Journal of *Marine Science and Engineering*

MDPI

Article

Multi-Attention Pyramid Context Network for Infrared Small Ship Detection

Feng Guo [1,2], Hongbing Ma [1,2,3,*], Liangliang Li [4,*], Ming Lv [1,2] and Zhenhong Jia [1,2]

[1] School of Computer Science and Technology, Xinjiang University, Urumqi 830046, China; peerlessgf@163.com (F.G.)
[2] Key Laboratory of Signal Detection and Processing, Xinjiang University, Urumqi 830046, China
[3] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[4] School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China
* Correspondence: hbma@tsinghua.edu.cn (H.M.); leeliangliang@163.com (L.L.)

Abstract: In the realm of maritime target detection, infrared imaging technology has become the predominant modality. Detecting infrared small ships on the sea surface is crucial for national defense and maritime security. However, the challenge of detecting infrared small targets persists, especially in the complex scenes of the sea surface. As a response to this challenge, we propose MAPC-Net, an enhanced algorithm based on an existing network. Unlike conventional approaches, our method focuses on addressing the intricacies of sea surface scenes and the sparse pixel occupancy of small ships. MAPC-Net incorporates a scale attention mechanism into the original network's multi-scale feature pyramid, enabling the learning of more effective scale feature maps. Additionally, a channel attention mechanism is introduced during the upsampling process to capture relationships between different channels, resulting in superior feature representations. Notably, our proposed Maritime-SIRST dataset, meticulously annotated for infrared small ship detection, is introduced to stimulate advancements in this research domain. Experimental evaluations on the Maritime-SIRST dataset demonstrate the superiority of our algorithm over existing methods. Compared to the original network, our approach achieves a 6.14% increase in mIOU and a 4.41% increase in F1, while maintaining nearly unchanged runtime.

Keywords: infrared small target; maritime; small ship detection; attention; deep learning

Citation: Guo, F.; Ma, H.; Li, L.; Lv, M.; Jia, Z. Multi-Attention Pyramid Context Network for Infrared Small Ship Detection. *J. Mar. Sci. Eng.* **2024**, *12*, 345. https://doi.org/10.3390/jmse12020345

Academic Editor: Weicheng Cui

Received: 11 January 2024
Revised: 9 February 2024
Accepted: 15 February 2024
Published: 17 February 2024

1. Introduction

Infrared imaging technology has found widespread applications in both military and civilian domains due to its high imaging accuracy, superior concealment, extensive detection range, and resistance to electromagnetic interference. Particularly in the field of maritime target detection, infrared imaging technology has become a primary imaging modality and a key direction for overcoming existing guidance technology bottlenecks [1–3]. In comparison to visible light imaging, infrared imaging demonstrates significant advantages in penetrating smoke, working in all weather conditions, and being unaffected by adverse weather [4,5]. Compared to radar imaging, infrared imaging systems have a simpler structure, high resolution, excellent electromagnetic concealment, and resistance to echo and noise interference [6].

Despite the notable technical advantages of infrared imaging technology in acquiring images in maritime scenarios, the performance of maritime target detection is still constrained by the complexity of sea surface scenes [7]. Due to the long imaging distances, sea surface small ships lack effective features, resulting in targets often containing very few pixels, sometimes even appearing as patchy or point-like features. The complex background of sea surface scenes, including clouds, islands, and waves, poses significant challenges to small ship detection [8].

Currently, infrared small target detection algorithms can be broadly categorized into model-driven algorithms and data-driven algorithms. Model-driven algorithms can be further classified into the following: 1. Filter-Based Detection Algorithms: These algorithms rely on filtering techniques to highlight small targets based on pixel intensity differences and eliminate surrounding background noise interference. Examples include Top-hat [9], TDLMS [10], TTLDM [11], and others. 2. Human-Visual-System-Based Detection Algorithms: These algorithms are inspired by the human visual system, leveraging the ability of the human eye to rapidly locate regions of interest and identify target objects within them. This behavior is primarily based on the eye's ability to distinguish targets from the background using contrast rather than brightness, thus obtaining visually salient regions. Examples include LCM [12], DLCM [13], MPCM [14], DECM [15], and others. 3. Image-Data-Structure-Based Detection Algorithms: These algorithms integrate image data structures into infrared small target detection, utilizing the non-local self-similarity of the background and the sparse characteristics of the target in infrared images. In this context, background blocks belong to the same low-rank subspace, while the target is relatively small in the overall image size. Examples include IPI [16], RIPT [17], RPCA [18], SRWS [19], and others.

Model-driven algorithms rely on human-made prior assumptions [20], leading to the common issue of high false alarm rates, especially in complex backgrounds. In recent years, with the development of data-driven deep learning algorithms, deep learning has gradually been applied to small target detection. In recent years, researchers have been treating small object detection tasks as pixel segmentation tasks. MDvsFA cGan addresses the small target segmentation problem by dividing it into two sub-tasks: suppressing false positives and suppressing false alarms. It jointly solves these two sub-tasks through adversarial learning [21]. ACM adopts FPN [22] and U-Net [23] as backbone networks. In the encoder–decoder structure, it designs feature fusion modules for low and high semantics, obtaining more effective feature representations [24]. ALC-Net simulates local contrast through shift operations on semantic tensors to extract local information of the target [25]. IAA-Net generates rough potential target regions using RPN to suppress background and filter false alarm targets. It then models internal relationships between pixels through attention encoders, outputting attention-aware features. Finally, predictions are obtained by inputting attention-aware features into the classification head [26]. MTU-Net proposes a multi-level Trans-U-Net-based multi-level feature extraction module to adaptively extract multi-level remote features [27]. AGPC-Net designs attention-guided contextual blocks, perceptually capturing pixel correlations within and between blocks at different scales through local semantic association and global context attention. Subsequently, it fuses multi-scale contextual information to generate a context pyramid module for better feature representation [28].

We contribute to the research on infrared small ship detection at sea based on deep learning with the following key contributions:

1. Scale attention mechanisms in AGPC-Net for small target detection. Addressing the characteristics of small targets, we enhance the foundational network, AGPC-Net, by incorporating a scale attention mechanism after the feature pyramid, adjusting the weights of different scale feature maps.
2. Additionally, we add a channel attention mechanism during the upsampling process of AGPC-Net, facilitating the learning of relationships between channels and obtaining more effective feature representations.
3. Proposal of the Maritime-SIRST infrared small ship dataset in complex sea surface scenes based on satellite infrared band images. We present the Maritime-SIRST dataset, a comprehensive infrared small ship dataset derived from satellite infrared band images, designed to meet the requirements of our research and foster advancements in related fields.

The structure of this paper is as follows: In Section 2, we provide a brief review of related work. Section 3 provides a detailed description of the network architecture of

MAPC-Net, and introduces the Maritime-SIRST dataset proposed for infrared small ship detection on the sea surface. In Section 4, we present the results of comparative experiments and ablation studies, followed by discussions. Finally, in Section 5, we summarize the findings of this study.

2. Related Work

2.1. Infrared Small Target Detection Networks

Infrared images, compared to visible light images, generally contain less useful information [29,30]. In the early stages, the field predominantly relied on model-based algorithms [31], which exhibited poor performance. In recent years, with the advancement of deep learning and the successive release of public infrared small target datasets, deep learning-based infrared small target detection algorithms have made significant progress. However, most existing algorithms are developed based on datasets with relatively simple backgrounds. The application of infrared small ship detection on the sea surface is widespread, and the sea surface background is highly complex. Small ships often get submerged in strong noise and complex background clutter. This poses a significant challenge to stable detection, as separating small ships from complex background noise in infrared noisy images without generating false alarms is a challenge [32].

Current research on deep learning-based algorithms typically focuses on feature fusion, local information, feature pyramids, contextual information, etc. [33]. However, when dealing with complex sea surface backgrounds, the issue of high false alarm rates persists. To address this challenge, there is a need for a further exploration and development of more precise and robust infrared small target detection algorithms to meet the demands of sea surface small ship detection in complex backgrounds.

2.2. Attention Mechanism

The attention mechanism in deep learning is an approach that mimics the human visual and cognitive system, enabling neural networks to focus attention on relevant parts of input data. By incorporating attention mechanisms, neural networks can automatically learn and selectively attend to crucial information in the input, enhancing the model's performance and generalization capabilities [34]. With the development of attention mechanisms, various types have emerged, including self-attention [35], channel attention [36], spatial attention [37], scale attention [38], and more. Additionally, each attention mechanism has different implementation versions. There are also hybrid attention mechanisms that combine multiple types of attention, such as CBAM [39], BAM [40], and scSE [41], effectively integrating channel and spatial attention mechanisms. The diverse range of attention mechanisms yields varying improvements in model performance. Attention mechanisms prove to be helpful for computer vision tasks, making them a simple and effective strategy to enhance network performance by integrating suitable attention mechanisms into the network architecture.

2.3. Datasets for Infrared Small Targets

Currently, the lack of public datasets remains a significant constraint in the development of deep learning-based infrared small target detection technologies. In recent years, several public infrared small target datasets have been gradually proposed. According to our survey, Dai first released the SIRST dataset in 2021 [25], which was later supplemented with the SIRST-V2 dataset [42]. Subsequently, others proposed datasets such as IRSTD-1K [43], NUDT-SIRST [44], and SIRST-AUG [28]. However, these datasets primarily focus on land and sky scenes. In 2023, Wu proposed the NUDT-SIRST-SEA dataset, the first dataset specifically designed for infrared small ship detection in satellite-based maritime scenes [27]. This dataset was annotated using remote sensing satellite infrared band images. However, the majority of images in NUDT-SIRST-SEA have relatively simple backgrounds, with over 50% of the images containing no targets. Additionally, the dataset includes a mix of infrared images from coastal land scenes, contributing to an overall lower dataset quality.

3. Materials and Methods

3.1. MAPC-Net

We propose a new network architecture, MAPC-Net, based on AGPC-Net, as illustrated in Figure 1. The input is an image processed through the Res-Net, generating a spatial feature map X with dimensions H × W × C. Subsequently, the feature map X is input into the context pyramid module (CPM). CPM transmits the feature map X to multiple scales, S ∈ {S_1..... S_n}, of the Attention-Guided Context Block (AGCB). For each scale S_i, AGCB integrates contextual information, preserving key details for small targets, resulting in the feature map A^{S_i}. The A^{S_i} obtained from AGCB at multiple scales is then passed to the scale attention (LA) module, which automatically learns image-specific weights for each scale to calibrate features across different scales. Finally, the resulting feature map is connected with the original feature map X and convolved to produce the feature map C. During the upsampling process, two bilinear interpolations are conducted. In the asymmetric fusion module (AFM) following linear interpolation, semantic information of 1/4 and 1/2 sizes from lower layers is fused with deep semantic information. Additionally, the Squeeze-and-Excitation block (SE) is incorporated at each upsampling layer. The SE block adaptively adjusts channel weights in the feature map through squeeze and excitation operations to enhance the network's representational capacity at different depths, ultimately yielding the predicted feature map. Each module is detailed in the following sections:

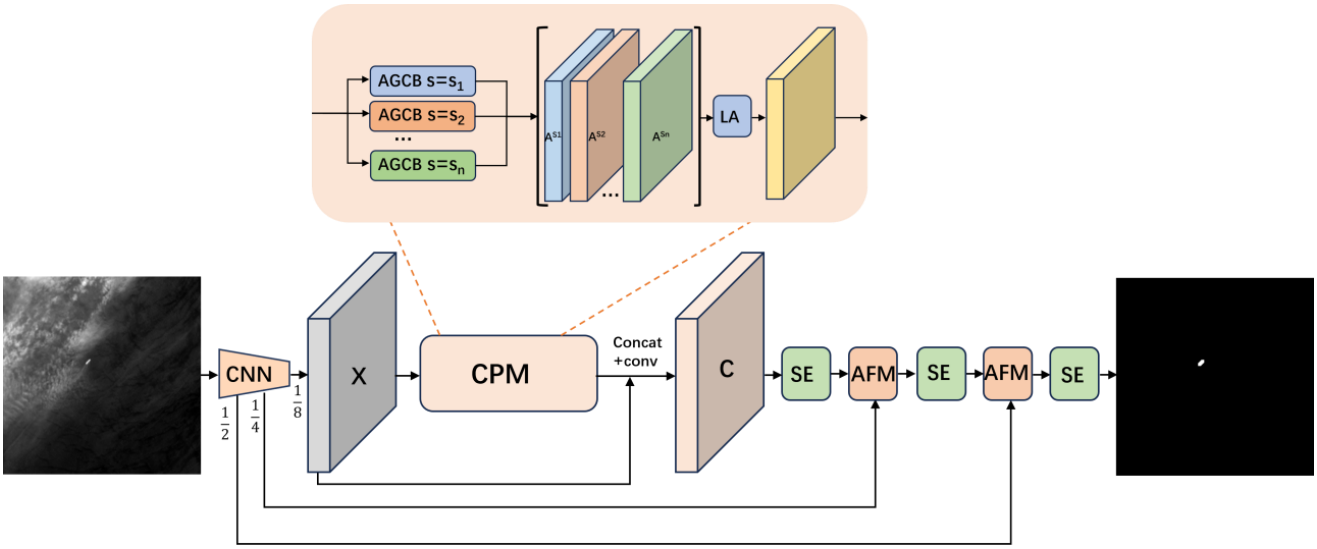


Figure 1. Structure of MAPC-Net.

3.1.1. Attention-Guided Context Block

As depicted in Figure 2, AGCB comprises the local semantic association (LSA) and the global context attention (GCA). For a given feature map X and scale S, the lower branch LSA divides the feature map into S × S blocks, each of size $\frac{H}{S} \times \frac{W}{S} \times C$, denoted as X_i, i ∈ {1, 2, . . ., S^2}. For each block X_i, a new block, P^i, is obtained after NL block processing, as shown in Equations (1) and (2):

$$P_k^i = \beta \sum_{j=1}^{HW/S^2} \omega_{kj}^i \Psi(X_k^i) + X_k^i \tag{1}$$

$$\omega_{kj}^i = \frac{\exp\left(\theta(X_k^i)^T \Phi(X_j^i)\right)}{\sum_{j=1}^{HW/S^2} \exp\left(\theta(X_k^i)^T \Phi(X_j^i)\right)} \tag{2}$$

where P_k^i represents the kth element, β is a learnable parameter, ω_{kj}^i denotes the element at the kth row and jth column of ω^i, and $\Psi(\cdot)$, $\theta(\cdot)$, and $\Phi(\cdot)$ all represent 1×1 convolutional layers. Finally, each block is reassembled into a feature map, P^i, according to a pre-defined order.

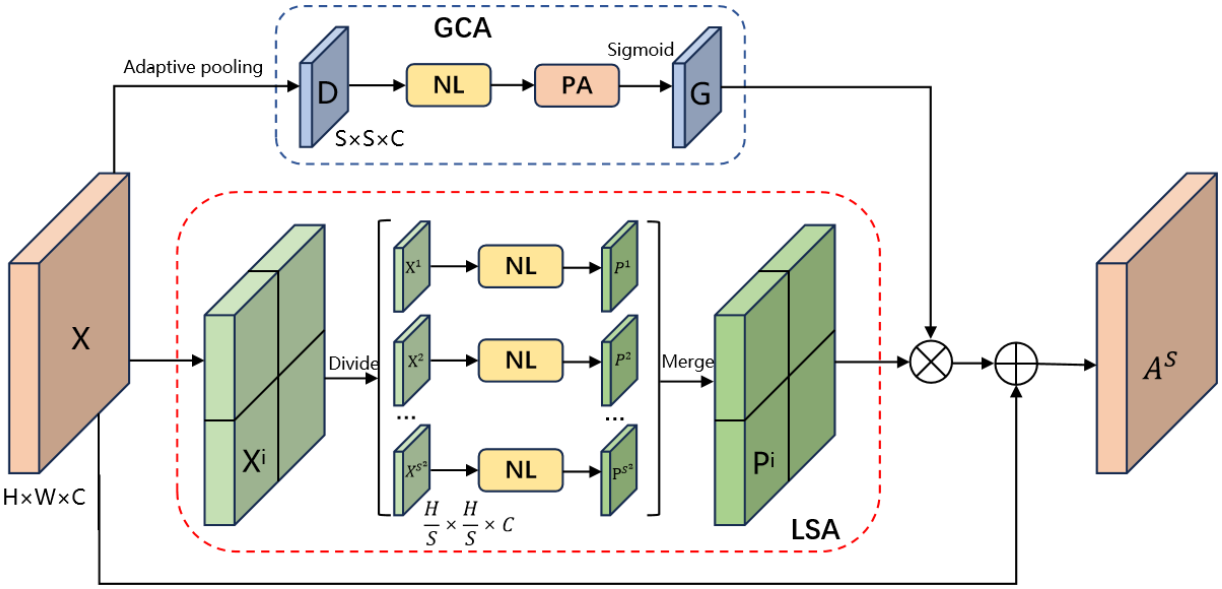


Figure 2. Structure of AGCB.

GCA serves as the upper branch of AGCB, aiming to estimate the dependencies between each P^i block. For each scale S, the feature map X undergoes adaptive pooling to obtain a feature map D of size S × S × C, where each point corresponds to the features of each block in LSA. Subsequently, NL blocks are applied to estimate the correlations at each position in the feature map D. To enhance pixel-level representation, the features are then passed to the pixel attention module (PA) for integrating channel information at each pixel. Finally, as shown in Equations (3)–(5), a guided map, G_i, is obtained through a sigmoid transformation. Here, G_i represents the ith element of G, PA denotes pixel attention, and δ is the sigmoid function.

$$G_i = \delta\left(PA\left(\beta\sum_{j=1}^{S^2}\omega_{mj}\Psi(D_m) + D_m\right)\right) \tag{3}$$

$$\omega_{mj} = \frac{1}{Z_m}exp\left(\theta(D_m)^T\varnothing(D_j)\right) \tag{4}$$

$$Z_m = \sum_{j=1}^{S^2} exp\left(\theta(D_m)^T\varnothing(D_j)\right) \tag{5}$$

In AGCB, P^i and G_i describe the semantic correlations at the local level and the associations between blocks, respectively. LSA computes the correlations between the current pixel and the rest of the pixels, estimating the probability that each pixel is part of the target at the local scale. Within a block, the target locations are highlighted accordingly. Subsequently, each pixel in GCA aggregates information contained in each block and estimates the probability of the target occurring therein. The resulting A^S, as shown in Equation (6), is a fusion of P^i and G_i, considering both the localized targets in P^i and the contextual background information in G_i.

$$A_P^S = P_k^i \times G_i \tag{6}$$

3.1.2. Scale Attention Module

After generating feature maps at different scales, the AGPC-Net is combined by simple stacking to form a feature pyramid. However, this straightforward stacking approach fails to fully exploit the characteristic of small-sized ships. To overcome this limitation, we add the Scale Attention Module behind the generated feature pyramid. This enables us to adaptively adjust the weights of feature maps at different scales, emphasizing the focus on effective scale feature maps while reducing attention to ineffective scale feature maps, thereby obtaining superior feature representation. By adding the Scale Attention Module, our network can more effectively utilize information from feature maps at different scales, enhancing the detection capability for targets of various sizes. This adaptive scale adjustment helps the network better adapt to the common small size of ships, thereby improving the performance of the network in infrared small ship detection tasks.

Figure 3 illustrates the LA [38], which can automatically learn specific weights for the feature maps at each scale to calibrate features at different scales. Firstly, a 1×1 convolution compresses feature maps at different scales into 4 channels, and the compression results from different scales are concatenated into a mixed feature map, F. Next, the combined features P_{avg}, P_{max}, and MLP are used to obtain coefficients for each channel. The scale attention coefficients are represented as $\gamma \in [0, 1]^{4\times1\times1}$. To allocate multi-scale attention weights at each pixel, an additional spatial attention block, LA*, is employed using $F\times\gamma$ as input, generating spatial attention coefficients $\gamma^* \in [0, 1]^{1\times H\times W}$. This makes $\gamma \times \gamma^*$ represent scale attention in the pixel direction. The final output of LA is as shown in Equation (7).

$$y_{LA} = F \times \gamma \times \gamma^* + F \times \gamma + F \tag{7}$$

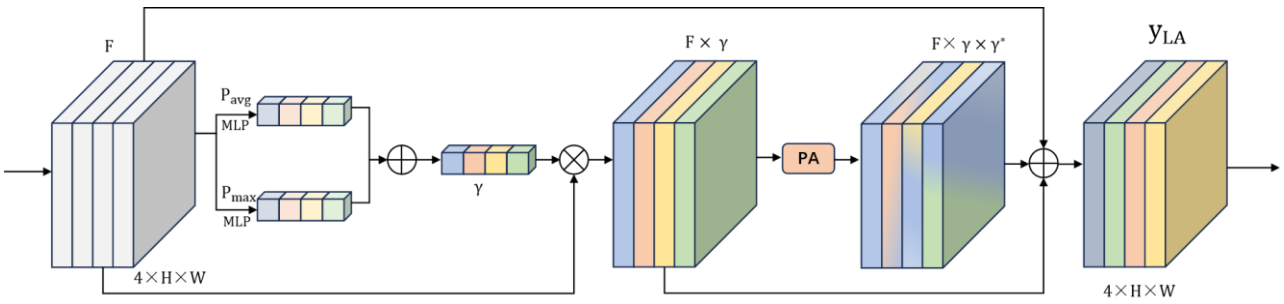


Figure 3. Structure of scale attention.

3.1.3. Squeeze-and-Excitation Block

We incorporated the Squeeze-and-Excitation Block [36] during the upsampling process, which is a channel attention mechanism capable of learning weights between different channels. Figure 4 illustrates the SE block. The SE block enhances the model's performance by adaptively adjusting channel weights in the feature map. The SE block initially compresses the input feature map of size $H \times W \times C$ in terms of spatial features, achieving global average pooling in the spatial dimension, resulting in a feature map of size $1 \times 1 \times C$. Through a fully connected (FC) layer, it learns and produces a feature map with channel attention, still possessing dimensions of $1 \times 1 \times C$. The channel attention feature map of size $1 \times 1 \times C$ is then multiplied channel-wise by the weight coefficients with the original input feature map of size $H \times W \times C$, ultimately yielding a feature map with channel attention.

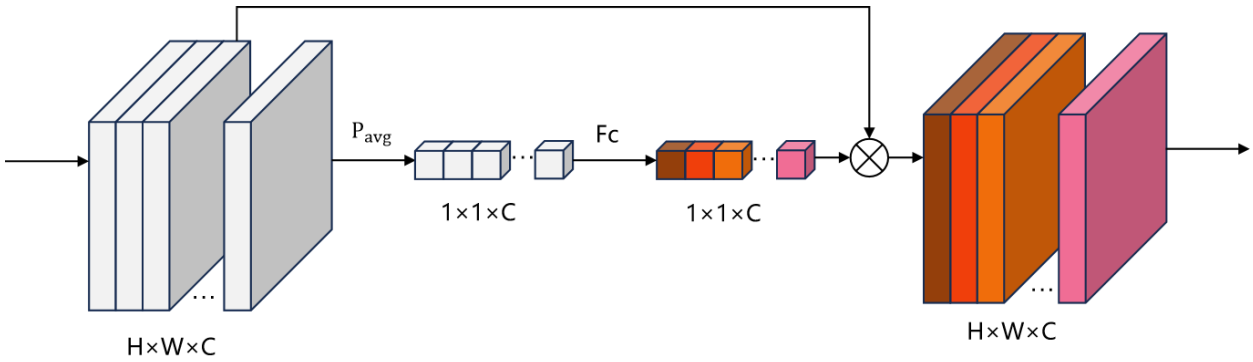


Figure 4. Structure of SE block.

3.2. Maritime-SIRST

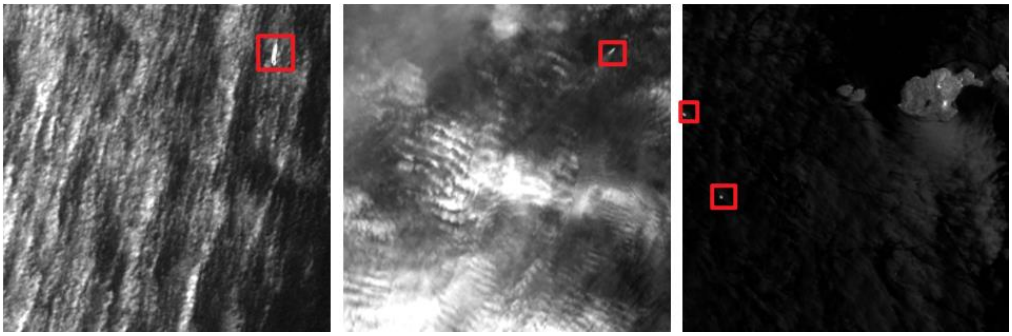
The quality of a dataset significantly influences the performance of deep learning algorithms. While there are existing infrared small target datasets, they are primarily based on land and sky backgrounds, failing to accurately represent real scenarios in a sea surface environment. Consequently, they are unsuitable for training and evaluating models for sea surface small ship detection. Additionally, some datasets employ simulation or synthetic techniques, making it challenging to assess the differences between simulated or synthetic images and real images. Discrepancies between simulated or synthetic targets and backgrounds may lead to overly optimistic detection rates, resulting in training outcomes that are overly optimistic.

Therefore, proposing a realistic infrared dataset for sea surface small ship detection is crucial for this research. Such a dataset would provide authentic sea surface backgrounds, encompassing various complex maritime conditions, including waves, islands, clouds, and more, closely mimicking real-world applications. This type of dataset not only holds paramount importance for this study but also contributes to advancing the field of infrared sea surface small ship detection. By utilizing a dataset based on real scenes, it becomes possible to more accurately evaluate algorithm performance in sea surface small ship detection tasks, enhancing algorithm robustness and reliability.

In this study, we leverage publicly available images from the Landsat-8 remote sensing satellite's near-infrared band to propose a high-quality infrared small ship detection dataset tailored for complex sea surface scenes, named Maritime-SIRST. In Landsat-8 remote sensing images, we selected infrared band images from different regions such as Asia, Africa, and North America, near ports and canals. The selected images cover the time span from 2013 to 2021. To enhance representativeness, we processed remote sensing images from different months and varying cloud cover ratios. We utilized the annotation tool LabelBee, an open-source product from SenseTime, to annotate the original images, generating labeled files in mask format. This dataset, dedicated to infrared sea surface small ships, utilizes authentic remote sensing satellite infrared images, resulting in a more diverse and complex background compared to other datasets. The Maritime-SIRST dataset comprises 566 images, featuring a total of 796 targets, with each image sized at 256 × 256 pixels. Table 1 provides a detailed distribution of the dataset. The sea surface backgrounds encompass various elements such as waves, complex clouds, islands, and ports, effectively covering a broad spectrum of maritime scenarios, with complex scenes accounting for 64.1% of the dataset. Additionally, over 90% of the targets are smaller than 0.35% of the image area. The dataset includes images with both targets and no targets, and approximately 27.7% of the images depict multiple targets, aligning with real-world scenarios where sea vessels often appear in groups. Figure 5 showcases selected images from the Maritime-SIRST dataset. The waves in the images are typically stripe-shaped, exhibiting a consistent direction and regular distribution. Cloud clusters, on the other hand, appear as block-shaped formations with a more chaotic distribution and fuzzy edges. Islands also manifest as block-shaped structures with clear and well-defined edges.

Table 1. The statistical data of Maritime-SIRST.

Features		Number	Ratio/%
Background	simple	203	35.9
	waves	62	11.0
	clouds	272	48.2
	islands, ports	28	4.9
Target size/pixel	<25	55	9.7
	25–81	319	56.4
	81–225	174	30.7
	>225	18	3.2
Target number	0	21	3.7
	1	388	68.6
	1–8	157	27.7



(**a**) (**b**) (**c**)

Figure 5. Partial image of Maritime-SIRST. (**a**) Image with ocean waves. (**b**) Image with complex clouds. (**c**) Image with islands.

4. Experiment and Discussion

4.1. Experimental Setting

To evaluate the performance of our model, we conducted comparative experiments with four traditional algorithms, including Top-hat [9], IPI [16], MPCM [14], and TTLDM [11], and four deep learning algorithms, including ACM-U-Net [24], ACM-FPN [24], IAA-Net [26], MTU-Net [27], and AGPC-Net [28]. All models were run using Python 3.8 on a computer with a 15vCPU AMD EPYC 7543 32-core Processor CPU and NVIDIA GeForce RTX 3090 GPU. Our model was trained with a batch size of 8, 60 epochs, an initial learning rate of 0.05, and the SoftIouLoss as the loss function. Other model training parameters followed the settings in the original papers.

The experiment utilized our self-proposed Maritime-SIRST dataset, which consists of real-world infrared small ship images in maritime scenarios, all with a size of 256 × 256 pixels. To enhance training stability, the training set was augmented with operations such as rotation and scaling, resulting in 752 images. The ratio of the training set, testing set, and validation set was approximately 7:1.5:1.5. The "rotation" operation refers to randomly rotating the image by an angle between 0 and 90 degrees, while the "scaling" operation involves randomly cropping a portion of the image and then resizing that cropped portion to the original image size.

4.2. Evaluation Metrics

We adopt the Precision (*Prec.*) [3], Recall (*Rec.*) [3], mIOU (mean Intersection over Union) [28], $F1$ score [28], AUC (Area Under the Curve) [28], and average algorithm execution time as evaluation metrics. The calculation formulas for Precision and Recall are provided in Equations (8) and (9):

$$Prec. = \frac{TP}{TP + FP} \tag{8}$$

$$Rec. = \frac{TP}{TP + FN} \tag{9}$$

In the formulas, TP represents the number of matched pixels detected as target pixels with true target pixels, FP represents the number of background pixels mistakenly detected as true target pixels, and FN represents the number of target pixels detected as background pixels.

$mIOU$ stands for mean Intersection over Union, and the calculation formula is given by Equation (10):

$$mIOU = \frac{1}{n} * \sum_{i=1}^{n} \frac{TP_i}{TP_i + FP_i + FN_i} \tag{10}$$

where n represents the total number of classes, TP_i denotes the true positives for the ith class (the number of correctly predicted positive pixels), FP_i represents the false positives for the ith class (the number of incorrectly predicted positive pixels), and FN_i represents the false negatives for the ith class (the number of incorrectly predicted negative pixels). For each class, the Intersection over Union (IoU) is calculated, and then the IoUs for all classes are summed and divided by the total number of classes to obtain the mean IoU value.

Precision and Recall are two performance metrics that are inversely related; usually, when Precision is high, Recall tends to be low, and vice versa. $F1$ is a metric that simultaneously considers both Precision and Recall, providing a balanced evaluation. The calculation formula is given in Equation (11):

$$F1 = \frac{2 * Prec. * Rec.}{Prec. + Rec.} \tag{11}$$

The ROC curve, short for Receiver Operating Characteristic curve, is a commonly used tool to evaluate the performance of binary classification models. The ROC curve is plotted with the true positive rate on the vertical axis and the false positive rate on the horizontal axis. The ROC curve illustrates the model's performance in classifying positives and negatives at different thresholds. AUC is widely used as a metric to evaluate model performance, with a higher AUC value indicating better performance.

4.3. Quantitative Results

Table 2 presents the comparative experimental results of our model with other models and algorithms on the Maritime-SIRST dataset. From the table, it can be observed that our model outperforms other models and algorithms, with a 5.79% increase in Precision, a 2.16% increase in Recall, a 6.14% increase in mIOU, a 4.41% increase in F1, and a 0.92% increase in AUC. The runtime is comparable to other deep learning algorithms. The experimental results demonstrate that our improved algorithm has achieved significant performance improvements.

Additionally, based on the Maritime-SIRST dataset, we plotted the ROC curves, as shown in Figure 6. In order to illustrate the performance differences of each algorithm, ROC curves with two different x-axis scales are presented separately. From the ROC curves, it can be observed that the curve of our model is closer to the upper-left corner, indicating that our model outperforms other models and algorithms.

In remote sensing images, small ships often have small sizes, and infrared images contain less information compared to visible light images. Moreover, there is complex background interference on the sea surface, causing many algorithms suitable for simple background infrared small target detection to perform poorly, often resulting in high false alarm rates. However, by adding attention mechanisms, the network can more effectively focus on valid features in the image while attenuating irrelevant background features. Therefore, in the task of infrared small ship detection, enhancing the network's performance can be achieved by introducing appropriate attention mechanisms. The introduction of attention mechanisms allows the network to concentrate more on small ships, improving detection accuracy. By adaptively learning weight assignments, attention mechanisms can selectively amplify features helpful for target detection while suppressing attention to

background interference. This mechanism can effectively enhance the network's capability to detect small ships and reduce false alarm rates.

Table 2. Comparative analysis of segmentation accuracy of different algorithms.

Methods	Prec./%	Rec./%	mIOU/%	F1/%	AUC/%	Time/s
Top-hat	38.83	22.53	14.96	23.75	62.32	0.0019
IPI	39.65	38.26	18.32	26.89	67.78	7.780
MPCM	27.11	59.57	16.73	22.98	64.05	1.240
TTLDM	42.92	52.75	31.15	40.81	82.36	0.017
ACM-FPN	58.10	54.69	39.22	56.34	74.06	0.160
ACM-U-NET	58.15	47.87	35.60	52.51	78.29	0.160
IAA-NET	69.59	52.55	45.03	59.88	83.80	0.160
MTU-NET	70.46	85.62	63.01	77.31	93.91	0.125
AGPC-NET	71.60	85.29	63.73	77.85	93.63	0.130
Ours	77.39	87.78	69.87	82.26	94.83	0.135

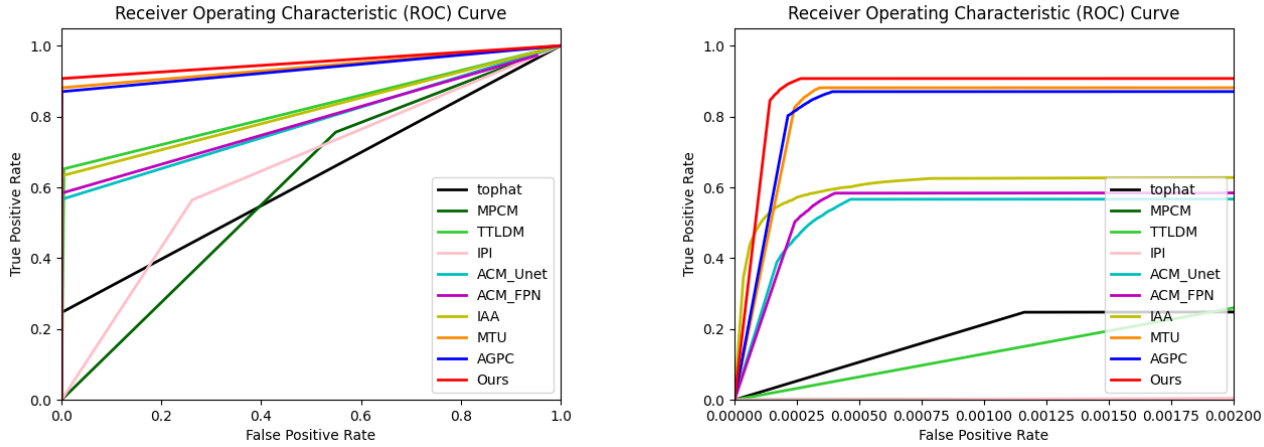
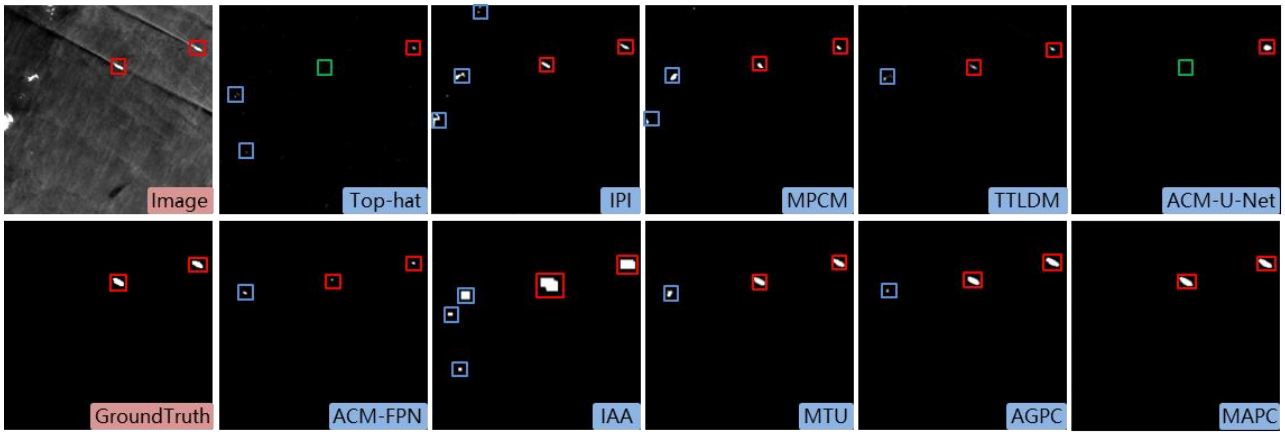


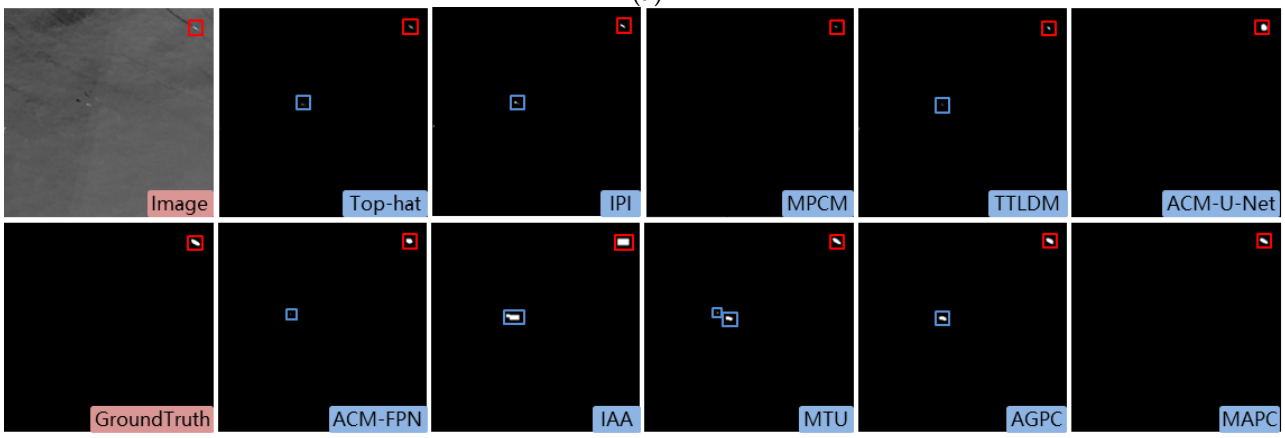
Figure 6. ROC curves of different methods on the Maritime-SIRST dataset.

4.4. Visual Results

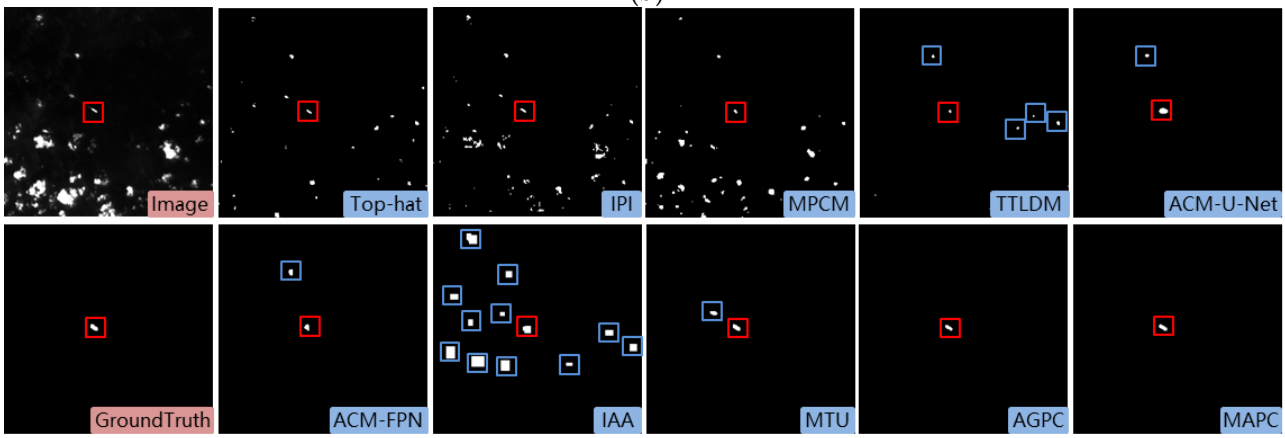
In order to visualize our segmentation results, we selected a subset of images, compared the segmentation results of all algorithms with the ground truth, and manually annotated them. Figure 7 displays the detection results of various algorithms compared in this paper (red boxes represent correctly detected targets, blue boxes represent false positives, and green boxes represent missed detections). From the detection results, it can be observed that traditional algorithms generally suffer from high false alarm rates, mistakenly identifying cloud clusters and islands as small ships, and also exhibiting instances of missed detections. Some deep learning algorithms also exhibit high false alarm rates, and the segmented targets differ significantly from the actual targets. This is mainly attributed to the complexity of the maritime scene, where clouds, waves, islands, and similar elements are prone to misclassification as targets. Overall, our algorithm demonstrates lower false alarm rates and missed detection rates, and in terms of segmentation results, our algorithm's outcomes are closer to the ground truth.



(**a**)



(**b**)



(**c**)

Figure 7. *Cont.*

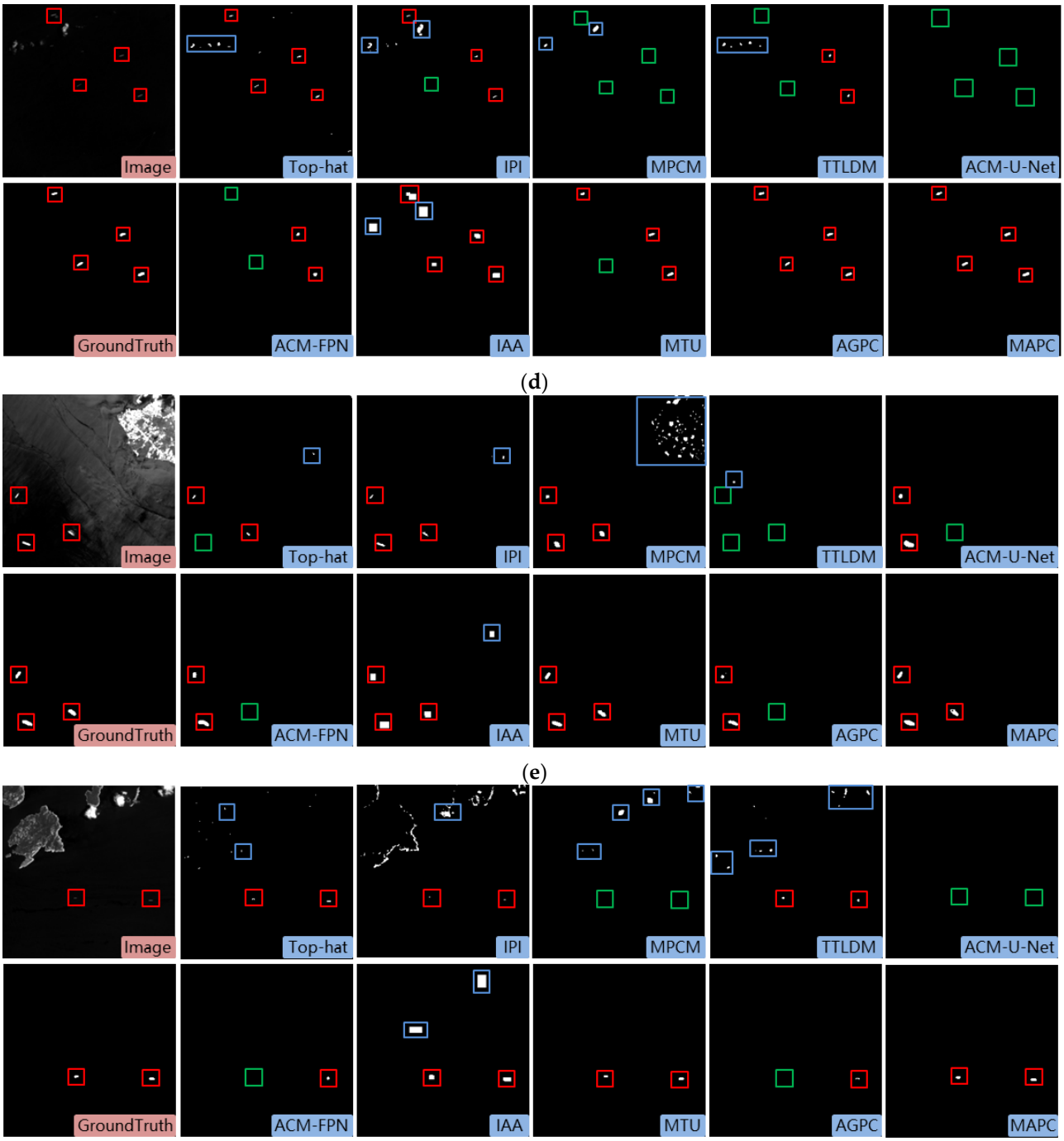


Figure 7. Partial image visualization results of different methods on Maritime-SIRST datasets. (**a**,**b**) showcase the image segmentation results with waves, figures (**c**,**d**) demonstrate the image segmentation results with cloud clusters, and figures (**e**,**f**) illustrate the image segmentation results with islands. The red box, the blue box, and the green box represent the correct detection box, the false detection box, and the missed detection box, respectively. Some algorithms exhibit an excessive number of false alarms; hence, annotations are omitted in the images.

4.5. Ablation Study

To validate the effectiveness of the LA module and SE block in enhancing the performance of the original network, ablation experiments were conducted on the Maritime-SIRST dataset. Table 3 presents the results of the ablation experiments, demonstrating that adding

the LA module to the base network increased Precision, Recall, mIOU, F1, and AUC by 0.43%, 1.12%, 0.97%, 0.72%, and 0.56%, respectively, compared to the base network. Similarly, incorporating the SE block into the base network led to increases of 4.09%, 1.54%, 4.16%, 3.03%, and 0.85% in Precision, Recall, mIOU, F1, and AUC, respectively. For MAPC-Net, the improvements were 5.79%, 2.16%, 6.14%, 4.41%, and 1.20% across the mentioned metrics compared to the base network. The ablation experiments indicate that the LA module, by learning the weights between different scales of the feature pyramid, and the SE block, by learning the weights between different channels, obtained better feature representations, resulting in a significant enhancement in network performance.

Table 3. Ablation study of the LA and SE block.

Module	Prec./%	Rec./%	mIOU/%	F1/%	AUC/%
AGPC	71.60	85.29	63.73	77.85	93.63
AGPC + LA	72.03	86.41	64.70	78.57	94.19
AGPC + SE	75.69	86.83	67.89	80.88	94.48
AGPC + LA + SE	77.39	87.78	69.87	82.26	94.83

Different attention modules yield varying performance improvements for different tasks and models. We explored several attention modules (GAM [45], CA [46], CBAM [39], SE [36]) during the upsampling stage, and the experimental data in Table 4 reveal that the SE block contributes the most to the comprehensive performance improvement in the model. Therefore, the SE block was adopted during the upsampling stage of the model to obtain better feature representations.

Table 4. Ablation study of different attention modules in upsampling.

Module	Prec./%	Rec./%	mIOU/%	F1/%	AUC/%
AGPC + GAM	74.35	84.26	65.29	79.00	93.44
AGPC + CA	70.44	87.94	64.24	78.22	94.21
AGPC + CBAM	62.08	86.96	56.80	72.45	94.71
AGPC + SE	75.69	86.83	67.89	80.88	94.48

5. Conclusions

To tackle the intricate challenge of infrared small ship detection in complex maritime scenes, we extend the existing AGPC-Net architecture, introducing multiple attention mechanisms to create the innovative network MAPC-Net. Experimental results affirm MAPC-Net's strong performance in infrared small ship detection on the sea surface. Specifically, MAPC-Net integrates a scale attention mechanism after the feature pyramid, facilitating adaptive learning of weights for diverse scale feature maps. This adaptive learning enhances the network's ability to utilize information from varying scales, thereby improving detection capabilities for targets of different sizes. In the upsampling stage, MAPC-Net incorporates SE blocks between each layer, fostering the learning of relationships between different channels and enhancing the network's feature representation capabilities. We have proposed Maritime-SIRST, a dataset tailored for infrared small ship detection on the sea surface. This dataset encompasses diverse scenarios found in maritime backgrounds, such as waves, clouds, and islands, rendering it more representative of real-world applications. Experimental comparisons on the Maritime-SIRST dataset reveal that MAPC-Net surpasses traditional algorithms and other deep learning methods. Through ablation experiments, we validate the efficacy of adding attention modules, demonstrating that these mechanisms enable the network to more accurately locate and identify infrared small ships on the sea surface.

The infrared small ship detection algorithm is typically deployed on embedded devices, imposing high demands on the algorithm's complexity. Although the proposed

network in this paper achieves high accuracy, it also comes with high complexity. Therefore, to achieve widespread application in practical engineering, it is necessary to explore lightweight techniques for the network. The next steps could involve model pruning, distillation, and the use of depth-wise separable convolutions as alternatives to conventional convolutions to conduct lightweight research on the model.

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

Funding: This research was funded by the National Science Foundation of China, grant number 62261053, and the Cross-Media Intelligent Technology Project of Beijing National Research Center for Information Science and Technology (BNRist), grant number BNR2019TD01022.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets analyzed or generated in this study are available from the authors upon reasonable request.

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

References

1. Tang, J.; Deng, C.; Huang, G.-B.; Zhao, B. Compressed-Domain Ship Detection on Spaceborne Optical Image Using Deep Neural Network and Extreme Learning Machine. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1174–1185. [CrossRef]
2. Wang, X.; Peng, Z.; Kong, D.; He, Y. Infrared Dim and Small Target Detection Based on Stable Multisubspace Learning in Heterogeneous Scene. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 5481–5493. [CrossRef]
3. Gao, Z.; Zhang, Y.; Wang, S. Lightweight Small Ship Detection Algorithm Combined with Infrared Characteristic Analysis for Autonomous Navigation. *J. Mar. Sci. Eng.* **2023**, *11*, 1114. [CrossRef]
4. Lu, C.; Qin, H.; Deng, Z.; Zhu, Z. Fusion2Fusion: An Infrared–Visible Image Fusion Algorithm for Surface Water Environments. *J. Mar. Sci. Eng.* **2023**, *11*, 902. [CrossRef]
5. Li, L.; Lv, M.; Jia, Z.; Ma, H. Sparse Representation-Based Multi-Focus Image Fusion Method via Local Energy in Shearlet Domain. *Sensors* **2023**, *23*, 2888. [CrossRef] [PubMed]
6. Deng, H.; Sun, X.; Liu, M.; Ye, C.; Zhou, X. Small Infrared Target Detection Based on Weighted Local Difference Measure. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4204–4214. [CrossRef]
7. Wang, N.; Li, B.; Wei, X.; Wang, Y.; Yan, H. Ship Detection in Spaceborne Infrared Image Based on Lightweight CNN and Multisource Feature Cascade Decision. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 4324–4339. [CrossRef]
8. Cao, Z.; Kong, X.; Zhu, Q.; Cao, S.; Peng, Z. Infrared Dim Target Detection via Mode-K1k2 Extension Tensor Tubal Rank under Complex Ocean Environment. *ISPRS J. Photogramm. Remote Sens.* **2021**, *181*, 167–190. [CrossRef]
9. Bai, X.; Zhou, F. Analysis of New Top-Hat Transformation and the Application for Infrared Dim Small Target Detection. *Pattern Recognit.* **2010**, *43*, 2145–2156. [CrossRef]
10. Cao, Y.; Liu, R.; Yang, J. Small Target Detection Using Two-Dimensional Least Mean Square (TDLMS) Filter Based on Neighborhood Analysis. *Int. J. Infrared Millim. Waves* **2008**, *29*, 188–200. [CrossRef]
11. Mu, J.; Li, W.; Rao, J.; Li, F.; Wei, H. Infrared Small Target Detection Using Tri-Layer Template Local Difference Measure. *Opt. Precis. Eng.* **2022**, *30*, 869–882. [CrossRef]
12. Chen, C.P.; Li, H.; Wei, Y.; Xia, T.; Tang, Y.Y. A Local Contrast Method for Small Infrared Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 574–581. [CrossRef]
13. Pan, S.D.; Zhang, S.; Zhao, M.; An, B.W. Infrared Small Target Detection Based on Double-Layer Local Contrast Measure. *Acta Photonica Sin.* **2020**, *49*, 0110003.
14. Wei, Y.; You, X.; Li, H. Multiscale Patch-Based Contrast Measure for Small Infrared Target Detection. *Pattern Recognit.* **2016**, *58*, 216–226. [CrossRef]
15. Bai, X.; Bi, Y. Derivative Entropy-Based Contrast Measure for Infrared Small-Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2452–2466. [CrossRef]
16. Gao, C.; Meng, D.; Yang, Y.; Wang, Y.; Zhou, X.; Hauptmann, A.G. Infrared Patch-Image Model for Small Target Detection in a Single Image. *IEEE Trans. Image Process.* **2013**, *22*, 4996–5009. [CrossRef]
17. Dai, Y.; Wu, Y. Reweighted Infrared Patch-Tensor Model with Both Nonlocal and Local Priors for Single-Frame Small Target Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3752–3767. [CrossRef]

18. Wang, C.; Qin, S. Adaptive Detection Method of Infrared Small Target Based on Target-Background Separation via Robust Principal Component Analysis. *Infrared Phys. Technol.* **2015**, *69*, 123–135. [CrossRef]
19. Zhang, T.; Peng, Z.; Wu, H.; He, Y.; Li, C.; Yang, C. Infrared Small Target Detection via Self-Regularized Weighted Sparse Model. *Neurocomputing* **2021**, *420*, 124–148. [CrossRef]
20. Hou, Q.; Zhang, L.; Tan, F.; Xi, Y.; Zheng, H.; Li, N. ISTDU-Net: Infrared Small-Target Detection U-Net. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 7506205. [CrossRef]
21. Wang, H.; Zhou, L.; Wang, L. Miss Detection vs. False Alarm: Adversarial Learning for Small Object Segmentation in Infrared Images. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 8509–8518.
22. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
23. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2015; Volume 9351, pp. 234–241.
24. Dai, Y.; Wu, Y.; Zhou, F.; Barnard, K. Asymmetric Contextual Modulation for Infrared Small Target Detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Virtual Conference, 5–9 January 2021; pp. 950–959.
25. Dai, Y.; Wu, Y.; Zhou, F.; Barnard, K. Attentional Local Contrast Networks for Infrared Small Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 9813–9824. [CrossRef]
26. Wang, K.; Du, S.; Liu, C.; Cao, Z. Interior Attention-Aware Network for Infrared Small Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5002013. [CrossRef]
27. 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*, 5601015. [CrossRef]
28. Zhang, T.; Li, L.; Cao, S.; Pu, T.; Peng, Z. Attention-Guided Pyramid Context Networks for Detecting Infrared Small Target Under Complex Background. *IEEE Trans. Aerosp. Electron. Syst.* **2023**, *59*, 4250–4261. [CrossRef]
29. Sun, Y.; Yang, J.; An, W. Infrared Dim and Small Target Detection via Multiple Subspace Learning and Spatial-Temporal Patch-Tensor Model. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 3737–3752. [CrossRef]
30. Pan, P.; Wang, H.; Wang, C.; Nie, C. ABC: Attention with Bilinear Correlation for Infrared Small Target Detection. In Proceedings of the 2023 IEEE International Conference on Multimedia and Expo (ICME), Brisbane, Australia, 10–14 July 2023; pp. 2381–2386.
31. Kou, R.; Wang, C.; Yu, Y.; Peng, Z.; Yang, M.; Huang, F.; Fu, Q. LW-IRSTNet: Lightweight Infrared Small Target Segmentation Network and Application Deployment. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5621313. [CrossRef]
32. Huang, S.; Liu, Y.; He, Y.; Zhang, T.; Peng, Z. Structure-Adaptive Clutter Suppression for Infrared Small Target Detection: Chain-Growth Filtering. *Remote Sens.* **2019**, *12*, 47. [CrossRef]
33. Kou, R.; Wang, C.; Peng, Z.; Zhao, Z.; Chen, Y.; Han, J.; Huang, F.; Yu, Y.; Fu, Q. Infrared Small Target Segmentation Networks: A Survey. *Pattern Recognit.* **2023**, *143*, 109788. [CrossRef]
34. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All you Need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. Available online: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf (accessed on 8 February 2024).
35. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-Local Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803.
36. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
37. Mnih, V.; Heess, N.; Graves, A. Recurrent Models of Visual Attention. *Adv. Neural Inf. Process. Syst.* **2014**, *27*. Available online: https://proceedings.neurips.cc/paper_files/paper/2014/file/09c6c3783b4a70054da74f2538ed47c6-Paper.pdf (accessed on 8 February 2024).
38. Gu, R.; Wang, G.; Song, T.; Huang, R.; Aertsen, M.; Deprest, J.; Ourselin, S.; Vercauteren, T.; Zhang, S. CA-Net: Comprehensive Attention Convolutional Neural Networks for Explainable Medical Image Segmentation. *IEEE Trans. Med. Imaging* **2020**, *40*, 699–711. [CrossRef] [PubMed]
39. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
40. Park, J.; Woo, S.; Lee, J.-Y.; Kweon, I.S. BAM: Bottleneck Attention Module. *arXiv* **2018**, arXiv:1807.06514.
41. Roy, A.G.; Navab, N.; Wachinger, C. Recalibrating Fully Convolutional Networks with Spatial and Channel "Squeeze and Excitation" Blocks. *IEEE Trans. Med. Imaging* **2018**, *38*, 540–549. [CrossRef]
42. Dai, Y.; Li, X.; Zhou, F.; Qian, Y.; Chen, Y.; Yang, J. One-Stage Cascade Refinement Networks for Infrared Small Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5000917. [CrossRef]
43. Zhang, M.; Zhang, R.; Yang, Y.; Bai, H.; Zhang, J.; Guo, J. ISNet: Shape Matters for Infrared Small Target Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 877–886.
44. Li, B.; Xiao, C.; Wang, L.; Wang, Y.; Lin, Z.; Li, M.; An, W.; Guo, Y. Dense Nested Attention Network for Infrared Small Target Detection. *IEEE Trans. Image Process.* **2022**, *32*, 1745–1758. [CrossRef]

45. Liu, Y.; Shao, Z.; Hoffmann, N. Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions. *arXiv* **2021**, arXiv:2112.05561.
46. Hou, Q.; Zhou, D.; Feng, J. Coordinate Attention for Efficient Mobile Network Design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual Conference, 19–25 June 2021; pp. 13713–13722.