*applied sciences*

MDPI

*Article*

# YOLO-SWD—An Improved Ship Recognition Algorithm for Feature Occlusion Scenarios

**Ruyan Zhou, Mingkang Gu and Haiyan Pan ***

College of Information Technology, Shanghai Ocean University, Shanghai 201306, China; ryzhou@shou.edu.cn (R.Z.); m220901556@st.shou.edu.cn (M.G.)
* Correspondence: hy-pan@shou.edu.cn

**Abstract:** Ship detection and recognition hold significant application value in both military and civilian domains. With the continuous advancement of deep learning technologies, multi-category ship detection and recognition methods based on deep learning have garnered increasing attention. However, challenges such as feature occlusion caused by interfering objects, cloudy and foggy weather leading to feature loss, and insufficient accuracy in remote sensing imagery persist. This study aims to enhance the accuracy and robustness of ship recognition by improving deep learning-based object detection models, enabling the algorithm to perform ship detection and recognition tasks effectively in feature-occluded scenarios. In this research, we propose a ship detection and recognition algorithm based on YOLOv11. YOLOv11 possesses stronger feature extraction capabilities and its multi-branch structure effectively captures features of targets at different scales. Three improved modules are introduced: the DLKA module enhances the perception of local details and global context through dynamic deformable convolution and large receptive field attention mechanisms; the CKSP module improves the model's ability to extract target boundaries and shapes; and the WTHead enhances the diversity and robustness of feature extraction. Comparative experiments with classical object detection models on visible and SAR datasets, which include a variety of feature occlusion scenarios, show that our proposed model achieved the best results across multiple metrics, specifically, our method achieved a mAP of 83.9%, surpassing the second-best result by 2.7%.

**Keywords:** ship detection; ship recognition; deep learning; remote sensing imagery

check for updates

Academic Editor: Douglas O'Shaughnessy

Received: 7 February 2025
Revised: 27 February 2025
Accepted: 28 February 2025
Published: 4 March 2025

**Citation:** Zhou, R.; Gu, M.; Pan, H. YOLO-SWD—An Improved Ship Recognition Algorithm for Feature Occlusion Scenarios. *Appl. Sci.* **2025**, *15*, 2749. https://doi.org/10.3390/app15052749

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Ships play a critical role in human marine activities. With the increasing growth of coastal activities, issues such as coastal transportation, national defense security, and resource protection in coastal areas have drawn significant attention [1,2]. Against this backdrop, ship detection and recognition have become important research topics. In recent years, high-resolution remote sensing imagery has provided a crucial data foundation for monitoring and analyzing ship targets in marine environments by capturing clear surface images over vast spatial ranges at short time intervals [3–5].

Under current technological trends, deep learning-based detection and recognition methods have demonstrated strong potential. Deep learning, particularly the rise of convolutional neural networks (CNNs), has brought revolutionary changes to remote sensing ship target detection and recognition. Benefiting from their powerful feature extraction capabilities, deep learning models can automatically learn multi-level feature representations of ships, thereby effectively improving detection accuracy [4,6–8]. Meanwhile, optimizations in ship target detection frameworks and the introduction of high-performance models

have provided new research directions and practical references for enhancing the accuracy and practicality of remote sensing ship target detection and recognition [6,9–18].

In recent years, numerous improved methods for ship detection and recognition networks have emerged. Zhao et al. [19] utilized Transformers to focus on global image features and enhanced ship target detection capabilities in SAR image datasets through pseudo-label-based feature clustering. Xu et al. [20] proposed the NWD-RKA strategy to replace standard IoU-threshold-based detectors, providing sufficient supervisory information for network training and improving detection capabilities for small ships. Zhang et al. [21] suppressed sea clutter in SAR images using adaptive processing of frequency-domain information. Zhang et al. [22] introduced scattering features into SAR ship recognition tasks, fusing scattering and deep features to enrich ship feature diversity. Shang et al. [23] incorporated prior knowledge of ship shapes into the network, providing shape-aligned rectangular detection boxes to overcome the limitations of traditional square convolutions, thereby improving ship recognition accuracy.

For images with feature occlusion, Sun et al. [24] achieved satisfactory results in classifying seven types of military ships using Gaofen-3 SAR satellite imagery through a dual-branch feature extraction and fusion architecture that separately extracts discriminative local fine-grained features and global semantic information. Wang et al. [25] improved target localization in complex backgrounds by introducing an asymmetric pyramid non-local block. Wang et al. [26] employed polarization self-attention and a weighted bidirectional feature pyramid network to preserve long-range dependencies and high-resolution channel-spatial features, effectively addressing ship detection challenges in sea fog scenarios. Yan et al. [27] designed a Perlin noise dehazing hybrid module and a target-oriented structural similarity loss function to remove clouds and restore ship texture details, improving ship detection accuracy in cloud-interfered remote sensing images.

In summary, current ship detection and recognition algorithms have made significant progress, but several challenging issues remain unresolved. Most existing models focus on detecting offshore ship targets in open seas, where ship features are distinct. However, as shown in Figure 1, in scenarios where ships are occluded by shoreline terrain, port structures, or other vessels, traditional detection techniques face significant challenges [28]. Furthermore, while some researchers have improved partial occlusion scenarios (e.g., cloud occlusion), they have not addressed general ship target occlusion. For instance, models optimized for cloud occlusion often underperform in near-shore occlusion scenarios. Due to the diverse morphological features of occluded targets, effectively extracting residual features and accurately identifying their categories remains a challenge.

Therefore, building on existing research, this paper explores efficient detection and recognition methods for occluded ship targets. By improving and optimizing general object detection algorithms, we aim to enhance model robustness in complex scenarios and improve accuracy for occluded ship targets. The main contributions of this paper are as follows:

- A new ship detection and recognition dataset is constructed. This dataset integrates existing international ship satellite imagery and newly annotated data, covering six categories and scenarios such as cloud interference, dense ships, and port docks, providing a data foundation for ship-related research.
- A novel robust and general-purpose ship detection and recognition model is designed to address the challenge of large differences in occluded ship features across scenarios. The model is validated in cloud-interfered, dense ship, and port dock scenarios, demonstrating stronger robustness and higher accuracy.
- For feature occlusion and mutual interference, we first design a new StripPooling-based convolutional block in the backbone network to enhance perception of target

boundaries and shapes during convolution. Dynamic deformable convolution and large-receptive-field attention mechanisms are introduced to strengthen the perception of local details and global context. Finally, wavelet convolution is employed to enhance local detail feature perception, improving feature extraction capabilities in complex scenarios.

The remainder of this paper is organized as follows: Section 2 describes the sources and construction process of the new dataset. Section 3 details the selection of the new model and the principles of each module. Section 4 elaborates on the experimental process and results. Section 5 discusses the experiments and Section 6 summarizes the limitations of this study and future research directions.

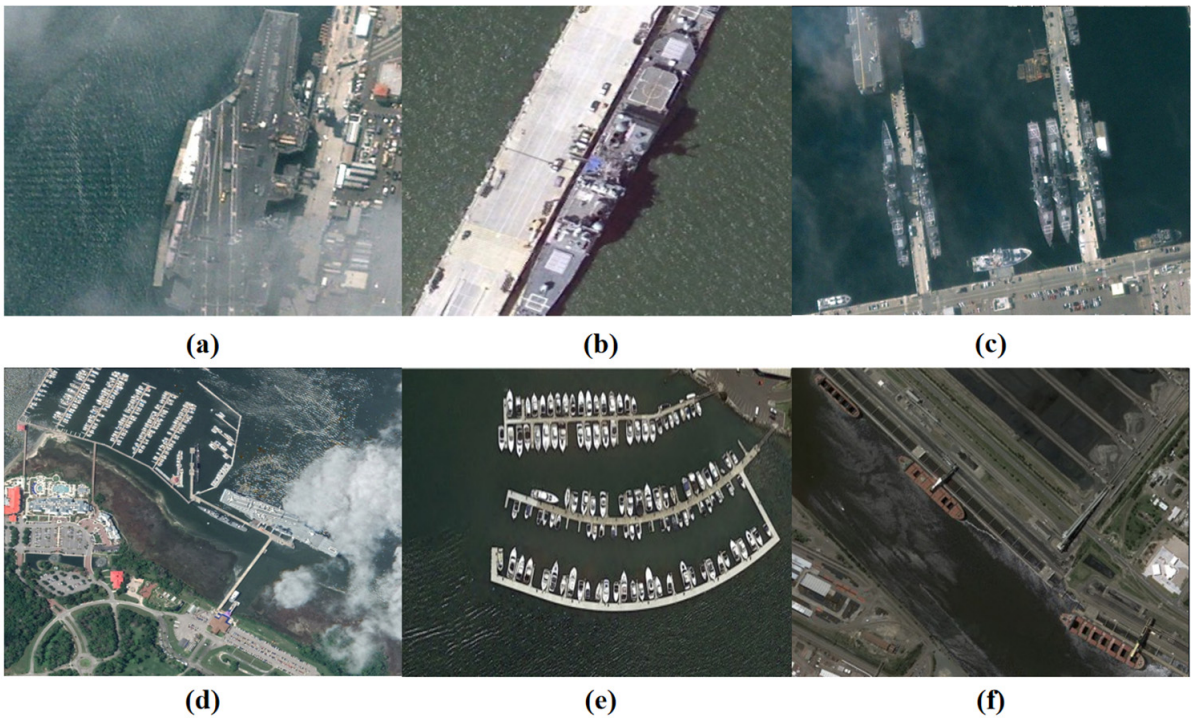



**Figure 1.** Common feature occlusion phenomena in remote sensing ship imagery. while (**a**,**c**,**d**) depict scenarios with cloud or fog occlusion; (**b**,**e**,**f**) illustrate occlusions caused by other ships or port obstacles.

## 2. Datasets

At present, numerous publicly available ship recognition datasets exist, primarily categorized into optical remote sensing image datasets and synthetic aperture radar (SAR) image datasets. For example, the FGSCR dataset [29], widely used for ship recognition tasks, and the OpenSARShip dataset [30], mainly designed for ship detection, contain approximately 9000 high-resolution remote sensing images and around 2000 SAR images, respectively. However, existing datasets predominantly consist of simple scenarios, with occluded ship scenes being relatively rare. Therefore, this paper constructs a new ship dataset to address this limitation.

The ship dataset established in this study incorporates images from publicly available visible light and SAR ship datasets. Visible light images are sourced from the HRSC2016 dataset [31], FGSC-23 dataset, SPOT satellite images provided by Kaggle, and other open datasets. High-resolution SAR images are derived from the OpenSARShip and SSDD datasets [32]. After filtering out abnormal data and removing non-interesting scenes and targets, all images are uniformly re-annotated using the labelme 1.0 software. Additionally, we further select qualified ship targets and scenarios from Sentinel-2, Sentinel-1 satellite

slices, and Google Earth imagery. These images are manually annotated in labelme to mark ship target locations and categories before being added to the dataset. The final dataset comprises 5392 multi-source (optical and SAR) ship images, covering scenarios such as ports, near-shore environments, and cloudy weather. The dataset includes six categories: military vessels (e.g., aircraft carriers, cruisers, destroyers, helicopter carriers, civilian ships, and special-purpose vessels). The sample sizes of the six types of ships are as follows: 2196 aircraft carrier samples, 1941 cruiser samples, 2521 destroyer samples, 561 helicopter carrier samples, 261 special-purpose vessels, and 6232 civilian ship samples. Examples from our dataset and some other publicly available datasets are shown in Figure 2. Complex scenarios with occluded ship features account for over 60% of the dataset. To validate the robustness of the proposed method, the dataset is divided into training and testing sets in an 8:2 ratio.

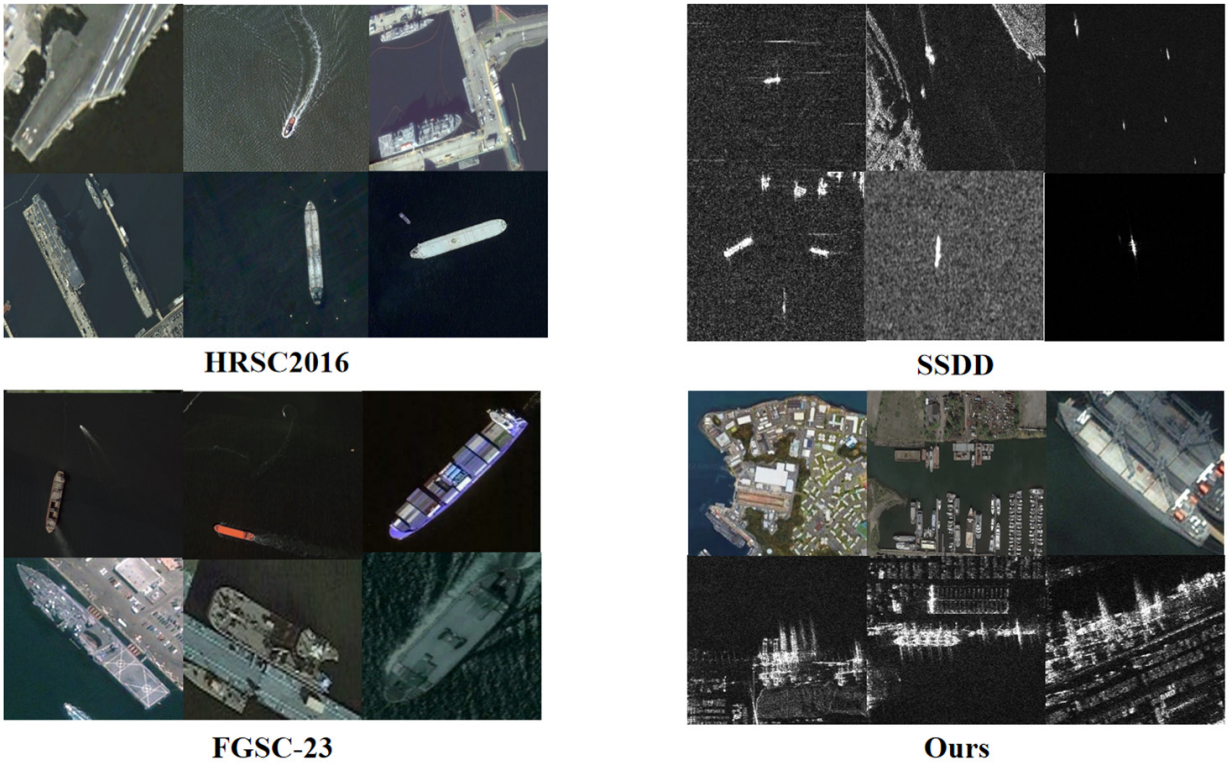



**Figure 2.** Examples of internationally common ship datasets and images from the proposed dataset.

## 3. Methods

In this section, we first introduce the overall structure and principles of the proposed model, followed by detailed descriptions of its key improvement modules and loss function.

### *3.1. Overall Pipeline*

The YOLO series is a classic single-stage object detection algorithm that has garnered significant attention since its introduction. YOLOv11 achieves remarkable performance improvements over its predecessors in object detection, primarily through three aspects: model architecture optimization, multi-scale feature fusion mechanisms, and enhanced training strategies [33]. First, YOLOv11 adopts a novel backbone network architecture with lightweight design, balancing model complexity and feature extraction capability. This enables efficient capture of local details and global contextual information while reducing computational overhead. Second, the multi-scale feature fusion mechanism is further optimized through cross-stage partial connections (CSP) and enhanced feature pyramid

networks (FPNs), improving detection capabilities for small targets and robustness for multi-scale targets in complex scenes.

Additionally, YOLOv11 employs dynamic label assignment and mosaic augmentation during training, allowing the model to more effectively learn feature distributions of challenging samples and improve adaptability to occlusion and dense target scenarios. Combined with an optimized loss function, YOLOv11 achieves a better balance between accuracy and speed, particularly excelling in edge device deployment. Overall, these advancements establish YOLOv11 as a mainstream method in object detection tasks, offering high efficiency and precision. Given its superior detection accuracy and robustness, we select YOLOv11 as the foundation to further address challenges in ship detection, such as small targets, crowding, and occlusion.

In this study, to tailor YOLOv11 for ship recognition tasks, we enhance the network by incorporating three key modules: the spatial pooling convolution block (CKSP), wavelet convolution detection head (WTHead), and dynamic large-kernel attention (DLKA). These additions enrich target feature extraction and improve the detection of occluded ships. The enhanced network is referred to as "YOLO-SWD", which signifies three core innovations integrated into the YOLO framework:

S (strip pooling): Enhances spatial context modeling by aggregating long-range dependencies in horizontal and vertical directions.

W (wavelet convolution): Combines wavelet transforms with convolutional layers to preserve high-frequency details.

D (DLKA): Employs dynamic local-kernel attention to adaptively focus on critical regions while suppressing background noise.

The full network architecture is illustrated in Figure 3.

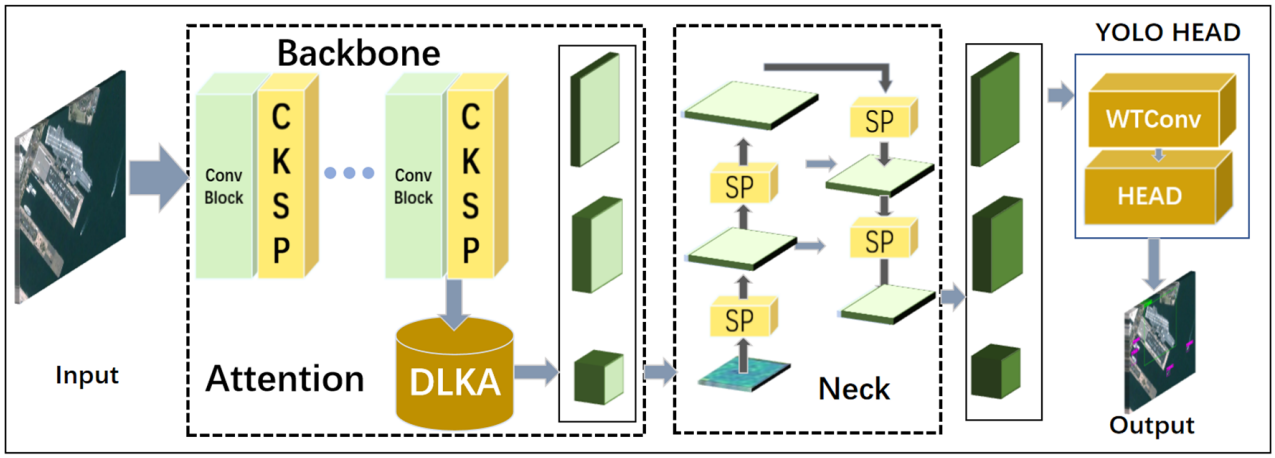



**Figure 3.** Overall architecture of the proposed model.

### *3.2. CKSP Convolution Module Based on StripPooling*

In YOLOv11, the C3K2 module serves as the core feature extraction and fusion unit, designed based on the CSP architecture to enhance model expressiveness and computational efficiency. The CSP module splits input feature maps into two branches: one undergoes deep feature extraction via bottleneck layers while the other retains original information via skip connections. This design reduces computational costs, enhances gradient flow, and improves training effectiveness. The C3K2 module leverages the advantages of CSP architecture by combining deep processing of partial features with direct transmission of others, achieving efficient feature fusion. This not only improves detection performance in complex scenarios but also maintains low computational costs, enabling YOLOv11 to balance real-time processing with strong detection capabilities.

To further enhance YOLOv11's ability to detect occluded targets in complex scenes, this paper integrates the StripPooling [34] module into the C3K2 component. StripPooling captures global contextual information along the horizontal and vertical directions using elongated pooling kernels ($1 \times N$ and $N \times 1$), effectively aggregating long-range dependencies and local context. For input feature map X, StripPooling operates as follows: (1) Extract horizontal and vertical features to capture long-range relationships and local context. (2) Apply 1D convolution to pooled features to extract adjacent feature information. (3) Fuse horizontal and vertical features via summation to generate a feature map enriched with global priors. (4) Refine the fused features using $1 \times 1$ convolution and a sigmoid function to balance feature weights. (5) Multiply the refined features with the original input to produce the final output. The structure diagram of CKSP is shown in Figure 4.

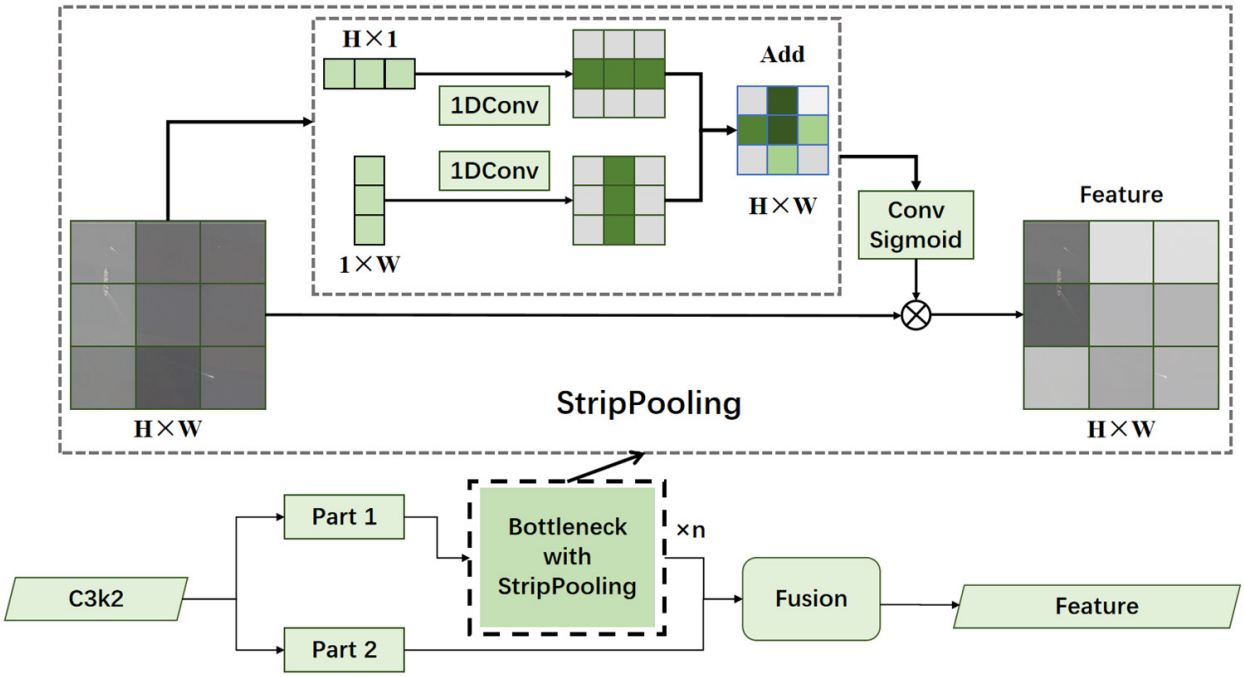



**Figure 4.** Structure of the CKSP module.

After integrating StripPooling, the C3K2 module adjusts input channels via standard convolution, then splits features into a main branch (with StripPooling-enhanced bottlenecks for global context) and an auxiliary branch (preserving original details). The two branches are fused along the channel dimension, followed by $1 \times 1$ convolution for further integration. This significantly enhances YOLOv11's robustness and accuracy for small and occluded targets.

### *3.3. Backbone Network with DLKA Attention Mechanism*

In ship detection tasks, targets exhibit diverse sizes and shapes and mutual occlusion in complex scenes often results in only partial visibility of targets. Traditional convolutions, constrained by fixed sampling patterns, struggle to adaptively adjust to irregular or occluded regions, leading to reduced detection accuracy. To address this, the dynamic large-kernel attention (DLKA) [35,36] module combines deformable convolution and large-kernel convolution, enabling adaptive feature extraction for occluded targets.

Deformable convolution dynamically learns spatial offsets during training, allowing the network to sample features from non-grid locations tailored to target morphology. This is particularly effective for irregular or partially occluded regions. Large-kernel convolution

expands the receptive field to capture broader contextual information, which is critical for long-range dependency modeling and cross-region feature fusion.

The DLKA module is integrated into the PSABlock at the end of the backbone network. The structure diagram of DLKA is shown in Figure 5. The PSABlock employs parallel processing branches to handle multi-scale and multi-region features, leveraging diverse convolution kernel sizes, attention mechanisms, and dilation rates. Within this framework, DLKA introduces two deformable convolution branches:

(1) $5 \times 5$ deformable convolution: Captures medium-range contextual information with a balance between detail preservation and computational efficiency.
(2) $7 \times 7$ dilated deformable convolution (dilation rate = 3): Achieves an expanded receptive field to aggregate long-range dependencies while maintaining resolution.

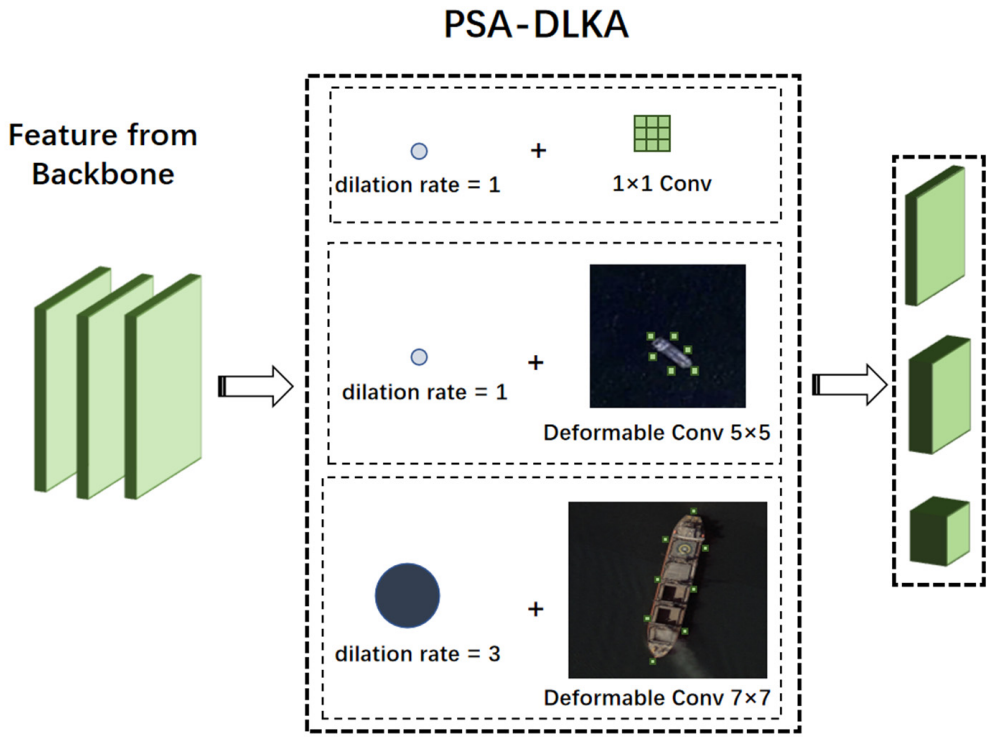



**Figure 5.** Structure of the DLKA module.

These branches operate in parallel, adaptively fusing local details (e.g., ship edges and occluded parts) with global context (e.g., maritime backgrounds and distant targets). By learning dynamic offsets, DLKA enables the network to focus sampling on discriminative regions, overcoming the rigidity of traditional convolutions. This design ensures robust detection of multi-scale occluded ships in complex marine environments while maintaining network lightweightness.

### *3.4. WTHead with Wavelet Convolution*

Wavelet transformation (WT) [37] is a multi-resolution analysis method capable of decomposing input images into subbands of different frequency ranges, modeling both low-frequency and high-frequency information. The low-frequency subband (LL) typically contains global contours and large-scale structures, while the high-frequency subbands (LH, HL, and HH) encapsulate finer details such as textures and edges. Wavelet convolution (WTConv) explicitly integrates wavelet decomposition and reconstruction into convolutional networks. Unlike fixed-size convolutions, WTConv partitions input features in the frequency domain, enabling the network to distinguish high-frequency edges from

low-frequency backgrounds more effectively. This enhances the detection of small targets and the extraction of intricate texture details.

In the head section of YOLOv11, Depthwise Separable Convolution (DWConv) is conventionally used for feature processing and fusion due to its parameter efficiency and fast inference speed. To further optimize this, we replace DWConv with WTConv [38]. The structure diagram of WTConv is shown in Figure 6. The implementation involves the following steps:

(1) Wavelet decomposition: The input features undergo multi-level wavelet decomposition, producing the LL subband and three high-frequency subbands (LH, HL, and HH).
(2) High-frequency processing: Convolution operations are applied to the high-frequency subbands, followed by multiplication with learnable scaling coefficients. These scaled subbands are stored alongside the LL subband.
(3) Reconstruction: The stored LL and high-frequency subbands are inversely combined layer-by-layer to reconstruct features to their original resolution.
(4) Residual fusion: The reconstructed wavelet features are added to the original features via residual connections, ensuring seamless integration of multi-frequency information.

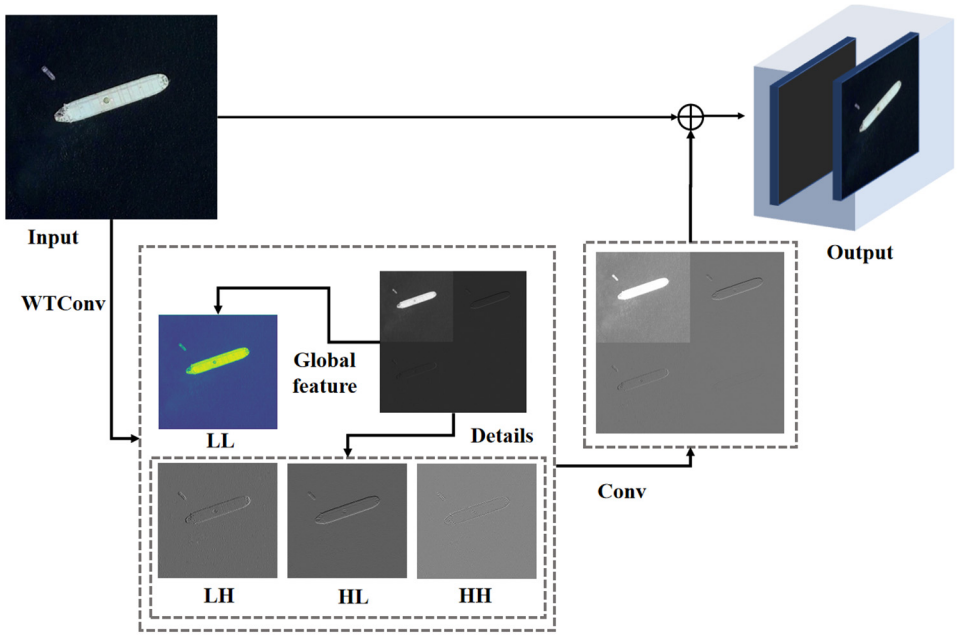



**Figure 6.** Structure of the WTConv module.

Although wavelet decomposition and reconstruction introduce additional computational steps, WTConv remains lightweight compared to traditional large-kernel convolutions or repeated stacking of complex layers. This design maintains computational efficiency while significantly improving the detection accuracy for small targets and high-detail features (e.g., ship components).

### *3.5. Loss Function*

The loss function for object detection and recognition comprises three components: regression loss, classification loss, and confidence loss. Regression loss is typically measured using IoU-based metrics (e.g., GIoU loss), while classification and confidence losses often

employ Cross-Entropy Loss or Focal Loss. For regression loss, we utilize the Generalized Intersection over Union (GIoU) loss [39], defined as

$$L_{Giou} = 1 - IOU + \frac{|C \setminus (A \cup B)|}{|C|} \quad (1)$$

where IOU represents the intersection-over-union ratio between the predicted bounding box A and the ground truth box B; C denotes the smallest enclosing convex shape covering both A and B. Giou extends the standard IOU by introducing a penalty term for non-overlapping boxes, guiding the predicted box to converge toward the ground truth even when no overlap exists initially.

In ship detection tasks, class imbalance is prevalent, where the number of samples across classes (e.g., background vs. rare ship types) varies significantly. Traditional Cross-Entropy Loss tends to overfit majority classes (e.g., background) while underperforming on minority classes (e.g., specific ship targets). To address this, we introduce Adaptive Threshold Focal Loss (ATFL) [40] into YOLOv11's loss computation. The ATFL formula is defined as

$$ATFL = \begin{cases} -(\lambda - p_t)^{-\ln(p_t)} \log(p_t) & p_t \leq 0.5 \\ -(1 - p_t)^{-\ln(\hat{p_c})} \log(p_t) & p_t > 0.5 \end{cases} \quad (2)$$

where $\gamma$ is a modulating factor. Derived from Focal Loss, ATFL assigns higher weights to hard-to-classify samples and lower weights to easy samples, mitigating class imbalance. Unlike standard Focal Loss, ATFL incorporates an adaptive threshold mechanism that dynamically adjusts the focal parameters based on class distribution during training. Specifically, ATFL calculates the sample proportion for each class and adaptively scales the modulating factor, ensuring minority classes receive greater emphasis in loss computation. By integrating ATFL into the loss module, the model prioritizes feature learning for minority classes while maintaining stable overall detection performance.

### 3.6. Evaluation Metrics

Common performance metrics for object detection and recognition algorithms include Precision, Recall, Average Precision (AP), and mean Average Precision (mAP). These metrics are derived from four fundamental components: True Positives (TPs): Correctly predicted positive samples; True Negatives (TNs): Correctly predicted negative samples; False Positives (FPs): Incorrectly predicted positive samples (false alarms); and False Negatives (FNs): Incorrectly predicted negative samples (missed detections).

The formulas for these metrics are as follows:

$$Accuracy = (TP + TN)/(TP + FN + FP + TN) \quad (3)$$

$$Precision = TP/(TP + FP) \quad (4)$$

$$Recall = TP/(TP + FN) \quad (5)$$

$$AP = \int_0^1 Precision(Recall) dRecall \quad (6)$$

$$mAP = \frac{\sum_{i=0}^{n} AP(i)}{n} \quad (7)$$

Precision represents the ratio of correctly identified true ship targets to the total number of detected ships. The false alarm rate is defined as 1−Precision, reflecting the proportion of incorrect detections among all predictions. Recall measures the ratio of correctly identified true ship targets to the total number of actual ship targets in the dataset. The miss rate (1−Recall) quantifies the proportion of undetected true targets.

Average Precision (AP) is determined by calculating the area under the Precision–Recall curve for each class, which integrates both Precision and Recall values across varying confidence thresholds. AP comprehensively evaluates the performance of an object detection network for a specific class. To assess overall model performance, the mean Average Precision (mAP) is computed by averaging AP values across all classes. Higher AP and mAP values indicate superior detection accuracy and robustness of the network.

## 4. Results and Analysis

### *4.1. Experimental Environment*

The proposed model is implemented based on the Darknet framework. All neural network structures required for training and experiments are deployed on a server equipped with an RTX 4090D GPU (24 GB), an AMD EPYC 9754 128-Core Processor CPU, and 60 GB of RAM. The environment utilizes CUDA 12.0 and Python 3.6, ensuring strong compatibility with GPU platforms.

In this study, the Adam (Adaptive Moment Estimation) optimizer is employed for model training. The Adam optimizer minimizes the discrepancy between network predictions and ground truth values during training. Its hyperparameters—$\beta_1$, $\beta_2$ and $\varepsilon$—were set as 0.9, 0.999 and $10^{-8}$, respectively. The initial learning rate was set to $10^{-6}$. The training process runs for 300 epochs to obtain final results. The Precision–Recall (P-R) curves and F1–score curves are illustrated in Figure 7. The accuracy metrics Average Precision (AP) and mean Average Precision (mAP) are used to evaluate the precision of ship recognition.

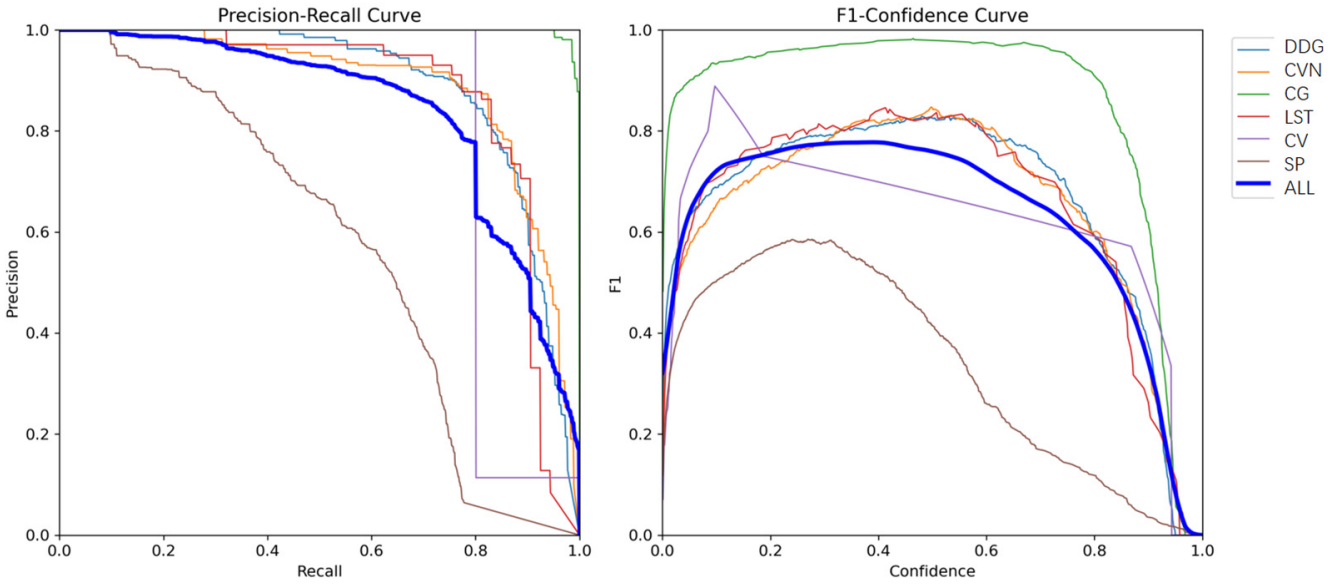



**Figure 7.** Training mAP curves and loss function curves.

### *4.2. Experimental Results*

We compared the improved YOLO-SWD with recent YOLO series algorithms, including YOLOv8 [41], YOLOv9 [42], YOLOv10 [43], Faster-RCNN [44], CenterNet [45], and YOLOv11. Considering hardware limitations and inference speed, all YOLO networks were configured with similar parameter scales: YOLOv9 adopts the "t" structure, while other YOLO versions use the "n" structure.

In near-shore occlusion scenarios, object detection algorithms face challenges such as densely docked ships, port facility occlusion, and weather interference. In these scenarios, ship targets are frequently occluded by structures or weather conditions or submerged in the background, significantly increasing detection difficulty. The visualization results

are shown in Figure 8. Based on the experimental data and visualizations, we analyze the performance of each algorithm as follows:

(1) Under weather interference: For large ship categories such as CVNs (aircraft carriers), DDGs (destroyers), CGs (cruisers), and LSTs (large ships), cloud and fog occlusion often obscure parts of the ship structures, preventing models from learning complete features. As shown in Figure 8(9), YOLOv8, YOLOv10, and YOLOv11 fail to detect the cloud-occluded carrier, while YOLOv9, Faster-RCNN, and our method succeed, demonstrating our algorithm's ability to learn local features and reduce missed detections. Additionally, as shown in Figure 8(8), when cloud shapes resemble ship contours, only YOLOv11 and our method correctly identify the ship near the carrier. In summary, our algorithm achieves the best performance under weather interference, primarily due to the enhanced local edge and contextual feature extraction capabilities of the improved modules, which are absent in other algorithms.

(2) In near-shore scenarios: For medium and small ship categories such as CVs (civilian vessels), DDGs, and CGs, their features are easily occluded by dock facilities, as shown in Figure 8(1,4). Notably, our method exhibits the lowest miss rate overall, outperforming YOLOv9, Faster-RCNN, and YOLOv11, which struggle when large and medium targets coexist. YOLOv8, CenterNet, and YOLOv10 perform the worst, with severe miss rates. For the CV category, due to the small size of targets and dense overlapping (over 60% overlap in docking areas), miss rates are particularly severe. As shown in Figure 8(2), CV-class ships are densely packed, occluding each other's edges. The low resolution (20–30 pixels per target) further exacerbates miss rates. All the algorithms exhibit significant miss rates, but our method, CenterNet, and YOLOv11 achieve the best overall performance, with no widespread misses, while YOLOv8 and YOLOv10 perform the worst. As shown in Figure 8(11,12), in noisy images, all the algorithms exhibited a certain degree of false positives and false negatives, with YOLOv8 performing the worst, being almost unable to detect any targets. However, thanks to the improved modules that enhance the model's ability to extract ship features more accurately, our model had no false negatives and exhibited the fewest missed detections, achieving the best performance. When port facilities dominate the interference (Figure 8(3,6,10)), our method and CenterNet achieve the best overall performance, with minimal misses despite occasional false positives. YOLOv9 and YOLOv11 follow with more misses, while YOLOv8 and YOLOv10 nearly fail to detect CV-class ships under port interference.

From Table 1, the performance gaps among algorithms in the CV category are significant: YOLOv8 (Precision 68.5%/Recall 45.9%) and YOLOv9 (Precision 70.8%/Recall 47.4%) perform poorly. Faster-RCNN (Precision 78.2%/Recall 33.8%) achieves the highest precision but suffers from the lowest Recall. YOLOv10 (Precision 68.2%/Recall 47.2%) and YOLOv11 (Precision 79.6%/Recall 51.8%) show moderate improvements. Our method achieves a Precision of 68.5%, Recall of 53.5%, and F1 of 60.1%, outperforming other versions, particularly in recall rate, indicating reduced miss rates. However, CVs remain a challenging category due to severe near-shore occlusion, small target sizes, and inherent model limitations in complex environments. Despite this, our model maintains balanced Precision and Recall, demonstrating advantages in minimizing both false alarms and misses.

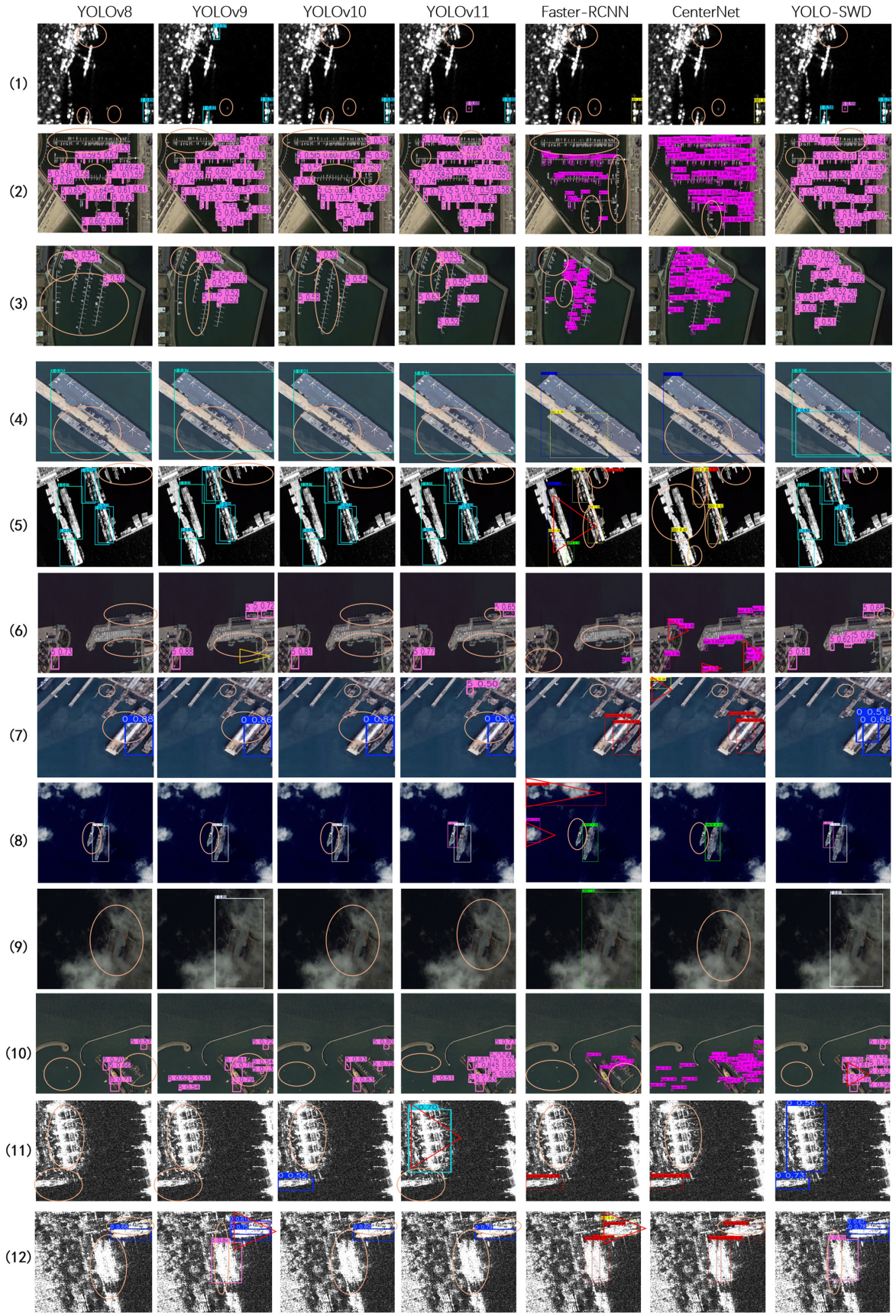



**Figure 8.** Qualitative comparison results (1–12) (magenta: CV, navy blue: DDG, white: CVN, light green: LST, light blue: CG; the orange and blue ellipses and red triangles represent missed detections and wrong detections, respectively).

**Table 1.** Quantitative comparison of experimental results.

| Method | Category | Precision (%) | Recall (%) | F1 | AP (%) | MAP (%) |
|---|---|---|---|---|---|---|
| YOLOv8 | DDG | 81.8 | 86.6 | 84.1 | 89.6 | 80.5 |
| | CVN | 95.6 | 99.0 | 97.3 | 99.4 | |
| | CG | 76.9 | 88.9 | 82.5 | 91.1 | |
| | LST | 76.6 | 79.2 | 77.9 | 84.6 | |
| | CV | 68.5 | 45.9 | 54.9 | 55.2 | |
| | SP | 79.6 | 60.0 | 68.4 | 62.8 | |
| YOLOv9 | DDG | 84.2 | 83.4 | 83.8 | 89.9 | 81.0 |
| | CVN | 95.1 | 99.5 | 97.3 | 99.3 | |
| | CG | 77.4 | 89.2 | 82.9 | 82.1 | |
| | LST | 88.2 | 84.3 | 86.2 | 87.2 | |
| | CV | 70.8 | 47.4 | 56.8 | 56.9 | |
| | SP | 47.0 | 60.0 | 52.7 | 60.6 | |
| YOLOv10 | DDG | 81.5 | 85.0 | 83.2 | 88.8 | 80.1 |
| | CVN | 94.3 | 96.6 | 95.4 | 98.9 | |
| | CG | 73.1 | 86.4 | 79.2 | 87.6 | |
| | LST | 79.0 | 77.4 | 78.2 | 78.2 | |
| | CV | 68.2 | 47.2 | 55.8 | 55.8 | |
| | SP | 100 | 53.1 | 69.4 | 71.3 | |
| YOLOv11 | DDG | 80.6 | 88.5 | 84.4 | 90.5 | 81.2 |
| | CVN | 96.5 | 97.6 | 97.0 | 99.1 | |
| | CG | 74.0 | 90.4 | 81.4 | 91.0 | |
| | LST | 84.8 | 79.2 | 81.9 | 84.6 | |
| | CV | 79.6 | 51.8 | 62.7 | 57.8 | |
| | SP | 100 | 58.8 | 74.1 | 66.5 | |
| Faster-RCNN | DDG | 75.7 | 76.6 | 76.1 | 80.6 | 71.0 |
| | CVN | 89.2 | 96.4 | 92.6 | 95.8 | |
| | CG | 69.2 | 74.3 | 71.6 | 79.9 | |
| | LST | 74.1 | 65.3 | 69.4 | 75.8 | |
| | CV | 78.2 | 33.8 | 47.1 | 43.9 | |
| | SP | 100 | 33.3 | 40.0 | 50.1 | |
| CenterNet | DDG | 89.8 | 80.6 | 84.9 | 90.5 | 75.7 |
| | CVN | 96.4 | 94.5 | 95.4 | 97.3 | |
| | CG | 80.2 | 73.2 | 76.5 | 81.9 | |
| | LST | 77.8 | 68.8 | 73.0 | 69.2 | |
| | CV | 73.7 | 51.3 | 60.4 | 56.7 | |
| | SP | 100 | 25 | 49.9 | 58.4 | |
| YOLO-SWD | DDG | 83.7 | 85.4 | 84.5 | 90.7 | 83.9 |
| | CVN | 95.4 | 98.6 | 96.9 | 99.3 | |
| | CG | 74.7 | 90.9 | 82.0 | 91.9 | |
| | LST | 81.9 | 79.2 | 80.5 | 85.3 | |
| | CV | 68.5 | 53.5 | 60.1 | 58.1 | |
| | SP | 100 | 71.7 | 83.5 | 78.1 | |

Comprehensive analysis reveals that from YOLOv8 to YOLO-SWD, all algorithms exhibit gradual optimization across most categories, particularly in large target categories (DDG and CVN), but significant gaps persist in small-target (CV) and low-sample (SP) categories. A detailed class-specific analysis follows:

In the DDG category, CenterNet leads all models with 89.8% Precision, slightly outperforming our method (83.7%). Other models, such as YOLOv8 (81.8%) and YOLOv10 (81.5%), lag behind. In recall, YOLOv11 leads (88.5%), but our method balances Precision

and Recall, achieving the highest AP (90.7%). In the CV category, all models exhibit low AP, while our method maintains balanced Precision and Recall, minimizing false alarms and misses.

In the CG category, YOLOv9 and YOLOv8 perform similarly, with Precision rates of 77.4% and 76.9%, respectively, ranking highest among all algorithms. However, our method leads in Recall (90.9%), with YOLOv11 following at 90.4%. YOLOv9 achieves the best overall AP (92.1%), while our method achieves 91.9%, ranking second. For CVN targets, most algorithms exceed 95% Recall and 99% AP (except YOLOv10, Faster-RCNN, and CenterNet), with minimal performance gaps due to large target sizes and distinct features.

For the LST and SP categories, which have smaller sample sizes, algorithm performance varies widely. In LST, YOLOv9 achieves the highest Recall (84.3%) and AP (87.2%), demonstrating strong small-sample feature capture. Our method achieves an AP of 85.3%, surpassing the others, indicating improved small-sample recognition capabilities. For SP (smallest sample size), YOLOv9 performs worst (Precision < 50%), while YOLOv10, YOLOv11, Faster-RCNN, CenterNet and our method achieve 100% Precision but low Recall rates, indicating high-confidence threshold filtering at the cost of Recall. Our method achieves the highest Recall (71.7%) and AP (78.1%), demonstrating robust feature extraction for small-sample categories. The key to achieving higher precision with our model lies in mechanisms such as strip pooling and DLKA attention, which allow the model to extract more key features from the limited small-sample categories. Additionally, the ATFL loss function adaptively assigns greater weights to the small-sample categories, ensuring the model places more focus on these categories, which in turn leads to higher precision.

Further analysis of AP metrics shows that YOLO-SWD excels in all categories, achieving the best results in three of six categories and second-best in the remaining three, with minimal gaps to the top performers. This highlights its significant precision advantages in complex near-shore scenarios, effectively addressing occlusion and visual diversity.

Finally, comprehensive comparison of mAP across models shows that our YOLO-SWD method achieves the highest 83.9%, significantly outperforming the others. This conclusively demonstrates its superior comprehensive performance in occlusion-heavy and complex scenarios. YOLO-SWD's improved network architecture and optimized feature extraction capabilities enable better adaptation to dynamic changes and multi-scale targets in near-shore occlusion environments.

## 5. Discussion

To evaluate the contributions of the three key modules, this paper designs progressive ablation experiments, with results shown in Table 2. Starting from the baseline model (original YOLOv11), modules are incrementally added (Schemes 1–3) to validate their individual and synergistic effects by comparing changes in mean Average Precision (mAP) and class-wise AP.

In Scheme 1, the original attention mechanism is replaced with the dynamic local-kernel attention (DLKA) module to enhance feature selection for ship targets. The DLKA module works by dynamically adjusting the attention weights applied to local regions in the image. This allows the model to focus on the most critical regions of the image by adaptively scaling attention across different spatial locations. As a result, DLKA helps the model to more effectively capture fine-grained details of the ship structures, leading to an improvement in the detection of mid-sized ships. Table 2 shows that the mAP increases from 81.2% to 82.5%, confirming DLKA's effectiveness in focusing on crucial regions of interest. Furthermore, small-sample categories (e.g., LST and SP) show significant improvements in Recall and AP, indicating DLKA's ability to better capture features and enhance recognition

for these targets. However, for other categories, there is a slight decrease in AP, which could be attributed to the model overly prioritizing small-sample classes, potentially limiting its ability to generalize to broader feature extraction.

**Table 2.** Quantitative comparison of ablation experimental results.

| Method | DLKA | CKSP | WTConv | Category | Precision (%) | Recall (%) | AP (%) | mAP (%) |
|---|---|---|---|---|---|---|---|---|
| Original | | | | DDG | 80.6 | 88.5 | 90.5 | 81.2 |
| | | | | CVN | 96.5 | 97.6 | 99.1 | |
| | | | | CG | 74.0 | 90.4 | 91.0 | |
| | | | | LST | 84.8 | 79.2 | 84.6 | |
| | | | | CV | 79.6 | 51.8 | 57.8 | |
| | | | | SP | 100 | 58.8 | 66.5 | |
| Scheme 1 | √ | | | DDG | 76.9 | 88.1 | 89.8 | 82.5 |
| | | | | CVN | 95.6 | 98.6 | 99.4 | |
| | | | | CG | 70.4 | 94.5 | 90.5 | |
| | | | | LST | 73.7 | 83.0 | 87.1 | |
| | | | | CV | 59.4 | 56.0 | 55.6 | |
| | | | | SP | 100 | 54.7 | 72.6 | |
| Scheme 2 | √ | √ | | DDG | 79.1 | 89.3 | 90.6 | 83.1 |
| | | | | CVN | 94.1 | 98.1 | 99.3 | |
| | | | | CG | 75.7 | 94.3 | 91.7 | |
| | | | | LST | 85.3 | 81.1 | 83.2 | |
| | | | | CV | 68.6 | 49.0 | 54.5 | |
| | | | | SP | 100 | 71.2 | 79.6 | |
| Scheme 3 | √ | √ | √ | DDG | 83.7 | 85.4 | 90.7 | 83.9 |
| | | | | CVN | 95.4 | 98.6 | 99.3 | |
| | | | | CG | 74.7 | 90.9 | 91.9 | |
| | | | | LST | 81.9 | 79.2 | 85.3 | |
| | | | | CV | 68.5 | 53.5 | 58.1 | |
| | | | | SP | 100 | 71.7 | 78.1 | |

In Scheme 2, the introduction of both DLKA and CKSP results in further performance gains, with mAP rising to 83.1%. CKSP (Contextual Kernel Spatial Pooling) improves the model's ability to capture critical spatial and structural features, especially for targets that require fine structural information. The key mechanism used in CKSP is StripPooling, which operates by pooling features along specific regions while maintaining the structural integrity of the target. StripPooling preserves spatial details by aggregating information in localized "strips" of the feature map, ensuring that the spatial relationships within a target are well captured. This approach improves the detection of targets with complex structures, as seen in SP's AP increase from 72.6% to 79.6%. CKSP's ability to focus on spatial relationships and finer details leads to better performance for such targets. Categories that previously experienced AP declines in Scheme 1 recover slightly due to CKSP's enhanced feature extraction. However, performance fluctuates for small-sample categories (e.g., LST and CV), highlighting the challenge of balancing the preservation of local details with general feature learning.

Finally, in Scheme 3, integrating all three modules (DLKA, CKSP, and WTConv) achieves the optimal performance, with mAP reaching 83.9%. WTConv (wavelet transform convolution) enhances convolutional feature extraction by introducing multi-resolution capabilities, allowing the model to capture features at multiple scales. This helps to improve detection across various target sizes, especially those with diverse spatial characteristics. The synergistic effect of DLKA, CKSP, and WTConv enables the model to perform better at

both the local and global levels. DLKA focuses on critical regions, CKSP enhances spatial and structural comprehension through StripPooling, and WTConv allows for more robust feature extraction across scales. As a result, the overall detection accuracy and stability improve, with consistent improvements in categories like DDG, CG, LST, and CV. Despite a slight drop in SP's AP (from 79.6% to 78.1%), the overall performance demonstrates that the three modules complement each other, providing a comprehensive solution for enhanced feature capture, spatial understanding, and convolutional expression.

In summary, each module contributes uniquely to the model's performance: DLKA improves fine-grained feature capture by focusing on critical regions, CKSP enhances spatial structure understanding through StripPooling, and WTConv boosts convolutional feature extraction across scales. Together, they provide complementary benefits that result in superior detection accuracy and model performance. In industries such as maritime surveillance, autonomous navigation, and satellite imagery analysis, the proposed algorithm can be leveraged to detect and recognize ships in diverse environments, including those with varying target sizes, complex structures, and occlusions. The ability to effectively handle small-sample categories and preserve spatial details ensures that the model performs well even in challenging scenarios, thus enhancing the overall reliability and performance of ship detection systems.

In addition, as shown in Table 3, compared to existing methods limited to single data sources, our approach demonstrates cross-domain versatility by integrating multi-sensor data, At the same time, our framework comprehensively improves the ship interference and occlusion issues in SAR and optical images through physics-based detailed feature learning and cross-sensor adaptability, providing new insights for ship detection and recognition in occluded scenarios under all-weather, multi-source imagery.

**Table 3.** Comparison table with some other researchers' studies.

| Method | Datasets Source | Methods | Achievement |
| --- | --- | --- | --- |
| Sun et al. [24] | Gaofen-3 SAR satellite imagery | (1) Leveraging enhanced regional-global feature extraction modules;<br>(2) Adaptive fusion mechanisms. | Address SAR scattering interference and feature confusion. |
| Wang et al. [25] | AIR–SARShip-1 dataset | Integrating anchor optimization, asymmetric pyramidal nonlocal blocks with SimAM attention, channel shuffling for feature enhancement, and Ghost convolution to balance performance and resource demands. | Address SAR ship detection complex inshore environments—multi-scale target variability, background interference, small-target oversight, and computational inefficiency. |
| Yan et al. [27] | DOTA dataset (optical) | Perlin noise fog-mixing module is used to generate realistic cloud data. High-quality prior knowledge is then applied to specifically restore the ship regions in the cloud images. Additionally, a target-oriented structural similarity loss is employed to enhance the network's ability to focus on cloud-occluded areas. | Address the issues of blurred ship boundaries and target occlusion caused by cloud cover in optical remote sensing images. |

**Table 3.** *Cont.*

| Method | Datasets Source | Methods | Achievement |
| --- | --- | --- | --- |
| Ours | Gaofen-2<br>Gaofen-3<br>RadarSat-2<br>Sentinel-1<br>Sentinel-2<br>Jilin-1<br>SPOT<br>Google Earth | (1) DLKA introduces dynamic deformable convolution and large-receptive-field attention mechanisms, enhancing local detail and global contextual modeling;<br>(2) StripPooling improves perception of targets in complex scenes;<br>(3). WTConv strengthens feature extraction diversity and robustness. | Enhance model robustness in complex scenarios and improve accuracy for occluded ship targets in multi-source images caused by weather and interference obstruction. |

## 6. Conclusions and Future Work

This paper addresses the challenges of small-target crowding and occlusion in complex ship scenarios by proposing an improved ship detection and recognition algorithm based on YOLOv11. The main contributions include the following: (1) DLKA introduces dynamic deformable convolution and large-receptive-field attention mechanisms, enhancing local detail and global contextual modeling; (2) StripPooling improves the perception of targets in complex scenes. (3) WTConv strengthens feature extraction diversity and robustness.

Experimental results indicate that the improved YOLOv11 model—YOLO-SWD—further enhances recognition accuracy for ships in occluded scenarios within remote sensing imagery and strengthens the anti-interference capability of the YOLO-based ship detection algorithm in complex environments.

This study investigates ship detection and recognition in remote sensing images but still faces unresolved issues requiring further research:

(1) Persistent ship tracking: Research on continuous ship tracking in image sequences requires further exploration.
(2) Rotatable bounding boxes: Current detection algorithms do not account for bounding box rotation.
(3) Model lightweighting: The current architecture remains relatively complex, posing deployment challenges.

Future work should focus on integrating rotatable detection boxes to improve ship localization accuracy, especially in scenarios with varying ship orientations and dense distributions. Additionally, the current algorithm should be extended to include ship tracking functionality, enabling the model to track ships in complex occlusion environments. This can be achieved by incorporating multi-frame information, feature association, and temporal consistency to ensure accurate tracking in dynamic and obstructed settings. Furthermore, simplifying the model to reduce computational complexity while maintaining performance will be essential for real-time deployment on resource-constrained devices.

**Author Contributions:** Conceptualization, R.Z. and M.G.; methodology, R.Z. and M.G.; software, M.G.; validation, R.Z.; formal analysis, R.Z.; investigation, M.G.; resources, M.G.; data curation, M.G.; writing—original draft preparation, M.G. and R.Z.; writing—review and editing, M.G. and H.P; visualization, R.Z.; supervision, R.Z. and H.P.; project administration, R.Z. and H.P.; funding acquisition, R.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The dataset is currently non-public due to ongoing research. We plan to release it after further improvements addressing existing limitations in this work.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Zhu, C.; Zhou, H.; Wang, R.; Guo, J. A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 3446–3456. [CrossRef]
2. Paolo, F.S.; Kroodsma, D.; Raynor, J.; Hochberg, T.; Davis, P.; Cleary, J.; Halpin, P. Satellite mapping reveals extensive industrial activity at sea. *Nature* **2024**, *625*, 85–91. [CrossRef]
3. Leira, F.S.; Johansen, T.A.; Fossen, T.I. Automatic detection, classification and tracking of objects in the ocean surface from UAVs using a thermal camera. In Proceedings of the 2015 IEEE Aerospace Conference, Big Sky, MT, USA, 7–14 March 2015; pp. 1–10.
4. Wang, Y.; Ning, X.; Leng, B.; Fu, H. Ship detection based on deep learning. In Proceedings of the 2019 IEEE International Conference on Mechatronics and Automation (ICMA), Tianjin, China, 4–7 August 2019; pp. 275–279.
5. Hong, Z.; Yang, T.; Tong, X.; Zhang, Y.; Jiang, S.; Zhou, R.; Han, Y.; Wang, J.; Yang, S.; Liu, S. Multi-scale ship detection from SAR and optical imagery via a more accurate YOLOv3. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 6083–6101. [CrossRef]
6. Leclerc, M.; Tharmarasa, R.; Florea, M.; Boury-Brisset, A.; Kirubarajan, T.; Duclos-Hindié, N. Ship classification using deep learning techniques for maritime target tracking. In Proceedings of the 2018 21st International Conference on Information Fusion (FUSION), Cambridge, UK, 10–13 July 2018; pp. 737–744.
7. Chen, Z.; Chen, D.; Zhang, Y.; Cheng, X.; Zhang, M.; Wu, C. Deep learning for autonomous ship-oriented small ship detection. *Saf. Sci.* **2020**, *130*, 104812. [CrossRef]
8. Escorcia-Gutierrez, J.; Gamarra, M.; Beleño, K.; Soto, C.; Mansour, R.F. Mansour, Intelligent deep learning-enabled autonomous small ship detection and classification model. *Comput. Electr. Eng.* **2022**, *100*, 107871. [CrossRef]
9. Xu, Y.; Lang, H. Distribution shift metric learning for fine-grained ship classification in SAR images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 2276–2285. [CrossRef]
10. Raj, J.A.; Idicula, S.M.; Paul, B. One-shot learning-based SAR ship classification using new hybrid Siamese network. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]
11. Xiong, G.; Xi, Y.; Chen, D.; Yu, W. Dual-polarization SAR ship target recognition based on mini hourglass region extraction and dual-channel efficient fusion network. *IEEE Access* **2021**, *9*, 29078–29089. [CrossRef]
12. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
13. Tang, P.; Wang, X.; Bai, X.; Liu, W. Multiple instance detection network with online instance classifier refinement. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2843–2851.
14. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
15. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Proceedings, Part I 14, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
16. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
17. Yasir, M.; Shanwei, L.; Mingming, X.; Jianhua, W.; Hui, S.; Nazir, S.; Zhang, X.; Colak, A.T.I. YOLOv8-BYTE: Ship tracking algorithm using short-time sequence SAR images for disaster response leveraging GeoAI. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *128*, 103771. [CrossRef]
18. Wu, T.; Li, B.; Luo, Y.; Wang, Y.; Xiao, C.; Liu, T.; Yang, J.; An, W.; Guo, Y. MTU-Net: Multilevel TransUNet for space-based infrared tiny ship detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–15. [CrossRef]
19. Zhao, S.; Luo, Y.; Zhang, T.; Guo, W.; Zhang, Z. A domain specific knowledge extraction transformer method for multisource satellite-borne SAR images ship detection. *ISPRS J. Photogramm. Remote Sens.* **2023**, *198*, 16–29. [CrossRef]
20. Xu, C.; Wang, J.; Yang, W.; Yu, H.; Yu, L.; Xia, G.-S. Detecting tiny objects in aerial images: A normalized Wasserstein distance and a new benchmark. *ISPRS J. Photogramm. Remote. Sens.* **2022**, *190*, 79–93. [CrossRef]
21. Zhang, L.; Liu, Y.; Zhao, W.; Wang, X.; Li, G.; He, Y. Frequency-adaptive learning for SAR ship detection in clutter scenes. *IEEE Trans. Geosci. Remote. Sens.* **2023**, *61*, 1–14. [CrossRef]
22. Zhang, X.; Feng, S.; Zhao, C.; Sun, Z.; Zhang, S.; Ji, K. MGSFA-Net: Multi-scale global scattering feature association network for SAR ship target recognition. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 4611–4625. [CrossRef]

23. Shang, Y.; Pu, W.; Liao, D.; Ji, Y.; Wu, C.; Huang, Y.; Zhang, Y.; Wu, J.; Yang, J.; Wu, J. SA 2 Net: Ship Augmented Attention Network for Ship Recognition in SAR Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 10036–10049. [CrossRef]
24. Sun, Z.; Leng, X.; Zhang, X.; Xiong, B.; Ji, K.; Kuang, G. Ship Recognition for Complex SAR Images via Dual-Branch Transformer Fusion Network. *IEEE Geosci. Remote. Sens. Lett.* **2024**, *21*, 1–5. [CrossRef]
25. Wang, Z.; Hou, G.; Xin, Z.; Liao, G.; Huang, P.; Tai, Y. Detection of SAR Image Multiscale Ship Targets in Complex Inshore Scenes Based on Improved YOLOv5. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **2024**, *17*, 5804–5823. [CrossRef]
26. Wang, N.; Wang, Y.; Feng, Y.; Wei, Y. MDD-ShipNet: Math-Data Integrated Defogging for Fog-Occlusion Ship Detection. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 15040–15052. [CrossRef]
27. Yan, Y.; He, Y.; Su, N.; He, G.; Fu, H. PNBT-CR: A Cloud Removal Method for Ship Detection. *IEEE Geosci. Remote. Sens. Lett.* **2024**, *21*, 1–5. [CrossRef]
28. Tian, Y.; Meng, H.; Yuan, F.; Ling, Y.; Yuan, N. Vision transformer with enhanced self-attention for few-shot ship target recognition in complex environments. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–12. [CrossRef]
29. Di, Y.; Jiang, Z.; Zhang, H. A public dataset for fine-grained ship classification in optical remote sensing images. *Remote. Sens.* **2021**, *13*, 747. [CrossRef]
30. Huang, L.; Liu, B.; Li, B.; Guo, W.; Yu, W.; Zhang, Z.; Yu, W. OpenSARShip: A dataset dedicated to Sentinel-1 ship interpretation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *11*, 195–208. [CrossRef]
31. Liu, Z.; Yuan, L.; Weng, L.; Yang, Y. A high resolution optical satellite image dataset for ship recognition and some new baselines. *Int. Conf. Pattern Recognit. Appl. Methods* **2017**, *2*, 324–331.
32. Zhang, T.; Zhang, X.; Li, J.; Xu, X.; Wang, B.; Zhan, X.; Xu, Y.; Ke, X.; Zeng, T.; Su, H.; et al. SAR ship detection dataset (SSDD): Official release and comprehensive data analysis. *Remote. Sens.* **2021**, *13*, 3690. [CrossRef]
33. Alkhammash, E.H. Multi-Classification Using YOLOv11 and Hybrid YOLO11n-MobileNet Models: A Fire Classes Case Study. *Fire* **2025**, *8*, 17. [CrossRef]
34. Hou, Q.B.; Zhang, L.; Cheng, M.M.; Feng, J.S. Strip Pooling: Rethinking Spatial Pooling for Scene Parsing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
35. Feng, X.; Hu, J.; Wu, W.; Fan, S. Dynamic Large-Small Kernel Convolutional Neural Network for Pansharpening. *IEEE Geosci. Remote. Sens. Lett.* **2023**, *20*, 1–5. [CrossRef]
36. Chen, Y.P.; Dai, X.Y.; Liu, M.C.; Chen, D.D.; Yuan, L.; Liu, Z.C. Dynamic Convolution: Attention Over Convolution Kernels. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
37. Zhang, D. Wavelet Transform. In *Fundamentals of Image Data Mining*; Texts in Computer Science; Springer: Berlin/Heidelberg, Germany, 2019.
38. Finder, S.E.; Amoyal, R.; Treister, E.; Freifeld, O. Wavelet Convolutions for Large Receptive Fields. In *Computer Vision–ECCV 2024*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2024.
39. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.
40. Yan, W.; Qian, Y.; Wang, C.; Yang, M. Threshold-Adaptive Unsupervised Focal Loss for Domain Adaptation of Semantic Segmentation. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 752–763. [CrossRef]
41. Sohan, M.; Sai Ram, T.; Reddy, R.; Venkata, C. A Review on YOLOv8 and Its Advancements. In *Data Intelligence and Cognitive Informatics, ICDICI*; Springer: Singapore, 2023.
42. Wang, C.Y.; Yeh, I.H.; Mark Liao, H.Y. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. In *Computer Vision–ECCV*; Springer: Cham, Switzerland, 2024.
43. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G.G. YOLOv10: Real-Time End-to-End Object Detection. *Adv. Neural Inf. Process. Syst.* **2025**, *37*, 107984–108011.
44. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [CrossRef]
45. Duan, K.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. Centernet: Keypoint triplets for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, South Korea, 27 October–2 November 2019.


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.