next, the three metrics needed to calculate $A(C)$ are described. TPA (true positive associations) is the number of true positive matches between the two tracks. Any remaining detections in the predicted track (which are either matched to other ground-truth tracks or none at all) are False Positive Associations (FPA), while any remaining detections in the ground-truth track would be False Negative Associations (FNA). Finally, TP refers to the number of correctly matched targets, FN represents actual targets mistakenly predicted as negative, and FP is the number of samples erroneously predicted as positive by the model.

Regarding MOTA calculation (Eq. (6)), FP denotes the cumulative false detections within a specific frame t, while FN indicates the total number of detections that were missed in the same frame. IDS is used to quantify the number of times that a tracked object identification number changes during the course of tracking within frame t, while g_t represents the count of actual targets present in frame t.

For the MOTP calculation in Eq. (6), i identifies the currently considered detection target. The term ct refers to the number of successful matches between identified targets and their estimated positions in frame t, and $d_{i,t}$ measures the distance between a particular detection and its estimated position at frame t.

IDF1 is calculated in Eq. (6) from IDTP, IDFP and IDFN metrics. IDTP is the complete count of targets accurately tracked without any change in their identification. IDFP represents the total count of targets mistakenly tracked without any change in their identification. IDFN is defined as the total count of targets lost during tracking without any change in their identification.

Furthermore, we also evaluate the model performance using two additional metrics: the number of Identity Switches (IDS) and Frames Per Second (FPS). Higher values of HOTA, MOTA, MOTP, IDF1, and FPS, along with lower values of IDS, indicate better model performance.

3.7. Results of ship tracking

To assess the effectiveness of the proposed buffer IoU (BIoU) method for multi-ship tracking, we conducted experiments using BIoU as a matching metric, comparing the obtained results to those obtained without including buffers (i.e., using IoU) for the ship tracking dataset. In both cases, the YOLOv8n enhanced lightweight model described and validated in the last sections was used to accomplish ship detection on each SAR image frame. Table 5 summarises the results of these experiments, demonstrating that employing BIoU as a criterion of matching provides notable benefits in terms of accuracy and consistency when monitoring multiple vessels. As shown in Fig. 12, visual comparisons

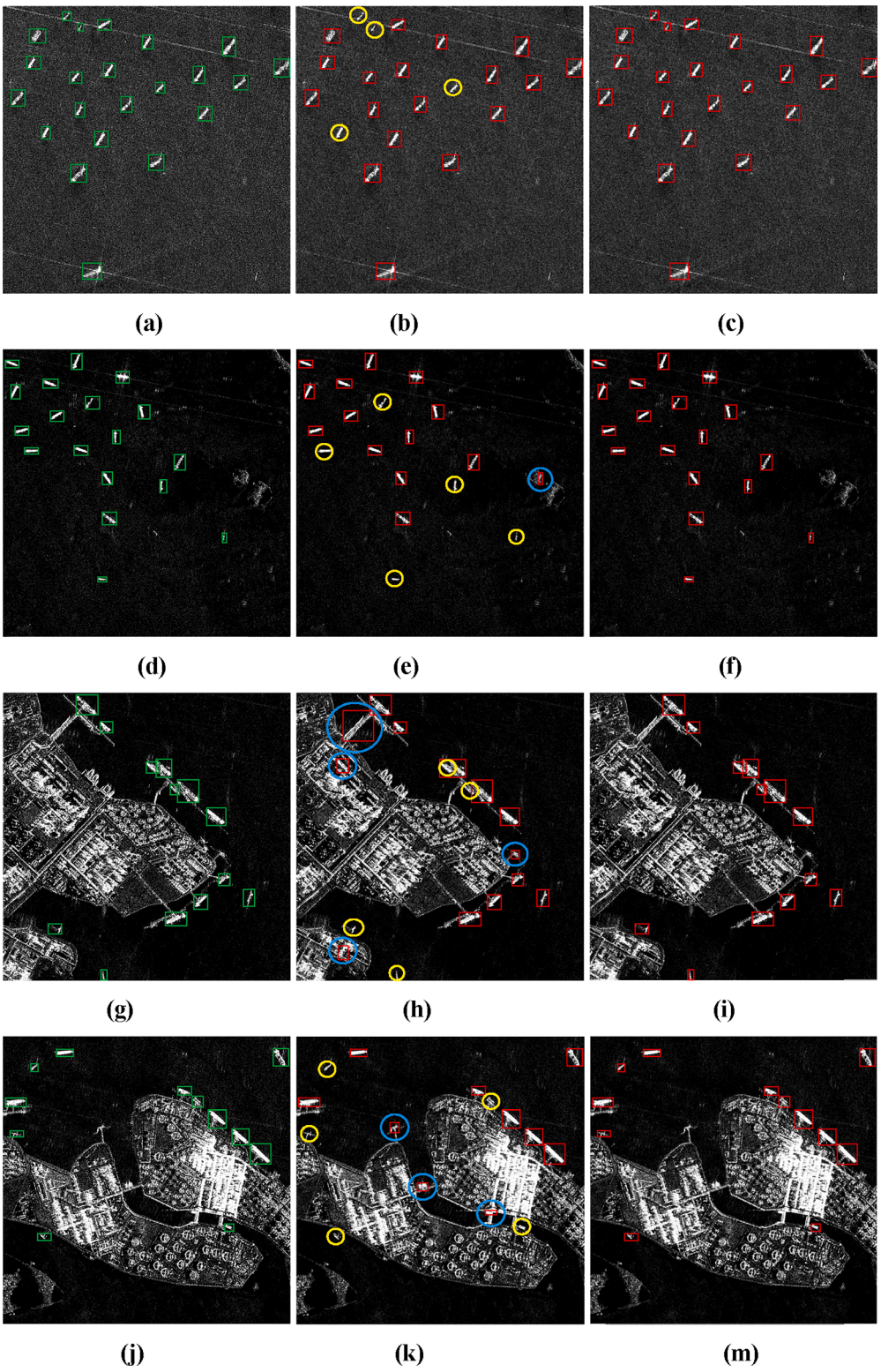


Fig. 11. Comparison detection results of the different models: (a, d, g, j) ground truth (green boxes), (b, e, h, k) baseline model, and (c, f, I, m) display YOLOShipTracker model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

further confirm the quantitative analysis in Table 5. In fact, Fig. 12(a) depicts some ID changes and track breaks when the buffer zone extension was not applied during the tracking process. However, the tracking ID remains consistent when BIoU is employed, indicating higher tracking stability. Furthermore, Fig. 12(b) shows that the number of missed detections decreases when BIoU is used for tracking. It is important to highlight that BIoU significantly widens the area surrounding the object's bounding box, enhancing the chances of trajectories and detection boxes in nearby frames intersecting or aligning. This allows for accurate matching even if the position or shape of the target changes.

To evaluate the effectiveness of our proposed approach in tracking

Table 3
Performance and specifications comparison between several baseline models and the proposed tracking model.

Model	P(%)	R(%)	F1 (%)	mAP50 (%)	FLOPs (G)	Params (M)	Model Size (MB)	FPS
RetinaNet (Lin et al., 2017)	89.1	56.8	69.4	63.7	170.0	139.1	524.6	64.8
EfficientDet-D1(Tan et al., 2019)	68.7	82.3	74.9	74.1	105.3	29.7	55.2	58.9
Faster R-CNN (Girshick, 2015)	69.2	76.6	72.8	64.8	369.0	136.7	522.3	38.7
SDD (Liu et al., 2015)	72.1	88.6	79.8	85.6	60.7	23.7	94.0	242.0
YOLOv5 (Zhao et al., 2023)	89.2	77.9	83.1	89.5	23.8	9.1	17.6	239.1
YOLOv7 (Wang et al., 2022)	87.0	79.3	82.9	88.2	44.2	16.2	32.1	252.5
Our Model	**92.9**	**90.7**	**91.8**	**95.4**	**4.5**	**2.05**	**3.92**	**339.7**

Table 4
Results of the ablation study for the YOLOShipTracker model.

Baseline model	HGNetv2	Slimneck	Efficient head	Distillation	P(%)	R(%)	F1 (%)	mAP50 (%)	FLOPs (G)	Params (M)	Model size (MB)
YOLOv8n	×	×	×	×	89.2	85.6	87.4	90.1	8.3	3.02	5.63
YOLOv8n	✓	×	×	×	87.8	90.6	89.1	91.9	6.4	2.62	4.65
YOLOv8n	✓	✓	×	×	82.1	87.7	84.9	91.3	6.5	2.17	5.43
YOLOv8n	✓	✓	✓	×	84.3	90.1	87.1	91.6	4.5	2.05	3.98
YOLOv8n	✓	✓	✓	✓	**92.9**	**90.7**	**91.8**	**95.4**	**4.5**	2.05	3.92

Table 5
Comparison results between IoU and BIoU ship tracking approaches.

Model	HOTA (%)	MOTA (%)	MOTP (%)	IDF1 (%)	FPS
IoU	70.8	78.1	86.6	79.3	79
BIoU	**73.6**	**80.9**	**87.5**	**82.1**	**82**

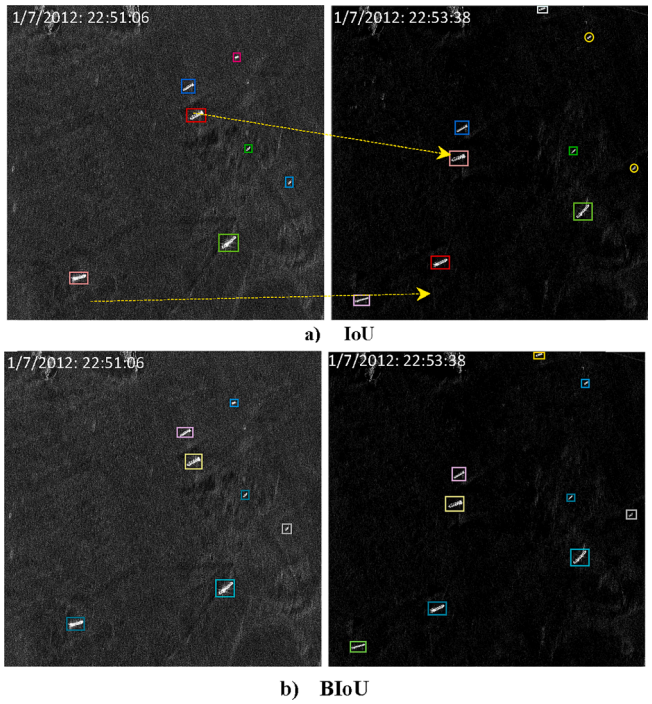


Fig. 12. Comparison of tracking results: a) IOU tracking results. b) BIoU tracking results.

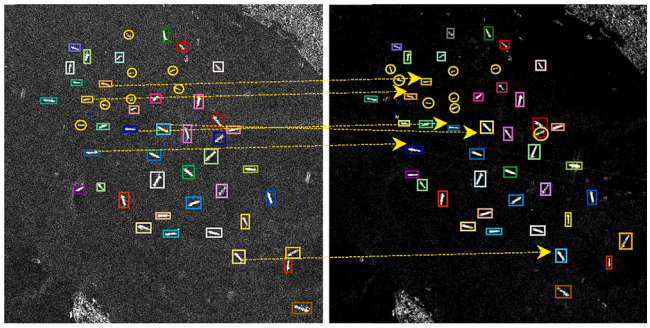
multiple ships, a comparative analysis with existing multi-object tracking (MOT) algorithms using YOLOv8n for ship detection was conducted. The comparative results are detailed in Table 6, showcasing the performance of various MOT models. Among the evaluated algorithms, such as DeepSort (Veeramani et al., 2018) and StrongSort (Du et al., 2023), those featuring re-identification components generally exhibit reduced processing speeds. Specifically, DeepSort employs a

Table 6
Comparison of ship tracking results from various algorithms.

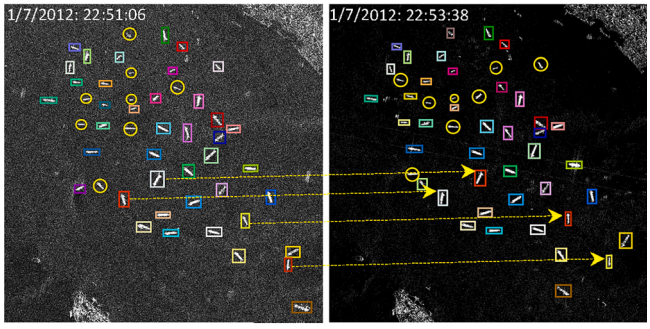
Model	HOTA (%)	MOTA (%)	MOTP (%)	IDF1 (%)	IDS	FPS
DeepSort (Veeramani et al., 2018)	67.4	79.1	86.0	75.3	**47**	27
StrongSort (Du et al., 2023)	71.9	79.2	85.4	79.1	91	19
OC Sort (Cao et al., 2023)	69.4	79.3	85.1	75.2	103	50
ByteTrack (Zhang et al., 2021)	70.6	76.9	82.9	80.0	111	53
YOLOShipTracker	**72.8**	**79.9**	**87.9**	**80.7**	79	**81**

combination of cascaded matching and appearance feature associations. However, it faces challenges in accurately distinguishing between targets that look similar, mainly due to the limited processing power available for extracting appearance features, which affects its tracking effectiveness. The OC-SORT algorithm, introduced by (Cao et al., 2023), enhances tracking accuracy through a motion-model strategy and by overcoming some of the deficiencies observed in SORT algorithms and in the Kalman filter. However, our approach outperforms OC-SORT in terms of overall tracking and accuracy performance. It notably achieves the highest scores in HOTA (73.6 %), MOTP (87.5 %), and IDF1 (82.1 %) metrics among the state-of-the-art methods we examined, demonstrating exceptional accuracy and stability in tracking targets.

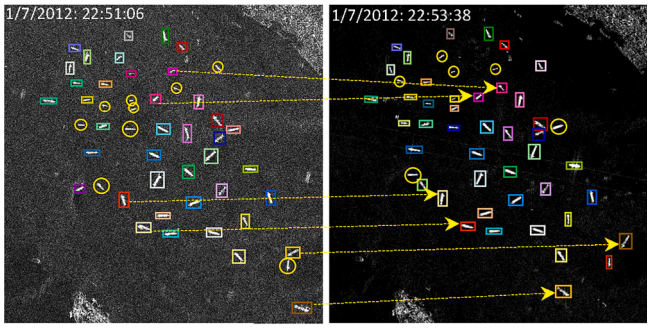
Furthermore, our technique attains a processing speed of up to 82 FPS, marking a 50 % enhancement compared to the ByteTrack algorithm (Zhang et al., 2021). This is crucial for achieving real-time ship tracking in natural scenarios. The efficiency of the YOLOShipTracker algorithm proposed is further demonstrated through visualization findings in Fig. 13, where it can be made out that, despite challenges such as occlusion and lighting interference, it exhibits accurate tracking. In contrast to DeepSort, our approach exhibits enhanced performance by reducing missed tracking instances, demonstrating its improved ability to match detections with trajectories with the inclusion of a buffer. To better visualize missed detections and instances where ship IDs have been swapped, we have highlighted missed detections with yellow circles and identified ships with swapped IDs by using dotted arrays. These visual cues provide a better understanding of tracking performance and potential vessel tracking issues. In summary, the tracking results presented here qualify the proposed model as presenting exceptional accuracy and stability for tracking vessels in complex maritime environments.



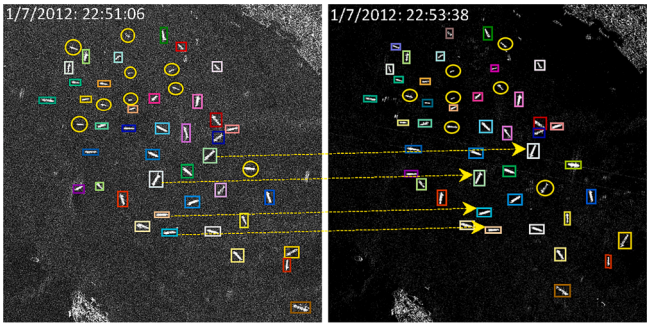
a) DeepSort



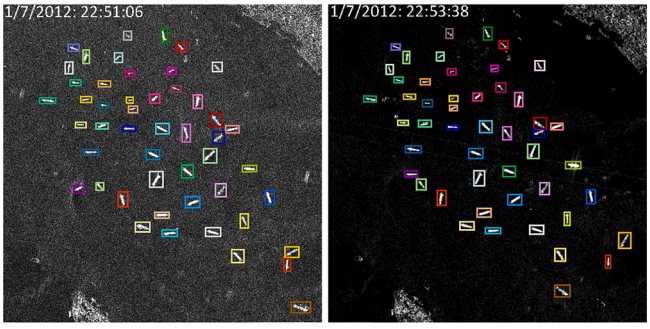
b) StrongSort



c) OC Sort



d) ByteTrack



e) YOLOShipTracker

Fig. 13. Comparative illustration of tracking results for various tracking algorithms: a) DeepSort, b) StrongSort, c) OC Sort, d) ByteTrack, and e) YOLOShipTracker. Missed tracking targets are highlighted in the figure with a yellow circle, while dotted array represents the ID switch. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We further evaluated our MOT algorithm to deepen our insight into its robustness, especially under challenging conditions that include non-ship elements. Fig. 14 highlights the capability of our YOLOShipTracker algorithm to consistently and accurately track ships across SAR short-time sequence images, ensuring no lapses in ship detection. Importantly, our algorithm remains effective in complex environments, capable of handling diverse vessel sizes, and amidst non-vessel distractions. This adaptability is critical, illustrating the effectiveness of our method in varied and cluttered maritime backgrounds, which are common in real-world scenarios. The results from these tests not only confirm the reliability of our algorithm, but also its relevance to enhancing maritime tracking technologies.

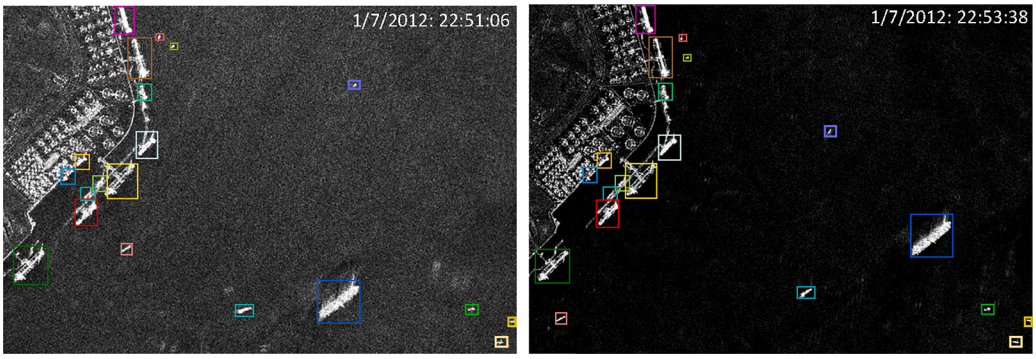


Fig. 14. Tracking results of the proposed model in a complex environment.

4. Discussion

This study leveraged two comprehensive datasets of SAR images from RADARSAT-2 and TerraSAR-X satellites, focusing on the Malacca Strait (Malay Peninsula) and the Zhejiang region (China). The model was trained over 300 epochs, with hyperparameters refined iteratively for optimal performance. The metrics comprehensively evaluate the model's accuracy, efficiency, and real-time detection capabilities. Precision and Recall are fundamental metrics in deep learning object detection, and the F1-Score combines these to provide a balanced measure. The mAP at a 50 % IoU threshold evaluated detection accuracy across all object categories, although in this study it directly relates to ship detection. In addition, metrics like FLOPs, Parameters, Model Size, and FPS assess the model's computational demand, parameter count, and processing speed. Our results indicated a better detection performance and a more balanced precision-recall trade-off. This is crucial for reliable ship detection in SAR images, where high confidence is essential. The proposed model effectively identifies all vessels, whereas the baseline model suffers from false detections and missing vessels near the coast. The superior performance of the proposed model in challenging conditions underscores its robustness and reliability, while a thorough comparison with several state-of-the-art models also demonstrates its superior effectiveness. In fact, our model outperformed other SOTA models in terms of key metrics like FLOPs, reduced number of parameters, model size, and processing speed. Specifically, our model achieves 4.5 GFLOPs, 2.05 million parameters, a model size of 3.9 MB, and up to 339.7 FPS, significantly improving the performance of other models such as YOLOv5, YOLOv7 and YOLOv8n. These results highlight the proposed model's balance between accuracy and efficiency, making it suitable for edge devices and real-time detection.

An ablation study is crucial for understanding the contributions of individual components within a complex system. By systematically removing or adding specific modules, we can isolate their effects on the overall performance, leading to more informed design and optimization decisions. In our study, the ablation experiments focused on evaluating the impact of various lightweight modules on key metrics such as FLOPs, number of parameters, and model size. The results highlighted the performance differences when individual modules—HGNetv2, Slimneck, Efficient Head, and Knowledge Distillation—are included or excluded from the baseline YOLOv8 model. In this sense, the HGNetv2 module enhanced the model's efficiency without significantly altering its structural parameters. Moreover, adding the Slimneck module primarily improved the Recall and F1 score, indicating its effectiveness in capturing more true positives while maintaining a lightweight footprint. While the individual impact of the Efficient Head module was less pronounced, it contributed to the overall performance gains when combined with other modules, particularly regarding mAP50 and Precision. Notably, the Knowledge Distillation module did not affect the structural aspects of the model but significantly enhanced recognition performance, underscoring its role in refining the model's predictive capabilities. When all modules were integrated, the proposed model achieved a remarkable 2.3 % increase in Precision, a 7 % rise in Recall, a 3.3 % boost in F1 score, and a 4.5 % improvement in mAP50. These findings confirm the goodness of our model's architecture to improve ship detection accuracy in SAR images, underscoring the advantages offered by each constituent element. Moreover, metrics including MOTA, HOTA, IDF1, and MOTP thoroughly evaluated tracking precision and consistency. Additionally, we considered the number of Identity Switches (IDS) and Frames Per Second (FPS) to gauge the model's real-time tracking capabilities.

The experiments above validate our model's effectiveness, demonstrating its capability to accurately identify ships in various scenarios depicted in SAR images. We have optimized the model's architecture by integrating multiple lightweight modules, achieving enhanced accuracy. This outcome aligns with our expectations, as we selected modules that are experts in extracting intricate details from SAR images. Our proposed model outperforms the original YOLOv8n model across all evaluation metrics, showcasing superior performance in both accuracy and structural assessments. Visualizations indicate that our model effectively focuses on target vessels in diverse situations, thereby minimizing false positives and negatives. Its simplified and efficient structure contributes to superior detection results, even with limited training data.

It is worth noting that Depthwise Convolution (DWConv) is prioritized over traditional convolutional operations in the backbone architecture of the proposed tracking model in order to reduce the model parameter count significantly. As it is widely known, DWConv conducts convolutions independently across each channel of the input feature maps, ensuring the preservation of fine-grained details within individual channels, which is crucial for maintaining the fidelity of feature information across the network. However, we recognize that DWConv lacks direct interaction between channels, which could impact the model's feature extraction capability. To address this issue, one potential solution is to incorporate pointwise convolutions (PWConv) following DWConv operations (Zhang et al., 2020). PWConv performs 1x1 convolutions across all channels, effectively reintroducing inter-channel communication and ensuring that the features extracted by DWConv are combined in a meaningful way. While we have not implemented this approach in the current model, it remains a viable option for future optimizations to further enhance feature extraction capabilities.

On the other hand, our experiments demonstrated the superiority of using BIoU over traditional IoU for matching metrics in ship tracking. BIoU vs. IoU showed significantly enhanced tracking performance, with higher HOTA, MOTA, MOTP, and IDF1 scores. BIoU expanded the space around bounding boxes, facilitating more accurate matching even when the target's position or shape changed, leading to fewer ID changes and track breaks. Visual comparisons (Fig. 12) reinforced these findings, showing more stable and accurate tracking with BIoU, particularly in complex maritime environments. Finally, when comparing our proposed YOLOShipTracker model with other SOTA algorithms, several key advantages are observed, such as higher accuracy, stability, and real-time capability. Thus, with a processing speed of up to 82 FPS, our model significantly outperformed other SOTA algorithms in terms of real-time processing capability, making it well-suited for practical applications. In this way, our model consistently outperformed DeepSort, StrongSort, OC-SORT, and ByteTrack regarding accuracy and processing speed. Visualizations in Fig. 13 further highlighted our model's robustness in tracking ships amidst occlusion and varying lighting conditions, with fewer missed detections and ID swaps. Our additional assessments in challenging environments (Fig. 14) also demonstrated the robustness and adaptability of YOLOShipTracker, being able to provide consistent tracking accuracy even amidst the presence of non-ship objects, affirming its efficacy in real-world settings characterized by cluttered and varied backgrounds.

5. Conclusion and future work

The proposed YOLOShipTracker model, through strategic modifications of the original YOLOv8n model, achieves significant reductions in computational complexity and model size while enhancing ship detection accuracy and real-time processing capabilities. The comprehensive evaluation and comparison with SOTA object detection models concluded its robustness and efficiency, making it an excellent choice for practical applications in ship detection using SAR imagery.

This study also underscored the significant advancements made by integrating lightweight modules and employing innovative tracking metrics like BIoU. Our proposed YOLOShipTracker model not only enhances detection accuracy and tracking stability, but also excels in real-time processing, making it a robust solution for maritime surveillance and SAR image analysis. By systematically evaluating and comparing our approach with existing SOTA object tracking algorithms, we have demonstrated its superior performance, paving the way for its application in complex and dynamic maritime environments. In this sense, the

YOLOShipTracker model can be considered for real-time deployment on edge devices, suggesting its suitability for practical applications. Future efforts will involve testing our model on multiple platforms to further validate it.

Regarding further works, we propose to investigate other potential lightweight modules that could be integrated into the model to further reduce the computational burden while maintaining or improving detection accuracy. We also recommend extending the model to handle various SAR imaging conditions, such as different resolutions, noise levels, and environmental factors (e.g., weather conditions). This would involve training and testing the model on a more diverse dataset to ensure robustness across different scenarios. Researchers may implement advanced data augmentation techniques tailored to SAR images, such as SAR simulation and data synthesis, to improve the model's generalization capability and performance on unseen data. We also suggest exploring the integration of data from other sensors (e.g., optical images, AIS data) to complement SAR imagery. This multimodal approach could enhance detection and tracking performance by providing additional context and information. Perhaps researchers can develop collaborative tracking systems that combine multiple models or systems to improve overall tracking accuracy and reliability. This could include ensemble methods or fusion algorithms that aggregate predictions from different models. Finally, this study recommends establishing frameworks for monitoring and maintaining the long-term performance of the implemented model. This would include periodic retraining, updating, and validation to ensure the model remains effective as conditions and data evolve over time. By pursuing these future research directions, we can further improve the robustness, accuracy, and applicability of the YOLOShipTracker model, making it an even more powerful tool for maritime surveillance and SAR image analysis.

Funding

This research work was supported by National Key Research and Development Program of China (2017YFC1405600).

CRediT authorship contribution statement

Muhammad Yasir: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Shanwei Liu:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition. **Saied Pirasteh:** Writing – review & editing, Data curation, Conceptualization. **Mingming Xu:** Writing – review & editing, Validation, Formal analysis, Data curation. **Hui Sheng:** Writing – review & editing, Validation, Data curation, Conceptualization. **Jianhua Wan:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Data curation. **Felipe A.P. de Figueiredo:** Writing – review & editing, Data curation. **Fernando J. Aguilar:** Conceptualization, Data curation, Validation, Visualization, Writing – review & editing. **Jonathan Li:** Conceptualization, Data curation, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Aceto, G., Ciuonzo, D., Montieri, A., Pescape, A., 2019. Mobile encrypted traffic classification using deep learning: experimental evaluation, lessons learned, and challenges. IEEE Trans. Netw. Service Manag. 16, 445–458. https://doi.org/10.1109/TNSM.2019.2899085.

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., Farhan, L., 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J. Big Data 8, 53. https://doi.org/10.1186/s40537-021-00444-8.

Bewley, A., Ge, Z., Ott, L., Ramos, F., Upcroft, B., 2016. Simple online and realtime tracking. In: 2016 IEEE Int. Conf. Image Proc. (ICIP). IEEE, pp. 3464–3468. doi: 10.1109/ICIP.2016.7533003.

Cao, J., Pang, J., Weng, X., Khirodkar, R., Kitani, K., 2023. Observation-Centric SORT: Rethinking SORT for Robust Multi-Object Tracking, in: 2023 IEEE/CVF Conf. Comput. Vis. Patt. Recog. (CVPR). IEEE, pp. 9686–9696. doi: 10.1109/CVPR52729.2023.00934.

Chen, L., Ai, H., Zhuang, Z., Shang, C., 2018. Real-Time Multiple People Tracking with Deeply Learned Candidate Selection and Person Re-Identification. In: 2018 IEEE International Conference on Multimedia and Expo (ICME). IEEE, pp. 1–6. https://doi.org/10.1109/ICME.2018.8486597.

Chen, X., Wu, H., Han, B., Liu, W., Montewka, J., Liu, R.W., 2023. Orientation-aware ship detection via a rotation feature decoupling supported deep learning approach. Eng. Appl. Artif. Intell. 125, 106686 https://doi.org/10.1016/j.engappai.2023.106686.

Chen, X., Dou, S., Song, T., Wu, H., Sun, Y., Xian, J., 2024. Spatial-temporal ship pollution distribution exploitation and harbor environmental impact analysis via large-scale AIS data. J. Mar. Sci. Eng. 12, 960. https://doi.org/10.3390/jmse12060960.

Chen, J., Yu, H., 2023. Wide-beam SAR autofocus based on blind resampling. Sci. China Info. Sci. 66, 140304 https://doi.org/10.1007/s11432-022-3574-7.

Chu, Q., Ouyang, W., Li, H., Wang, X., Liu, B., Yu, N., 2017. Online Multi-object Tracking Using CNN-Based Single Object Tracker with Spatial-Temporal Attention Mechanism. In: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, pp. 4846–4855. https://doi.org/10.1109/ICCV.2017.518.

Dechesne, C., Lefèvre, S., Vadaine, R., Hajduch, G., Fablet, R., 2019. Ship Identification and Characterization in Sentinel-1 SAR Images with Multi-Task Deep Learning. Remote Sens. (Basel) 11, 2997. https://doi.org/10.3390/rs11242997.

Du, Y., Zhao, Z., Song, Y., Zhao, Y., Su, F., Gong, T., Meng, H., 2023. StrongSORT: Make DeepSORT Great Again. IEEE Trans. Multimedia 25, 8725–8737. https://doi.org/10.1109/TMM.2023.3240881.

Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A., 2010. The Pascal Visual Object Classes (VOC) Challenge. Int. J. Comput. Vis. 88, 303–338. https://doi.org/10.1007/s11263-009-0275-4.

Gao, G., Bai, Q., Zhang, C., Zhang, L., Yao, L., 2023. Dualistic cascade convolutional neural network dedicated to fully PolSAR image ship detection. ISPRS J. Photogram. Remote Sens. 202, 663–681. https://doi.org/10.1016/j.isprsjprs.2023.07.006.

Gao, F., He, Y., Wang, J., Hussain, A., Zhou, H., 2020. Anchor-free convolutional network with dense attention feature aggregation for ship detection in SAR images. Remote Sens. (Basel) 12, 2619. https://doi.org/10.3390/rs12162619.

Girshick, R., 2015. Fast r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448.

Hinton, G., Vinyals, O., Dean, J., 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.

Humayun, M.F., Nasir, F.A., Bhatti, F.A., Tahir, M., Khurshid, K., 2024. YOLO-OSD: optimized ship detection and localization in multiresolution SAR satellite images using a hybrid data-model centric approach. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 17, 5345–5363. https://doi.org/10.1109/JSTARS.2024.3365807.

Kumar, L.A., Renuka, D.K., 2023. Deep Learning Approach for Natural Language Processing, Speech, and Computer Vision. CRC Press, Boca Raton. doi: 10.1201/9781003348689.

Li, H., Li, J., Wei, H., Liu, Z., Zhan, Z., Ren, Q., 2024. Slim-neck by GSConv: a lightweight-design for real-time detector architectures. J Real Time Image Process 21, 62. https://doi.org/10.1007/s11554-024-01436-6.

Li, Q., Min, R., Cui, Z., Pi, Y., Xu, Z., 2019. Multi-scale ship detection based on dense attention pyramid network in SAR images. In: IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 1–4. https://doi.org/10.1109/IGARSS40859.2019.8999648.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollar, P., 2017. Focal Loss for Dense Object Detection. In: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, pp. 2999–3007. https://doi.org/10.1109/ICCV.2017.324.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C., 2015. SSD: Single shot multibox detector. In: Computer Vision–ECCV 2015: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2015, Proceedings, Part I 14, pp. 21–37. doi: 10.1007/978-3-319-46448-0_2.

Liu, S., Kong, W., Chen, X., Xu, M., Yasir, M., Zhao, L., Li, J., 2022. Multi-scale ship detection algorithm based on a lightweight neural network for spaceborne SAR images. Remote Sens. (Basel) 14, 1149. https://doi.org/10.3390/rs14051149.

Ma, W., Yang, X., Zhu, H., Wang, X., Yi, X., Wu, Y., Hou, B., Jiao, L., 2024. NRENet: neighborhood removal-and-emphasis network for ship detection in SAR Images. Int. J. Appl. Earth Obser. Geoinfo. 131, 103927 https://doi.org/10.1016/j.jag.2024.103927.

Mao, W., Wang, X., Liu, G., Zhang, R., Shi, Y., Pirasteh, S., 2022. Estimation and compensation of ionospheric phase delay for multi-aperture InSAR: An Azimuth Split-Spectrum Interferometry Approach. IEEE Trans. Geosci. Remote Sens. 60, 1–14. https://doi.org/10.1109/TGRS.2021.3095272.

Mao, W., Wang, X., Liu, G., Pirasteh, S., Zhang, R., Lin, H., Xie, Y., Xiang, W., Ma, Z., Ma, P., 2023. Time series InSAR ionospheric delay estimation, correction, and ground deformation monitoring with reformulating range split-spectrum interferometry. IEEE Trans. Geosci. Remote Sens. 61, 1–18. https://doi.org/10.1109/TGRS.2023.3298919.

Miao, T., Zeng, H., Yang, W., Chu, B., Zou, F., Ren, W., Chen, J., 2022. An improved lightweight RetinaNet for ship detection in SAR images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15, 4667–4679. https://doi.org/10.1109/JSTARS.2022.3180159.

O'Shea, T., Hoydis, J., 2017. An introduction to deep learning for the physical layer. IEEE Trans. Cogn. Commun. Netw. 3, 563–575. https://doi.org/10.1109/TCCN.2017.2758370.

Popp, N., Metzen, J.H., Hein, M., 2024. Zero-Shot Distillation for Image Encoders: How to Make Effective Use of Synthetic Data. arXiv preprint arXiv:2404.16637.

Ren, X., Bai, Y., Liu, G., Zhang, P., 2023. YOLO-Lite: an efficient lightweight network for SAR ship detection. Remote Sens. (Basel) 15, 3771. https://doi.org/10.3390/rs15153771.

Smith, J.S., Tian, J., Halbe, S., Hsu, Y.C., Kira, Z., 2023. A closer look at rehearsal-free continual learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 2410-2420.

Tan, M., Pang, R., Le, Q.V., 2019. Efficientdet: Scalable and efficient object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 10781-10790.

Tang, H., Gao, S., Li, S., Wang, P., Liu, J., Wang, S., Qian, J., 2024. A lightweight SAR image ship detection method based on improved convolution and YOLOv7. Remote Sens. (Basel) 16, 486. https://doi.org/10.3390/rs16030486.

Veeramani, B., Raymond, J.W., Chanda, P., 2018. DeepSort: deep convolutional networks for sorting haploid maize seeds. BMC Bioinf. 19, 289. https://doi.org/10.1186/s12859-018-2267-2.

Wang, X., Zhang, W., Chu, Y., Liu, P., Yin, Q., Li, Q., 2021. Research on Knowledge Distillation Algorithm of Object Detection. In: 2021 IEEE International Conference on Progress in Informatics and Computing. IEEE, pp. 87–93. doi: 10.1109/PIC53636.2021.9687066.

Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M., 2022. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 7464–7475.

Wang, C., Su, W., Gu, H., 2020. Two-stage ship detection in synthetic aperture radar images based on attention mechanism and extended pooling. J. Appl. Remote Sens. 14 https://doi.org/10.1117/1.JRS.14.044522.

Wei, S., Su, H., Ming, J., Wang, C., Yan, M., Kumar, D., Shi, J., Zhang, X., 2020. Precise and robust ship detection for high-resolution SAR imagery based on HR-SDNet. Remote Sens. (Basel) 12, 167. https://doi.org/10.3390/rs12010167.

Wen, X., Zhang, S., Wang, J., Yao, T., Tang, Y., 2024. A CFAR-enhanced ship detector for SAR images based on YOLOv5s. Remote Sens. (Basel) 16, 733. https://doi.org/10.3390/rs16050733.

Wojke, N., Bewley, A., Paulus, D., 2017. Simple online and realtime tracking with a deep association metric. In: 2017 IEEE International Conference on Image Processing. IEEE, pp. 3645–3649. doi: 10.1109/ICIP.2017.8296962.

Xiong, B., Sun, Z., Wang, J., Leng, X., Ji, K., 2022. A lightweight model for ship detection and recognition in complex-scene SAR images. Remote Sens. (Basel) 14, 6053. https://doi.org/10.3390/rs14236053.

Xu, X., Zhang, X., Zhang, T., 2022. Lite-YOLOv5: a lightweight deep learning detector for on-board ship detection in large-scene sentinel-1 SAR Images. Remote Sens. (Basel) 14, 1018. https://doi.org/10.3390/rs14041018.

Yang, F., Odashima, S., Masui, S., Jiang, S., 2023. Hard to Track Objects with Irregular Motions and Similar Appearances? Make It Easier by Buffering the Matching Space. In: 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). IEEE, pp. 4788–4797. doi: 10.1109/WACV56688.2023.00478.

Yasir, M., Liu, S., Mingming, X., Wan, J., Pirasteh, S., Dang, K.B., 2024a. ShipGeoNet: SAR Image-Based Geometric Feature Extraction of Ships Using Convolutional Neural Networks. IEEE Trans. Geosci. Remote Sens. 62, 1–13. https://doi.org/10.1109/TGRS.2024.3352150.

Yasir, M., Shanwei, L., Mingming, X., Jianhua, W., Hui, S., Nazir, S., Zhang, X., Tugsan Isiacik Colak, A., 2024b. YOLOv8-BYTE: Ship tracking algorithm using short-time sequence SAR images for disaster response leveraging GeoAI. Int. J. Appl. Earth Obser. Geoinfo. 128, 103771 https://doi.org/10.1016/j.jag.2024.103771.

Yasir, M., Shanwei, L., Mingming, X., Jianhua, W., Nazir, S., Islam, Q.U., Dang, K.B., 2024c. SwinYOLOv7: robust ship detection in complex synthetic aperture radar images. Appl. Soft Comput. 160, 111704 https://doi.org/10.1016/j.asoc.2024.111704.

Yu, W., Wang, Z., Li, J., Luo, Y., Yu, Z., 2022. A lightweight network based on one-level feature for ship detection in SAR images. Remote Sens. (Basel) 14, 3321. https://doi.org/10.3390/rs14143321.

Zha, C., Min, W., Han, Q., Xiong, X., Wang, Q., Xiang, H., 2023. SAR ship detection based on salience region extraction and multi-branch attention. Int. J. Appl. Earth Obser. Geoinfo. 123, 103489 https://doi.org/10.1016/j.jag.2023.103489.

Zhang, P., Lo, E., Lu, B. 2020a. High performance depthwise and pointwise convolutions on mobile devices. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 04, pp. 6795–6802.

Zhang, T., Zhang, X., Liu, C., Shi, J., Wei, S., Ahmad, I., Zhan, X., Zhou, Y., Pan, D., Li, J., Su, H., 2021b. Balance learning for ship detection from synthetic aperture radar remote sensing imagery. *ISPRS J. Photogram. Remote Sens*182, 190–207. doi: 10.1016/j.isprsjprs.2021.10.010.

Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., Wang, X., 2021. ByteTrack: Multi-Object Tracking by Associating Every Detection Box.

Zhang, X., Li, Y., Li, F., Jiang, H., Wang, Y., Zhang, L., Zheng, L., Ding, Z., 2024. Ship-Go: AR S ip mages In ainting via instance-to-image enerative Diffusion M dels. ISPRS J. Photogram. Remote Sens. 207, 203–217. https://doi.org/10.1016/j.isprsjprs.2023.12.002.

Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Wang, X., 2022b. Bytetrack: Multi-object tracking by associating every detection box. In: European Conference on Computer Vision. Springer Nature Switzerland, Cham, pp. 1–21.

Zhang, J., Xing, M., Sun, G.-C., Bao, Z., 2022a. Integrating the reconstructed scattering center feature maps with deep CNN feature maps for automatic SAR target recognition. IEEE Geosci. Remote Sens. Lett. 19, 1–5. https://doi.org/10.1109/LGRS.2021.3054747.

Zhang, Y., Xing, M., Zhang, J., Sun, G.-C., Xu, D., 2023b. Robust multi-ship tracker in SAR imagery by fusing feature matching and modified KCF. IEEE Geosci. Remote Sens. Lett. 20, 1–5. https://doi.org/10.1109/LGRS.2023.3251975.

Zhang, B., Xu, G., Zhou, R., Zhang, H., Hong, W., 2023a. Multi-channel back-projection algorithm for mmWave automotive MIMO SAR imaging with Doppler-division multiplexing. IEEE J. Sel. Top. Signal Proc. 17, 445–457. https://doi.org/10.1109/JSTSP.2022.3207902.

Zhang, T., Zhang, X., Shi, J., Wei, S., 2020. HyperLi-Net: a hyper-light deep learning network for high-accurate and high-speed ship detection from synthetic aperture radar imagery. ISPRS J. Photogram. Remote Sens. 167, 123–153. https://doi.org/10.1016/j.isprsjprs.2020.05.016.

Zhang, T., Zhang, X., Liu, C., Shi, J., Wei, S., Ahmad, I., Zhan, X., Zhou, Y., Pan, D., Li, J., Su, H., 2021a. Balance learning for ship detection from synthetic aperture radar remote sensing imagery. ISPRS J. Photogram. Remote Sens. 182, 190–207. https://doi.org/10.1016/j.isprsjprs.2021.10.010.

Zhao, Y., Ju, Z., Sun, T., Dong, F., Li, J., Yang, R., Fu, Q., Lian, C., Shan, P., 2023. TGC-YOLOv5: An Enhanced YOLOv5 Drone Detection Model Based on Transformer, GAM & CA Attention Mechanism. Drones 7, 446. https://doi.org/10.3390/drones7070446.

Zhou, K., Zhang, M., Wang, H., Tan, J., 2022. Ship detection in SAR images based on multi-scale feature extraction and adaptive feature fusion. Remote Sens. (Basel) 14, 755. https://doi.org/10.3390/rs14030755.

Zhu, M., Hu, G., Zhou, H., Wang, S., Feng, Z., Yue, S., 2022. A ship detection method via redesigned FCOS in Large-Scale SAR Images. Remote Sens. (Basel) 14, 1153. https://doi.org/10.3390/rs14051153.