remote sensing

MDPI

Article

TFCD-Net: Target and False Alarm Collaborative Detection Network for Infrared Imagery

Siying Cao [1,2], Zhi Li [1,2], Jiakun Deng [1,2], Yi'an Huang [1,2] and Zhenming Peng [1,2,*]

[1] School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; caosiying3008@std.uestc.edu.cn (S.C.); zhili8888@std.uestc.edu.cn (Z.L.); dengjiakun@std.uestc.edu.cn (J.D.); huangyian@std.uestc.edu.cn (Y.H.)

[2] Laboratory of Imaging Detection and Intelligent Perception, University of Electronic Science and Technology of China, Chengdu 611731, China

* Correspondence: zmpeng@uestc.edu.cn; Tel.: +86-028-8320-8946

Abstract: Infrared small target detection (ISTD) plays a crucial role in both civilian and military applications. Detecting small targets against dense cluttered backgrounds remains a challenging task, requiring the collaboration of false alarm source elimination and target detection. Existing approaches mainly focus on modeling targets while often overlooking false alarm sources. To address this limitation, we propose a Target and False Alarm Collaborative Detection Network to leverage the information provided by false alarm sources and the background. Firstly, we introduce a False Alarm Source Estimation Block (FEB) that estimates potential interferences present in the background by extracting features at multiple scales and using gradual upsampling for feature fusion. Subsequently, we propose a framework that employs multiple FEBs to eliminate false alarm sources across different scales. Finally, a Target Segmentation Block (TSB) is introduced to accurately segment the targets and produce the final detection result. Experiments conducted on public datasets show that our model achieves the highest and second-highest scores for the IoU, Pd, and AUC and the lowest Fa among the DNN methods. These results demonstrate that our model accurately segments targets while effectively extracting false alarm sources, which can be used for further studies.

Keywords: infrared small target detection; false alarm source; collaborative modeling; clutter suppression; deep learning

Citation: Cao, S.; Li, Z.; Deng, J.; Huang, Y.; Peng, Z. TFCD-Net: Target and False Alarm Collaborative Detection Network for Infrared Imagery. *Remote Sens.* **2024**, *16*, 1758. https://doi.org/10.3390/rs16101758

Academic Editors: Chein-I Chang and Gemine Vivone

Received: 16 January 2024
Revised: 10 May 2024
Accepted: 11 May 2024
Published: 15 May 2024

1. Introduction

Infrared Search and Tracking (IRST) systems play a crucial role in various civilian and military applications [1]. They are widely utilized for tasks like pinpointing heat sources during firefighting operations and detecting abnormalities in medical applications [2,3]. These systems make use of emitted or reflected infrared radiation from objects to accomplish target detection, tracking, and identification [4,5]. Their capacity to identify targets becomes particularly valuable in situations where visual identification is hindered or impractical. Examples include scenarios involving targets with camouflage, distant targets, or challenging weather conditions [6].

Detecting infrared small targets presents a significant challenge in IRST applications. These targets are characterized by their small size (typically less than 9×9 pixels or constituting less than 0.15% of the field of view) and lack of detailed texture and shape information [7]. Consequently, distinguishing small infrared targets from complex backgrounds becomes challenging, as these backgrounds often contain elements (such as complex terrains, man-made structures, and clouds) that reflect solar radiation in patterns resembling the targets [8–10]. As shown in Figure 1, as the complexity of the scenes increases, an increasing number of false alarm sources emerge, leading to reduced saliency of the targets.

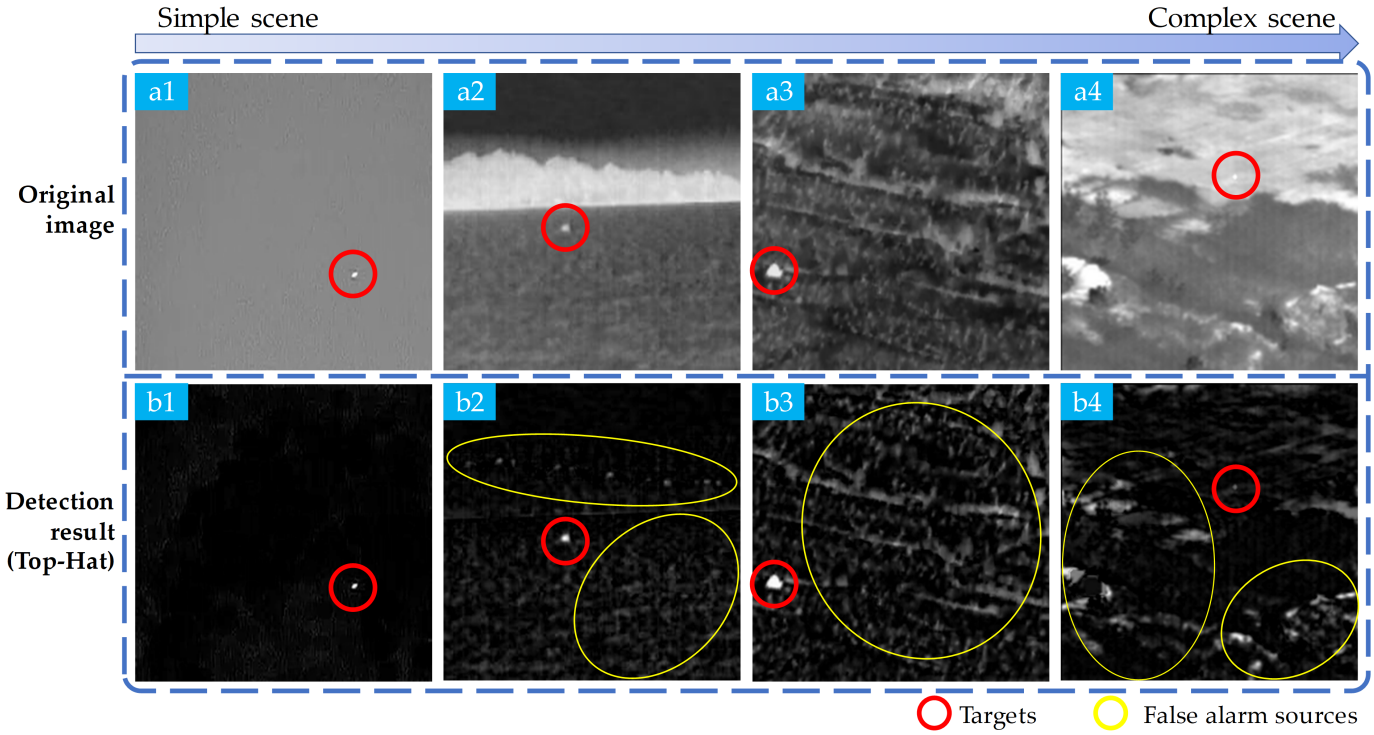

Figure 1. Images (**1**) to (**4**) depict four scenes with backgrounds ranging from simple to complex. Row (**a**) shows the original images; row (**b**) displays the results processed by the Top-Hat algorithm. It is evident that, as the scenes become more complex, an increasing number of false alarm sources appear, resulting in the targets being less salient.

Numerous algorithms have been developed for infrared small target detection (ISTD). These algorithms can be broadly grouped into two categories: multi-frame and single-frame methods [11].

Multi-frame methods detect targets by leveraging the relative motion between targets and background across frames in a sequence [12–14]. They assume a relatively static background and require the accumulation of information from multiple frames to determine the targets' location. The advantage of multi-frame methods is that temporal information is used to enhance detection; however, these methods have some limitations: (1) multi-frame methods naturally do not apply to single-frame scenarios; (2) their effectiveness is hindered when the relative motion assumption is not met; (3) in practical applications, there is a strong demand for fast detection using as few frames as possible. Moreover, multi-frame detection can be approached by detecting targets in each frame using single-frame methods and, then, analyzing the trajectories of the detected targets [15]. Therefore, studying single-frame detection methods is prevalent, and we focus on the single-frame method in this paper.

Single-frame detection methods can be grouped into two categories: non-deep learning and deep learning methods. Non-deep learning methods can be further divided into background-suppression methods, target enhancement methods, image structure-based methods, and classifier-based methods. Background-suppression methods segment target regions by subtracting an estimated background from the input image [16–18]. Target-enhancement methods employ calculations of local contrast and saliency to search for or amplify the target regions [19–24]. Image-structure-based methods assume a mathematical model of low-rank background and sparse targets and solve for the target region through optimization techniques [25–28]. Classifier-based methods are typically combinations of candidate extraction, feature extraction, and feature classification [29,30]. These non-deep learning methods use interpretable mathematical models to model the target and background; however, their modeling heavily relies on prior knowledge, so their effectiveness

is often compromised in complex scenes, as manually designing constraints and features to accurately discern between targets and false alarms can become problematic.

Deep learning methods automate the process of extracting features for targets using deep neural networks (DNNs). Among them, target-detection networks estimate bounding boxes to approximate the location of small targets [31–34], while target-segmentation networks address target detection as a semantic segmentation problem [35–39], aiming to accurately identify the location of small targets at the pixel level. The target-segmentation approach is gaining popularity in research due to its ability to more precisely identify regions containing small targets [7], indicating a promising direction for future progress. Although deep learning methods have achieved impressive progress in ISTD [40–45], there are notable drawbacks. Firstly, existing methods mainly concentrate on modeling the targets, disregarding the modeling of false alarm sources, which also carry useful information, while in practical scenarios, the detection of infrared small targets against densely cluttered backgrounds requires a combined effort to eliminate false alarm sources and detect the targets in order to minimize the false alarm rate. Additionally, the opaque nature of DNNs is a significant limitation, for interpretability is crucial due to the fact that ISTD is a risk-sensitive task.

Taking into account the practical needs and the advantages and drawbacks of the aforementioned methods, the objectives of this research are to employ DNN techniques that make use of information from both targets and false alarm sources to achieve accurate and robust target segmentation, while maintaining interpretability.

To achieve these objectives, we present the Target and False Alarm Collaborative Detection Network for Infrared Imagery (TFCD-Net). Our network addresses the need to incorporate information from false alarm sources by utilizing specialized False Alarm Source Estimation Blocks (FEBs). In order to enhance interpretability, we have designed the overall framework of our network to resemble the background suppression process employed in non-deep learning methods. Specifically, the network first suppresses false alarm sources by subtracting the results of multiple FEBs from the input image and, then, proceeds to segment the targets using a Target Segment Block (TSB).

The major contributions of this research can be summarized as follows:

1. We propose a framework that effectively models both targets with TSB and false alarm sources with FEBs. This approach aims to address the challenges posed by complex and cluttered backgrounds while maintaining interpretability.
2. We introduce a dedicated FEB to estimate potential false alarm sources. By integrating multiple FEBs into the framework, false alarm sources are estimated and eliminated on multiple scales and in a blockwise manner. This block not only enhances the accuracy of our method, but also can serve as a preprocess module to suppress false alarm sources for other ISTD algorithms.
3. Extensive experiments on public datasets validated the effectiveness of our model compared to other state-of-the-art approaches. In addition to accurately detecting targets, our model can produce multi-scale false alarm source estimation results. These estimations are not just incidental outcomes; they can be used to generate false alarm source datasets that can contribute to further studies in the field.

2. Related Works

2.1. Image-Structure-Based ISTD

Numerous non-deep learning algorithms have been proposed for ISTD; among them, image structure-based methods have achieved high performance and provide solid mathematical foundation for ISTD. Image-structure-based methods typically represent infrared images using Equation (1):

$$f_D(x,y) = f_T(x,y) + f_B(x,y) + f_N(x,y), \tag{1}$$

where f_D, f_T, f_B, f_N, and (x,y) denote the original image, the target image, the background image, the noise, and the pixel location, respectively. This representation suggests that

an infrared image can be expressed as a superimposition of three components: target, background, and noise component. Therefore, ISTD is equivalent to separating the target component from the background and noise components.

To separate the target component, image-structure-based approaches assume that the background of an infrared image is low-rank, while the small target is sparse. By formulating constraint equations for the background and target, mathematical optimization techniques are employed to separate the input image into these two components. Then, target areas are extracted by binary segmentation. For example, in NRAM [46], the weighted l_1-norm is used to constrain the target, and the $l_{2,1}$-norm is used for noise; in the PSTNN [47], the partial sum of tensor nuclear norm joint weighted l_1-norm is used to constrain the background in the tensor space; in SRWS [48], a self-regularization item is used to constrain the background, and overlapping edge information is used to constrain the target. These constraint-optimization problems are then solved by applying the alternating direction method of multipliers.

Image-structure-based methods focus on discovering improved mathematical models to represent elements of an infrared image. These mathematical representations offer interpretability for the methods. However, a drawback of image-structure-based methods is that, due to their reliance on manual modeling, their robustness and generalizability are hindered in complex scenes.

2.2. Deep Learning-Based ISTD

In contrast to non-deep learning methods, DNN models are data-driven, and the features and relationships between the input and desirable output are automatically learned during training.

Deep learning-based ISTD methods are developed from visible light image target detection and segmentation networks [49–53]. Efforts include applying UNet [54] with reduced input channels for ISTD. UNet is composed of a progressive downsampling encoder, a progressive feature concatenation, and an upsampling decoder. In ACM [35], an asymmetric contextual modulation module is used to exchange both high-level and low-level information to enhance the encoder–decoder framework. In ALCNet [43], multiscale local contrast blocks are used to extract local contrast feature maps, and bottom-up attentional modulation modules are used to encode low-level features into high-level features. In ISTDUNet [45], merge connections are employed to replace the skip connections between the encoder and decoder to improve UNet. In DNANet [44], dense nested attention blocks are used to extract and fuse features at multiple layers. In RDIAN [40], special convolution layers that resemble the local-contrast-extraction process are designed to reduce the parameter size and achieve fast detection.

Although automatic feature extraction and relationship mapping enable DNN models to achieve high accuracy and generalization, a limitation of existing deep learning methods is that they often focus solely on designing structures to model the target component. However, there has been limited effort in modeling and reducing false alarm sources, which actually contain valuable information. By combining knowledge from image-structure-based methods and making use of false alarm source modeling, the performance and interpretability of deep learning-based ISTD methods can be further improved.

3. Proposed Method

3.1. Overall Framework

The overall framework of the proposed TFCD-Net is depicted in the upper section of Figure 2. It is designed to suppress false alarm sources in a multi-stage process before segmenting the target. The suppression is achieved through the implementation of multiple FEBs, each functioning as a residual block [55], where the input is subtracted by the output, then the result serves as the input for subsequent blocks. Although the number of these suppression stages can be varied, experiments have shown that 2 to 3 stages are usually sufficient to meet the requirements of subsequent high-performance target segmentation. A series connection configuration of the blocks is illustrated in the upper section of Figure 3.

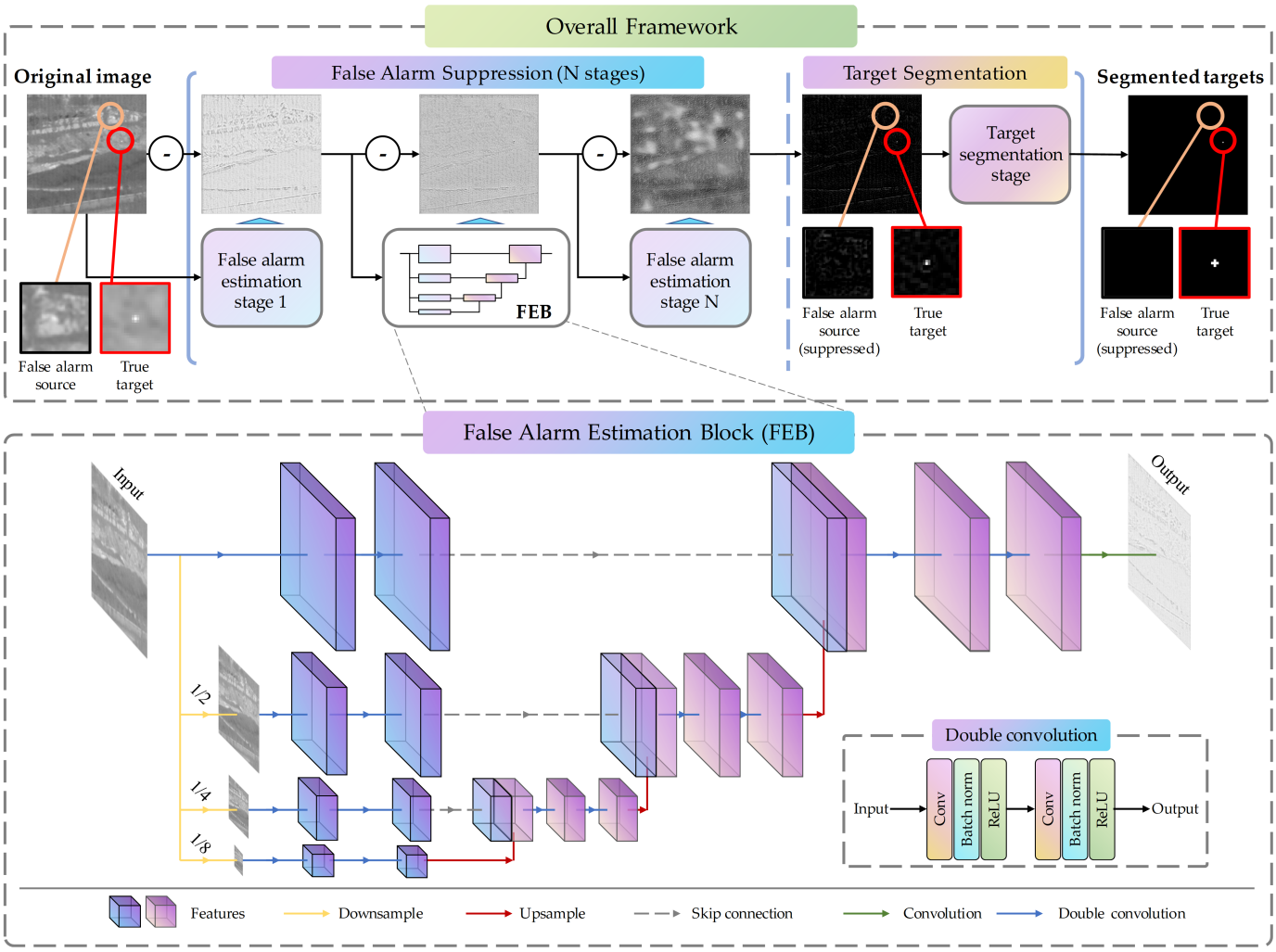

Figure 2. Architecture of the proposed TFCD-Net consisting of two main stages. In the false alarm suppression stage, multiple FEBs are employed to estimate false alarm sources stepwisely. The false alarm-subtracted result is input into the target segmentation stage to detect the targets. The lower section depicts the FEB.

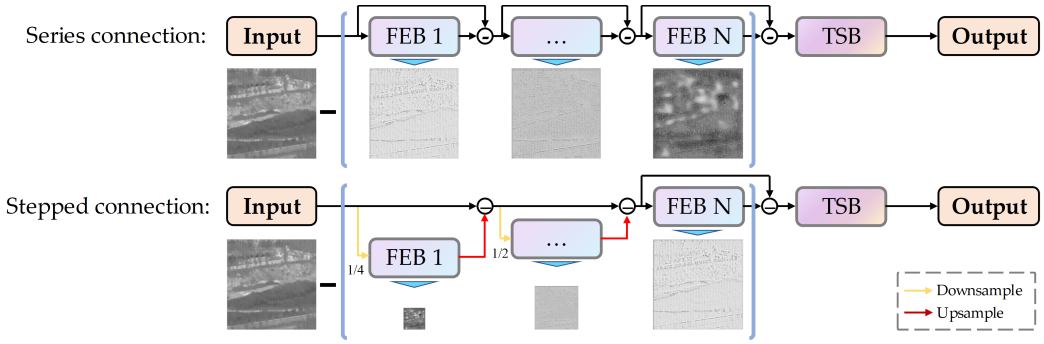

Figure 3. Overall framework utilizing different connection configurations. The upper section shows the series configuration with 3 FEBs, and the lower section shows the stepped configuration with 3 FEBs. The number of FEBs can be adjusted according to the requirements.

We investigated an alternative multiscale, multi-stage structure, as shown in the lower section of Figure 3. Although this configuration was found to be less effective in Section 4.2, we believe that retaining the information could be beneficial for future studies. In this configuration, three FEBs operate at different scales: 1/4, 1/2, and the original scale. The formula is as follows:

$$O_{\mathrm{FEB}_k} = \mathrm{Up}_{2(n-k)\times}^{\mathrm{bilinear}}\left(\mathrm{FEB}\left(\mathrm{MaxPool}_{2(n-k)}(I_{\mathrm{FEB}_k})\right)\right), \tag{2}$$

where $\mathrm{Up}_x^{\mathrm{bilinear}}$ is factor x bilinear upsampling, $\mathrm{MaxPool}_x$ is factor x max pooling, $k = 1, 2, \ldots, n-1$, and n is the number of FEBs. This design strategy is aimed at enabling the network to suppress false alarms progressively from coarse to fine levels. In this framework, the output of each FEB is upsampled back to the original scale before it is subtracted by the output of the previous stage, ensuring that no original image information is lost during the downsampling process.

The role of the FEB is to suppress false alarm sources, and it does not precisely extract the shape of the targets. To accurately segment targets from the false alarm-source-suppressed result, a semantic segmentation network is employed as the segmentation head. While there are various choices for the segmentation head, we utilized the same architecture as the FEB for the TSB, but the output of the block is processed through a sigmoid activation function to ensure the output pixel values are confined between 0 and 1. Unlike the TSB outputs, FEBs do not restrict their outputs with an activation function. Instead, false alarm subtraction is performed using a linear rectification function (ReLU) [56] as the activation function to constrain negative values. The overall process is shown in Equation (3):

$$\begin{cases} O_1 = \mathrm{ReLU}(\mathrm{FEB}_1(I_1)) \\ O_k = \mathrm{ReLU}(O_{k-1} - \mathrm{FEB}_k(O_{k-1})), \quad k = 2, 3, \ldots, n \\ O_{\mathrm{final}} = \mathrm{Sigmoid}(O_n - \mathrm{TSB}(O_n)) \end{cases} \tag{3}$$

where I_1 is the original image, O_k is the output of the k-th stage, FEB_k is the k-th FEB with the aforementioned downsampling and upsampling if in stepped connection as in Equation (2), and O_{final} is the final target segmentation result.

3.2. False Alarm Source Estimation Block

The FEB is an important component of the proposed TFCD-Net framework, designed to accurately estimate false alarm sources. As shown in Equation (1), the estimation of false alarm sources involves the subtraction of the target component from the background. Given the small size of infrared small targets, this estimation process is comparable to image denoising, which aims at removing small objects from the image.

Typically, denoising networks use stacked convolutional layers or encoder–decoder architectures similar to UNet [57,58]. The multiscale structures are beneficial as they capture background patterns in a wider area, making the estimation more complete. In consideration of these studies, we introduce a multiscale architecture for modeling false alarm sources, as shown in the lower section of Figure 2.

Within the FEB, the original image is first downsampled by $1/2$, $1/4$, and $1/8$, resulting in a total of 4 scales of input images as follows.

$$O_{\mathrm{pool,k}} = \mathrm{MaxPool}_{2(k-1)}(I_{\mathrm{FEB}}), \tag{4}$$

where $O_{\mathrm{pool,k}}$ is the downsampled image of the k-th scale, $k = 2, 3, \ldots, n$. I_{FEB} is the input of the FEB. $\mathrm{MaxPool}_{2(k-1)}$ performs max pooling of factor $2(k-1)$.

Subsequently, two double-convolution operations are performed on each scale, forming the encoder component of the FEB. The double-convolution block is depicted in the bottom-right corner of Figure 2, consisting of two sets of convolutional layers, batch normalization layers, and ReLU activations. The formula is as follows:

$$O_{dc} = \mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv}_{3\times3}(\mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv}_{3\times3}(I_{dc})))))), \tag{5}$$

where O_{dc} and I_{dc} are the output and input of the double-convolution block.

The decoder component of the FEB uses a UNet-like decoder structure for progressive upsampling and feature fusion: for each scale, upsampling by a factor of 2 is initially performed, followed by feature concatenation with the previous level. The formula is as follows:

$$I_{\text{de}} = O_k \oplus \text{Up}_{2\times}^{\text{bilinear}}(O_{k+1}), \tag{6}$$

where O_k and O_{k+1} are the k- and $(k+1)$-th scale, $k = 1, 2, \ldots, n-1$. $\oplus$ represents the concatenation operation. $\text{Up}_{2\times}^{\text{bilinear}}$ indicates bilinear upsampling by a factor of 2.

The concatenated feature is then forwarded through two double-convolution operations. This upward feature fusion continues until the original image size is reached, and the output image is obtained through a 1-channel 1×1 convolution operation to produce a single-channel output. The FEB uses max poolings for downsampling, bilateral filters for upsampling, and 64-channel 3×3 convolution kernels for all double-convolution operations. There is no activation function used after the final convolution layer, as activations between blocks are placed in false alarm subtractions, as in Equation (3).

3.3. Loss Function

In the field of ISTD, the commonly used loss function is the soft Intersection over Union (IoU) loss function [59], which is a loss function based on the IoU. The IoU is a common evaluation metric for image-segmentation tasks, used to measure the overlap between predicted and ground truth segmentations, defined as the ratio of the intersection of the prediction and the ground truth to their union. However, since the IoU is not differentiable, it is hard to use directly as a loss function in the training process of deep learning models. Instead, a variant of the IoU, the soft IoU, is used to design the loss function. The soft IoU is a differentiable approximation of the IoU and is defined in Equation (7):

$$L_{SI}(P, Y) = 1 - \frac{\sum_{i=1}^{N} P_i \cdot Y_i + \varepsilon}{\sum_{i=1}^{N} P_i + \sum_{i=1}^{N} Y_i - \sum_{i=1}^{N} P_i \cdot Y_i + \varepsilon}, \tag{7}$$

where P and Y, respectively, denote the predicted output of the network and the ground truth, ε represents the smoothing factor, and N indicates the total number of samples. The incorporation of the smoothing factor is a common strategy to prevent the denominator of the loss function, which includes division, from becoming zero. To ensure that the inclusion of the smoothing factor does not significantly affect the actual loss value, ε is typically chosen to be very small.

The soft IoU loss provides several advantages in infrared small target segmentation. Unlike the mean-squared error (MSE) and cross-entropy (CE) losses, which calculate the loss function for each pixel independently, the IoU-based loss takes into account the spatial arrangement of pixels. This approach places greater emphasis on accurately locating targets. Moreover, in tasks involving the segmentation of small targets in infrared images, the number of background pixels greatly outweighs the number of target pixels, resulting in a class imbalance. By utilizing IoU-based calculations, the influence of this imbalance is reduced, thereby enhancing the network's IoU performance.

However, in practical applications, the soft IoU loss function also has limitations, such as maximizing the IoU may not necessarily result in clear edges. To address unclear edges, a weighted binary cross-entropy (WBCE) loss function designed for single-class small target segmentation is formulated as in Equation (8):

$$L_{WBCE}(P, Y, W) = -\frac{1}{N} \sum_{i=1}^{N} W_i [Y_i \cdot \log(P_i) + (1 - Y_i) \cdot \log(1 - P_i)], \tag{8}$$

where P and Y, respectively, denote the predicted output of the network and the ground truth, W represents the weight map, and N indicates the total number of samples. The weights W are applied on a per-pixel basis to the binary cross-entropy (BCE) loss, an approach that serves to emphasize specific regions. The weights are designed as Equation (9):

$$W = D(C(Y), E), \tag{9}$$

where C represents the Canny edge detection operator, D denotes the morphological dilation operation, and E is a square structural element measuring 2×2.

The total loss function of this method is as given in Equation (10):

$$L(P, Y, W) = L_{SI}(P, Y) + \alpha L_{WBCE}(P, Y, W), \tag{10}$$

where α is the weight coefficient.

The idea behind the loss function design is to use the soft IoU function as the core and improve the performance of the loss function through edge loss to achieve stable segmentation of small targets.

4. Experiments

4.1. Settings

In the experiments, two public datasets were used: the NUAA-SIRST dataset [35] and the NUDT-SIRST dataset [44]. The NUAA-SIRST dataset consists of 427 images; the NUDT-SIRST dataset contains 1327 images. These images represent various common infrared scenes, such as clouds, sea surfaces, urban environments, and ground scenes. They are relevant for both terrestrial and aerial ISTD tasks. The resolution of all images in the datasets is 256 × 256 pixels. We evenly divided the datasets into training and testing sets, with a split ratio of 50%.

The performance of the algorithms and networks was evaluated using various metrics. The key evaluation metrics used were the probability of detection (Pd), false alarm rate (Fa), intersection over union (IoU), and receiver operating characteristic (ROC). The definitions of these metrics are as follows.

Probability of detection (Pd): Measures the ability of detecting true targets on the target level. Defined as the ratio of the number of correctly detected targets over true targets as follows:

$$Pd = \frac{T_{correct}}{T_{All}}, \tag{11}$$

where $T_{correct}$ is the number of correctly detected targets and T_{All} is the number of true targets. A higher score indicates better detection capability.

False alarm rate (Fa): Measures the rate of detecting false targets on the pixel level. Defined as follows:

$$Fa = \frac{FP}{FP + TN}, \tag{12}$$

where FP and TN are pixel-level false positive and true negative detections, respectively. A lower score indicates fewer false positive detections.

Intersection over union (IoU): Measures the accuracy of detection on the pixel level. Defined as the area of overlap between the predicted and the ground truth targets divided by the area of their union as follows:

$$IoU = \frac{A_I}{A_U}, \tag{13}$$

where A_I is the area of intersection and A_U is the area of union. Area is calculated as the number of pixels. A higher score indicates higher accuracy segmenting the targets.

F1-score: Measures the balanced score considering both precision and recall as follows:

$$\begin{cases} Precision = TP/(TP + FP) \\ Recall = TP/(TP + FN) \\ F1 = 2(Precision \cdot Recall)/(Precision + Recall) \end{cases} \tag{14}$$

where TP, FP, and FN are the pixel-level true positive, false positive, and false negative detections, respectively. A higher score indicates higher overall detection performance.

Receiver operating characteristic (ROC): Measures the robustness of detection. Defined as the curve of the false positive rate (FPR) to the true positive rate (TPR). To improve the effectiveness of evaluation, a series of 3D ROC curve-derived evaluation metrics proposed

in [60,61] were used. Among the metrics, higher area under the curve (AUC) scores (except the lower score for $AUC_{F,\tau}$) indicate better robustness.

For the model setup, the network structures displayed in Figure 4 were employed in the ablation study; structure A3 in Figure 4 was employed in the comparative experiments. Regarding the loss function, the coefficient α was set to 0 during the initial 50 training epochs and was adjusted to 0.2 thereafter. The optimization algorithm used was Adaptive Moment Estimation (ADAM). The initial learning rate was set at 0.001 and scheduled to decay by a factor of 0.1 every 50 epochs. The training regimen spanned 200 epochs with a batch size of 8.

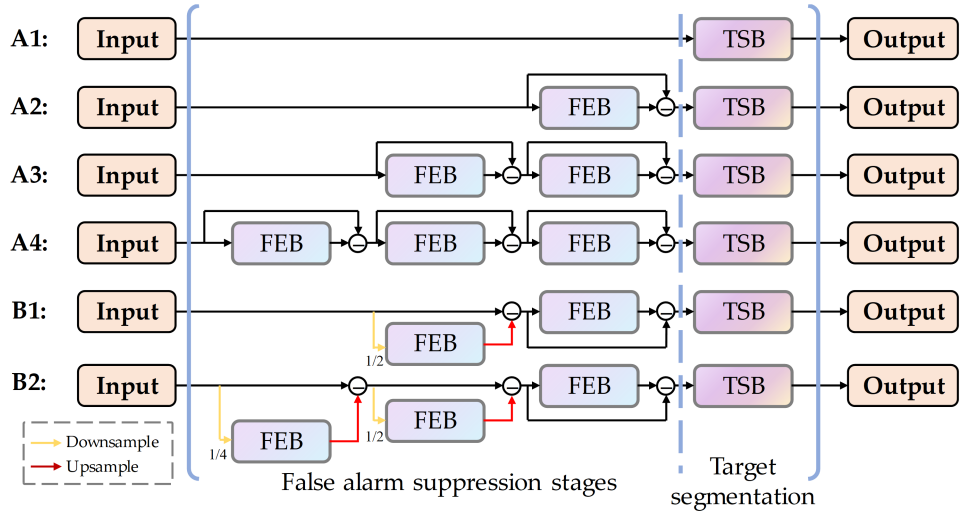

Figure 4. Structural configurations of overall framework used in ablation experiments.

Illustrative samples from the datasets are showcased in Figures 5 and 6, providing visual context to the types of infrared images used in the experiments.

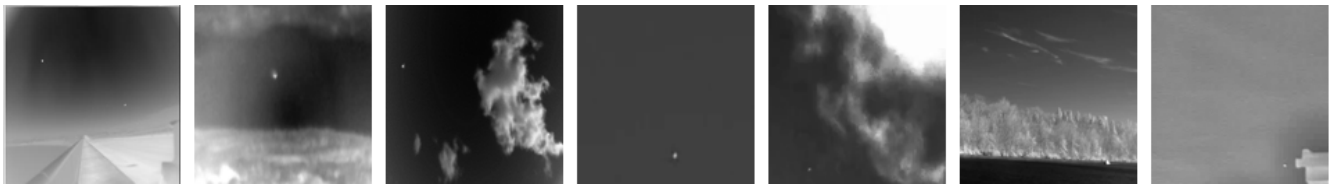

Figure 5. Sample images of NUAA-SIRST dataset.

Figure 6. Sample images of NUDT-SIRST dataset.

4.2. Ablation Experiments

The overall framework of the network was subjected to ablation testing, which included evaluating six different structural configurations, as shown in Figure 4. Type-A networks are represented in a serial connection, composed of 0, 1, 2, or 3 FEBs coupled with a single TSB. Type-B networks feature a stair-step connection consisting of either 2 or 3 FEBs and a single TSB. The intent behind these experiments was to identify the optimal number of stages for background suppression and to validate the effectiveness of the stair-step connection approach.

The six aforementioned network structures were trained and tested on two datasets: NUAA-SIRST and NUDT-SIRST. The outcomes of these trials are presented in Table 1.

Table 1. Ablation results of different structures on 2 datasets.

Structure	NUAA-SIRST			NUDT-SIRST		
	IoU $(\times 10^{-2})$	Pd $(\times 10^{-2})$	Fa $(\times 10^{-6})$	IoU $(\times 10^{-2})$	Pd $(\times 10^{-2})$	Fa $(\times 10^{-6})$
A1	65.49	92.02	88.9	78.23	90.79	41.09
A2	67.13	94.3	108.94	91.19	97.04	18.67
A3	68.65	94.30	103.52	92.66	97.46	16.63
A4	66.95	93.92	104.45	92.43	97.88	17.61
B1	62.06	91.63	120.15	75.11	93.76	68.85
B2	58.86	95.06	148.97	54.34	84.34	149.49

The results outlined in Table 1 indicate that employing a stair-step connection (Type-B networks) does not enhance network performance. Contrarily, an increase in the number of stages correlated with a decline in the performance metrics, a trend that was particularly pronounced with the NUDT-SIRST dataset. This dataset typically features smaller targets, and it is hypothesized that the Type-B network's upsampling and downsampling procedures might introduce disturbances to the edge characteristics of these targets. Though Type-B structures have shown less effectiveness, we consider that retaining the records in this section could be a reference and beneficial for future studies.

For Network Type-A, incorporating two FEBs resulted in the best performance. Increasing the number of blocks beyond that did not lead to further enhancements in network capacity. Therefore, based on these findings, we chose the network structure A3 for comparative testing in this section.

4.3. False Alarm Source Estimation Capability

Experiments were carried out to validate the ability of the proposed model to estimate false alarm sources. To conduct a detailed analysis, the A4 framework was selected based on its performance in the ablation experiments. The various stages of the model's input and output, depicted in Figure 7, were examined. These stages include the network input, which corresponds to the original image (S1in), and the network output (S4out).

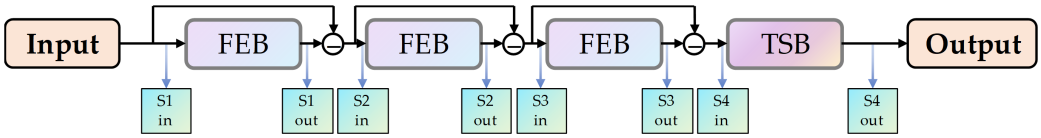

Figure 7. The sampling points for each block, indicated by blue arrows.

The experiments were performed on multiple scenes, with results depicted in Figures 8–11. Figures 8 and 10 show two dense clutter scenes from the NUDT-SIRST dataset. In these figures, the first row demonstrates the inputs of each stage, while the second row shows the corresponding outputs, with red circles indicating the targets. The 3D visualization of these scenes are, respectively, presented in Figures 9 and 11.

From Figures 8 and 10, it is evident that the false alarm sources were progressively suppressed from S1in to S4in; meanwhile, the target regions were preserved. This was achieved by the gradual suppression of non-target edge areas, making targets more salient and easier for the final TSB to process. The outputs from S1out to S3out correspond to the outputs from the FEBs. It can be observed that the suppression of false alarm sources progresses from finer to coarser details and from higher to lower frequencies. This suggests that each stage of the network estimates and suppresses the most salient non-target edge regions. By S3 stage, where most high-frequency regions have already been suppressed, the FEB begins estimating the low-frequency fluctuations in the background. It should be noted that the FEBs do not suppress the target area at any stage, as is evident from S1out, S2out, and S3out, thus ensuring the preservation and increasing saliency of the target. This characteristic is particularly evident in the images of S2in, S3in, and S4in depicted in Figure 9.

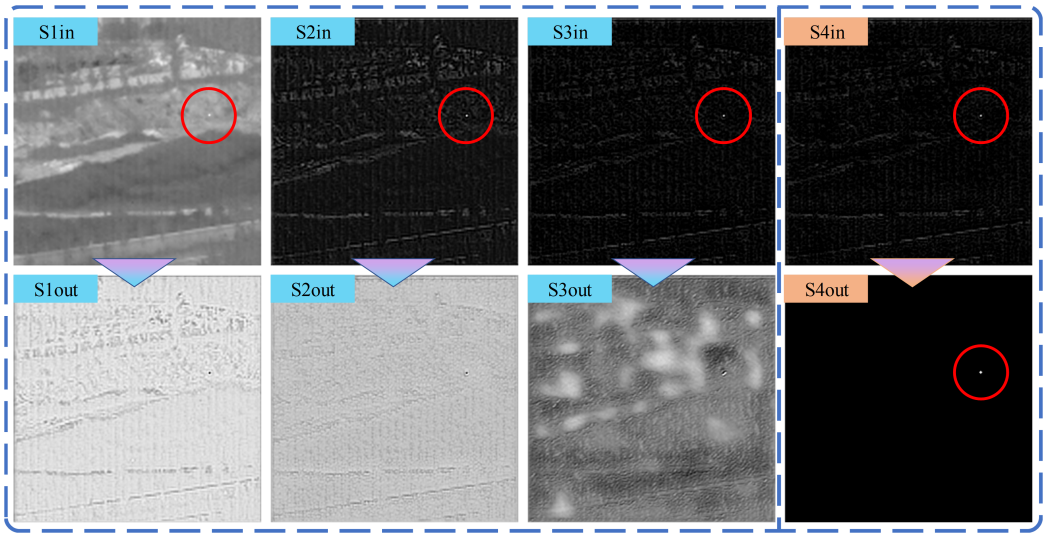

Figure 8. Inputs and outputs of each block for scene 1.

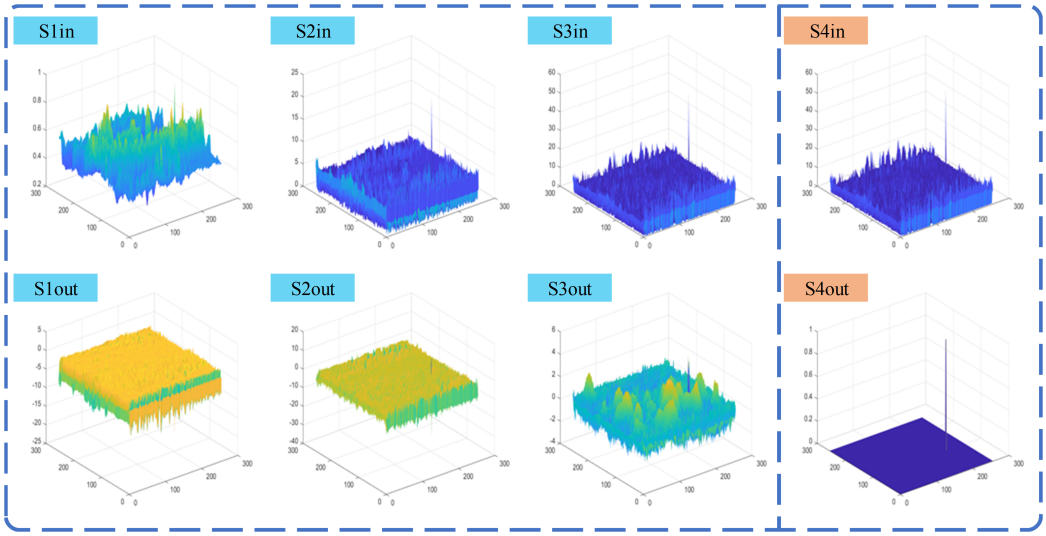

Figure 9. Three–dimensional visualization of inputs and outputs of each block for scene 1.

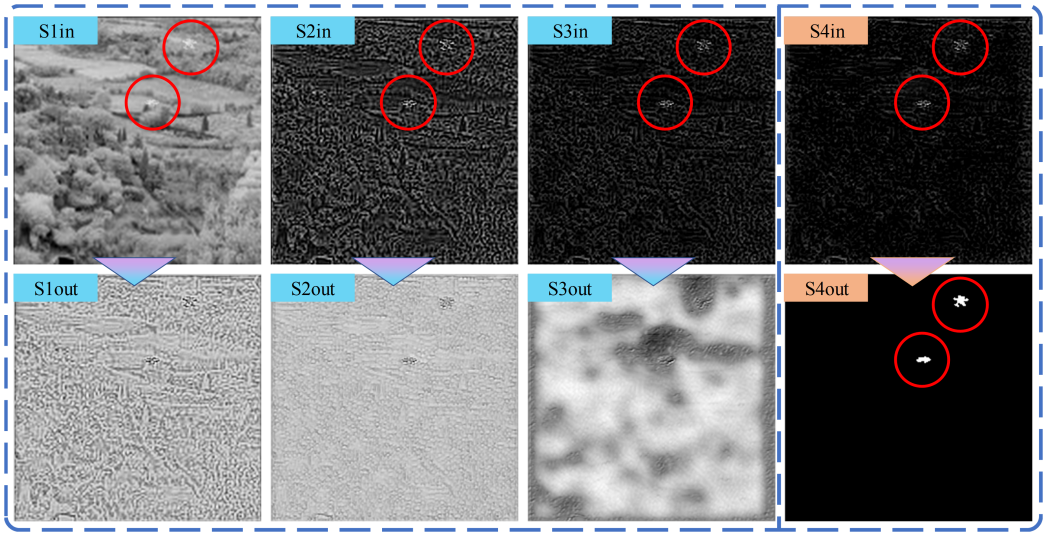

Figure 10. Inputs and outputs of each block for scene 2.

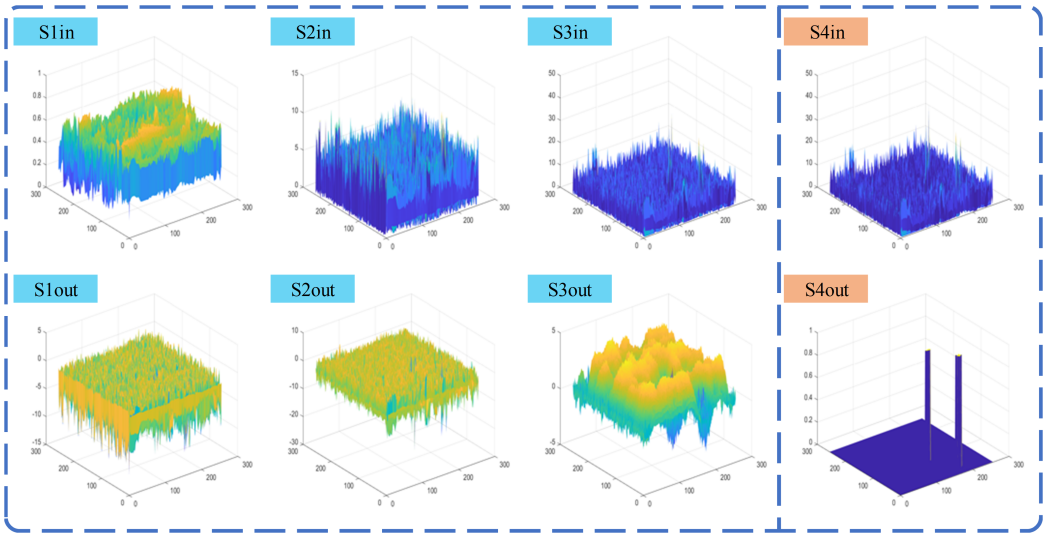

Figure 11. Three–dimensional visualization of inputs and outputs of each block for scene 2.

4.4. Comparative Experiments

The proposed TFCD-Net was evaluated and compared with state-of-the-art methods, including the optimization-based methods NRAM [46], the PSTNN [47], and SRWS [48], as well as the deep learning methods ACM [35], ALCNet [43], RDIAN [40], UNet [54], ISTDUNet [45], and DNANet [44].

For the non-deep learning methods, the original parameters from their respective publications were employed. The DNN models were trained using the ADAM optimizer with a batch size of 8, an initial learning rate of 0.001 for 200 epochs, and a learning rate decay by a factor of 0.1 every 50 epochs. The soft IoU loss was used for training the DNN models.

The performance of the proposed TFCD-Net and the comparison methods is demonstrated on six representative scenes from the NUAA-SIRST and NUDT-SIRST datasets. The output results and 3D visualizations for these scenes are depicted in Figures 12–17. It can be observed that, in scenes 1 to 5, the three non-deep learning methods failed to detect some targets, highlighting the insufficient stability of manually modeling in complex environments. This was particularly evident when the assumptions of sparsity for targets and low rank for backgrounds were not met, such as in heavily cluttered scenes and scenes with larger targets. Figure 14 showcases a complex scene where most comparison methods failed to detect the target. While the ISTDUNet and DNANet models detected the target, they lacked precision in segmentation. In contrast, the proposed model was capable of accurately segmenting the target. In Figure 16, larger targets posed a challenge for models like ACM and ALCNet, which produced imprecise contours.

In Figure 17, RDIAN, UNet, ISTDUNet, and DNANet segmented the single target into multiple parts, potentially affecting precise localization and subsequent identification in practical applications; for instance, UNet's detection was more than three pixels away from the true center of the target, which is significant given the small size of the targets. The proposed TFCD-Net performed well across all six scenes, especially on the more complex NUDT-SIRST dataset, as shown in Figures 12–14.

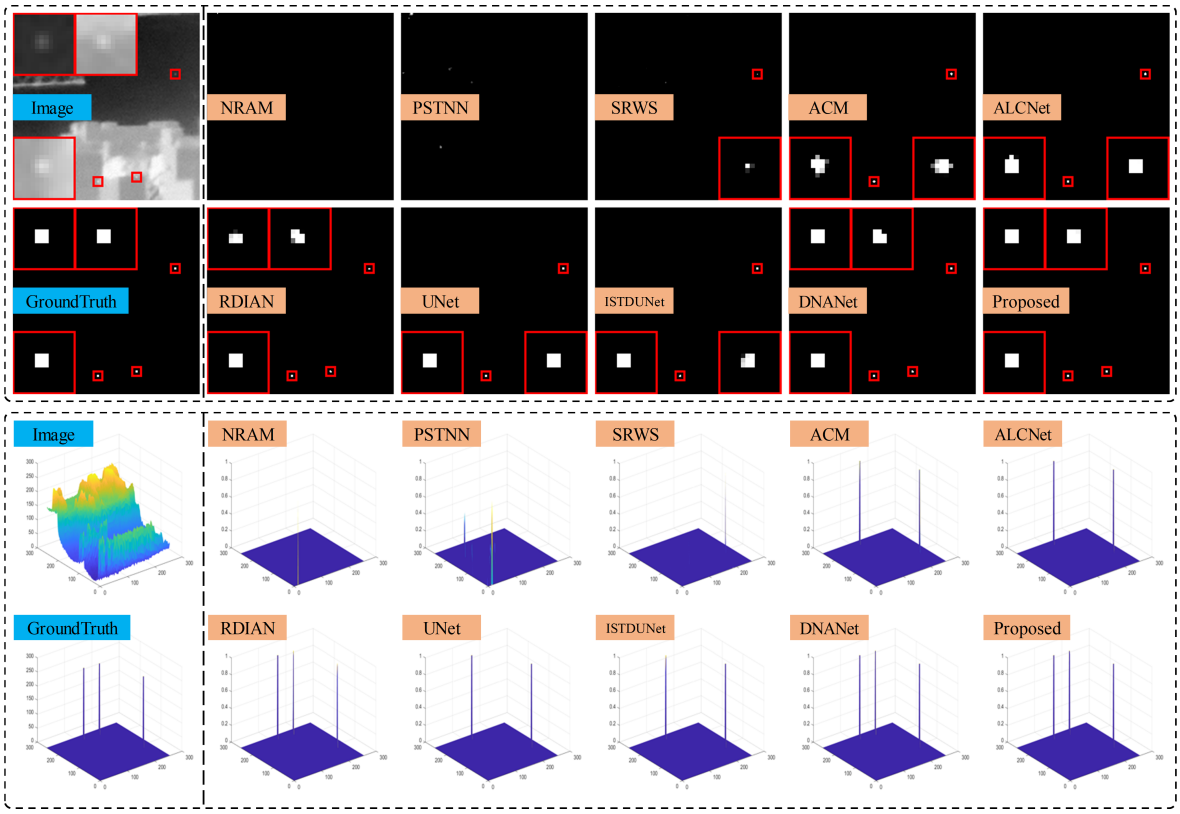

Figure 12. Results and 3D visualizations of different methods for scene 1. The red boxes identify the target area and zoom in for display.

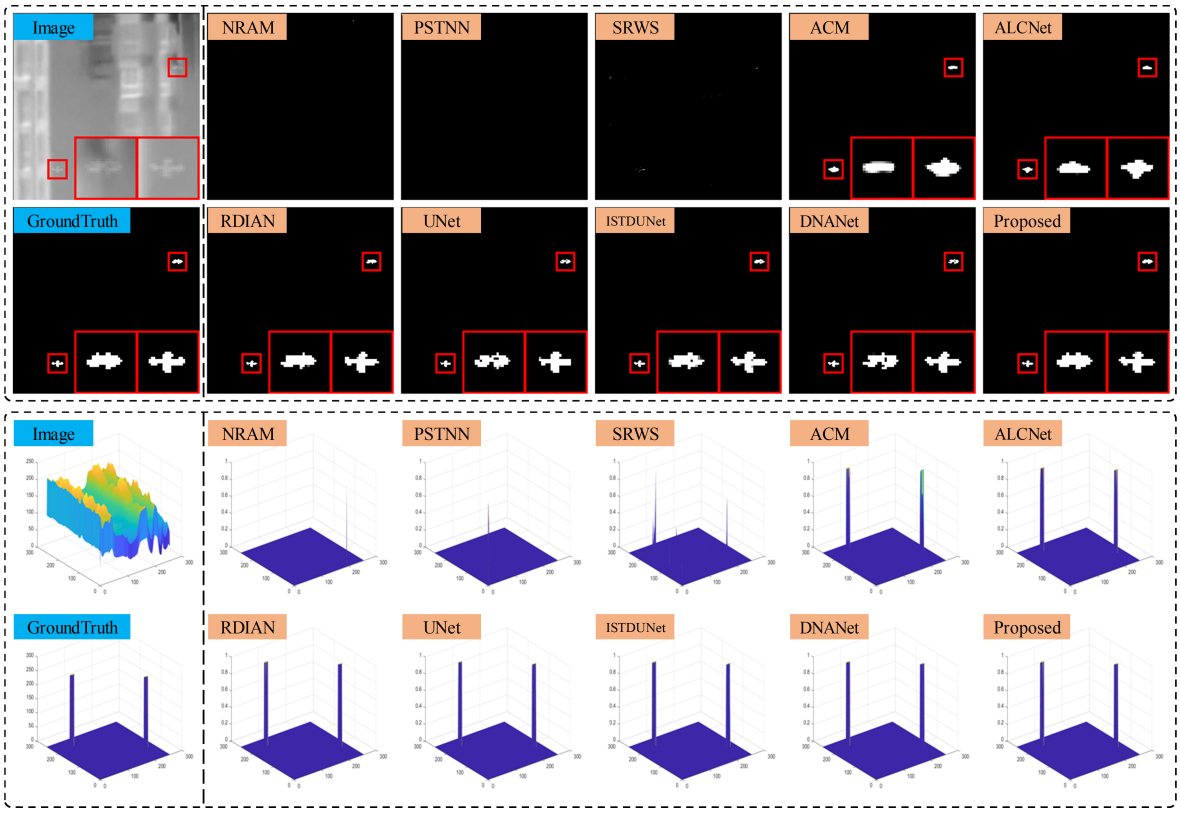

Figure 13. Results and 3D visualizations of different methods for scene 2. The red boxes identify the target area and zoom in for display.

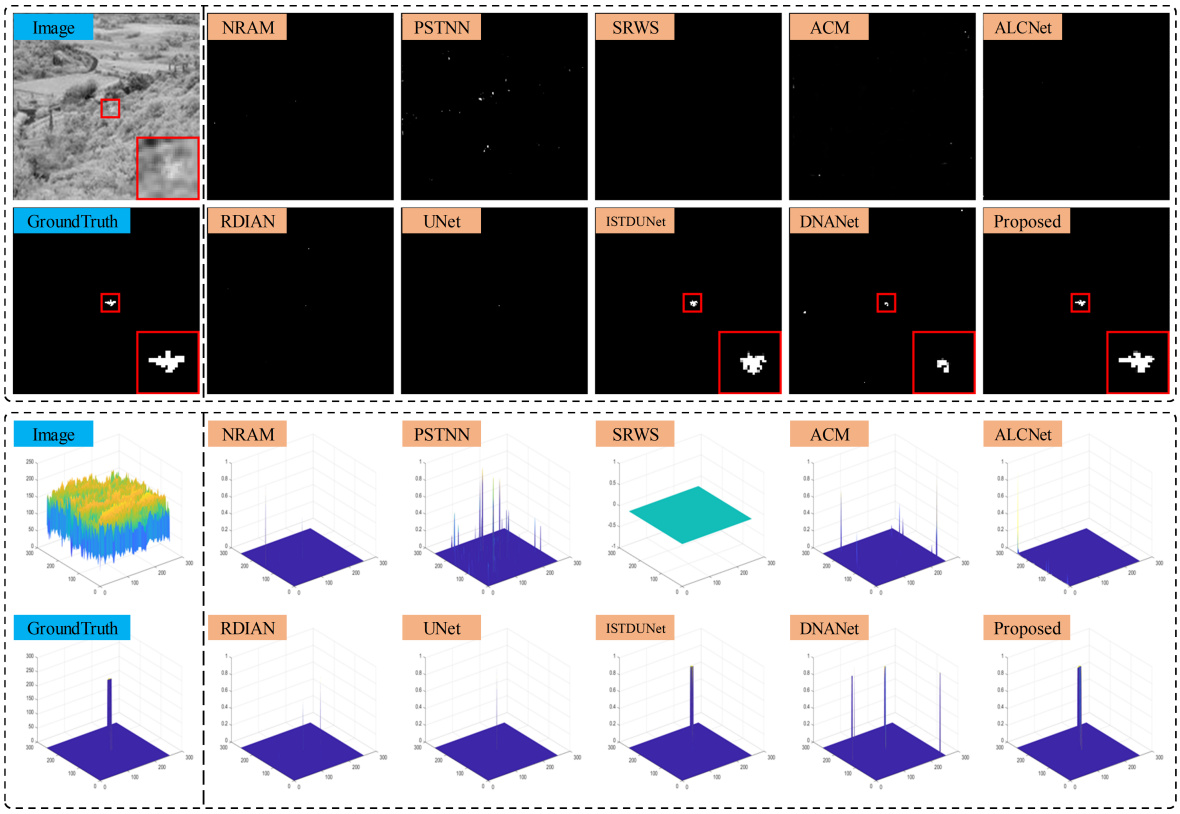

Figure 14. Results and 3D visualizations of different methods for scene 3. The red boxes identify the target area and zoom in for display.

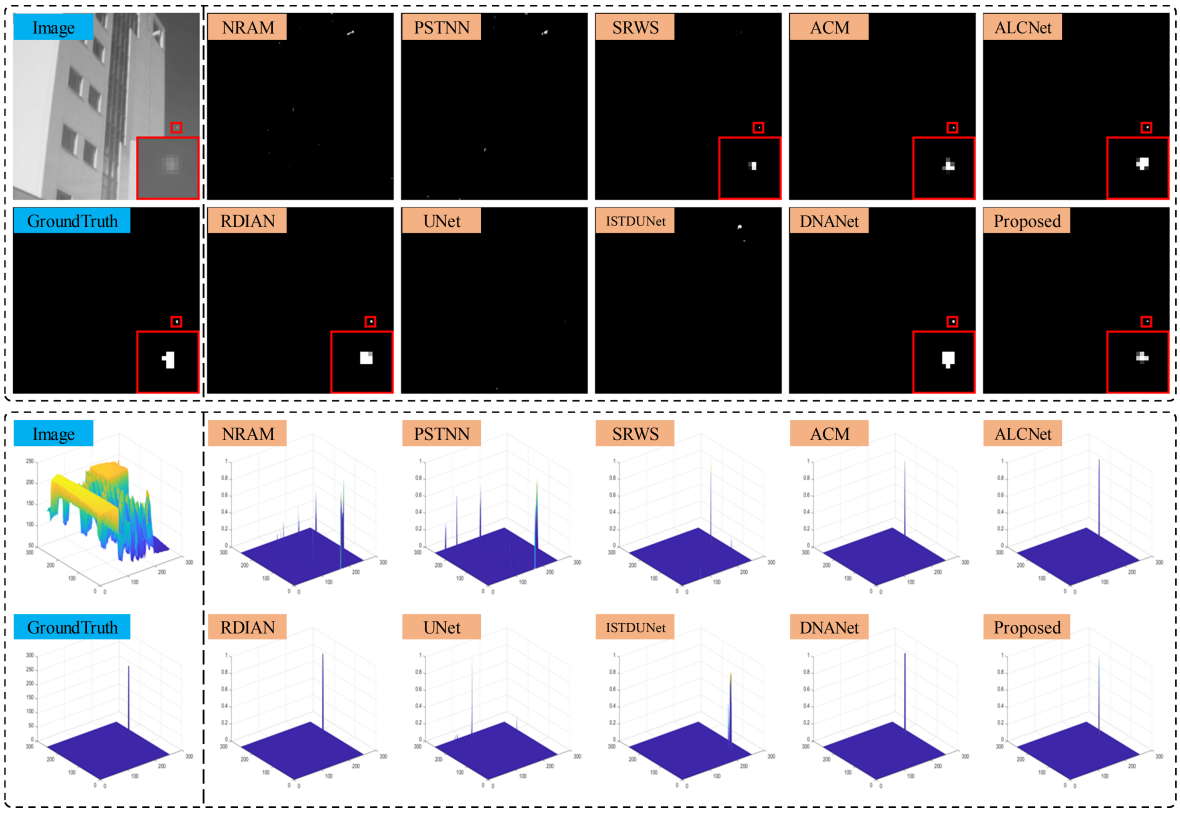

Figure 15. Results and 3D visualizations of different methods for scene 4. The red boxes identify the target area and zoom in for display.

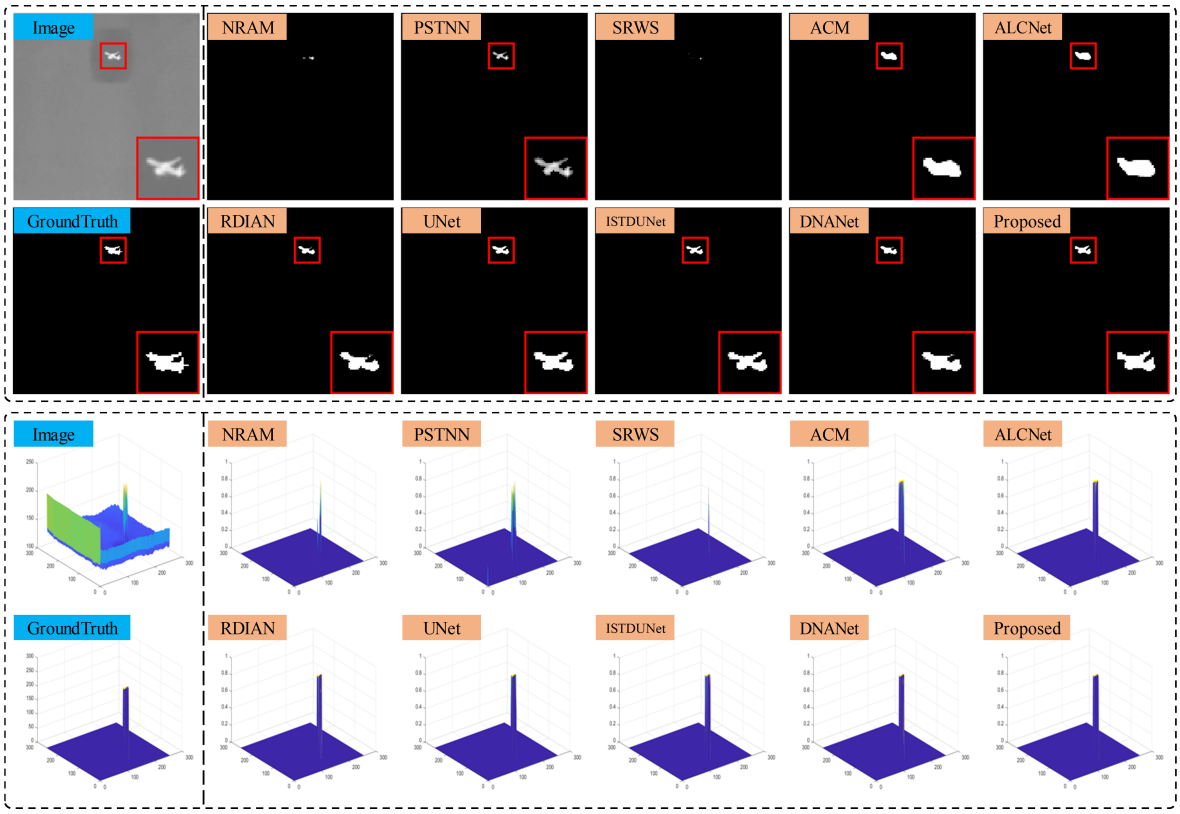

Figure 16. Results and 3D visualizations of different methods for scene 5. The red boxes identify the target area and zoom in for display.

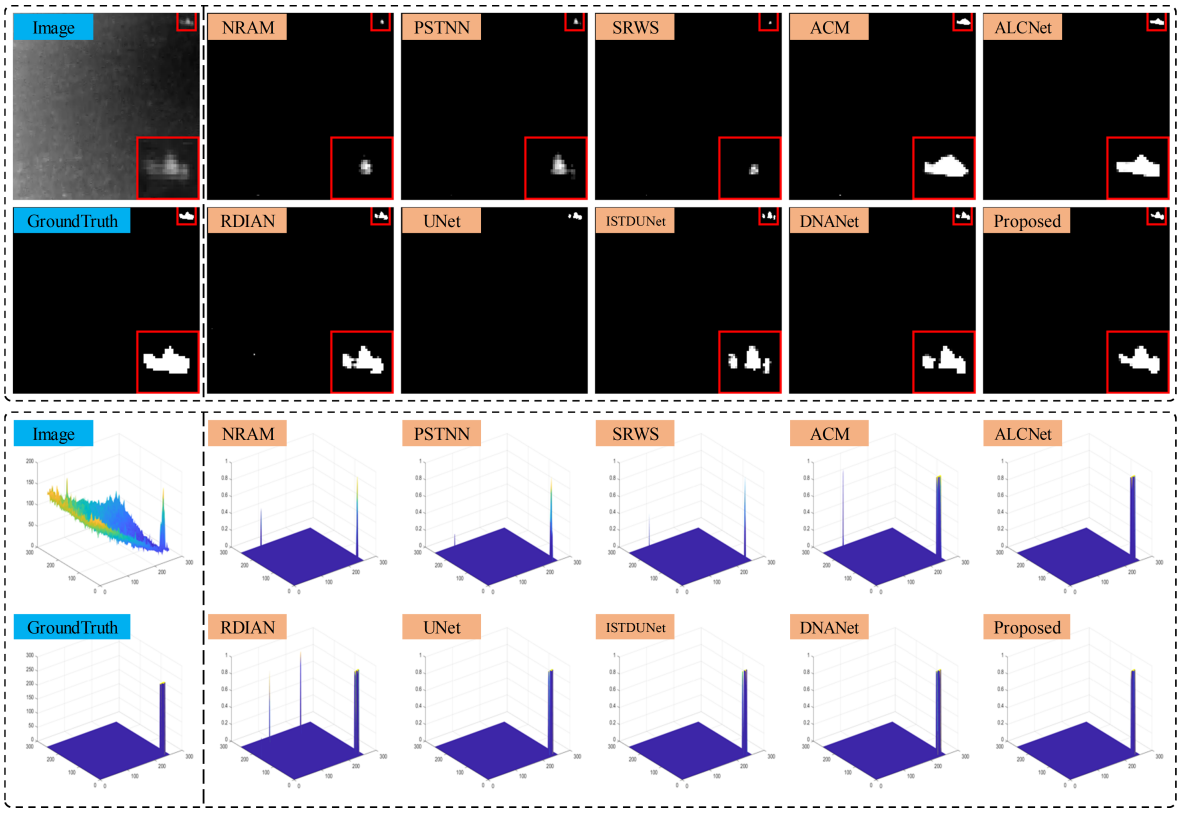

Figure 17. Results and 3D visualizations of different methods for scene 6. The red boxes identify the target area and zoom in for display.

For a quantitative assessment, Tables 2 and 3 present the comparative test results of the proposed TFCD-Net and other methods on the NUAA-SIRST and NUDT-SIRST datasets.

Table 2. Results achieved on NUDT-SIRST dataset.

Method	IoU ($\times 10^{-2}$)	Pd ($\times 10^{-2}$)	Fa ($\times 10^{-6}$)	$F1$ ($\times 10^{-2}$)
NRAM	11.4	58.52	23.45	20.47
PSTNN	21.69	68.04	216.64	35.65
SRWS	8.69	66.35	9.27	15.99
ACM	67.65	95.77	138.66	80.71
ALCNet	69.93	94.92	118.31	82.31
RDIAN	86.93	97.25	34.19	93.01
UNet	89.84	96.4	19.89	94.65
ISTDUNet	89.73	97.88	29.76	94.58
DNANet	91.63	97.46	22.74	95.63
Proposed	92.66	97.99	17.01	96.19

Table 3. Results achieved on NUAA-SIRST dataset.

Method	IoU ($\times 10^{-2}$)	Pd ($\times 10^{-2}$)	Fa ($\times 10^{-6}$)	$F1$ ($\times 10^{-2}$)
NRAM	26.17	81.75	10.27	41.49
PSTNN	41.69	84.79	56.51	58.84
SRWS	12.36	84.79	4.00	22.00
ACM	63.49	92.78	113.08	77.67
ALCNet	64.52	93.54	117.79	78.43
RDIAN	70.46	93.54	95.89	82.67
UNet	68.28	93.16	98.39	81.15
ISTDUNet	66.66	92.78	104.24	79.99
DNANet	69.23	93.16	104.38	81.86
Proposed	69.38	93.16	91.82	81.92

From Tables 2 and 3, it can be observed that, for the Pd value, the proposed TFCD-Net achieved the highest score on the NUDT-SIRST dataset, and the second-best on the NUAA-SIRST dataset, comparable to the performance of DNANet, confirming its effectiveness in detecting small targets alongside the leading methods.

For the IoU and $F1$ scores, the proposed method achieved the highest on the NUDT-SIRST dataset and the second-best on the NUAA-SIRST dataset, second to RDIAN, which demonstrates a slightly more precise target segmentation. The RDIAN model, with its MPCM-inspired convolutional kernel design, showed limited generalizability on the NUDT-SIRST dataset, where the proposed TFCD-Net maintained its high performance.

For the Fa score, the proposed method outperformed all other deep learning methods on both datasets, demonstrating its superior capability in suppressing false alarms. It is important to note that the SRWS algorithm, while not a DNN-based approach, showed a significantly lower Fa score, which may be attributed to its lower IoU and Pd values, suggesting a tendency to output smaller target areas, which can also be inferred from the results in Figures 12–17.

The robustness of the methods was analyzed by plotting the 3D ROC-derived curves shown in Figures 18 and 19 and calculating the AUC values as presented in Tables 4 and 5. The proposed TFCD-Net achieved the best score on the NUDT-SIRST dataset and showed competitive AUC scores among DNN approaches on the NUAA-SIRST dataset. A low score for $AUC_{F,\tau}$ and a high score for other AUC metrics suggest effective background suppression and strong target responses, which indicate the robustness of the proposed TFCD-Net.

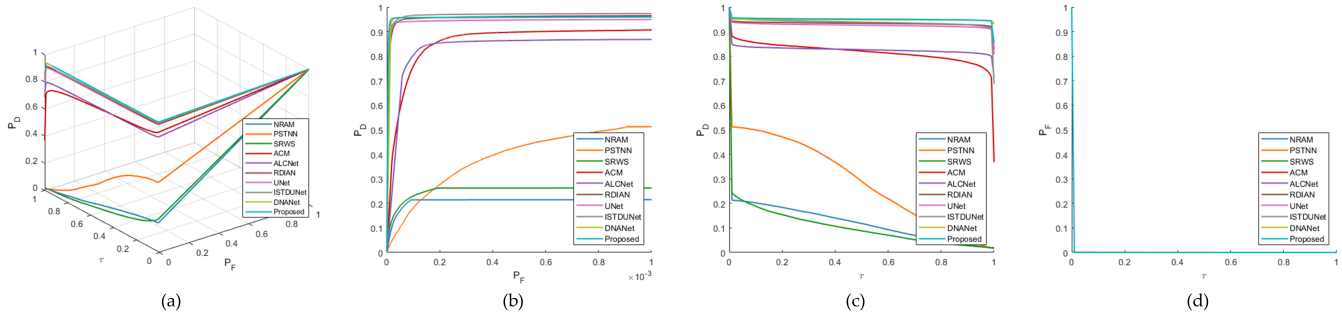

Figure 18. 3D ROC curves with corresponding 2D ROC curves of different methods on NUDT-SIRST dataset. (**a**) 3D ROC curves. (**b**) 2D ROC curves of (P_D, P_F). (**c**) 2D ROC curves of (P_D, τ). (**d**) 2D ROC curves of (P_F, τ).

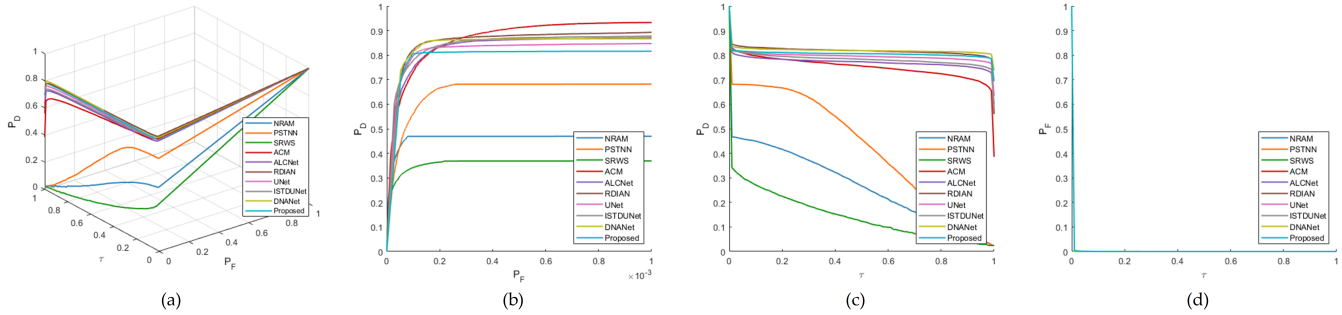

Figure 19. 3D ROC curves with corresponding two-dimensional ROC curves of different methods on NUAA-SIRST dataset. (**a**) 3D ROC curves. (**b**) 2D ROC curves of (P_D, P_F). (**c**) 2D ROC curves of (P_D, τ). (**d**) 2D ROC curves of (P_F, τ).

Table 4. Three-dimensional ROC-derived AUC results achieved on NUDT-SIRST dataset.

Method	$AUC_{D,F}$	$AUC_{D,\tau}$	$AUC_{F,\tau}$	AUC_{TD}	AUC_{BS}	AUC_{SNPR}	AUC_{TDBS}	AUC_{ODP}
NRAM	0.6067	0.1209	0.0050	0.7276	0.6118	24.0264	0.1159	0.7226
PSTNN	0.7556	0.2860	0.0053	1.0416	0.7609	53.7055	0.2807	1.0362
SRWS	0.6210	0.1019	0.0050	0.7229	0.6261	20.2999	0.0969	0.7179
ACM	0.9406	0.8161	0.0051	1.7566	0.9457	158.6904	0.8109	1.7515
ALCNet	0.9233	0.8261	0.0051	1.7494	0.9285	161.3891	0.8210	1.7443
RDIAN	0.9713	0.9331	0.0050	1.9045	0.9764	185.3606	0.9281	1.8994
UNet	0.9691	0.9255	0.0050	1.8946	0.9741	184.3357	0.9205	1.8896
ISTDUNet	0.9778	0.9369	0.0050	1.9148	0.9828	186.2700	0.9319	1.9097
DNANet	0.9757	0.9483	0.0050	1.9240	0.9807	188.8106	0.9433	1.9190
Proposed	0.9782	0.9504	0.0050	1.9287	0.9833	189.3753	0.9454	1.9237

Table 5. Three-dimensional ROC-derived AUC results achieved on NUAA-SIRST dataset.

Method	$AUC_{D,F}$	$AUC_{D,\tau}$	$AUC_{F,\tau}$	AUC_{TD}	AUC_{BS}	AUC_{SNPR}	AUC_{TDBS}	AUC_{ODP}
NRAM	0.7335	0.2656	0.0050	0.9990	0.7385	52.8825	0.2605	0.9940
PSTNN	0.8406	0.4227	0.0051	1.2632	0.8457	82.9856	0.4176	1.2581
SRWS	0.6709	0.1429	0.0050	0.8139	0.6759	28.5292	0.1379	0.8089
ACM	0.9146	0.7524	0.0051	1.6670	0.9197	147.1186	0.7473	1.6619
ALCNet	0.9053	0.7717	0.0051	1.6770	0.9104	150.7933	0.7666	1.6719
RDIAN	0.9227	0.8171	0.0051	1.7398	0.9278	160.3318	0.8120	1.7347
UNet	0.9092	0.7951	0.0051	1.7043	0.9143	155.9492	0.7900	1.6992
ISTDUNet	0.9125	0.7829	0.0051	1.6955	0.9176	153.3739	0.7778	1.6904
DNANet	0.9178	0.8191	0.0051	1.7369	0.9229	160.4584	0.8140	1.7318
Proposed	0.9098	0.8056	0.0051	1.7154	0.9151	154.0000	0.8003	1.7102

Further analysis of Tables 2–5 and Figures 18 and 19 indicates that the performance of deep learning methods significantly exceeds that of non-deep learning methods in all metrics except the Fa value. Non-deep learning model-driven algorithms are constrained by the need to manually model small target features, applying constraints such as shape or sparsity and extracting specific components from images. While such designs do not rely on large datasets, they are limited in their scope due to their modeling bias. In contrast, DNN models learn features that minimize the loss function through the backpropagation algorithm with appropriate loss function settings. The proposed TFCD-Net, using FEBs for progressive false alarm suppression combined with a TSB for target segmentation, achieves correct target detection and precise segmentation across various conditions.

For the evaluation of the complexity, Table 6 illustrates a comparison of DNN models on two datasets. The experiments were performed on an RTX 3090 GPU using Python. Our model exhibited a medium number of parameters compared to the other models. Notably, the inference speed of our model was 4.203 ms per image, and the training durations were 3.9463 s per epoch on the NUAA-SIRST dataset and 12.4757 s per epoch on the NUDT-SIRST dataset. Though our approach did not surpass the speed of ACM, ALCNet, and RDIAN, it was significantly faster than models such as UNet, ISTDUNet, and DNANet in both the training and inference times. These results show that our model achieves high performance with a comparatively modest parameter count, making it suitable for real-time applications due to its fast inference time.

Table 6. Comparison of complexity of DNN models on 2 datasets.

Method	Params ($\times 10^6$)	Inference (ms)	Training on NUAA (s/epoch)	Training on NUDT (s/epoch)
ACM	0.3978	3.905	1.5274	4.5036
ALCNet	0.4270	3.894	1.4335	4.7769
RDIAN	0.2166	2.757	2.6016	8.2245
UNet	34.5259	2.116	4.1787	13.0852
ISTDUNet	2.7519	13.489	6.4446	18.6608
DNANet	4.6966	15.819	8.4540	26.3606
Proposed	1.4501	4.203	3.9463	12.4757

5. Conclusions

In this paper, we introduce TFCD-Net for detecting small targets in infrared imagery. To reduce the false alarm rate, we utilized FEBs to model and estimate false alarm sources. For enhanced interpretability, we propose a framework that resembles the background suppression process utilized in non-deep learning approaches. The experimental results demonstrate that our model outperforms other state-of-the-art methods, achieving the highest and second-highest scores in the IoU, Pd, and AUC, while attaining the lowest Fa among the DNN methods. The high performance of our network is achieved through the collaboration of a multi-stage progressive suppression of false alarm sources using FEBs, as well as target segmentation with a TSB. The multi-stage framework of TFCD-Net, which remains end-to-end, not only provides a path for improving the performance of existing algorithms, but also the false alarm sources estimated by the FEBs on multiple scales provide valuable data for subsequent studies. For example, FEBs can act as preprocessing modules to suppress complex backgrounds in both current and future algorithms for detecting small targets in infrared imagery. Furthermore, the estimated false alarm sources can serve as samples to generate datasets for training models specialized in false alarm source detection, filling the current gap in the field, and they can be used to augment existing datasets to increase the diversity of object types.

As limitations of our method, due to its multi-stage structure, the model has more parameters compared to lighter models, leading to longer training times. Also, while FEBs are intended to suppress false alarm sources, there may be instances where true targets are suppressed.

Future efforts could focus on enhancing the architecture to improve robustness and interpretability, and implementing the framework to detect targets in datasets acquired from additional sensor types or containing multiple target categories.

Author Contributions: Conceptualization, S.C.; methodology, S.C.; validation, S.C., Z.L. and J.D.; formal analysis, S.C.; data collection, S.C.; writing—original draft preparation, S.C.; writing—review and editing, Z.L. and Y.H.; visualization, Z.L.; supervision, Z.P.; project administration, Z.P.; funding acquisition, Z.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the Natural Science Foundation of Sichuan Province of China under Grant 2022NSFSC40574 and Grant 2023NSFSC0508 and in part by the National Natural Science Foundation of China under Grant 61775030 and Grant 61571096.

Data Availability Statement: Publicly available datasets were analyzed in this study. The NUAA-SIRST dataset can be found here: https://github.com/YimianDai/sirst, accessed on 11 April 2023; the NUDT-SIRST dataset can be found here: https://github.com/YeRen123455/Infrared-Small-Target-Detection, accessed on 11 April 2023.

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

References

1. Prasad, D.K.; Rajan, D.; Rachmawati, L.; Rajabally, E.; Quek, C. Video Processing From Electro-Optical Sensors for Object Detection and Tracking in a Maritime Environment: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 1993–2016. [CrossRef]
2. Kim, S.; Lee, J. Scale invariant small target detection by optimizing signal-to-clutter ratio in heterogeneous background for infrared search and track. *Pattern Recognit.* **2012**, *45*, 393–406. [CrossRef]
3. Filizzola, C.; Corrado, R.; Marchese, F.; Mazzeo, G.; Paciello, R.; Pergola, N.; Tramutoli, V. RST-FIRES, an exportable algorithm for early-fire detection and monitoring: Description, implementation, and field validation in the case of the MSG-SEVIRI sensor. *Remote Sens. Environ.* **2016**, *186*, 196–216. [CrossRef]
4. Thanh, N.T.; Sahli, H.; Hao, D.N. Infrared Thermography for Buried Landmine Detection: Inverse Problem Setting. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 3987–4004. [CrossRef]
5. Kou, R.K.; Wang, C.P.; Peng, Z.M.; Zhao, Z.H.; Chen, Y.H.; Han, J.H.; Huang, F.Y.; Yu, Y.; Fu, Q. Infrared small target segmentation networks: A survey. *Pattern Recognit.* **2023**, *143*, 25. [CrossRef]
6. Rawat, S.S.; Verma, S.K.; Kumar, Y. Review on recent development in infrared small target detection algorithms. In Proceedings of the International Conference on Computational Intelligence and Data Science (ICCIDS), Chennai, India, 20–22 February 2020; Volume 167, pp. 2496–2505. [CrossRef]
7. Zhang, T.F.; Li, L.; Cao, S.Y.; Pu, T.; Peng, Z.M. Attention-Guided Pyramid Context Networks for Detecting Infrared Small Target under Complex Background. *IEEE Trans. Aerosp. Electron. Syst.* **2023**, *59*, 4250–4261. [CrossRef]
8. Deshpande, S.D.; Er, M.H.; Venkateswarlu, R.; Chan, P. Max-mean and max-median filters for detection of small targets. In *Signal and Data Processing of Small Targets 1999*; Drummond, O.E., Ed.; International Society for Optics and Photonics (SPIE): Bellingham, WA, USA, 1999; Volume 3809, pp. 74–83. [CrossRef]
9. Liu, Y.H.; Peng, Z.M.; Huang, S.Q.; Wang, Z.R.; Pu, T. River detection using LBP and morphology in infrared image. In Proceedings of the 9th International Symposium on Advanced Optical Manufacturing and Testing Technologies (AOMATT)—Optoelectronic Materials and Devices for Sensing and Imaging, Chengdu, China, 26–29 June 2018; Volume 10843. [CrossRef]
10. Xiao, S.Y.; Peng, Z.M.; Li, F.S. Infrared Cirrus Detection Using Non-Convex Rank Surrogates for Spatial-Temporal Tensor. *Remote Sens.* **2023**, *15*, 21. [CrossRef]
11. Kong, X.; Yang, C.P.; Cao, S.Y.; Li, C.H.; Peng, Z.M. Infrared Small Target Detection via Nonconvex Tensor Fibered Rank Approximation. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 21. [CrossRef]
12. Marvasti, F.S.; Mosavi, M.R.; Nasiri, M. Flying small target detection in IR images based on adaptive toggle operator. *IET Comput. Vis.* **2018**, *12*, 0327. [CrossRef]
13. Hu, Y.X.; Ma, Y.P.; Pan, Z.X.; Liu, Y.H. Infrared Dim and Small Target Detection from Complex Scenes via Multi-Frame Spatial-Temporal Patch-Tensor Model. *Remote Sens.* **2022**, *14*, 36. [CrossRef]
14. Wang, Y.; Cao, L.H.; Su, K.K.; Dai, D.; Li, N.; Wu, D. Infrared Moving Small Target Detection Based on Space-Time Combination in Complex Scenes. *Remote Sens.* **2023**, *15*, 25. [CrossRef]
15. Ding, L.H.; Xu, X.; Cao, Y.; Zhai, G.T.; Yang, F.; Qian, L. Detection and tracking of infrared small target by jointly using SSD and pipeline filter. *Digit. Signal Process.* **2021**, *110*, 9. [CrossRef]
16. Tom, V.T.; Peli, T.; Leung, M.; Bondaryk, J.E. Morphology-based algorithm for point target detection in infrared backgrounds. In *Signal and Data Processing of Small Targets 1993*; Drummond, O.E., Ed.; International Society for Optics and Photonics (SPIE): Bellingham, WA, USA, 1993; Volume 1954, pp. 2–11. [CrossRef]
17. Deng, L.Z.; Zhu, H.; Zhou, Q.; Li, Y.S. Adaptive top-hat filter based on quantum genetic algorithm for infrared small target detection. *Multimed. Tools Appl.* **2018**, *77*, 10539–10551. [CrossRef]

18. Bai, X.Z.; Zhou, F.G. Analysis of new top-hat transformation and the application for infrared dim small target detection. *Pattern Recognit.* **2010**, *43*, 2145–2156. [CrossRef]
19. Kim, S.; Yang, Y.; Lee, J.; Park, Y. Small Target Detection Utilizing Robust Methods of the Human Visual System for IRST. *J. Infrared Millim. Terahertz Waves* **2009**, *30*, 994–1011. [CrossRef]
20. Wei, Y.; You, X.; Li, H. Multiscale patch-based contrast measure for small infrared target detection. *Pattern Recognit.* **2016**, *58*, 216–226. [CrossRef]
21. Qi, S.X.; Xu, G.J.; Mou, Z.Y.; Huang, D.Y.; Zheng, X.L. A fast-saliency method for real-time infrared small target detection. *Infrared Phys. Technol.* **2016**, *77*, 440–450. [CrossRef]
22. Chen, C.L.P.; Li, H.; Wei, Y.T.; Xia, T.; Tang, Y.Y. A Local Contrast Method for Small Infrared Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 574–581. [CrossRef]
23. Han, J.H.; Moradi, S.; Faramarzi, I.; Zhang, H.H.; Zhao, Q.; Zhang, X.J.; Li, N. Infrared Small Target Detection Based on the Weighted Strengthened Local Contrast Measure. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 1670–1674. [CrossRef]
24. Lu, R.T.; Yang, X.G.; Li, W.P.; Fan, J.W.; Li, D.L.; Jing, X. Robust Infrared Small Target Detection via Multidirectional Derivative-Based Weighted Contrast Measure. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 5. [CrossRef]
25. Gao, C.Q.; Meng, D.Y.; Yang, Y.; Wang, Y.T.; Zhou, X.F.; Hauptmann, A.G. Infrared Patch-Image Model for Small Target Detection in a Single Image. *IEEE Trans. Image Process.* **2013**, *22*, 4996–5009. [CrossRef]
26. Dai, Y.; Wu, Y.; Song, Y.; Guo, J. Non-negative infrared patch-image model: Robust target-background separation via partial sum minimization of singular values. *Infrared Phys. Technol.* **2017**, *81*, 182–194. [CrossRef]
27. Sun, Y.; Yang, J.; An, W. Infrared small target detection based on reweighted infrared patch-image model and total variation regularization. In *Image and Signal Processing for Remote Sensing XXV*; Bruzzone, L.; Bovolo, F., Eds.; International Society for Optics and Photonics (SPIE): Bellingham, WA, USA, 2019; Volume 11155, p. 111551F. [CrossRef]
28. Zhu, H.; Ni, H.; Liu, S.; Xu, G.; Deng, L. TNLRS: Target-Aware Non-Local Low-Rank Modeling With Saliency Filtering Regularization for Infrared Small Target Detection. *IEEE Trans. Image Process.* **2020**, *29*, 9546–9558. [CrossRef] [PubMed]
29. Bi, Y.G.; Bai, X.Z.; Jin, T.; Guo, S. Multiple Feature Analysis for Infrared Small Target Detection. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1333–1337. [CrossRef]
30. Cao, S.Y.; Deng, J.K.; Luo, J.H.; Li, Z.; Hu, J.S.; Peng, Z.M. Local Convergence Index-Based Infrared Small Target Detection against Complex Scenes. *Remote Sens.* **2023**, *15*, 18. [CrossRef]
31. Li, S.S.; Li, Y.J.; Li, Y.; Li, M.J.; Xu, X.R. YOLO-FIRI: Improved YOLOv5 for Infrared Image Object Detection. *IEEE Access* **2021**, *9*, 141861–141875. [CrossRef]
32. Ma, J.Y.; Tang, L.F.; Xu, M.L.; Zhang, H.; Xiao, G.B. STDFusionNet: An Infrared and Visible Image Fusion Network Based on Salient Target Detection. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 13. [CrossRef]
33. Ryu, J.; Kim, S. Heterogeneous Gray-Temperature Fusion-Based Deep Learning Architecture for Far Infrared Small Target Detection. *J. Sens.* **2019**, *2019*, 4658068. [CrossRef]
34. McIntosh, B.; Venkataramanan, S.; Mahalanobis, A. Infrared Target Detection in Cluttered Environments by Maximization of a Target to Clutter Ratio (TCR) Metric Using a Convolutional Neural Network. *IEEE Trans. Aerosp. Electron. Syst.* **2021**, *57*, 485–496. [CrossRef]
35. Dai, Y.M.; Wu, Y.Q.; Zhou, F.; Barnard, K. Asymmetric Contextual Modulation for Infrared Small Target Detection. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2021; pp. 949–958. [CrossRef]
36. Huang, L.; Dai, S.; Huang, T.; Huang, X.; Wang, H. Infrared small target segmentation with multiscale feature representation. *Infrared Phys. Technol.* **2021**, *116*, 103755. [CrossRef]
37. Zhao, B.; Wang, C.; Fu, Q.; Han, Z. A Novel Pattern for Infrared Small Target Detection With Generative Adversarial Network. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 4481–4492. [CrossRef]
38. Chen, F.; Gao, C.; Liu, F.; Zhao, Y.; Zhou, Y.; Meng, D.; Zuo, W. Local Patch Network With Global Attention for Infrared Small Target Detection. *IEEE Trans. Aerosp. Electron. Syst.* **2022**, *58*, 3979–3991. [CrossRef]
39. 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*, 1–13. [CrossRef]
40. Sun, H.; Bai, J.X.; Yang, F.; Bai, X.Z. Receptive-Field and Direction Induced Attention Network for Infrared Dim Small Target Detection With a Large-Scale Dataset IRDST. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 13. [CrossRef]
41. Hou, Q.Y.; Wang, Z.P.; Tan, F.J.; Zhao, Y.; Zheng, H.L.; Zhang, W. RISTDnet: Robust Infrared Small Target Detection Network. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 5. [CrossRef]
42. Wang, H.; Zhou, L.P.; 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 (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 8508–8517. [CrossRef]
43. Dai, Y.M.; Wu, Y.Q.; Zhou, F.; Barnard, K. Attentional Local Contrast Networks for Infrared Small Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 9813–9824. [CrossRef]
44. Li, B.Y.; Xiao, C.; Wang, L.G.; Wang, Y.Q.; Lin, Z.P.; Li, M.; An, W.; Guo, Y.L. Dense Nested Attention Network for Infrared Small Target Detection. *IEEE Trans. Image Process.* **2023**, *32*, 1745–1758. [CrossRef] [PubMed]

45. Hou, Q.Y.; Zhang, L.W.; Tan, F.J.; Xi, Y.Y.; Zheng, H.L.; Li, N. ISTDU-Net: Infrared Small-Target Detection U-Net. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 5. [CrossRef]
46. Zhang, L.D.; Peng, L.B.; Zhang, T.F.; Cao, S.Y.; Peng, Z.M. Infrared Small Target Detection via Non-Convex Rank Approximation Minimization Joint l(2,1) Norm. *Remote Sens.* **2018**, *10*, 26. [CrossRef]
47. Zhang, L.D.; Peng, Z.M. Infrared Small Target Detection Based on Partial Sum of the Tensor Nuclear Norm. *Remote Sens.* **2019**, *11*, 34. [CrossRef]
48. Zhang, T.F.; Peng, Z.M.; Wu, H.; He, Y.M.; Li, C.H.; Yang, C.P. Infrared small target detection via self-regularized weighted sparse model. *Neurocomputing* **2021**, *420*, 124–148. [CrossRef]
49. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 779–788. [CrossRef]
50. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Computer Vision—ECCV 2016, PT I*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing AG: Cham, Switzerland, 2016; Volume 9905, pp. 21–37. [CrossRef]
51. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 3431–3440. [CrossRef]
52. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1440–1448. [CrossRef]
53. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]
54. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, 5–9 October 2015; Volume 9351, pp. 234–241. [CrossRef]
55. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J.; Ieee. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
56. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on International Conference on Machine Learning, Madison, WI, USA, 21–24 June 2010; pp. 807–814.
57. Zhang, K.; Zuo, W.M.; Chen, Y.J.; Meng, D.Y.; Zhang, L. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [CrossRef] [PubMed]
58. Tripathi, P.C.; Bag, S. CNN-DMRI: A Convolutional Neural Network for Denoising of Magnetic Resonance Images. *Pattern Recognit. Lett.* **2020**, *135*, 57–63. [CrossRef]
59. Rahman, M.A.; Wang, Y. Optimizing Intersection-Over-Union in Deep Neural Networks for Image Segmentation. In *Advances in Visual Computing*; Springer: Cham, Switzerland, 2016; pp. 234–244.
60. Chang, C.I. An Effective Evaluation Tool for Hyperspectral Target Detection: 3D Receiver Operating Characteristic Curve Analysis. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5131–5153. [CrossRef]
61. Luo, Y.; Li, X.; Chen, S.; Xia, C.; Zhao, L. IMNN-LWEC: A Novel Infrared Small Target Detection Based on Spatial–Temporal Tensor Model. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–22. [CrossRef]