remote sensing

MDPI

Article

An Infrared Maritime Small Target Detection Algorithm Based on Semantic, Detail, and Edge Multidimensional Information Fusion

Jiping Yao, Shanzhu Xiao *, Qiuqun Deng, Gongjian Wen, Huamin Tao and Jinming Du

National Key Laboratory of Science and Technology on Automatic Target Recognition, Collage of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China; yaojiping@nudt.edu.cn (J.Y.)
* Correspondence: xiaoshanzhu@nudt.edu.cn

Abstract: The infrared small target detection technology has a wide range of applications in maritime defense warning and maritime border reconnaissance, especially in the maritime and sky scenes for detecting potential terrorist attacks and monitoring maritime borders. However, due to the weak nature of infrared targets and the presence of background interferences such as wave reflections and islands in maritime scenes, targets are easily submerged in the background, making small infrared targets hard to detect. We propose the multidimensional information fusion network(MIFNet) that can learn more information from limited data and achieve more accurate target segmentation. The multidimensional information fusion module calculates semantic information through the attention mechanism and fuses it with detailed information and edge information, enabling the network to achieve more accurate target position detection and avoid detecting one target as multiple ones, especially in high-precision scenes such as maritime target detection, thus effectively improving the accuracy and reliability of detection. Moreover, experiments on our constructed dataset for small infrared targets in maritime scenes demonstrate that our algorithm has advantages over other state-of-the-art algorithms, with an IoU of 79.09%, nIoU of 79.43%, F1 score of 87.88%, and AuC of 95.96%.

Keywords: infrared maritime images; small infrared target detection; feature fusion; gradient information; attention mechanism

Citation: Yao, J.; Xiao, S.; Deng, Q.; Wen, G.; Tao, H.; Du, J. An Infrared Maritime Small Target Detection Algorithm Based on Semantic, Detail, and Edge Multidimensional Information Fusion. *Remote Sens.* **2023**, *15*, 4909. https://doi.org/10.3390/rs15204909

Academic Editors: Yanni Dong, Xiaochen Yang and Qian Du

Received: 29 August 2023
Revised: 5 October 2023
Accepted: 9 October 2023
Published: 11 October 2023

1. Introduction

The automatic monitoring and detection of targets on the sea surface have significant scientific research and practical application significance in maintaining national sovereignty and safeguarding maritime rights and interests. Compared to radar imaging methods, infrared imaging is a passive imaging method with strong resistance to smoke interference, longer detection distance, and wider temporal applicability. Infrared small target detection technology has been widely used in fields such as maritime defense and maritime surveillance in sea-sky scenes.

However, in general, targets are far away from the observation equipment, and infrared small targets occupy very few pixels in the image, appearing as patches or even dots, lacking effective shape features [1], as well as lacking texture, color, and shape features of common objects [2]. At the same time, in practical applications, sea surface scenes are complex and often contain static or slowly varying clutter, such as sea-sky lines, cloud clusters, and islands, as well as dynamic clutter, such as fish-scale light and sun glint. In harsh environments, strong wave clutter may also occur, making targets easily submerged in complex backgrounds. Based on these circumstances, infrared small target detection has great application value and research value but faces significant challenges.

Infrared small target detection technology can be divided into Track Before Detect (TBD) [3–6] and Detect Before Track (DBT) according to the order of utilizing prior information [7]. TBD can handle low signal-to-noise ratio situations but generally requires high computational complexity, making it difficult to detect high-speed targets. On the other hand, DBT algorithms generally have lower complexity and can meet real-time requirements, making them widely deployed on hardware platforms and widely applied in various fields. The quality of DBT algorithms is determined by single-frame infrared small target detection algorithms, making them a research hotspot. Currently, mainstream single-frame infrared small target detection algorithms can be divided into deep learning-based algorithms and traditional feature extraction-based algorithms. Traditional algorithms can be further divided into filter-based algorithms [8,9], human vision-inspired algorithms [10,11], and optimization-based methods [12–14]. Filter-based algorithms assume that there is a contrast difference between the target and the background, making high-contrast regions more likely to be targeted. In optimization-based methods, the target class is treated as a sparse matrix, while the background class is treated as a low-rank matrix. The goal is to continuously optimize the separation of low-rank and sparse matrices to achieve target detection. These methods heavily rely on handcrafted features, and their performance weakens when the pre-designed features do not match the actual scenarios.

Unlike traditional algorithms that rely on handcrafted features, deep learning methods rely on a large amount of data to learn target features [15–17]. It has achieved remarkable results in terms of speed and accuracy in object detection on general datasets. Dai et al. proposed the ACM [18], which combines high-level semantic information with low-level details and uses the U-net and FPN network structures suitable for extracting multiscale features to detect small infrared targets. Although these methods have achieved significant achievements, they still face some challenges when applied to the detection of small infrared targets in marine and aerial scenes [19]. Due to the sensitivity of the data, efficiently utilizing the limited data to ensure the effectiveness of the algorithm in similar scenarios becomes a key task. Therefore, we must delve into the image information that can effectively describe the targets. More importantly, the targets and backgrounds in the marine environment are complex and diverse. The direct application of models trained on general object detection datasets or specific non-marine small object detection datasets to marine target detection still needs to be verified for its effectiveness.

This paper explores an infrared small target detection algorithm in marine and aerial scenes that utilizes edge information, detail information, and semantic information for multidimensional information fusion. Specifically, it consists of three information processing channels. Firstly, the input infrared image is processed by a multiscale object detection FPN network to extract semantic information. Then, the input infrared image is simultaneously processed by the edge information extraction module and the detail information extraction module to extract edge information and detail information, respectively. As shown in Figure 1, the three types of information, with semantic information as the main component, detail information, and edge information as the auxiliary, are fused and processed by the FCN module to obtain the detection results. Additionally, considering that the weights of different layers in the fusion process of FPN cross-level connections are different, we design a multiscale information fusion module to address this issue. Experimental results on our dataset demonstrate that our algorithm outperforms state-of-the-art algorithms in terms of IoU, nIoU, F1, and AuC metrics.

The main contributions of this study are as follows:

a. We propose an infrared small target detection model that demonstrates excellent performance on a dataset specifically designed for infrared small targets in maritime and aerial scenes. We introduce an edge information extraction module, which not only compensates for the loss of target information caused by downsampling but also provides edge information to enable more precise target detection;

b. We draw inspiration from the deeplab network structure and introduce shallow feature maps with richer detailed information in the last stage to further reduce the loss caused by downsampling;
c. We propose a fusion mechanism that combines semantic information with detail and edge information. This mechanism first extracts semantic information from the FPN baseline network and then organically integrates all three components using an attention mechanism;
d. Experimental results on the dataset compared with other state-of-the-art algorithms demonstrate the excellent performance of our algorithm. It can effectively extract and learn the features of the targets.

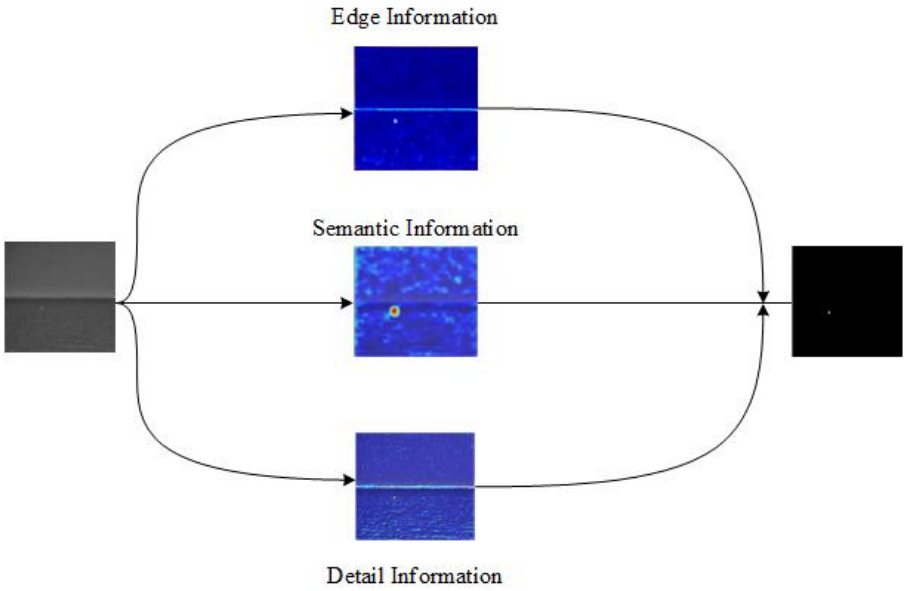

Figure 1. Main motivation.

The organizational structure of the entire article is as follows: In Section 2, the framework of the network and specific algorithm details are presented. Section 3 delves into our experimental details, results, and a comparative analysis of our algorithm against others. Section 4 comprehensively discusses and analyzes the experimental results. Finally, in Section 5, a comprehensive summary of the article is provided.

2. Materials and Methods

2.1. Related Work

2.1.1. Infrared Small Target Detection

Presently, the categorization of infrared small target detection algorithms in the context of maritime and sky scenes can be divided into single-frame-based detection methods and multi-frame-based detection methods. In the realm of multi-frame methods, approaches encompass particle filtering [5], Markov random fields [6], pipeline filtering [8], and dynamic programming [9], among others. Traditional single-frame-based infrared target detection algorithms can be further grouped into filter-based methods, visual saliency-based methods, and matrix decomposition-based methods.

Spatial filtering methods enhance the signal-to-noise ratio of infrared targets and suppress background noise by constructing filters. Notable algorithms include the Tophat [8] filter, Maxmean [9] filter, etc. These methods offer high real-time processing speed and simplicity. However, when encountering complex cloud layers or sea clutter, they might diverge from the underlying algorithm assumptions, leading to elevated false alarms.

Saliency detection methods emphasize the differences in grayscale values between objects and backgrounds akin to spatial filtering. This category includes Gaussian difference filters, Gaussian Laplacian filters, Gabor filters, second-order directional derivative filters,

and more. In 2013, Chen et al. proposed the Local Contrast Model (LCM) [10] based on the contrast differences between an image and its neighborhood. Although LCM is straightforward and effective, its effectiveness is limited to relatively bright targets against a background, and it might inadvertently enhance brighter noise, thereby resulting in a high false alarm rate. Subsequently, Wei et al. extended LCM to the MPCM model [11], which can detect both bright and dark targets but lacks adaptive threshold selection. Nevertheless, in complex scenarios, the prevalence of background interference often leads to high false alarm rates in these methods.

In recent years, low-rank sparse decomposition has garnered significant attention in the field of infrared small target detection due to its promising background suppression capabilities. IPI [12] is an exemplary model that treats targets as sparse matrices and backgrounds as low-rank components, constructing target patch models using local patches. Given that small targets occupy only a fraction of the entire image, this sparse hypothesis is suitable for diverse scenes. However, IPI preserves strong edge features of targets, resulting in time-consuming processing. Many researchers assume that targets come from multiple subspaces, leading to approaches like LRR [20], SMSL [21], and SRWS [13]. Zhang et al. [14] introduced the Partial Sum of the Tensor Nuclear Norm (PSTNN) model, employing the tensor nuclear norm and weighted L1 norm to suppress background effectively while retaining targets.

With the advancement of deep learning, researchers have increasingly turned to data-driven methods to address infrared weak small target detection challenges. Prominent algorithms include Fast RCNN [22–24], YOLO [25–30], as well as the Transformer initially applied in natural language processing. Unet and FPN are common network models in the field of infrared target detection, integrating high-level semantic information and low-level structural details. In 2021, Dai et al. introduced the ACM model [18], which combines non-local attention mechanisms with bidirectional pathways for cross-layer fusion. ISNet [19] employs Taylor finite difference edge blocks and bidirectional attention aggregation modules to ensure precise edge capture. AGPC [31] integrates an Attention-Guided Context Block (AGCB) to discern pixel correlations within and across distinct scales, and it further incorporates a Context Pyramid Module (CPM) for multiscale context information fusion. Li et al. [32] proposed a Dense Nested Attention Network (DNANet), where tri-directional interaction modules are densely nested, enhancing various depths of the UNet architecture. This augmentation significantly bolsters the model's performance in small target detection. However, the application of deep learning in this domain remains in the developmental phase due to the inherent sensitivity of infrared data.

2.1.2. Attention Mechanism

Attention mechanisms in deep learning refer to methods that redirect focus toward crucial regions within an image while disregarding irrelevant portions [33]. Attention mechanisms can be conceptualized as dynamic selection processes. These processes adaptively weight features based on their significance within the input, thus facilitating the generation of weighted features. Due to its excellent performance, attention mechanisms have been widely applied in neural networks, and various attention mechanisms are constantly evolving, such as SE-net [34], CBAM [35], EMANet [36], CCNet [37], and HamNet [38]. SEnet generates optimal feature maps through squeeze and excitation. CBAM enhances feature maps from both channel-wise and spatial-wise perspectives using channel attention modules and spatial attention modules. As the name suggests, the channel attention mechanism computes the significance level of each channel, making it commonly applied within convolutional neural networks. Among the established channel attention methods, the SENet model stands out. By learning inter-channel relationships, SENet enhances the network's expressive capability in feature representation, subsequently boosting model performance. The spatial attention mechanism shares a similar essence with the channel attention mechanism, with the former concentrating on capturing spatial importance by introducing attention modules. This enables the model to adaptively learn attention weights

for distinct regions. This approach allows the model to emphasize crucial image areas while disregarding less significant regions. Notably, the Convolutional Block Attention Module (CBAM) is a prime example, seamlessly integrating channel and spatial attention. CBAM aims to augment the convolutional neural network's capacity to focus on image content. EMANet approaches self-attention from an Expectation Maximization (EM) perspective. It introduces EM attention, utilizing the EM algorithm to derive a concise set of basis vectors rather than utilizing all points as reconstruction bases. In the context of CCNet, self-attention operations are conceptualized as graph convolutions. Instead of densely connected graphs handled by self-attention, CCNet introduces sparsely connected graphs. To achieve this, it introduces cross attention, involving both row and column attention mechanisms to capture global information comprehensively.

2.1.3. Edge Information

Edge features in images have always been the focus of scholars' attention [39]. Recently, researchers have been exploring the integration of edge information to tackle challenges in semantic segmentation. For instance, RPCNet [40] introduced an iterative pyramid context module, merging semantic edge detection and semantic segmentation within a unified multi-task learning framework. GSCNN [41] integrates shape flow to explicitly extract edge information, embedding it into the regular flow features. DFN [42] devised a depth-supervised edge network for enhancing the prediction of semantic edge. However, works such as [41,43] often solely utilize the deepest layer features for representing both semantic segmentation heads and edge detection heads, overlooking the potential of hierarchical convolution features at different stages. Considering that high-level features aid in classifying object categories and low-level features preserve fine image structures, it is imperative to comprehensively exploit hierarchical features across multiple stages to enhance both semantic segmentation and boundary detection.

2.2. Method

In this section, we describe MIFNet in detail. The following subsections detail the overall architecture, edge information extractor, multiscale information fusion module, multiple information fusion module, and training loss function of the proposed MIFNet.

2.2.1. Overall Architecture

The overall framework of the infrared maritime small target detection network based on edge information assistance is shown in Figure 2. Firstly, in one branch, the infrared image is fed into the FPN backbone network, while in another branch, the infrared image is input into the edge information extraction module. At the same time, the infrared image is fed into the detail module to enhance the extraction of detailed information. The main network adopts the mainstream FPN architecture, with Resnet-20 as the backbone network. The detailed structure of the baseline network is shown in Table 1. Features are extracted from different layers to obtain x2, x3, and x4. The input infrared image undergoes stage one processing to obtain feature map x2. X2 is then fused with x6 through the MSF (multiscale information fusion module) module. x2 undergoes stage two processing to obtain feature maps x3 and x4, which are then fused into x5 through the MSF module. The MSF module is detailed in Section 2.1.3. In the MSF module, high-level feature maps with rich semantic information are processed by the Channel Attention and Pixel Attention modules to generate a weight that guides the fusion of different scale information contained in high-level and low-level feature maps extracted from the previous stage. Additionally, high-level feature maps with semantic information processed by the Channel Attention module can make the feature maps pay more attention to regions of interest in images.

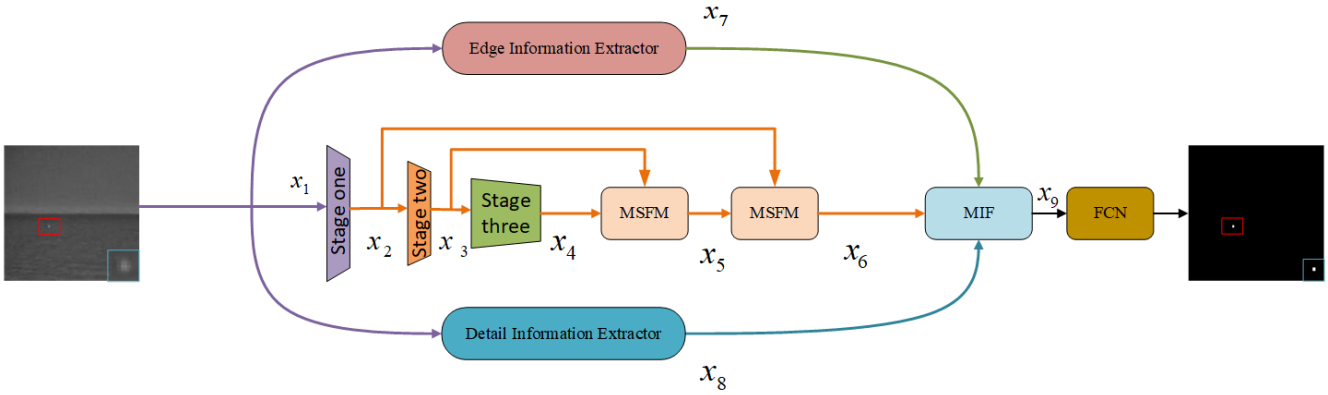

Figure 2. Overall Architecture.

Table 1. BNet backbones.

Stage	Output	Backbone
Stage one	$H \times W$	$\begin{bmatrix} 3 \times 3\text{conv}, 16 \\ 3 \times 3\text{conv}, 16 \end{bmatrix} \times b$
Stage two	$\frac{H}{2} \times \frac{W}{2}$	$\begin{bmatrix} 3 \times 3\text{conv}, 32 \\ 3 \times 3\text{conv}, 32 \end{bmatrix} \times b$
Stage three	$\frac{H}{4} \times \frac{W}{4}$	$\begin{bmatrix} 3 \times 3\text{conv}, 64 \\ 3 \times 3\text{conv}, 64 \end{bmatrix} \times b$

In addition, the input infrared image is processed by the edge information extraction module and the detail information extraction module, respectively, to obtain the edge and detail information of the infrared image. Finally, in the decision stage, the fusion is performed by the multidimensional information fusion module, and the binary image containing the target position information is obtained through the FCN module. Additionally, the input infrared image is individually processed by the edge information extraction module and the detail information extraction module, respectively, to obtain the edge information and the detail information of the infrared image. Then, in the decision stage, the fusion is performed by the multidimensional information fusion module, and the binary image containing the target position information is obtained through the FCN module. The pseudocode for the MIFNet is presented in Algorithm 1.

Algorithm 1: The Method Processing of an Image

Input: An Infrared Image
begin
Do abstract feature extraction
$X_2 = X_1 \otimes Resnet1$
$X_3 = X_2 \otimes Resnet2$
$X_4 = X_3 \otimes Resnet3$
$X_7 = X_1 \otimes SConv$
$X_8 = X_1 \otimes BConv$
End
Do FPN feature fusion
$X_5 = MSFM(X_4, X_3)$
$X_6 = MSFM(X_5, X_2)$
End
Do Multiple dimension information fusion
$X_9 = MIF(X_6, X_7, X_8)$
$result = FCN(X_9)$
End
Output: Binary Mask Image

MIFNet adopts FPN as the network framework and uses Resnet-20 as the backbone network to extract semantic features, as shown in Table 1. As illustrated in Figure 2, when an infrared image with a height of H and width of W is input into the network, it goes through the detail information extraction module, edge information extraction module, and semantic information extraction module separately. After processing and fusion, it outputs a binary mask image with target value 1. Due to its ability to simultaneously fuse high-level features with low resolution and high semantic information, as well as low-level features with high resolution but low semantic information, FPN has been widely applied in the field of object detection. In FPN, the high-resolution information from lower layers is transmitted to higher layers through cross-layer connections, enabling the network to integrate features at different scales and reduce information loss during downsampling processes.

2.2.2. Edge Information Extractor

During the downsampling process, the resolution of the image gradually decreases. We observed that edge information can be combined with deep learning methods to complement the loss of image resolution.

For all input infrared images, the processing flow of the edge information branch is as follows:

$$x_{i.v} = Conv2d(x_i, kernel_v) \tag{1}$$

$$x_{i,h} = Conv2d(x_h, kernel_h) \tag{2}$$

where *Conv2d* represents the two-dimensional convolution operation, x_i is the i-th input channel, and $kernel_h$ and $kernel_v$ are the vertical and horizontal convolution kernels. In this paper, the configuration of the convolution kernel is as shown in Equation (1):

$$kernel_h = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{3}$$

$$kernel_v = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{4}$$

After calculating the horizontal and vertical gradients separately, the final gradient is calculated using the formula shown in Equation (2):

$$\chi_i = \sqrt{\chi_{i,v}^2 + \chi_{i,h}^2} \tag{5}$$

2.2.3. Multiscale Information Fusion Module

In the semantic information extraction module, we use cross-layer connections to fuse feature maps of different scales. Inspired by the non-local idea [18], we employ a non-modulation demodulation fusion mechanism to fuse information from different layers. Inspired by BiseNet [44] and DMANet [45], we adopt a two-branch architecture to use semantic information. MSF (Multiscale Information Fusion) module consists of Channel Attention and Pixel Attention, and the final fusion is performed by elementwise addition. The structure of Module MSFM is depicted in Figure 3.

Compared to other fusion modules, our MSFM module tends to focus more on semantic information, with shallow-level details serving as supplementary components. Due to the typically small proportion of infrared targets in images, they often lack detailed information and texture. Therefore, we have devised the Channel Attention mechanism to enhance targets by reinforcing channel-related information, which enhances the network's

overall attention toward the targets. As shown in the diagram, the formula for Channel Attention can be summarized as follows:

$$MSFM(F_D, F_S) = (\alpha \otimes (CA(F_D))) \oplus (((1-\alpha) \otimes (F_S) \otimes CA(X_l))) \quad (6)$$

$$F_M = CA(F_D) == \sigma(\mathcal{B}(\mathbf{W}_2\delta(\mathcal{B}(\mathbf{W}_1 F_D)))) \quad (7)$$

$$\alpha = PA(F_M) = \sigma((\mathbf{W}\delta(\beta(\mathbf{W}F_M)))) \quad (8)$$

σ, δ, β, W, $\oplus$, $\otimes$ represents the Sigmoid function, Rectified Linear Unit(ReLU), Batch Normalization(BN), Convolution function, add function, and element-wise multiplication, respectively. F_D represents the deep feature map containing semantic information, F_S represents the shallow feature map containing detail information, CA represents the channel attention mechanism, PA represents the pixel attention mechanism, α represents the dynamic weights obtained by fusing guidance information extracted from the deep feature map, and F_M represents F_D after channel attention.

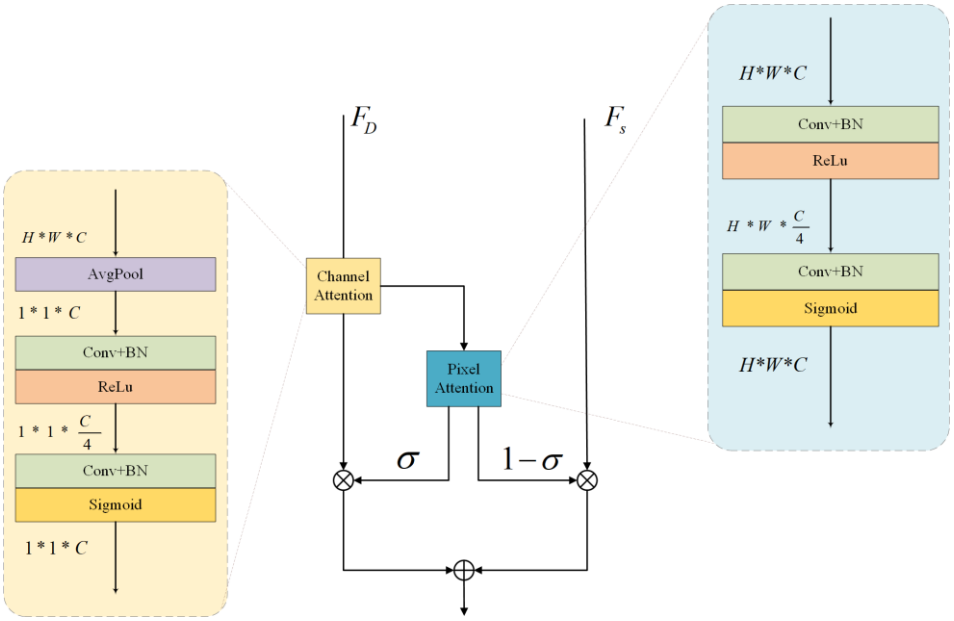

Figure 3. Multiscale Information Fusion Module.

2.2.4. Multiple Information Fusion Module

The MIF (Multiple Information Fusion) module is designed to guide the fusion of edge features and enrich detailed information using semantic information. As shown in the figure, the context branch contains rich semantic information but lacks specific spatial and geometric information. On the other hand, the detailed branch preserves relatively rich spatial details. Additionally, due to the differences between infrared targets and the environment, we introduce edge information as a supplement. The structural layout of Module MIF aligns with the visual representation presented in Figure 4.

The extracted semantic information is guided by weights obtained through global average pooling, convolution blocks, batch normalization, and the Sigmoid function. Based on these guided weights, different weights are assigned and multiplied into the edge information feature map and detail information feature map. The formula is as follows:

$$MIF(F_B, F_S, F_D) = Conv((\alpha \otimes F_B) \oplus (((1-\alpha) \otimes (F_D)))) \quad (9)$$

$$\alpha = SW(F_S) = \sigma((\beta(\mathbf{W}f_D))) \quad (10)$$

The formula for average pooling is as follows:

$$f_D = \mathcal{F}_{ap}(F_D) = \frac{1}{H \times W} \sum_{i=1,j=1}^{H,W} F[:, i, j] \tag{11}$$

σ, β, W, $\oplus$, $\otimes$ represents the Sigmoid function, Batch Normalization(BN), Convolution function, add function, and element-wise multiplication, respectively. F_B represents edge information, F_S represents semantic information, F_D represents detail information, SW represents the mechanism for generating semantic information guided weights, and α represents dynamic weights fused from the guided information extracted from semantic information.

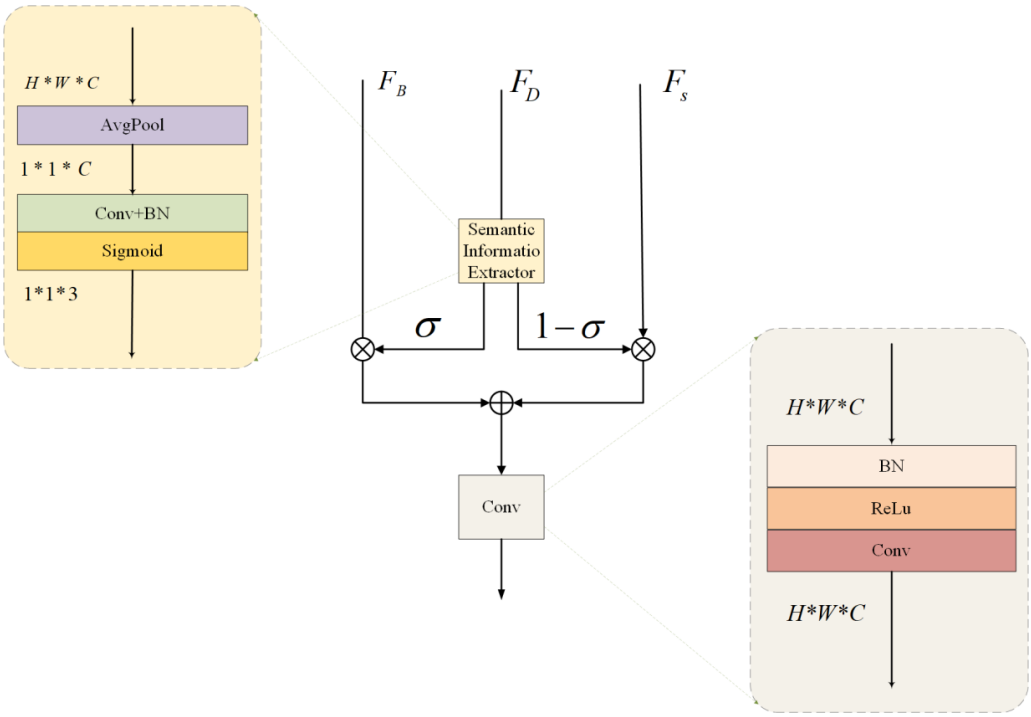

Figure 4. Multiple dimension Fusion Module.

2.2.5. Loss Function

Due to the issue of positive-negative sample imbalance between infrared small targets and background, we employ a soft margin loss function during the network training process. The formula for this loss function is defined as follows:

$$\updownarrow_{\text{soff}-\text{IoU}}(x, s) = \frac{\sum_{i,j} x_{i,j} \cdot S_{i,j}}{\sum_{i,j} s_{i,j} + x_{i,j} - x_{i,j} \cdot s_{i,j}} \tag{12}$$

$s \in \mathbb{R}^{H \times W}$ represents the predicted target score map, while $x \in \mathbb{R}^{H \times W}$ represents the annotated ground truth target image.

3. Results

In this section, a series of experiments were conducted to validate the effectiveness of our method. Firstly, we briefly introduced the evaluation metrics used to validate the algorithm. Then, the dataset used for validation was described. Subsequently, the algorithm was tested qualitatively and quantitatively. Finally, a series of ablation experiments were conducted to demonstrate the effectiveness of the added module. In the following equation, N represents the total number of samples, and TP, FP, TN, FN, T, and P, respectively,

represent true positives, false positives, true negatives, false negatives, true positives, and positives.

3.1. Evaluation Metrics

(1) Intersection over Union (IoU) is a pixel-level evaluation metric used to assess the contour description capability of an algorithm [18]. It is calculated as the ratio of the intersection area between the predicted region and the ground truth label to the union area of the two, as shown below:

$$IoU = \frac{TP}{T + P - TP} \tag{13}$$

(2) Normalized Intersection over Union (nIoU) is a metric specifically designed to evaluate the performance of infrared small target detection [18], aiming to avoid the influence of large target segmentation on the evaluation metric. It is defined as follows:

$$nIoU = \frac{1}{N}\sum_{i}^{N}\frac{TP[i]}{T[i] + P[i] - TP[i]} \tag{14}$$

(3) The Receiver Operating Characteristic (ROC) curve is used to describe the relationship between the True Positive Rate (*TPR*) and the False Positive Rate (*FPR*) [19].

$$TPR = \frac{TP}{TP + FN} \quad FPR = \frac{FP}{FP + TN} \tag{15}$$

(4) The F1-measure [46] is a key metric used to evaluate algorithm performance. It calculates the harmonic mean of precision and recall, effectively capturing both dimensions. This metric encompasses the inherent trade-off relationship between precision and recall, thus forming a comprehensive performance evaluation scale that balances the two key evaluation metrics.

$$precision = \frac{TP}{TP + FP} \quad recall = \frac{FP}{TP + FN} \tag{16}$$

$$Fmeasure = \frac{2precsion \times recall}{precsion + recall} \tag{17}$$

3.2. Experiment Settings and Dataset

Dataset: In order to compare infrared small target detection algorithms in maritime scenes, we have constructed an infrared ship target dataset in maritime scenes. This dataset consists of six scenes, and the detailed information of the dataset is shown in Table 2. The total number of images used for training is 210, and there are 215 images used for testing. The dataset images have a size of 640 × 512 and overall appear dark, with small targets located in complex backgrounds. The backgrounds are blurry and primarily consist of sea clutter, islands, and other interference.

Table 2. Information of the test Dataset.

	Target Size	Target Category	Background Type
(a)	5×7	One small target with low local contrast	Calm sea
(b)	$4 \times 8, 5 \times 6$	Two small targets with low local contrast	Floating interface
(c)	$5 \times 5, 5 \times 6, 5 \times 7, 5 \times 8$	Three small targets with low local contrast	Calm sea
(d)	$9 \times 9, 5 \times 6, 5 \times 6$	Three small targets with low local contrast	Wave clutter
(e)	$7 \times 7, 3 \times 6, 9 \times 9, 4 \times 8, 4 \times 7$	Four small targets with low local contrast	Wave clutter
(f)	$4 \times 6, 4 \times 5, 5 \times 6$	Three small targets with low local contrast	Dynamic camera

Experimental Details: ACMUNet [18], ACMFPN [18], AGPCNet [33], and our method are trained on an NVIDIA GTX 1080Ti with 12 GB of memory. We use Python as the programming language, Pycharm version 2022 as the editor, and PyTorch 1.8.0 as the deep learning framework. We choose Adagrad as the optimizer with a learning rate update strategy of CosineAnnealingLR. The initial learning rate is set to 0.05, the batch size is

set to 4, and the total number of training epochs is set to 200. Traditional algorithms such as Tophat [8], Maxmean [9], MPCM [11], IPI [12], SRWS [13], and PSTNN [14] are implemented on an Intel i7-12700 CPU with 32 GB of memory using Matlab R2021b. The specific experimental parameters for traditional algorithms are shown in Table 3 [47].

Table 3. Computational complexity and running time of deep learning algorithms.

	ACMFPN	ACMUnet	AGPC	Ours
FLOPs	564.537 M	1.003 G	86.362 G	2.013 G
Params	386.615 K	519.271 K	12.360 M	397.666 K

3.3. Equations Comparision to the State-of-the-Art Method

To validate the effectiveness of our method, we compare our method with other advanced traditional and deep learning algorithms in both qualitative and quantitative aspects. The results are shown in Table 4, Figures 5–10.

Table 4. Parameter settings.

Method	Hyper-Parameters Settings
MaxMean	$\text{Patchsize}_{median} = 3 \times 3$
Tophat	Patch size $= 3 \times 3$
MPCM	window size = {3, 5, 7, 9}
PTSNN	Patch size : 40, Slide step : 40, $\lambda = 0.6/\sqrt{\max(n_1, n_2) * n_3}$
IPI	Patch size : 50×50, Slide $= 10$, $\lambda = L/\sqrt{\max(m, n)}, L = 4.5, , \varepsilon = 10^{-7}$
SRWS	Patch size : 50, Slide step : 10, $\beta = 1/\sqrt{\max(m, n)}$

Quantitative results: Table 5 presents the results of different algorithms. It is evident that our algorithm performs among the top algorithms tested and achieves the best detection results. From the quantitative results, we can observe that deep learning algorithms outperform traditional algorithms. The reasons for the suboptimal performance of the traditional algorithm are as follows:

1. Reliance on Manual Feature Design: Most traditional algorithms heavily rely on manually designed features. However, in maritime and aerial scenes, interference from factors such as reflected light from waves and islands can render the designed features incapable of meeting the requirements for distinguishing targets from the background. As a result, the comparative metrics are lower.

2. Fragmented Results in Visual Analysis: By combining the visualized results for analysis, it can be observed that the traditional detection algorithm exhibits a fragmented phenomenon in the results for ships, failing to recognize ships as a whole entity. This may result in lower detection metrics.

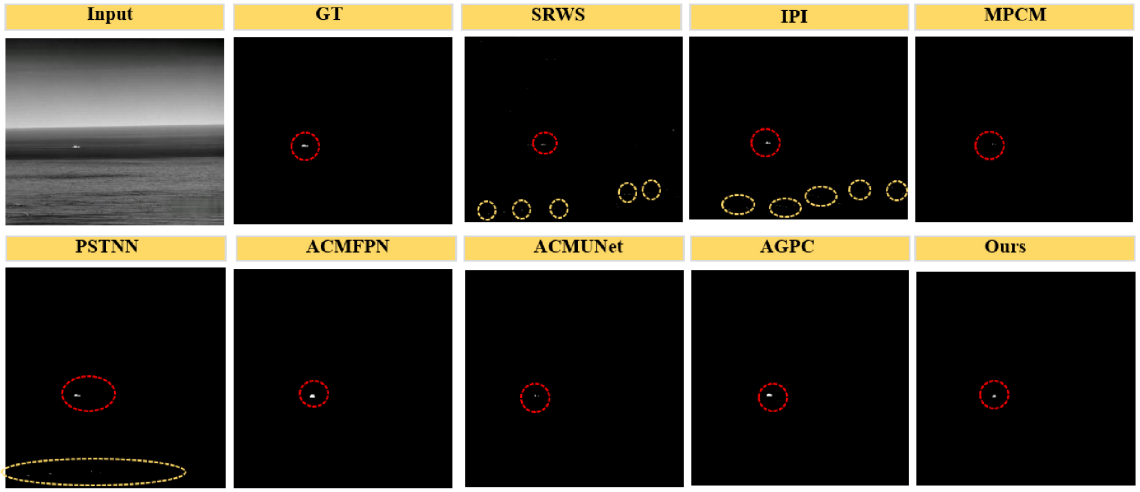

Figure 5. Detection results of infrared small targets in Scene 1 using different detection methods.

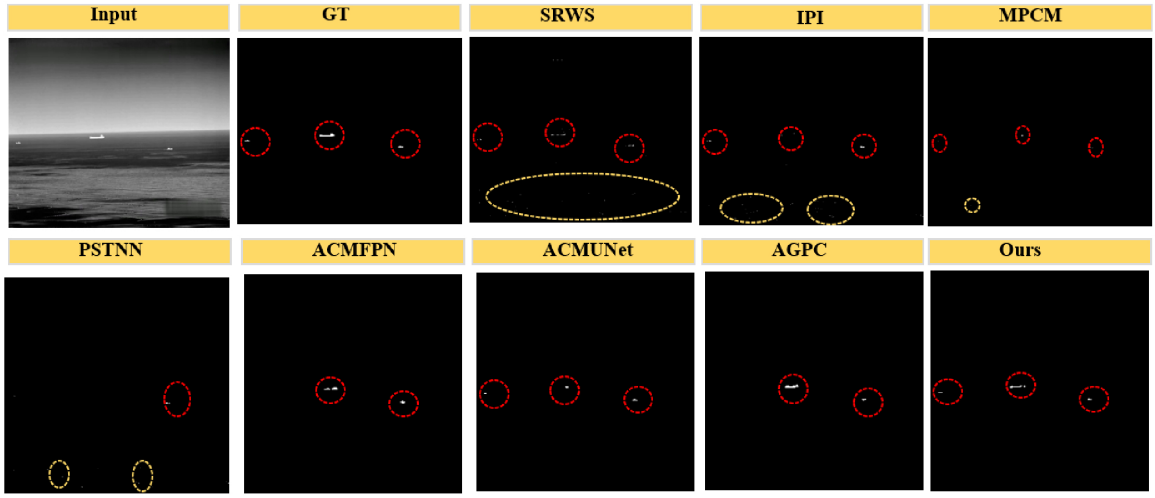

Figure 6. Detection results of infrared small targets in Scene 2 using different detection methods.

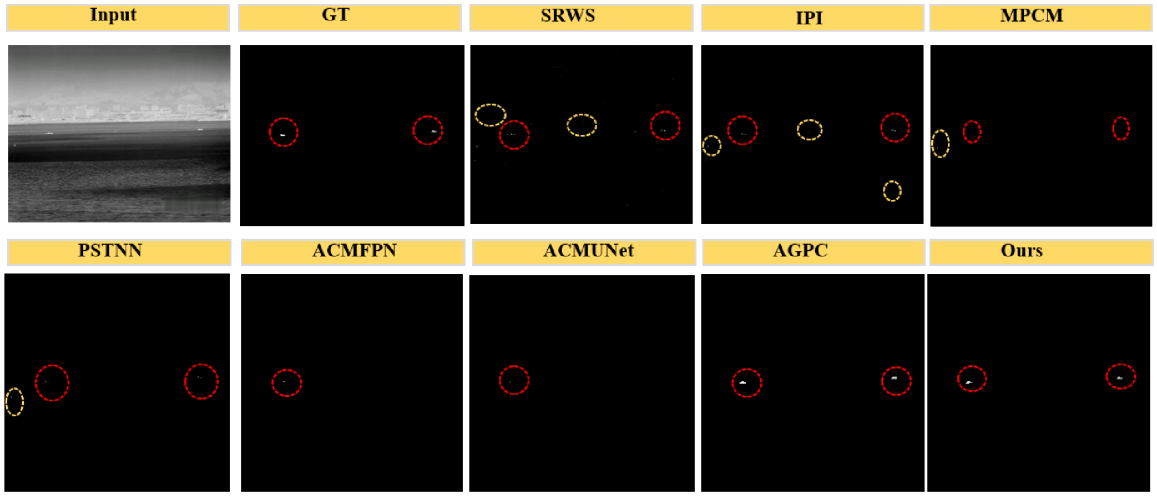

Figure 7. Detection results of infrared small targets in Scene 3 using different detection methods.

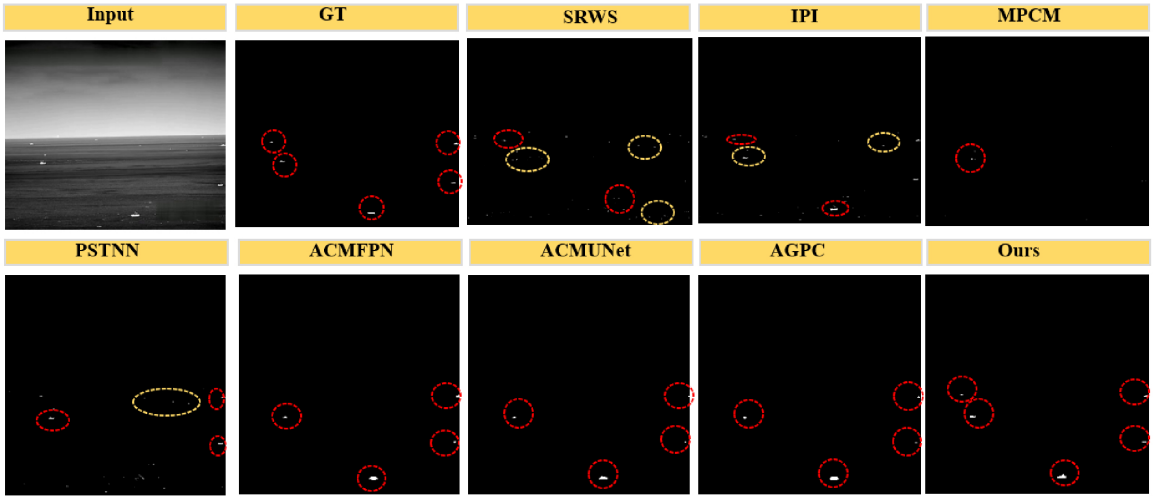

Figure 8. Detection results of infrared small targets in Scene 4 using different detection methods.

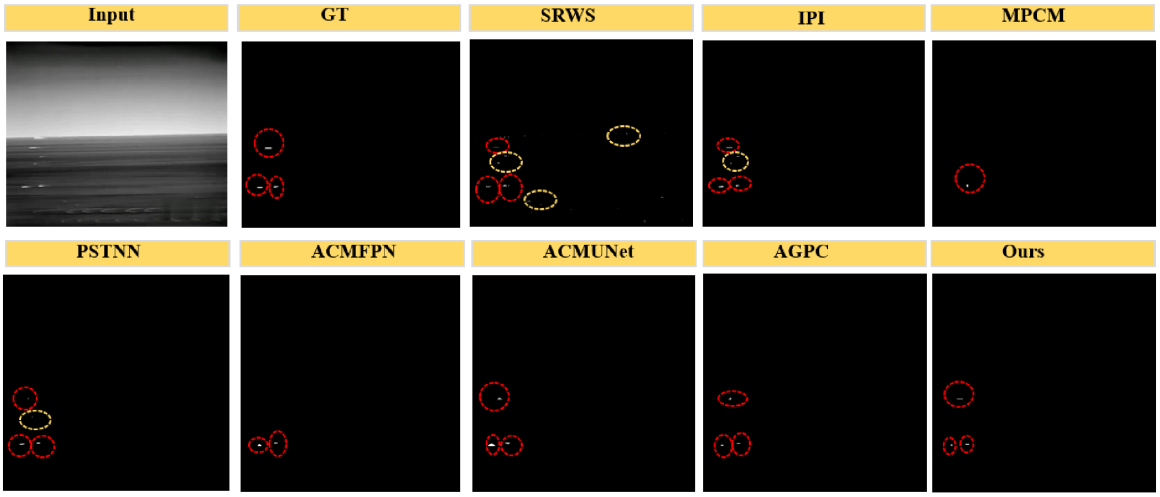

Figure 9. Detection results of infrared small targets in Scene 5 using different detection methods.

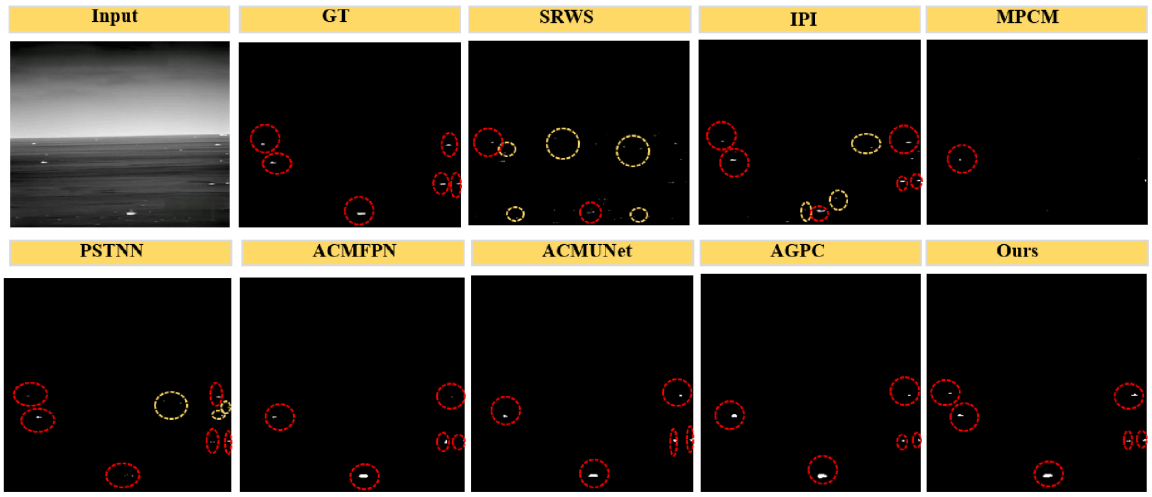

Figure 10. Detection results of infrared small targets in Scene 6 using different detection methods.

Table 5. Comparison with other state-of-the-art methods.

Method	IoU	nIoU	F1	AuC
Maxmean	0.12	3.54	0.23	54.45
Tophat	26.33	26.7	41.69	63.44
MPCM	11.58	12.49	20.75	55.80
IPI	48.05	48.17	64.91	77.23
PSTNN	43.51	44.22	60.64	74.52
SRWS	26.39	27.71	41.76	26.39
ACMFPN	72.66	72.97	84.19	95.75
ACMUnet	72.55	73.24	83.3	95.50
AGPC	77.61	78.13	83.39	95.58
Ours	79.09	79.43	87.88	95.96

Qualitative results: Figures 5–10 present the qualitative results of different detection methods on the infrared small target dataset in maritime and aerial scenes. The true targets are indicated by red bounding boxes in the images, while the false alarms are indicated by yellow bounding boxes. By comparing the detection results of each algorithm with the ground truth (GT) of the input image, it can be observed that the algorithms have missed detections.

As shown in Figure 5, the traditional algorithm SRWS fails to detect the targets, while Tophat, Maxmean, and PSTNN are able to detect the targets but with a large number of false alarms. On the other hand, deep learning algorithms such as ACMFPN, ACMUnet, AGPC, MPCM, and our algorithm successfully detect the targets. In Figure 6, SRWS, IPI, MPCM, PSTNN, ACMFPN, ACMUnet, and our algorithm are able to detect the targets. However, SRWS, IPI, and PSTNN algorithms have a large number of false alarms and can only detect the edges of the targets. In this case, both ACMFPN and AGPC algorithms have some missed detections., while ACMFPN can only detect the edges of the targets. ACMUnet and our algorithm successfully detect the targets, but ACMUnet's detection is incomplete. In Figure 7, SRWS, IPI, MPCM, PSTNN, ACMFPN, ACMUnet, and our algorithm are able to detect the two targets successfully but with a large number of false alarms. As for the deep learning detection results, ACMFPN and ACMUnet have some missed detections, while AGPC and our algorithm successfully detect the targets. In Figure 8, SRWS, IPI, MPCM, PSTNN, ACMUnet, AGPC, GNet, and other deep learning algorithms are able to detect the targets, but with some missed detections, while GNet successfully detects all the targets. SRWS, IPI, and PSTNN also have relatively high false alarm rates. In Figure 9, SRWS, IPI, PSTNN, ACMUnet, AGPC, GNet, and other algorithms are able to detect the targets, but SRWS, IPI, and PSTNN have relatively high false alarm rates, and ACMFPN has some missed detections. In Figure 10, SRWS, IPI, PSTNN, ACMFPN, ACMUnet, AGPC, GNet, and other algorithms are able to detect the targets, but SRWS, IPI, and PSTNN have relatively high false alarm rates, while ACMFPN, ACMUnet, and AGPC have some missed detections.

Figure 11 shows the ROC curves of the detection results obtained by different testing algorithms on the dataset. The larger the area under the ROC curve (AUC), the better the algorithm's performance. It can be observed that deep learning algorithms demonstrate better performance on the ROC curves compared to traditional algorithms. Additionally, our algorithm achieves top performance in terms of the AuC metric among the tested algorithms. The 3D visualization of the detection results simultaneously presents a more vivid display of the superior performance of our proposed algorithm (Appendix A Figure A1).

3.4. Ablation Study

In this phase, we designed a series of experiments to validate the potential benefits of the proposed network modules and design choices, ensuring the rationality of the contributions made by the components in our proposed model.

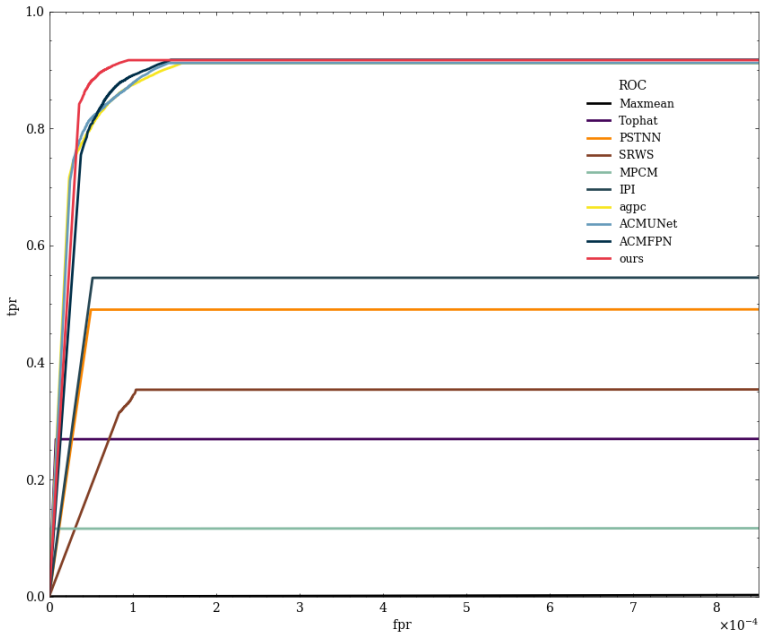

Figure 11. ROC curve.

1. Influence of edge information.

As shown in Table 6, incorporating edge information leads to an improvement of 1.8% and 2.19% in network IoU and nIoU, respectively. Since infrared small targets often occupy fewer pixels in the image, it is crucial to preserve and utilize the effective features of small targets in the network. Therefore, on the one hand, in order to provide the model with more dimensional information, and on the other hand, to compensate for the information loss caused by the downsampling process, we introduce edge information. Experimental results demonstrate that the inclusion of edge information enhances the network's performance and provides more features of infrared small targets.

Table 6. Ablation study on the module.

Method	IoU	nIoU
Baseline	72.66	72.97
Baseline + fusion	76.19	76.28
Baseline + sobel + fusion	77.99	78.47
Baseline + x1 + sobel + fusion	79.09	79.43

2. Influence of detailed information.

With the increase in network depth and the improvement in downsampling frequency, the resolution of the image decreases. In order to alleviate the information loss caused by downsampling operations, we introduce the finer details present in the shallow feature maps during the final fusion stage. As shown in Table 6, the introduction of detail information leads to an increase of 0.2% and 0.12% in network IoU and nIoU, respectively. The experimental results effectively demonstrate that the introduction of detail information in the final stage preserves some original features of the infrared small targets and compensates for the information loss caused by downsampling, thereby improving the network's performance.

3. Influence of Fusion module

As shown in Table 6, the introduction of the context interaction module resulted in an improvement of 3.53% and 3.31% in network IoU and nIoU, respectively. Compared to the fusion module in ACM, we focused more on extracting semantic information from high-level feature maps and guiding the fusion of high-level feature maps with semantic information and shallow-level feature maps with detailed information through the generation of weights. Additionally, an adversarial approach was employed to achieve an effective balance in the fusion. Furthermore, the Channel Attention mechanism was used to optimize the high-level feature maps. Experimental results on the infrared small target dataset in the maritime and sky scenes demonstrated that this fusion approach outperformed the fusion approach in ACM.

4. Influence of depth of layers

As shown in Table 7, we also considered the impact of network depth on the algorithm performance. When the number of Resnet blocks in the network was reduced from 4 to 3, IoU and nIoU decreased by 0.8% and 0.65%, respectively. Similarly, when the number of Resnet blocks in the network was reduced from 3 to 2, IoU and nIoU decreased by 0.32% and 0.38%, respectively. Furthermore, reducing the number of Resnet blocks in the network from 2 to 1 resulted in a decrease of 1.3% and 1.31% in IoU and nIoU, respectively. It can be observed that as the network depth decreases, the detection accuracy of the network also decreases.

Table 7. Ablation study on the depth of the network.

Depth	IoU	nIoU
1	76.60	77.09
2	77.90	78.40
3	78.22	78.78
4	79.09	79.43

4. Discussion

Comparing the visualization and quantitative results obtained by our proposed algorithm with other advanced algorithms, it is inevitable for model-driven algorithms relying on manually designed features to have false alarms and missed detections when facing complex background changes. Compared with other data-driven algorithms of the same type, our algorithm has stronger robustness and finer detection results. Meanwhile, the ablation experiments have demonstrated the effectiveness of the multiscale feature fusion module and multidimensional feature fusion module, which effectively utilizes multidimensional information in the process of target detection. In the future, we will explore the field of multidimensional information fusion, including how to represent each dimension of information and fusion methods.

5. Conclusions

This paper proposes an infrared maritime small target detection algorithm in maritime scenes. To overcome the issue of information loss during the typical downsampling process, we choose to simultaneously fuse detailed information and edge information. By comparing our algorithm with other state-of-the-art methods on the infrared small target detection dataset we established in maritime scenes, our algorithm performs at the top. Our algorithm achieves IoU, nIoU, F1, and AuC scores of 79.09%, 79.43%, 87.88%, and 95.96%, respectively. Visualizing the detection results in different scenes reveals that our algorithm has a lower false negative rate compared to other algorithms, indicating that our network can extract more effective information from the targets. Additionally, the results of the ablation experiments demonstrate that the adoption of attention mechanism-based cross-layer connection information fusion and the introduction of edge information and detail information contribute to improving the detection performance of the algorithm. In future work, we will explore the integration of different forms of information with deep learning techniques to address the issue of infrared small target detection. This article did not consider the algorithm performance on embedded platforms such as DSP, FPGA, ARM, etc., and the algorithm performance on embedded platforms will be further investigated in the future.

Author Contributions: Conceptualization, J.Y.; methodology, J.Y.; software, J.Y. and J.D.; validation, J.Y. and J.D.; formal analysis, J.Y., Q.D. and G.W.; investigation, J.Y. and S.X.; resources, J.Y.; data curation, J.Y. and J.D.; writing—original draft preparation, J.Y.; writing—review and editing, J.Y., S.X. and Q.D.; visualization, J.Y. and S.X.; supervision, G.W. and S.X.; project administration, H.T.; funding acquisition, H.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (61921001).

Data Availability Statement: Not applicable.

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

Appendix A. Three-Dimensional Visualization Results of Different Methods on 6 Test Images

Figure A1 presents the 3D intensity visualization results of different testing algorithms, and it can be observed that our algorithm's detection results are closer to the ground truth.

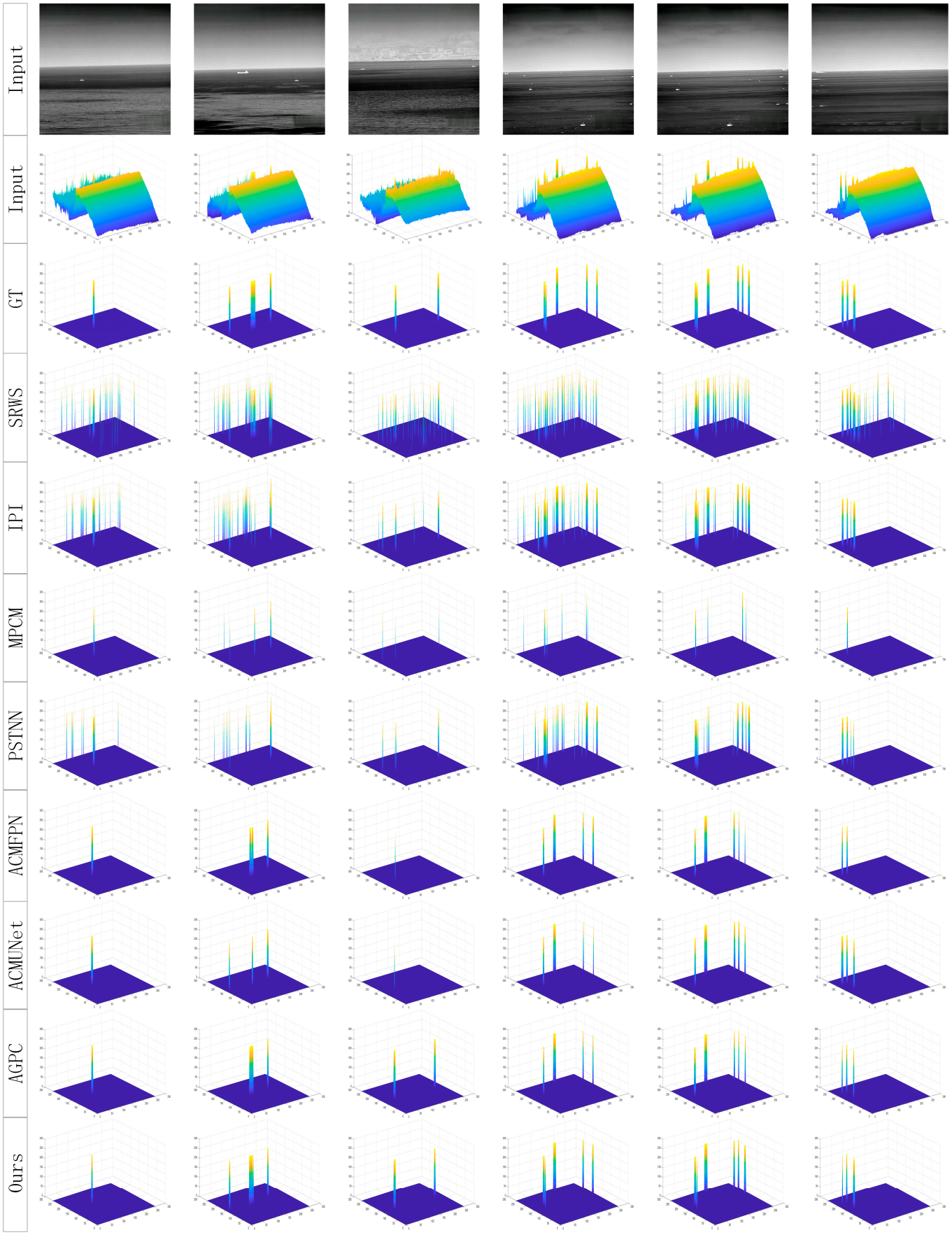

Figure A1. Three-dimensional visualization results of different methods on 6 test images.

References

1. Yang, C.; Ma, J.; Qi, S.; Tian, J.; Zheng, S.; Tian, X. Directional support value of gaussian transformation for infrared small target detection. *Appl. Opt.* **2015**, *54*, 2255–2265. [CrossRef] [PubMed]
2. Qi, H.; Mo, B.; Liu, F.; He, Y.; Liu, S. Small infrared target detection utilizing local region similarity difference map. *Infared Phys. Technol.* **2015**, *71*, 131–139. [CrossRef]
3. Pak, J. Visual odometry particle filter for improving accuracy of visual object trackers. *Electron. Lett.* **2020**, *56*, 884–887. [CrossRef]
4. Lin, G.; Fan, W. Unsupervised video object segmentation based on mixture models and saliency detection. *Neural Process. Lett.* **2020**, *51*, 657–674. [CrossRef]
5. Li, B.; Xu, Z.; Zhang, J.; Wang, X.; Fan, X. Dim-Small Target Detection Based on Adaptive Pipeline Filtering. *Math. Probl. Eng.* **2020**, *2020*, 8234349. [CrossRef]
6. Fu, J.; Zhang, H.; Luo, W.; Gao, X. Dynamic Programming Ring for Point Target Detection. *Appl. Sci.* **2022**, *12*, 1151. [CrossRef]
7. Zhao, M.; Li, W.; Li, L.; Hu, J.; Ma, P.; Tao, R. Single-Frame Infrared Small-Target Detection: A Survey. *IEEE Geosci. Remote Sens. Mag.* **2022**, *10*, 87–119. [CrossRef]
8. Rivest, J.F.; Fortin, R. Detection of dim targets in digital infrared imagery by morphological image processing. *Opt. Eng.* **1996**, *35*, 1886–1893. [CrossRef]
9. Deshpande, S.D.; Er, M.H.; Venkateswarlu, R.; Chan, P. Maxmean and max-median filters for detection of small targets. In *Signal and Data Processing of Small Targets*; International Society for Optics and Photonics: Bellingham, WA, USA, 1999; Volume 3809, pp. 74–83.
10. 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]
11. Wei, Y.; You, X.; Li, H. Multiscale patch-based contrast measure for small infrared target detection. *Pattern Recognit.* **2016**, *58*, 216–226. [CrossRef]
12. 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] [PubMed]
13. 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]
14. Zhang, L.; Peng, Z. Infrared small target detection based on partial sum of the tensor nuclear norm. *Remote Sens.* **2019**, *11*, 382. [CrossRef]
15. Haq, M.A.; Hassine, S.B.H.; Malebary, S.J.; Othman, H.A.; Tag-Eldin, E.M. 3D-cnnhsr: A 3-dimensional convolutional neural network for hyperspectral super-resolution. *Comput. Syst. Sci. Eng.* **2023**, *47*, 2689–2705.
16. Haq, M.A. CNN based automated weed detection system using uav imagery. *Comput. Syst. Sci. Eng.* **2022**, *42*, 837–849.
17. Stupariu, M.-S.; Cushman, S.A.; Pleşoianu, A.-I.; Pătru-Stupariu, I.; Fuerst, C. Machine learning in landscape ecological analysis: A review of recent approaches. *Landsc. Ecol.* **2022**, *37*, 1227–1250. [CrossRef]
18. 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, 5–9 January 2021; pp. 950–959.
19. 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 (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 877–886.
20. Li, M.; He, Y.J.; Zhang, J. Small infrared target detection based on low-rank representation. In *Image and Graphics*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 393–401.
21. 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]
22. Girshick, R.; Donahue, J.; Darrell, T. 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.
23. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
24. He, K.; Gkioxari, G.; Dollár, P. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
25. Redmon, J.; Divvala, S.; Girshick, R. 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.
26. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
27. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
28. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
29. Li, C.; Li, L.; Jiang, H. YOLOv6: A single-stage object detection framework for industrial applications. *arXiv* **2022**, arXiv:2209.02976.
30. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**, arXiv:2207.02696.

31. 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]
32. 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.* **2023**, *32*, 1745–1758. [CrossRef] [PubMed]
33. Lee, M. The geometry of feature space in deep learning models: A holistic perspective and comprehensive review. *Mathematics* **2023**, *11*, 2375. [CrossRef]
34. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation neworks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; pp. 7132–7141.
35. Woo, S.; Park, J.; Lee, J.; 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.
36. Li, X.; Zhong, Z.; Wu, J.; Yang, Y.; Lin, Z.; Liu, H. Expectation-maximization attention networks for semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019.
37. Huang, Z.; Wang, X.; Wei, Y.; Huang, L.; Shi, H.; Liu, W.; Huang, T.S. Ccnet: Criss-cross attention for semantic segmentation. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2020.
38. Abdulqader, D.A.; Hathal, M.S.; Mahmmod, B.M.; Abdulhussain, S.H.; Al-Jumeily, D. Plain, edge, and texture detection based on orthogonal moment. *IEEE Access* **2022**, *10*, 114455–114468. [CrossRef]
39. Geng, Z.; Guo, M.-H.; Chen, H.; Li, X.; Wei, K.; Lin, Z. Is attention better than matrix decomposition? In Proceedings of the International Conference on Learning Representations, Virtual, 3–7 May 2021.
40. Zhen, M.; Wang, J.; Zhou, L.; Li, S.; Shen, T.; Shang, J.; Fang, T.; Quan, L. Joint semantic segmentation and boundary detectionusing iterative pyramid contexts. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 13666–13675.
41. Takikawa, T.; Acuna, D.; Jampani, V.; Fidler, S. Gated-SCNN: Gatedshape CNNs for semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October 2019–2 November 2019; pp. 5229–5238.
42. Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; Sang, N. Learninga discriminative feature network for semantic segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 1857–1866.
43. Yuan, Y.; Xie, J.; Chen, X.; Wang, J. SegFix: Model-agnostic boundary refinement for segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, 23–28 August 2020; pp. 489–506.
44. Yu, C.; Wang, J.; Peng, C.; Gao, C.; Yu, G.; Sang, N. BiSeNet: Bilateral Segmentation Network for Real-Time Semantic Segmentation. In *Computer Vision—ECCV 2018, Proceedings of the 15th European Conference, Munich, Germany, 8–14 September 2018*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 334–349.
45. Sun, X.; Xie, Y.; Jiang, L.; Cao, Y.; Liu, B. DMA-Net: DeepLab With Multi-Scale Attention for Pavement Crack Segmentation. *IEEE Trans. Intell. Transport. Syst.* **2022**, *23*, 18392–18403. [CrossRef]
46. Faisal, M.M.; Mohammed, M.S.; Abduljabar, A.M.; Abdulhussain, S.H.; Mahmmod, B.M.; Khan, W.; Hussain, A. Object detection and distance measurement using AI. In Proceedings of the 2021 14th International Conference on Developments in eSystems Engineering (DeSE), Sharjah, United Arab Emirates, 7–10 December 2021; pp. 559–565.
47. Lv, G.; Dong, L.; Liang, J.; Xu, W. Novel Asymmetric Pyramid Aggregation Network for Infrared Dim and Small Target Detection. *Remote Sens.* **2022**, *14*, 5643. [CrossRef]