applied sciences

MDPI

Article

Dual Enhancement Network for Infrared Small Target Detection

Xinyi Wu [1], Xudong Hu [1], Huaizheng Lu [2], Chaopeng Li [1,*], Lei Zhang [1] and Weifang Huang [1]

[1] School of Ocean Information Engineering, Jimei University, Xiamen 361021, China; 202121306003@jmu.edu.cn (X.W.); huxudong@jmu.edu.cn (X.H.); 202111810011@jmu.edu.cn (L.Z.); 202121306075@jmu.edu.cn (W.H.)
[2] College of Computer Engineering, Jimei University, Xiamen 361021, China; luhuaizheng@jmu.edu.cn
* Correspondence: licp@jmu.edu.cn

Abstract: Infrared small target detection (IRSTD) is crucial for applications in security surveillance, unmanned aerial vehicle identification, military reconnaissance, and other fields. However, small targets often suffer from resolution limitations, background complexity, etc., in infrared images, which poses a great challenge to IRSTD, especially due to the noise interference and the presence of tiny, low-luminance targets. In this paper, we propose a novel dual enhancement network (DENet) to suppress background noise and enhance dim small targets. Specifically, to address the problem of complex backgrounds in infrared images, we have designed the residual sparse enhancement (RSE) module, which sparsely propagates a number of representative pixels between any adjacent feature pyramid layers instead of a simple summation. To handle the problem of infrared targets being extremely dim and small, we have developed a spatial attention enhancement (SAE) module to adaptively enhance and highlight the features of dim small targets. In addition, we evaluated the effectiveness of the modules in the DENet model through ablation experiments. Extensive experiments on three public infrared datasets demonstrated that our approach can greatly enhance dim small targets, where the average values of intersection over union (IoU), probability of detection (P_d), and false alarm rate (F_a) reached up to 77.33%, 97.30%, and 9.299%, demonstrating a performance superior to the state-of-the-art IRSTD method.

Keywords: infrared image; small target detection; sparse semantic propagation; spatial attention; feature enhancement

Citation: Wu, X.; Hu, X.; Lu, H.; Li, C.; Zhang, L.; Huang, W. Dual Enhancement Network for Infrared Small Target Detection. *Appl. Sci.* **2024**, *14*, 4132. https://doi.org/10.3390/app14104132

Academic Editor: Antonio Fernández-Caballero

Received: 8 April 2024
Revised: 5 May 2024
Accepted: 9 May 2024
Published: 13 May 2024

1. Introduction

Infrared imaging technology provides excellent concealment, good portability, and reliable detection of blind areas compared to radar imaging [1–3]. Compared to visible light imaging, infrared imaging technology offers numerous advantages. First of all, it uses infrared sensors for its strong night vision, which detects heat radiated by the target and considerably helps with night operations and search missions in dark areas. Second, even under low light conditions, infrared imaging can clearly see through atmospheric obstructions like smoke, fog, and clouds. However, infrared targets usually seem very small, often as small as one pixel in size, due to factors like lengthy imaging distances, air dispersion, and flash noise. Therefore, the detection of small targets in infrared images has become an important research direction, i.e., infrared small target detection (IRSTD). IRSTD has significant advantages in scientific research and military applications, including military early warning [4], missile tracking systems [5], and maritime surveillance [6]. Lack of information such as colour, shape, and texture makes IRSTD more complex, further making it a highly sought after and challenging task.

For most semantic segmentation, IRSTD has its unique features. Firstly, unlike normal small target detection, infrared images contain many complex backgrounds, and infrared small targets usually contain two features: “small” and “weak”, where “weak” means that the contrast between the target and the background is poor and that the target has

a low signal-to-noise ratio, and "small", which means that the target has fewer pixels and sometimes even contains only one pixel. Obtaining particular information about the target during detection is challenging. Therefore, it is necessary to reduce the interference of background noise while focusing on small targets during the detection process. In addition, due to the significant attenuation of infrared radiation energy with distance [7], infrared targets gradually become very dim. As a result, infrared targets are easily submerged in the background clutter, and their ambiguity makes segmentation difficult.

In this paper, we propose a dual enhancement network for background noise reduction and dim small target feature enhancement. Specifically, to minimise the interference of complex backgrounds, we propose the residual sparse enhancement (RSE) module, which shifts and spreads a number of representative pixels between any adjacent feature pyramid layers instead of simply summarising them. We achieve the suppression of ambiguous background noise and the enhancement of small target features by propagating sparse key pixels only on infrared small target features (as shown in Figure 1c), generating high-quality features that facilitate subsequent prediction. Thus, the problem of complex backgrounds and unusually small targets can be solved in this innovative way. Furthermore, considering that infrared small targets usually appear very faint and tiny, we propose a spatial attention enhancement (SAE) module to adaptively enhance dim small target features. Detection accuracy can be greatly improved by these enhanced features. Note that this task-specific module enhances the visibility of dim small targets while minimising background noise, thus creating a more distinct and focused visual presentation.

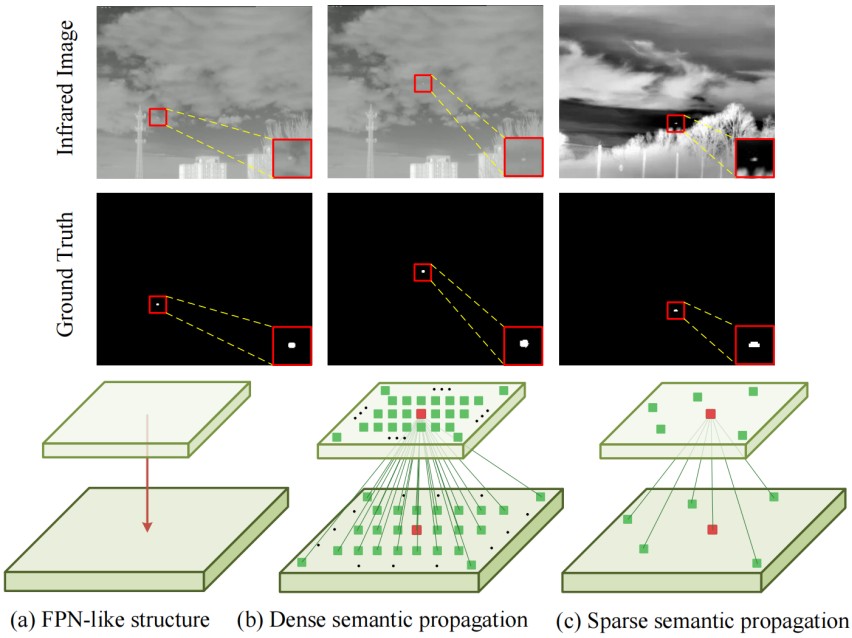

Figure 1. Illustration of an infrared small target detection example and our proposed module. The first two rows present the input image and ground truth with complex backgrounds and small targets. The third row indicates three ways of semantic propagation, consisting of FPN-like structure, dense semantic propagation, and our proposed sparse semantic propagation.

In summary, the contributions of this work are as follows:

- To address the problem of complex backgrounds in infrared images, we propose the residual sparse enhancement (RSE) module, which sparsely selects a number of representative pixels for semantic information propagation, thereby innovatively suppressing background noise.
- To address the problem of unusually faint and small infrared targets, we propose a task-specific module, the spatial attention enhancement (SAE) module, which adaptively

enhances and highlights dim and small target features, thus effectively improving the performance of dim and small target detection.
- Extensive experiments demonstrated that our method outperforms the state-of-the-art (SOTA) method, and ablation studies fully validate the effectiveness of each component of our proposed method.

2. Related Work

For infrared small target detection, both single-frame image-based detect before track (DBT) and sequential-frame image-based track before detect (TBD) [8] techniques are very important methods. The DTB technique can analyse single-frame infrared images and use target detection algorithms to achieve the localisation and identification of small targets, which is suitable for static or slow-moving scenarios, thus disregarding the information of time [9]. The TBD technique, on the other hand, is more suitable for dealing with motion changes in infrared small targets, and by analysing a continuous sequence of infrared images, the motion trajectory and appearance changes of small targets can be better captured. In practical applications, for scenarios that require high real-time requirements and limited resources, single-frame image-based detection can complete the target detection and tracking tasks more quickly and, at the same time, the progress of single-frame image detection algorithms will also directly affect the performance of DBT techniques.

Methods for traditional IR small target detection are mainly classified as the following: filter-based [10,11], local information-based [12,13], data structure-based [14,15], and deep learning-based methods [8,16,17], but there are some limitations in coping with complex backgrounds and small target detection. Firstly, unlike common small object detection, infrared images contain many complex backgrounds, and infrared small targets are very small and sparse, sometimes even containing only one pixel. Therefore, it is essential to reduce the interference of background noise and pay attention to small targets simultaneously in the detection process. With the development of deep learning technology, deep learning models are gradually replacing the traditional methods, and they can deal with the infrared small target detection problem in complex scenes more effectively. For most semantic segmentation, IRSTD has its unique characteristics, such as filter-based methods [18], low-rank-based methods [14,19], and data-driven methods [7,20–24], but few of them consider the challenges of complex backgrounds and fuzzy targets in infrared images. Wang et al. [7] proposed a deep adversarial learning framework called MDvsFA-cGAN that includes two generators and a discriminator to balance false detections and false alarms. Dai et al. [20] designed an asymmetric context modulation (ACM) module to better highlight small objects. Shortly thereafter, Dai et al. [21] proposed the attentional local contrast network (ALCNet) with a feature graph cyclic shifting scheme and bottom-up attentional modulation. Zhang et al. [22] proposed ISNet to detect precise shape information and effectively suppress complex background noise by aggregating different layers of edge information. Wu et al. [23] advocated for the UIU-Net to learn multi-level and multi-scale representations. Li et al. [24] offered a dense nested attention network (DNA-Net) to fully merge and utilise contextual information from small targets. Hou et al. proposed the RISTDnet network [25], which combines a hand-designed feature approach convolutional neural network to create a multi-sized feature extraction framework, introducing a feature mapping network, and the thresholding operation and segmentation techniques are able to extract the target accurately from the complex background. Rawat et al. proposed that a patch-based approach is more effective than other filter-based approaches [26].

This paper presents a new network, DENet, to address the IRSTD task. We believe that the intricacies of infrared camera equipment and environmental conditions significantly compound the difficulty of detecting small targets in infrared imagery, transcending the realm of conventional image-processing techniques alone. As a result, we should identify the global information and, specifically, improve the local information in accordance with the relevant occasions. We propose a DENet model that introduces a residual structure that suppresses the gradient vanishing problem in deep networks and improves the accuracy

of infrared, and which also introduces sparse enhancement, which enhances only representative features. In addition, this data-driven spatial attention improvement approach aims to efficiently deal with dim small targets by progressively learning more advanced and abstract feature representations using a multilevel convolutional process.

In order to adapt to more infrared small target detection scenarios, we designed the DENet model to allow for problem-specific optimisation, which can be more flexibly adapted to different detection tasks by adjusting the equilibrium parameters of the individual modules and the loss function.

3. The Proposed Dual Enhancement Network

For the infrared small target detection task, we propose a residual sparse enhancement (RSE) module and a spatial attention enhancement (SAE) module to cope with the complex background interference and adaptive enhancement of weak small targets. As shown in Figure 2, we named our approach dual enhancement network (DENet), since these two enhancement modules are placed on either side of the network. Specifically, the RSE module processes multi-level features as input, sparsely propagates semantic information, and outputs finely processed features, significantly reducing the effect of background noise. The SAE module, on the other hand, focuses on solving the problem of enhancing weak and small targets in infrared images and adaptively enhances the relevant features. Notably, the SAE module focuses on the weak and small targets while reducing the focus on the complex background, which greatly improves the overall segmentation performance. This dual enhancement strategy effectively solves the key problem in infrared small target detection and provides reliable technical support for improving the accuracy and robustness of target detection.

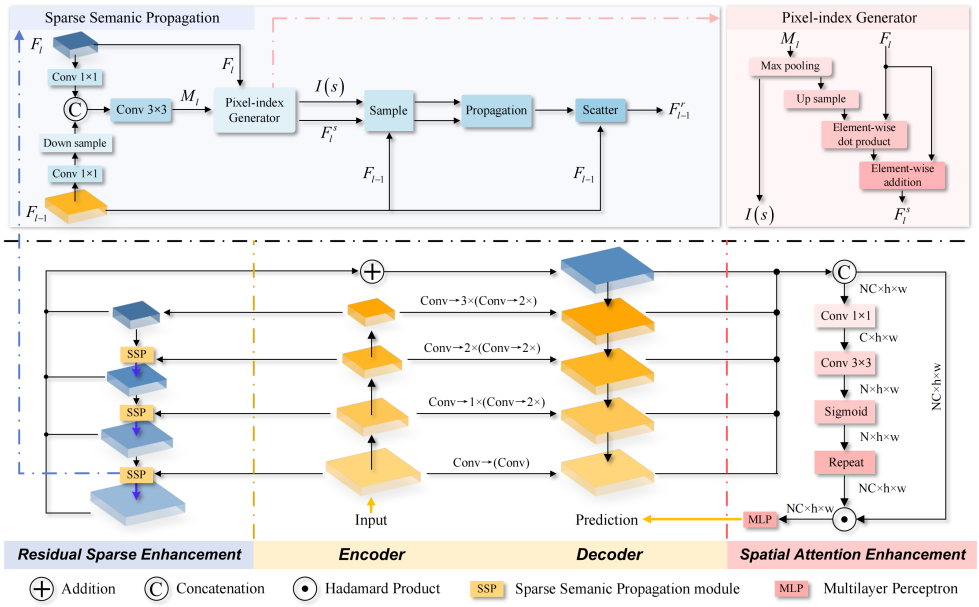

Figure 2. Framework of our DENet, which consists of two distinct designs: (1) residual sparse enhancement (RSE) module and (2) spatial attention enhancement (SAE) module. Above: Pipeline of sparse semantic propagation (SSP) module.

3.1. Residual Sparse Enhancement Module

Inspired by [27], we considered propagating global context information in a sparse manner, which only propagates selective key pixels between adjacent levels. To achieve this, we designed a sparse semantic propagation (SSP) module, and we appended it to the FPN [28] framework to capture multi-level feature advantages. In the SSP module, we first selected the key pixels and then spread sparse semantic information between adjacent levels. The following is the detailed process.

As is shown in the upper part of Figure 2, the most salient part of the input features and key pixel indexes can be extracted through the following two steps: (1) generate the salient map; (2) generate sampled pixel indexes. Then, sparse semantics are propagated between the adjacent levels shown in Equation (3). For the first step, we convert the channel dimensions of two input feature maps to be consistent, downsample the high-resolution feature F_{l-1} to match the low-resolution feature F_l, and obtain a new resized feature denoted as $\widetilde{F}_{l-1}$. Then, we concat the $\widetilde{F}_{l-1}$ and the converted F_l, and we exploit one 3 × 3 convolution following with the sigmoid function to generate the saliency map M_l.

The saliency map M_l can be defined as

$$M_l = \text{Sigmoid}\left(\text{conv}_l\left(\text{Concat}\left(F_l, \widetilde{F}_{l-1}\right)\right)\right), \tag{1}$$

For the second step, the pixel index generator takes F_l and M_l as inputs. To gain the most salient pixels, we employ adaptive max pooling on the saliency map M_l . Then, as is shown in Equation (2), we multiply the saliency map on F_l with the residual structure to enhance the saliency of the foreground objects:

$$F_l^s = \text{Maxpool}(M_l) \times F_l + F_l, \tag{2}$$

We choose the salient indexes from $Maxpool(M_l)$ and denote them as I_s for short. To sum up, the pixel index generator takes F_l and M_l as inputs and generates the salient indexes I_s and the salient map F_l^s for a sampler. The complete operation of the pixel index generator is illustrated in Figure 2. With the salient indexes I_s gained from the pixel index generator, we sample pixels from the input feature map $F_{l-1} \in R^{C\times H\times W}$ and salient map $F_l^s \in R^{C\times H/2\times W/2}$. For each selected pixel $\hat{p}$, we extract the pixel-wise feature f on both adjacent input features, and we then propagate those sampled pixels from the top to the bottom. The specific propagation process is shown in Equation (3):

$$f_{l-1}(\hat{p})^r = A(f_{l-1}(\hat{p}), f_l(\hat{p}))f_l(\hat{p}) + f_{l-1}(\hat{p}), \tag{3}$$

We begin by flattening F_l^s into a two-dimensional vector. Subsequently, we utilise the salient index I_s, generated by the pixel index generator, to gather high-resolution salient features from F_l^s. The resulting features are then reshaped to match the dimensions of I_s. Similarly, we flatten F_{l-1} into a 2D vector. The corresponding salient index I_{low}^s of F_{l-1} is determined based on the dimensional relationship between I_s and the upper and lower features. Utilizing this index, we gather the low-resolution salient features from F_{l-1}, adjusting their shapes to align with I_{low}^s. Finally, we transform both the high-resolution and low-resolution salient features into $f_l(\hat{p})$ and $f_{l-1}(\hat{p})$ through size transformation, where A represents the affinity function, f_l denotes the sampled feature at level l for the salient region, and f_{l-1}^r denotes the refined affinity feature. For A, we use pixel-wise matrix multiplication along with the softmax function for normalization. We adjust the high-resolution salient features according to the affinity matrix. Following the same operation as the previous work [29] mentioned, we employ the residual structure in Equation (3). The low-resolution salient features are added to the adjusted high-resolution salient features to obtain the final $f_{l-1}(\hat{p})^r$. We calculate the high-semantic sampled pixels through pixel-wise affinity according to the semantic similarity on the low-semantic sampled pixels, which can diminish the interference of redundant background information in the infrared scene. Finally, we scatter f_{l-1}^r into F_{l-1} according to the salient indexes I_s and obtain the refined feature F_{l-1}^r . Since the substitution is based on the salient index, only useful salient information is propagated from top to bottom. The sparse correlation modelling for the high-resolution salient feature $f_l(\hat{p})$ and the low-resolution salient feature $f_{l-1}(\hat{p})$ also results in less interference from background noise when adjusting $f_l(\hat{p})$ using the affinity matrix . Compared with the original features generated from the encoder, these refined features gain more salient information about small targets and less redundant information about complex backgrounds. Considering the importance of capturing contextual

information, we additionally insert the pyramid pooling module (PPM) [30] into the top layer between the encoder and the RSE module. Finally, the enhanced feature with less background noise, precise location information, and rich semantic information can be obtained by combining these refined features in an additive way.

3.2. Spatial Attention Enhancement Module

Because detection in the IRSTD task has the characteristic of targets being specially sparse and tiny, we put forward a spatial attention enhancement (SAE) module to enhance and highlight the dim small targets in the infrared images. Figure 2 illustrates the structure of SAE in detail. In brief, SAE takes the concated features as input and produces one spatial weight map for each feature, and then the concated features are fused with the spatial weight map by the Hadamard product operation. After that, we adopt the concatenation operation to fuse the features. Finally, an MLP is employed to generate the final prediction. With the feature enhancement in space, the problem of targets being dim and small is well solved, and the detection performance is also highly improved.

The SAE module aims to enhance spatial attention in infrared small target detection and works in detail as follows: firstly, it takes as input the feature maps extracted from the convolutional layers in front of the CNN, capturing the hierarchical features learnt by the network. Second, its core mechanism is the spatial attention mechanism, which consists of multiple sets of convolutional layers and activation functions that generate an attention map that emphasises important spatial regions in the feature map. Task-specific convolutional operations are also applied to further improve the detection performance. Thirdly, the attention map generated by the spatial attention mechanism is used to modulate the original feature map. This modulation is achieved through element-level multiplication or addition, where the attention maps are used as weights to amplify or suppress feature activations depending on their importance. Finally, the SAE module is seamlessly integrated into the CNN architecture, often as a component or add-on module in the residual block, enabling the network to learn spatially adaptive feature representations during training.

The SAE module we designed has several advantages in infrared small target segmentation. Its spatial attention mechanism directs the network's attention to relevant spatial regions, enhancing the identification of relevant features of small infrared targets. By emphasising information specific to the target, better segmentation performance is formed. Secondly, the adaptive feature enhancement of the SAE module helps the network to better deal with background clutter and noise in infrared images by suppressing irrelevant background features, thereby improving segmentation accuracy in complex scenes. Finally, the SAE module dynamically adjusts the feature representations according to the spatial attention map, which helps to achieve scaling and rotation invariance, and it is crucial for accurate segmentation of small targets in different scenes.

3.3. Loss Function

We adopted the binary cross-entropy (BCE) loss ℓ_{bce} and the soft intersection over union (Soft-loU) loss ℓ_{iou} to supervise the final prediction during the network training, and the total loss function ℓ_{all} is expressed as follows:

$$\ell_{\text{all}} = \lambda \ell_{\text{bce}} + (\lambda - 1)\ell_{\text{iou}} \tag{4}$$

The combined use of the BCE loss function and the Soft-loU loss function offers significant advantages for infrared small target detection. The BCE loss function aids in learning target boundaries and features effectively, while Soft-loU enhances stability and robustness by considering pixel-level matching, crucial for tasks with blurry boundaries or uncertainty. By integrating these two loss functions, the model can comprehensively address both object detection and pixel-level segmentation requirements, leading to improved segmentation accuracy and robustness. In summary, this combined approach represents a rational and effective strategy, leveraging the strengths of both loss functions to enhance model performance and generalisation ability in infrared small target detection tasks.

By adjusting the equilibrium parameter λ in the loss function, DENet has the capability to dynamically adjust the model's emphasis on accuracy and false alarm rate in accordance with diverse task requirements, thus optimising the overall performance, where λ= 0.8 denotes the balance parameter.

4. Experiment

We trained and tested our model on the NUAA-SIRST [20], NUDT-SIRST [24], and IRSTD-1K [22] datasets, and we followed references [24,31] to set the same train–test ratio. We used three semantic segmentations (i.e., IoU, P_d, F_a) as performance criteria. For IoU and Pd, larger values indicate higher performance. For Fa, smaller values indicate higher performance. To evaluate the segmentation performance, we compared our method with nine state-of-the-art (SOTA) algorithms, including three traditional methods (Top-Hat [18], IPI [14], PSTNN [19]) and six deep learning-based methods (MDvsFAcGAN [7], ACM [20], ALCNet [21], ISNet [22], UIU-Net [23], and DNA-Net [24]). We implemented our method with the PyTorch framework and used an NVIDIA GeForce RTX 3090 (24 GB) in our experiment. Before training, all input images were first resized to a resolution of 512 × 512. During the network training, the AdamW optimizer was adopted for optimization with the CosineAnnealingLR scheduler, and the weight decay was adjusted to 1 $\times 10^{-4}$. The batch size, learning rate, and total epoch number were set to be 6, 1 $\times 10^{-4}$, and 100, respectively. Additionally, specific hyperparameters were configured as follows: the random seed was set to 3407 to ensure reproducibility; the patch size was defined as 512 to facilitate efficient processing of input data. Moreover, to prevent overfitting and to stabilize training, the weight decay rate was set to 0.5, with weight decay applied every 15 epochs.

4.1. Datasets

Existing open-source datasets for infrared small target detection are scarce, and in this paper three currently used mainstream infrared small target detection datasets were selected for study, including NUAA-SIRST [20], NUDT-SIRST [24], and IRSTD-1K [22].

The NUAA-SIRST dataset is a publicly available single-frame dataset containing 427 infrared images covering a total of 480 targets. The dataset also includes infrared images at 950 nm wavelength, and the target labelling approach consists of five forms applicable to different detection models: image classification, instance segmentation, bounding box regression, semantic segmentation, and instance point recognition tasks. The training set is 50%, the validation set is 20%, and the test set is 30%.

NUDT-SIRST is a synthetic dataset containing 1327 images with a resolution of 256 × 256. Compared with the real dataset, NUDT-SIRST has the advantages of accurate labelling, a large number of target categories, rich target sizes, and diverse cluttered backgrounds.

The IRSTD-1K dataset provides 1000 real images with a variety of target shapes, different target sizes, and a rich clutter background with precise pixel-level annotations in the background. The dataset is divided into two folders, where IRSTD1k_Img contains the real images and IRSTD1k_Label contains the label mask.

4.2. Performance Comparisons

Table 1 indicates the detection accuracy of 10 infrared small target detection methods on the NUAA-SIRST, NUDT-SIRST, and IRSTD-1K datasets in terms of three semantic segmentation metrics. Figures 3–5 show a visual comparison of different IRSTD methods on different datasets. Obviously, our method surpasses the compared methods and achieved SOTA results on three public infrared datasets. There are two main reasons for this: (1) by propagating selective pixels between adjacent feature pyramid levels, the residual sparse enhancement (RSE) module can drastically suppress the interference of redundant background information; (2) by focusing on dim small targets through the spatial attention enhancement (SAE) module, the corresponding features are adaptively enhanced, and the segmentation performance is thus greatly improved. In addition, some visualization examples of different approaches have been shown in Figure 6. Specifically, our DENet achieves

optimal visual results, particularly for infrared images with complex backgrounds (the first row in Figure 6) and infrared targets with dim small characteristics (the second row in Figure 6). Compared with the SOTA methods, the better detection performance of our proposed method indicates its greater effectiveness in reducing background noise and enhancing infrared dim small targets. Moreover, to make an intuitive comparison, we plotted the receiver operating characteristic (ROC) curves of different methods on the datasets in Figure 7, further validating the superiority and effectiveness of our proposed method.

Table 1. Quantitative comparison in terms of IoU ($\times 10^2$), P_d ($\times 10^2$), and F_a ($\times 10^6$) values. The best results are shown in red. IoU , P_d, and F_a are expressed in percentages (%). The up arrow next to the assessment indicator indicates that a larger value is better, and the down arrow indicates that a smaller value is better.

Method	NUAA-SIRST [20]			NUDT-SIRST [24]			IRSTD-1K [22]			Average		
	$IoU\uparrow$	$P_d\uparrow$	$F_a\downarrow$	$IoU\uparrow$	$P_d\uparrow$	$F_a\downarrow$	$IoU\uparrow$	$P_d\uparrow$	$F_a\downarrow$	$IoU\uparrow$	$P_d\uparrow$	$F_a\downarrow$
Top-Hat [18]	7.143	79.84	1012	20.72	78.41	166.7	10.06	75.11	1432	12.64	77.79	870.2
IPI [14]	25.67	85.55	11.47	17.76	74.49	41.23	27.92	81.37	16.18	23.78	80.47	22.96
PSTNN [19]	22.40	77.95	29.11	14.85	66.13	44.17	24.57	71.99	35.26	20.61	72.02	36.18
MDvsFA [7]	61.70	91.44	21.11	43.32	86.90	131.1	33.26	86.36	66.50	46.09	88.23	72.90
ACM [20]	62.50	90.49	20.82	57.69	91.43	43.58	56.74	90.57	33.57	58.98	90.83	32.66
ALCNet [21]	69.49	93.92	38.90	64.62	91.53	39.97	63.45	92.26	16.63	65.85	92.57	31.83
ISNet [22]	71.11	92.78	40.57	69.09	94.39	55.30	63.05	93.27	33.35	67.75	93.48	43.07
UIUNet [23]	69.73	95.18	51.44	76.16	97.61	17.63	61.19	92.86	27.53	69.03	95.22	32.20
DNA-Net [24]	76.05	96.58	21.96	86.54	98.84	8.040	63.12	89.12	13.06	75.23	94.85	14.35
DENet (ours)	77.55	98.10	13.50	89.66	99.26	2.987	64.78	93.94	11.41	77.33	97.30	9.299

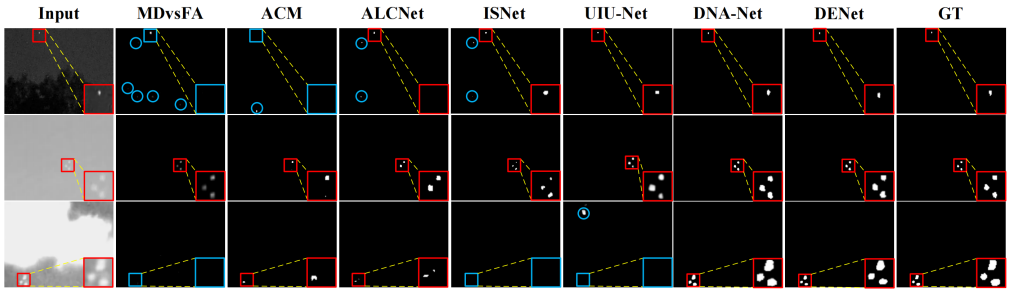

Figure 3. Visual comparisons of different IRSTD methods on the NUAA-SIRST dataset. The correctly detected targets, miss detection areas, and false alarm areas are highlighted by red rectangles, blue rectangles, and blue circles, respectively.

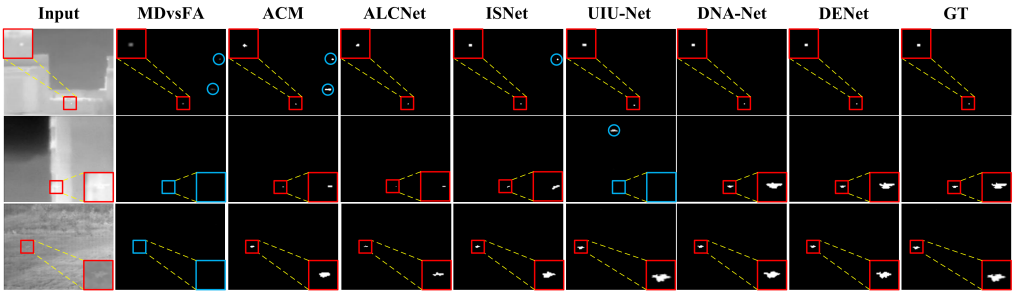

Figure 4. Visual comparisons of different IRSTD methods on the NUDT-SIRST dataset. The correctly detected targets, miss detection areas, and false alarm areas are highlighted by red rectangles, blue rectangles, and blue circles, respectively.

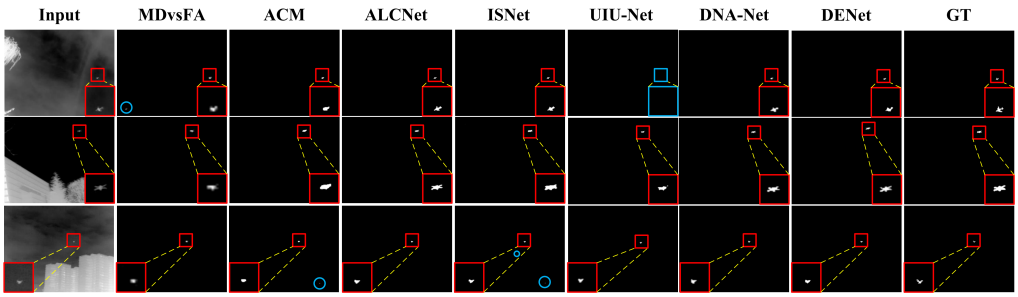

Figure 5. Visual comparisons of different IRSTD methods on the IRSTD-1K dataset. The correctly detected targets, miss detection areas, and false alarm areas are highlighted by red rectangles, blue rectangles, and blue circles, respectively.

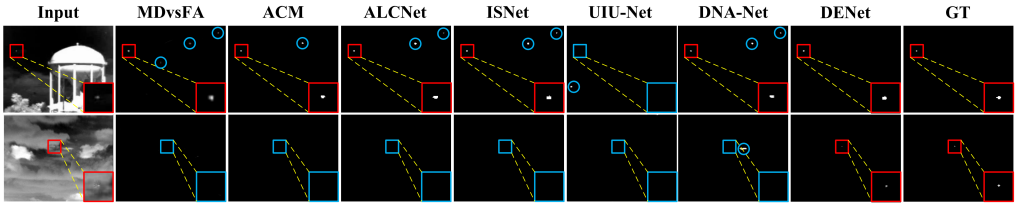

Figure 6. Qualitative results achieved by different IRSTD methods. The correctly detected targets, miss detection areas, and false alarm areas are highlighted by red rectangles, blue rectangles, and blue circles, respectively.

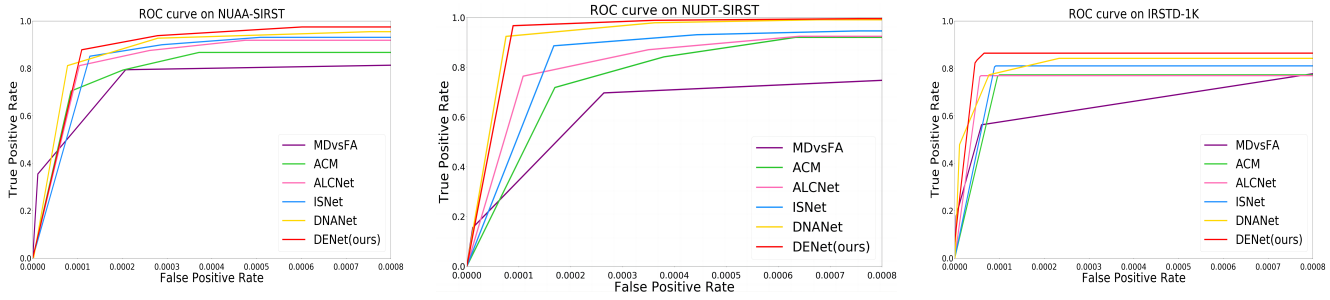

Figure 7. ROC curve comparisons on the NUAA-SIRST, NUDT-SIRST, and IRSTD-1K datasets.

4.3. Ablation Study

(1) *Effectiveness of each component*: We conducted ablation studies to assess the effectiveness of each component in our proposed DENet. The quantitative results are presented in Table 2, and the qualitative results are showcased in Figure 8. R, S, and SSP represent the RSE module, the SAE module, and the SSP module, respectively. In our method, SSP is inserted into R. We can observe that directly using baseline cannot obtain satisfactory results because it does not take the special characteristics (complex backgrounds, exceptionally dim and small targets) of the infrared images into account. The experiment results clearly demonstrate the effectiveness of each module in our proposed approach. As expected, the best results were obtained by simultaneously suppressing complex backgrounds and enhancing dim small targets in the infrared images.

Additionally, to further validate the effectiveness of SAE, we also conducted a comparison between "baseline + SAE" and "baseline + CA (Channel Attention) module [32]" on NUAA. The results reveal that the former achieved a better IoU/Pd/Fa (75.63/96.95/17.77), whereas the latter yielded a worse IoU/Pd/Fa (74.84/96.83/18.72), indicating the superiority and pertinence of the SAE module in the IRSTD task. Compared with the regular attention mechanism (CA), SAE places greater emphasis on dim small targets within intricate backgrounds and is capable of suppressing the interference of background clutter while highlighting the dim small targets.

Table 2. Ablation study of each component on the NUAA-SIRST, NUDT-SIRST, and IRSTD-1K datasets. R, S, and SSP represent the RSE module, the SAE module, and the SSP module, respectively. The best results are in red. IoU , P_d, and F_a are expressed in percentages (%). The up arrow next to the assessment indicator indicates that a larger value is better, and the down arrow indicates that a smaller value is better. A cross (×) indicates that the corresponding module is not used, a tick (✓) indicates that the corresponding module is used.

Method	Abalation Module			MUAA-SIRST [20]			NUDT-SIRST [24]			IRSTD-1K [22]		
	R	*S*	*SSP*	$IoU\uparrow$	$P_d\uparrow$	$F_a\downarrow$	$IoU\uparrow$	$P_d\uparrow$	$F_a\downarrow$	$IoU\uparrow$	$P_d\uparrow$	$F_a\downarrow$
baseline	×	×	×	74.33	95.06	19.67	85.48	98.62	7.882	61.13	90.59	18.37
w/o *R*	×	✓	×	75.63	96.90	17.77	86.91	98.91	4.667	62.58	91.93	17.30
w/o *SSP*	✓	✓	×	75.72	96.95	15.67	88.24	99.2	3.543	63.42	92.31	15.32
w/o *S*	✓	×	✓	75.95	96.96	15.13	88.60	99.05	3.125	63.81	92.94	11.41
DENet (Ours)	✓	✓	✓	77.55	98.10	13.50	89.66	99.26	2.987	64.78	93.94	10.65

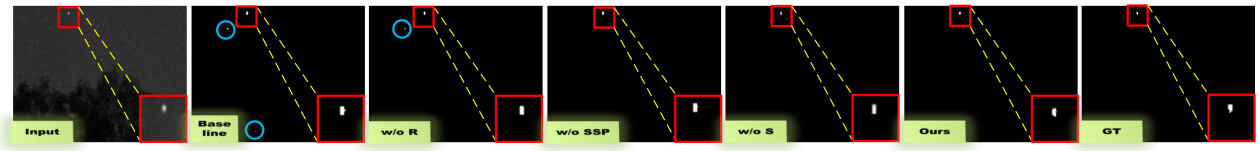

Figure 8. Ablation visual comparisons on the NUAA-SIRST dataset. For better visualization, the correct detected targets are enlarged in the right-bottom corner. The correctly detected targets and false alarm areas are highlighted by red rectangles and blue circles, respectively.

(2) *Impact of RSE*: Comparing the other modules, we found that RSE with SSP inserted has the largest contribution to the whole model, which can effectively suppress the background noise, and at the same time has a good segmentation effect and performs optimally in the NUDT-SIRST dataset. If only the SAE module is used to adaptively enhance the features of dim small targets, the targets are easily submerged in the noise. Therefore, the combination of the two is optimal.

(3) *Impact of SAE*: Although SAE can reduce noise and improve model performance to some extent, there are still some limitations to its optimisation capabilities. This is because its enhancement of a wide range of dim small targets requires careful parameter tuning for optimal results in complex visual tasks, and difficulties in parameter tuning may result in less than optimal model performance.

(4) *Impact of lambda*: The sensitive study of balance parameter λ is presented in Table 3. We investigated the effect of varying the value of λ from 0.1 to 0.9 on the NUAA-SIRST dataset, and set the default lambda as 0.8 with the best performance.

Table 3. Sensitive study of balance parameter λ. The up arrow next to the assessment indicator indicates that a larger value is better, and the down arrow indicates that a smaller value is better. The best values are marked in red.

λ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$IoU(\%)\uparrow$	76.23	77.13	75.62	76.83	75.46	77.28	76.56	77.55	71.63
$P_d(\%)\downarrow$	97.33	98.09	97.33	96.95	97.33	97.71	98.09	98.10	94.29
$F_a(\%)\uparrow$	20.36	17.73	24.62	21.36	22.17	18.44	21.77	13.50	21.91

5. Conclusions

In this paper, a dual enhancement network (DENet) is proposed to solve the problem of complex backgrounds and dim targets in infrared images. The DENet consists of two key modules: the residual sparse enhancement (RSE) module and the spatial attention enhancement (SAE) module. The RSE module effectively mitigates the impact of complex backgrounds by filtering out redundant information, thereby highlighting the salience

of the target region. Consequently, DENet successfully suppresses interference from complex backgrounds and improves target detection accuracy. Meanwhile, the SAE module automatically enhances the features of dim and small targets in infrared images to make them easier to detect and recognise. Using our DENet, the redundant information of complex backgrounds can be effectively filtered, and the contextual information of dim small targets can be well integrated and fully utilised to enhance the features. Our proposed DENet method demonstrated outstanding performance on three publicly available infrared datasets. It not only surpassed the state-of-the-art IRSTD method in terms of segmentation speed but also achieved superior detection results in visual analysis. In our future research, we will prioritize the exploration of novel approaches that integrate deep learning and image-processing techniques to further enhance the accuracy and efficiency of target detection in infrared images.

Author Contributions: Conceptualisation, X.W. and X.H.; methodology, H.L. and C.L.; software, L.Z. and W.H.; validation, C.L., X.W. and W.H.; formal analysis, X.W.; investigation, H.L.; resources, L.Z.; data curation, W.H.; writing—original draft preparation, X.W. and C.L.; writing—review and editing, X.W. and X.H.; visualisation, H.L.; supervision, X.H.; project administration, X.W.; funding acquisition, C.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Youth Program of National Natural Science Foundation of China (Grant No. 62106083) and by the Youth Program of the Natural Science Foundation of Fujian Province of China (Grant No. 2022J05162).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

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

References

1. 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]
2. Shao, X.; Fan, H.; Lu, G.; Xu, J. An improved infrared dim and small target detection algorithm based on the contrast mechanism of human visual system. *Infrared Phys. Technol.* **2012**, *55*, 403–408. [CrossRef]
3. Cheney, M.; Borden, B. *Fundamentals of Radar Imaging*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2009.
4. Ma, T.; Yang, Z.; Wang, J.; Sun, S.; Ren, X.; Ahmad, U. Infrared small target detection network with generate label and feature mapping. *IEEE Geosci. Remote. Sens. Lett.* **2022**, *19*, 6505405. [CrossRef]
5. 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.* **2020**, *59*, 4481–4492. [CrossRef]
6. Ying, X.; Wang, Y.; Wang, L.; Sheng, W.; Liu, L.; Lin, Z.; Zhou, S. Mocopnet: Exploring local motion and contrast priors for infrared small target super-resolution. *arXiv* **2022**, arXiv:2201.01014.
7. Wang, H.; Zhou, L.; Wang, L. Miss detection vs. false alarm: Adversarial learning for small object segmentation in infrared images. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 8509–8518.
8. 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]
9. Wang, X. Clutter-adaptive infrared small target detection in infrared maritime scenarios. *Opt. Eng.* **2011**, *50*, 067001. [CrossRef]
10. Peng, J.; Zhou, W. Infrared background suppression for segmenting and detecting small target. *Acta Electron. Sin.* **1999**, *27*, 47–51.
11. Azimi-Sadjadi, M.R.; Pan, H. Two-dimensional block diagonal LMS adaptive filtering. *IEEE Trans. Signal Process.* **1994**, *42*, 2420–2429. [CrossRef]
12. Chen, C.L.P.; Li, H.; Wei, Y.; Xia, T.; Tang, Y.Y. A local contrast method for small infrared target detection. *IEEE Trans. Geosci. Remote. Sens.* **2013**, *52*, 574–581. [CrossRef]
13. Deng, H.; Sun, X.; Liu, M.; Ye, C.; Zhou, X. Infrared small-target detection using multiscale gray difference weighted image entropy. *IEEE Trans. Aerosp. Electron. Syst.* **2016**, *52*, 60–72. [CrossRef]
14. 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]

15. 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]
16. Zhao, D.; Zhou, H.; Rang, S.; Jia, X. An adaptation of CNN for small target detection in the infrared. In Proceedings of the IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 669–672.
17. Wang, K.; Li, S.; Niu, S.; Zhang, K. Detection of infrared small targets using feature fusion convolutional network. *IEEE Access* **2019**, *7*, 146081–146092. [CrossRef]
18. Rivest, J.; Fortin, R. Detection of dim targets in digital infrared imagery by morphological image processing. *Opt. Eng.* **1996**, *35*, 1886–1893. [CrossRef]
19. Zhang, L.; Peng, Z. Infrared small target detection based on partial sum of the tensor nuclear norm. *Remote. Sens.* **2019**, *11*, 382. [CrossRef]
20. 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.
21. Dai, Y.; Wu, Y.; Zhou, F.; Barnard, K. Attentional local contrast networks for infrared small target detection. *IEEE Trans. Geosci. Remote. Sens.* **2021**, *59*, 9813–9824. [CrossRef]
22. Zhang, M.; Zhang, R.; Yang, Y.; Bai, H.; Zhang, J.; Guo, J. Isnet: Shape matters for infrared small target detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 877–886.
23. Wu, X.; Hong, D.; Chanussot, J. Uiunet: U-net in u-net for infrared small object detection. *IEEE Trans. Image Process.* **2022**, *32*, 364–376. [CrossRef] [PubMed]
24. Li, B.; Xiao, C.; Wang, L.; Wang, Y.; Lin, Z.; Li, M.; An, W.; Guo, Y. Dense nested attention network for infrared small target detection. *IEEE Trans. Image Process.* **2022**, *32*, 1745–1758. [CrossRef]
25. Hou, Q.; Wang, Z.; Tan, F.; Zhao, Y.; Zheng, H.; Zhang, W. RISTDnet: Robust infrared small target detection network. *IEEE Geosci. Remote. Sens. Lett.* **2021**, *19*, 7000805. [CrossRef]
26. Rawat, S.S.; Verma, S.K.; Kumar, Y. Review on recent development in infrared small target detection algorithms. *Procedia Comput. Sci.* **2020**, *167*, 2496–2505. [CrossRef]
27. Li, X.; He, H.; Li, X.; Li, D.; Cheng, G.; Shi, J.; Weng, L.; Tong, Y.; Lin, Z. Pointflow: Flowing semantics through points for aerial image segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 4217–4226.
28. Lin, T.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
29. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
30. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
31. Ying, X.; Liu, L.; Wang, Y.; Li, R.; Chen, N.; Lin, Z.; Sheng, W.; Zhou, S. Mapping degeneration meets label evolution: Learning infrared small target detection with single point supervision. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 15528–15538.
32. Hu, J.; Shen, L.; Sun, G. Squeeze-andexcitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.