DEFANet: Dual-Path Edge-Target Collaboration with Frequency-Aware Enhancement for Infrared Small Target Detection

Shuaiyuan Du, Yang Xiao, Zhiguo Cao*

School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, China
sydu@hust.edu.cn; Yang_Xiao@hust.edu.cn; zgcao@hust.edu.cn

Abstract

Infrared small target detection is challenging due to limited target size and low signal-to-noise ratio. Unlike common targets, infrared small targets contain a higher proportion of edge pixels and exhibit blurred boundaries due to diffraction and quantization artifacts, making boundaries uniquely valuable cues for target perception. However, existing methods often emphasize holistic modeling while underutilizing such informative boundary cues. Motivated by this observation, we propose a Dual-Path Edge-Guided Frequency-Aware Network (DEFANet), which enables edge-target collaborative modeling for enhanced feature representation. DEFANet features a dual-path design, consisting of a main branch for holistic target modeling and an edge branch for boundary transition perception. To facilitate interaction and enhance representation in both branches, we introduce two core modules: Frequency-Aware Dual Enhancement Module (FADE) and Edge-Guided Integration Module (EGI). FADE employs a Frequency-Decoupled Attention Enhancement Mechanism to enhance both branches in the frequency domain, strengthening holistic modeling in the main branch and boundary representation in the edge branch. EGI leverages a Dual-Path Group-Wise Guidance Mechanism to integrate enhanced edge features into the main branch, improving boundary perception. Extensive experiments on four public infrared small target datasets, MDvsFA, LAFT, SIRST, and SIATD, demonstrate that DEFANet achieves SOTA performance. Ablation studies further validate the effectiveness of DEFANet and the soundness of its design motivation.

Introduction

Infrared imaging offers strong anti-interference capability and reliable performance under both day and night conditions, making it an essential tool in modern long-range surveillance systems (Zhao et al. 2022; Zhang et al. 2025a; Pang et al. 2022). As one of its core tasks, infrared small target detection (ISTD) is widely applied in scenarios such as flight guidance (Wang et al. 2023; Guo et al. 2023), target warning (Ling et al. 2024; Zhao et al. 2023), ground surveillance (Xu et al. 2023b), maritime monitoring (Chen et al. 2025; Duan et al. 2024), and personnel search (Yang et al. 2025; Zhang et al. 2025b), and has received sustained research interest in recent years.

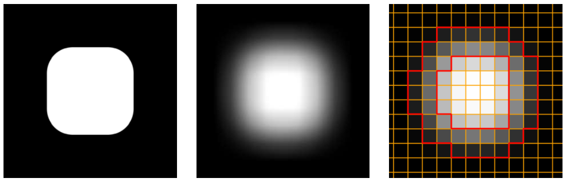

Figure 1: Illustration of diffraction and quantization effects in infrared small target imaging. Orange grids denote pixels, and the region enclosed by red boundaries indicates the transitional zone between the target and background.

Infrared small target detection approaches can generally be categorized into traditional approaches and deep learning-based approaches, depending on whether deep learning is employed for feature modeling. Traditional methods typically rely on handcrafted features based on prior knowledge (Wang et al. 1996; Chen et al. 2013; Aghaziyarati, Moradi, and Talebi 2019; Du, Wang, and Cao 2022), and thus exhibit strong dependence on specific scenes or target types, leading to limited adaptability and poor robustness in complex environments.

In contrast, deep learning methods adopt an end-to-end paradigm for automatic feature learning, enabling effective fusion of local details and global semantic information. In recent years, they have achieved remarkable progress in ISTD. Related studies have explored diverse strategies from the perspectives of target modeling, attention mechanisms, and multi-scale feature fusion. For example, MDvsFA is the first to treat missed detection and false alarm as complementary sub-tasks, constructing a dual-generator framework to optimize detection balance (Wang, Zhou, and Wang 2019). IAANet adopts a coarse-to-fine framework, where an intra-pixel attention mechanism is introduced in the fine segmentation stage to model pixel-wise correlations, improving detection performance in complex backgrounds (Wang et al. 2022). BPRNet employs an explicit multi-scale attention mechanism to enhance interactions between shallow and deep features, improving the exploitation of high-level representations and achieving a better balance between de-

---

[*] Corresponding author: Zhiguo Cao.

Network	MDvsFA	LAFT	SIRST	Mean	Δ Mean
Baseline	0.582	0.188	0.679	0.483	-
Baseline+Easy ECT	0.598	0.202	0.688	0.496	$+2.7\%$
Baseline+Easy ECT +Easy EGI	0.621	0.239	0.704	0.521	$+7.9\%$

Table 1: Effectiveness of edge information. "Easy ECT" denotes auxiliary joint learning, and "Easy EGI" refers to a simplified strategy that directly injects edge features into the main branch. "Δ" indicates the relative improvement.

tection precision and recall (Du, Wang, and Cao 2023).

Despite recent progress in ISTD, most methods primarily focus on the holistic features of the infrared small target, while paying limited attention to the edge regions and their potential value. In fact, edge regions, as an integral part of infrared small targets, highlight the gray-level transitions between the target and the background, offering unique spatial cues that differ from those in the central region. On the one hand, unlike common targets, infrared small targets are extremely small in size, and the edge pixels account for a large proportion of the target area, making them a critical component of the target itself. On the other hand, due to optical diffraction and quantization artifacts introduced during imaging, the boundaries of such targets often appear blurred and gradually blend into the background. As shown in Fig. 1, these proportionally large and characteristic edge regions constitute one of the key features of infrared small targets and hold potential for independent modeling.

To verify the modeling value of edge regions in infrared small targets, we introduce an auxiliary edge branch into a baseline network and assess its impact on the performance of the main segmentation branch. As shown in Table 1, the average F1 score shows a 2.7% relative improvement when jointly trained with the edge branch, even without feature interaction. When intermediate features from the edge branch are further fused into the main branch, the relative improvement increases to 7.9%. These simple experiments suggest that the edge regions of infrared small targets contain stable, perceptible gray-level transition features that warrant dedicated modeling to provide performance improvements.

These findings reveal the modeling potential of edge regions and motivate us to explicitly incorporate edge-aware mechanisms into network design. To this end, we propose a Dual-path Edge-guided Frequency-aware Network (DEFANet), which promotes collaborative modeling between edge and target features. Built upon the UNet framework, DEFANet introduces a dual-path architecture composed of a main branch for modeling the whole infrared small target and an edge branch focusing on edge information. To enhance the effectiveness of the main branch, we design a Frequency-Aware Dual Enhancement (FADE) Module, which leverages a frequency-decoupled attention enhancement mechanism to separately model and fuse high- and low-frequency components. This not only enhances the structural representation of the target within the main branch but also provides the edge branch with dedicated high-frequency cues. Furthermore, to enable edge–target interaction, we propose an Edge-Guided Integration Module (EGI), which employs a dual-path group-wise guidance mechanism to inject edge features into the main branch, thereby enhancing the main branch's awareness of boundary regions. These two modules work synergistically to realize the main philosophy of edge–target collaborative modeling and jointly improve detection performance. Unlike previous edge-based methods (e.g., ISNet (Zhang et al. 2022), GCI-Net (Zhang et al. 2024)) that explicitly reconstruct or regularize edges, DEFANet leverages infrared transition zones to achieve implicit edge–target collaboration, providing adaptive structural guidance during feature learning.

In summary, the main contributions of this study are as follows:

- We analyze the unique characteristics of edge regions in infrared small targets and demonstrate their significant value. Based on this observation, we propose the research motivation, edge–target collaborative modeling, aiming to enhance the whole target perception by explicitly modeling edge structures.
- To realize this motivation, we propose DEFANet, a dual-path architecture that integrates two key modules: FADE and EGI. FADE enhances high- and low-frequency features in decoder, thereby improving the target representation of the main branch while providing the edge branch with dedicated high-frequency cues. EGI injects edge information into the main branch, enhancing the main branch's awareness of boundary regions and promoting edge–target collaboration.
- DEFANet achieves SOTA performance on four public ISTD datasets, demonstrating strong robustness and generalization. Furthermore, extensive ablation and analysis experiments verify that the design of the proposed model aligns well with our motivation.

Related Work

Traditional Methods

Early ISTD methods enhance low-SNR infrared small targets by designing handcrafted features. Among them, local contrast-based techniques constitute one of the most representative methods, aiming to amplify target saliency by comparing regional responses. For example, LCM (Chen et al. 2013) compares the maximum contrast between target and background regions, while MPCM (Wei, You, and Li 2016) improves robustness at complex edges via multi-directional patch-based contrast.

In addition to contrast-based strategies, local statistical and filtering-based approaches have also shown effectiveness. LHM (Nie et al. 2018) uses local homogeneity to highlight anomalies, and WLDM (Deng et al. 2016) leverages entropy to suppress noise. MDRTH (Wang and Wang 2021) incorporates multi-orientation structuring to enhance directional sensitivity in detection.

Several clutter suppression methods further boost robustness in complex scenes. NTFRA (Kong et al. 2022) utilizes low-rank tensor modeling, FKRW (Qin et al. 2019) employs diffusion-based segmentation, and IWS (Zhang et al.

2025c) combines multi-subspace modeling with structure tensor analysis to enhance dim target sparsity.

Deep-Learning Methods

With the rise of deep learning, data-driven methods have significantly advanced ISTD by enabling more effective feature representation. These methods typically leverage convolutional neural networks (CNNs) to capture a broader range of features, balancing local and global information (Wu, Hong, and Chanussot 2023; Li et al. 2023). MDvsFA (Wang, Zhou, and Wang 2019) utilizes adversarial learning with shallow and deep subnetworks to balance recall and precision.

To further improve multi-scale feature modeling, Xu et al. (Xu et al. 2023a) incorporate grid resampling and target-background differentiation strategies. Liu et al. (Liu et al. 2023) enhance target-background discrimination by combining the Wasserstein distance with multi-channel fusion. DMFNet (Guo et al. 2024) introduces a dual-encoder structure with receptive field expansion and cross-layer fusion to boost detection sensitivity. PGDN-Net (Liu et al. 2025) integrates structural priors—including Riesz transform and tensor corner features—into a densely nested backbone to guide saliency-aware feature extraction.

Contextual modeling is addressed in ACM (Dai et al. 2021) via an asymmetric modulation block that facilitates long-range interaction between local and global contexts. ALCNet (Dai et al. 2022) complements coarse-to-fine fusion with cyclic feature shifts to enhance spatial continuity. Further improvements are made by IAANet (Wang et al. 2022), which introduces a two-stage coarse-to-fine pipeline combining RPN proposals and transformer-based global interaction. Recently, MiM-ISTD (Chen et al. 2024) leverages the Mamba-in-Mamba architecture to hierarchically model global-local dynamics using a state-space approach, achieving strong performance with reduced complexity.

Recent edge-based methods further explored edge information for ISTD. ISNet (Zhang et al. 2022) incorporates a Taylor finite difference–inspired edge block and a two-orientation attention aggregation module to reconstruct clear target shapes by aggregating multi-level edge cues. GCI-Net (Zhang et al. 2024) introduces a Gaussian curvature branch to extract structural curvature features and suppress speckle noise, assisted by a patch–group attention for cross-scale feature fusion.

Unlike these edge-based frameworks, our DEFANet does not explicitly reconstruct or regularize edge maps. Instead, it leverages infrared transition zones, allowing edge cues to participate directly in mid-layer feature generation through edge–target collaboration. This design models morphological transition bands rather than binary "hard" edges, enabling implicit multi-level fusion and adaptive guidance with greater learning flexibility. Consequently, DEFANet can achieve effective integration of boundary-aware structural information and semantic representation for robust infrared small target detection.

Methodology

Overview

To effectively detect infrared small targets and leverage their characteristic edge information, we propose a Dual-path Edge-guided Frequency-aware Network (DEFANet), a dual-path network based on the UNet framework, as shown in Fig. 2. It consists of a main segmentation branch and an auxiliary edge branch, and integrates two key modules, FADE and EGI, into the decoder to enhance collaboration between the two branches. The overall pipeline consists of four stages:

Input and Encoder. The input $q \in \mathbb{R}^{1\times H\times W}$ is processed by a shared encoder to extract multi-scale features $\{F_i\}$ for both branches.

Frequency-Aware Dual Enhancement (FADE). Each F_i is transformed to the frequency domain via DCT and split into high- and low-frequency components, F_i^{fh} and F_i^{fl}. The main branch applies attention and fuses both components, then uses IDCT to obtain enhanced features F_i'. The edge branch enhances only the high-frequency part to obtain E_i'.

Edge-Guided Integration (EGI). EGI takes F_i' and E_i' as input and uses a dual-path group-wise mechanism to generate the fused feature F_i'', enhancing the main branch's boundary awareness.

Decoding and Supervision. The fused features F_i'' are decoded to produce the final segmentation map $\hat{M}$, from which detection results are obtained via thresholding. In parallel, the edge branch outputs an edge map $\hat{E}$. During training, both outputs are supervised; in inference, only $\hat{M}$ is used.

Frequency-Aware Dual Enhancement (FADE)

Infrared small target information is mainly distributed in high-frequency bands, while backgrounds are concentrated in low-frequency bands. Based on this observation, incorporating frequency-domain information to enhance spatial features can improve the network's sensitivity to small targets. To this end, we introduce a Frequency-Aware Dual Enhancement (FADE) Module into the decoder, which extracts frequency-domain representations from the features of the main segmentation branch and applies them to enhance both the main and edge branches, as shown in Fig. 2. Specifically, the frequency features are decoupled into high- and low-frequency components. The high-frequency features are utilized to guide the structural modeling in the edge branch, while both components are retained in the main branch to represent the target's structure and semantic context.

Specifically, the feature map F_i is first transformed into the frequency domain using the discrete cosine transform (DCT), yielding F_i^f. We then separate it into a low-frequency component F_i^{fl} and a high-frequency component F_i^{fh}. In the edge branch, only the high-frequency component F_i^{fh} is utilized. We apply global average pooling to F_i^{fh} to obtain a channel descriptor of shape $(B, C, 1, 1)$,

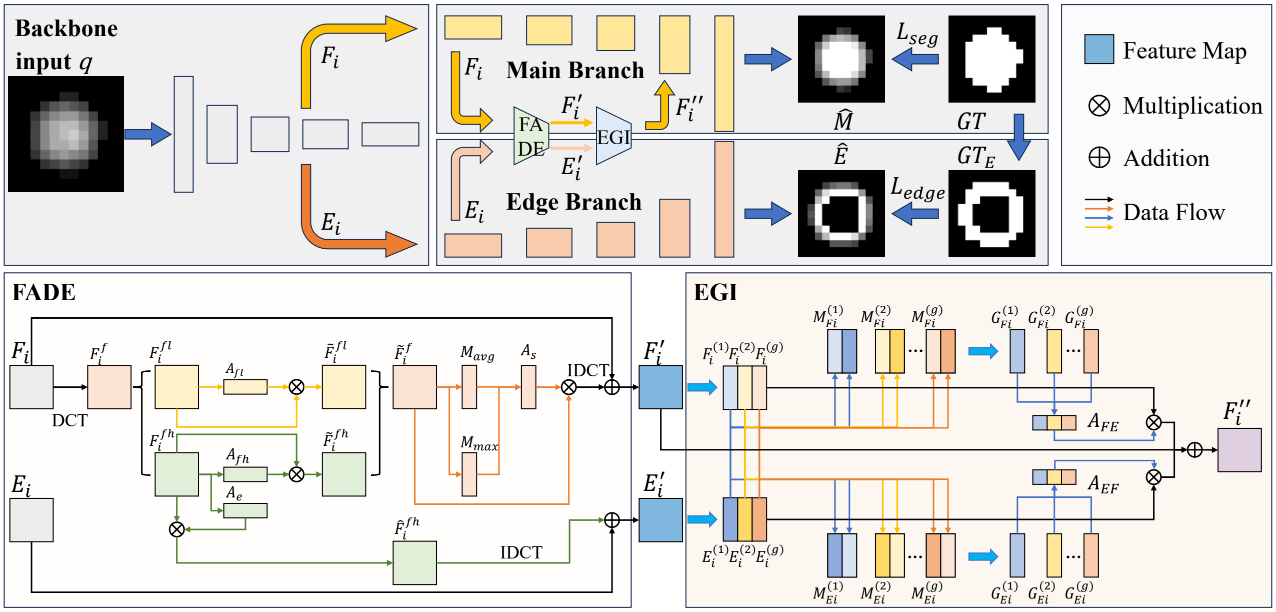

Figure 2: Pipeline of DEFANet. DEFANet adopts a dual-branch encoder–decoder architecture, where the main branch focuses on infrared small target detection, and the edge branch models boundary transition zones to guide the main branch. The decoder integrates two key modules: FADE enhances frequency-specific representations in both branches, while EGI performs dual-path group-wise guidance to fuse edge-aware features into the main branch for improved boundary perception.

which is passed through convolutional layers and ReLU to generate a channel attention vector A_e for modulating the channel-wise response:

$$A_e = \sigma(\mathrm{Conv}_2(\mathrm{ReLU}(\mathrm{Conv}_1(\mathrm{GAP}(F_i^{fh}))))), \tag{1}$$

where $\sigma(\cdot)$ denotes the Sigmoid function. The attention vector is then applied to the high-frequency feature to obtain $\hat{F}_i^{fh} = F_i^{fh} \cdot A_e$. Since $\hat{F}_i^{fh}$ only contains partial frequency components, we apply zero-padding to restore it to the original size, followed by an inverse DCT (IDCT) to map it back into the spatial domain as the enhanced edge feature. To retain the original edge feature information, we add a residual connection after the transformation:

$$E_i' = \mathrm{IDCT}(\mathrm{Pad}(\hat{F}_i^{fh})) + E_i, \tag{2}$$

where $\mathrm{Pad}(\cdot)$ denotes the zero-padding operation applied in the frequency domain.

In contrast to the edge branch, the main branch utilizes both the high-frequency component F_i^{fh} and the low-frequency component F_i^{fl}. We first compute two separate channel attention vectors, A_{fh} and A_{fl}, using the same mechanism as A_e, to modulate the channel-wise responses of F_i^{fh} and F_i^{fl}, respectively. The resulting weighted features are denoted as $\tilde{F}_i^{fh}$ and $\tilde{F}_i^{fl}$. These two components are then merged in the frequency domain to obtain a complete spectral representation:

$$\tilde{F}_i^{f} = \mathrm{Merge}(\tilde{F}_i^{fl}, \tilde{F}_i^{fh}). \tag{3}$$

Next, a spatial attention mechanism is applied to the merged frequency-domain feature to enhance the model's localization ability. Specifically, average pooling and max pooling are performed along the channel dimension to produce two spatial maps, which are concatenated and passed through a convolution layer to generate the spatial attention map A_s:

$$\begin{aligned} A_s &= \sigma(\mathrm{Conv}([M_{avg}; M_{max}])), \\ M_{avg} &= \mathrm{AvgPool}_{channel}(\tilde{F}_i^{f}), \\ M_{max} &= \mathrm{MaxPool}_{channel}(\tilde{F}_i^{f}). \end{aligned} \tag{4}$$

Finally, the enhanced frequency feature is mapped back to the spatial domain via inverse DCT (IDCT) and fused with the original main branch feature through a residual connection to obtain the enhanced representation:

$$F_i' = \mathrm{IDCT}(\tilde{F}_i^{f} \cdot A_s) + F_i. \tag{5}$$

Edge-Guided Integration (EGI)

Infrared small targets typically exhibit a blurred transitional structure at the boundaries, where the intensity gradually attenuates from the center to the background, as shown in Fig. 1. This boundary transition zone contains discriminative information beneficial for infrared small target detection. To exploit this, we propose an Edge-Guided Integration (EGI) module as shown in Fig. 2, which extracts features from the boundary transition region and injects them into the main branch in a guided manner to enhance its target modeling capability.

To facilitate the structural guidance of the main branch by edge information, we propose a Dual-Path Group-wise

Guidance Mechanism. This mechanism maintains the parallel structure of the two branches while performing grouped operations on their feature channels, enabling fine-grained interaction and flexible fusion of edge features.

Specifically, the main branch feature and the edge feature, $F_i', E_i' \in \mathbb{R}^{C\times H\times W}$, are divided into g channel groups.

$$\begin{aligned} F_i' &= [F_i^{(1)}; ...; F_i^{(g)}], \\ E_i' &= [E_i^{(1)}; ...; E_i^{(g)}]. \end{aligned} \tag{6}$$

For group j, we concatenate $F_i^{(j)}$ and $E_i^{(j)}$ along the channel dimension and apply convolution to extract the group-wise guided feature $G_{Fi}^{(j)}$.

$$G_{Fi}^{(j)} = \mathrm{Conv}(M_{Fi}^{(j)}) = [F_i^{(j)}; E_i^{(j)}]. \tag{7}$$

All $G_{Fi}^{(j)}$ are then concatenated along the channel dimension to form the integrated guidance feature, which is used to generate the attention vector A_{FE} for modulating the channel responses of the main branch feature F_i'.

In the edge branch, we adopt the same procedure as in the main branch: for each group, the edge feature $E_i^{(j)}$ is concatenated with the corresponding main branch feature $F_i^{(j)}$, and the guided feature $G_{Ei}^{(j)}$ is generated.

$$G_{Ei}^{(j)} = \mathrm{Conv}(M_{Ei}^{(j)}) = [E_i^{(j)}; F_i^{(j)}]. \tag{8}$$

All group-wise guided features are then concatenated to form the edge guidance vector A_{EF}, which is used to modulate the channel responses of the edge feature E_i'. Finally, the fused feature F_i'' is obtained by combining the residual connection from the main branch and the dual-path guidance results, as formulated below:

$$F_i'' = F_i' + A_{FE} \cdot F_i' + A_{EF} \cdot E_i'. \tag{9}$$

Additionally, we apply a supervision signal to the output of the edge branch, $\hat{E}$. Given that infrared small targets typically exhibit blurred and gradually attenuating boundaries, we construct a supervision label that focuses on the transitional region rather than directly extracting edges from the targets' GT. This encourages the network to attend to the boundary transition zone. Specifically, the supervision label is defined as the difference between the morphological dilation and erosion of the mask GT:

$$GT_E = \mathrm{Dilate}(GT) - \mathrm{Erode}(GT), \tag{10}$$

where $\mathrm{Dilate}(\cdot)$ and $\mathrm{Erode}(\cdot)$ denote the dilation and erosion operations, respectively. The edge branch is trained using the standard binary cross-entropy loss.

Training and Inference

During training, we apply supervision to both the segmentation map $\hat{M}$ predicted by the main branch and the edge map $\hat{E}$ predicted by the edge branch. Both outputs are optimized using the binary cross-entropy (BCE) loss. Specifically, the supervision label for the edge branch is constructed as the difference between the dilated and eroded versions of the original GT mask. The total loss is defined as:

$$\begin{aligned} L &= L_{seg} + \lambda L_{edge} \\ &= L_{BCE}(\hat{M}, GT) + \lambda L_{BCE}(\hat{E}, GT_E), \end{aligned} \tag{11}$$

where λ denotes the weighting factor for the edge loss. During inference, only the segmentation output $\hat{M}$ from the main branch is used to generate the final result.

Experiments

Setup

Datasets. Following the experimental protocol of BPR-Net (Du, Wang, and Cao 2023) and IAANet (Wang et al. 2022), we evaluate our method on four public datasets: MDvsFA (Wang, Zhou, and Wang 2019), LAFT (Hui et al. 2019), SIRST (Dai et al. 2022), and SIATD (Sun et al. 2021) under two scenarios. In the first scenario, the model is trained on the MDvsFA training set (10,000 images) and tested on MDvsFA (100 images), LAFT (140 images), and SIRST (427 images) to assess its cross-dataset generalization. In the second scenario, the model is trained and tested on the SIATD dataset, which contains 150,185 images with a resolution of 640×512. SIATD is evenly divided into training and testing subsets. It is characterized by urban environments and complex ground scenes, making it well-suited for evaluating model performance under challenging conditions.

Metrics. We adopt pixel-level F1 score, Intersection over Union (IoU), and normalized IoU (nIoU) (Du, Wang, and Cao 2023; Dai et al. 2022; Wang, Zhou, and Wang 2019; Wang et al. 2022) in Scenario 1 to measure the spatial overlap between the predicted regions and the ground truth (GT). The metrics are computed as follows:

$$\begin{aligned} &\mathrm{Prec} = \frac{tp}{tp+fp}, \mathrm{Rec} = \frac{tp}{tp+fn}, F1 = \frac{2\cdot\mathrm{Prec}\cdot\mathrm{Rec}}{\mathrm{Prec}+\mathrm{Rec}}, \\ &\mathrm{IoU} = \frac{1}{Q}\sum_{i=1}^{Q}\frac{tp_i}{tp_i+fp_i+fn_i}, \mathrm{nIoU} = \frac{1}{Q}\sum_{i=1}^{Q}\frac{tp_i}{t_i+fp_i}, \end{aligned} \tag{12}$$

where tp, fp, and fn denote the number of true positives, false positives, and false negatives at the pixel level, respectively. t_i represents the number of target pixels in the i-th image, and Q denotes the total number of images. In Scenario 2, we use the official F1 score provided by SIATD to ensure fair comparison (Sun et al. 2021).

Implementation Details. All experiments are conducted on a workstation equipped with an RTX 3090 GPU. The network is trained using SGD with an initial learning rate of 0.25, decayed to 0.2× at the 5th, 15th, and 25th epochs via StepLR. The batch size is 24, and the total epochs are 30. Hyper-parameters λ and g are set to 0.2 and 4, respectively. All reported results are averaged over three independent runs to ensure statistical reliability.

Method	DL	Scenario 1									Scenario 2
		MDvsFA			LAFT			SIRST			SIATD
		F1	IoU	nIoU	F1	IoU	nIoU	F1	IoU	nIoU	F1 (Prec, Rec)
MaxMean (Deshpande et al. 1999)	N	0.047	0.007	0.019	0.000	0.000	0.000	0.020	0.006	0.008	0.001 (0.001, 0.436)
LCM (Chen et al. 2013)		0.114	0.003	0.061	0.008	0.001	0.004	0.192	0.003	0.108	0.001 (0.000, 0.617)
MPCM (Wei, You, and Li 2016)		0.351	0.174	0.220	0.217	0.093	0.142	0.398	0.200	0.249	0.004 (0.002, 0.062)
LHM (Nie et al. 2018)		0.239	0.095	0.126	0.106	0.070	0.060	0.281	0.111	0.154	0.003 (0.001, 0.031)
FKRW (Qin et al. 2019)		0.311	0.145	0.194	0.168	0.064	0.106	0.439	0.203	0.292	0.006 (0.003, 0.034)
MDRTH (Wang and Wang 2021)		0.158	0.036	0.074	0.298	0.108	0.243	0.219	0.115	0.128	0.012 (0.006, 0.195)
NTFRA (Kong et al. 2022)		0.267	0.051	0.179	0.144	0.025	0.099	0.405	0.009	0.297	0.001 (0.000, 0.368)
MDvsFA (Wang, Zhou, and Wang 2019)	Y	0.594	0.454	0.427	0.157	0.047	0.097	0.706	0.519	0.578	0.715 (0.980, 0.562)
ACM (Dai et al. 2021)		0.578	0.399	0.402	0.198	0.081	0.125	0.646	0.435	0.501	0.038 (0.020, 0.799)
ALCNet (Dai et al. 2022)		0.585	0.413	0.414	0.202	0.092	0.122	0.652	0.401	0.508	0.567 (0.429, 0.840)
ISNet (Zhang et al. 2022)		0.643	0.498	0.478	0.233	0.088	0.118	0.735	0.528	0.569	0.766 (0.894, 0.670)
IAANet (Wang et al. 2022)		0.670	0.508	0.504	0.282	0.093	0.193	0.764	0.541	0.608	0.806 (0.987, 0.681)
MTUNet (Wu et al. 2023)		0.659	0.467	0.483	0.251	0.091	0.136	0.718	0.507	0.554	0.783 (0.692, 0.903)
DNANet (Li et al. 2023)		0.662	0.438	0.491	0.299	0.113	0.189	0.772	0.563	0.613	0.849 (0.948, 0.769)
UIUNet (Wu, Hong, and Chanussot 2023)		0.641	0.473	0.477	0.278	0.102	0.164	0.709	0.511	0.559	0.822 (0.899, 0.757)
BPRNet (Du, Wang, and Cao 2023)		0.678	0.487	0.522	0.362	0.257	0.276	0.783	0.630	0.634	0.941 (0.968, 0.915)
MSH (Liu et al. 2024)		0.617	0.441	0.463	0.189	0.078	0.099	0.734	0.558	0.589	0.789 (0.857, 0.730)
EGPNet (Li et al. 2024)		0.672	0.511	0.522	0.305	0.158	0.208	0.790	0.593	0.629	0.877 (0.913, 0.844)
SCTransNet (Yuan et al. 2024)		0.680	0.508	0.526	0.324	0.197	0.281	0.793	0.582	0.632	0.923 (0.909, 0.937)
MLPNet (Wang et al. 2025)		0.678	0.513	0.519	0.336	0.202	0.265	0.788	0.604	0.634	0.912 (0.886, 0.940)
DEFANet (Ours)		**0.703**	**0.558**	**0.531**	**0.393**	**0.261**	**0.288**	**0.809**	**0.639**	**0.659**	**0.953** (0.977, 0.930)

Table 2: Quantitative comparison of different methods on Scenario 1 (MDvsFA, LAFT, SIRST) and Scenario 2 (SIATD). "DL" indicates whether the method is based on deep learning.

Comparisons with SOTA Methods

Scenario 1. As shown in Table 2, our method achieves superior performance across all test sets (MDvsFA, LAFT, and SIRST) in Scenario 1. Compared to related approaches, the proposed joint modeling of target and boundary transition regions significantly enhances the model's adaptability to diverse data distributions. The improvements on LAFT and SIRST further demonstrate the strong generalization potential of our strategy.

Scenario 2. In Scenario 2 (SIATD), as reported in Table 2, the dataset features smaller targets and predominantly complex urban backgrounds, posing significant challenges. Experimental results indicate that traditional methods exhibit limited effectiveness on SIATD, and even some deep learning-based approaches struggle to converge. In contrast, our method still achieves the best results, further validating the stability and robustness of the proposed strategy under challenging conditions.

Ablation Studies

We conduct ablation studies on Scenario 1 (MDvsFA, LAFT, SIRST) as shown in Table 3. To evaluate the contribution of each module, we design multiple network configurations, including Baseline, Easy ECT, Easy EGI, EGI, and FADE, aiming to progressively analyze the impact of individual components on performance. Specifically, Easy ECT introduces an auxiliary edge branch during training for joint optimization, while Easy EGI further integrates edge features into the main segmentation branch via a simple self-attention module. It is worth noting that in the absence of the edge branch, FADE degenerates into a single-path enhancement module, enabling us to assess its standalone effectiveness. The reported performance improvements or drops in tables and text refer to relative gains or reductions compared to the corresponding baselines or configurations.

ID	Easy ECT	Easy EGI	EGI	FADE	MDvsFA	LAFT	SIRST	Mean
A1					0.582	0.188	0.679	0.483
A2	✓				0.598	0.202	0.688	0.496
A3		✓			0.621	0.239	0.704	0.521
A4			✓		0.668	0.324	0.748	0.580
A5				✓	0.643	0.271	0.741	0.552
A6	✓			✓	0.665	0.295	0.753	0.571
A7		✓		✓	0.686	0.357	0.779	0.607
A8			✓	✓	0.703	0.393	0.809	0.635

Table 3: Results of ablation studies.

Effectiveness of Dual-Path Group-Wise Guidance Mechanism. Compared to A1 (Baseline), A2 (Easy ECT) achieves a 2.7% average F1 improvement, indicating that simply involving edge supervision during training can enhance the backbone's feature extraction capacity. Further integrating edge features into the main branch via Easy EGI (A3) boosts performance to 7.9%, validating the contribution of edge cues to improving target modeling. Building upon this, A4 (EGI) adopts our proposed Dual-Path Group-wise Guidance Mechanism to better exploit the guidance from the edge branch, resulting in a substantial 20.1% F1 improvement.

Effectiveness of Frequency-Aware Enhancement. Without involving edge modeling, integrating the FADE

ID	MB	EB	MDvsFA	LAFT	SIRST	Mean	Δ Mean
B1	None	None	0.668	0.324	0.748	0.580	-
B2	None	H	0.679	0.342	0.778	0.600	+3.4%
B3	None	L	0.638	0.221	0.705	0.521	−10.1%
B4	None	H+L	0.674	0.332	0.761	0.589	+1.6%
B5	H	None	0.670	0.337	0.759	0.589	+1.5%
B6	L	None	0.666	0.331	0.739	0.579	−0.2%
B7	H+L	None	0.689	0.369	0.783	0.614	+5.8%
B8	H+L	H	0.703	0.393	0.809	0.635	+9.5%

Table 4: Results of the frequency decoupling strategy analysis. "MB" and "EB" denote the feature configurations of the main and edge branches, respectively. "H" and "L" represent the high-frequency and low-frequency components, respectively.

into the main branch (A5) yields a 14.3% average F1 improvement, confirming the effectiveness of frequency-domain information in ISTD. When combined with the edge branch, a clear trend of synergistic enhancement emerges: the performance gain from Easy ECT alone is 2.7% (A1 vs. A2), which increases to 3.9% when FADE is added (A5 vs. A6); similarly, the improvement from Easy EGI rises from 7.9% (A1 vs. A3) to 11.4% with FADE (A5 vs. A7). This consistent improvement across all datasets indicates that FADE not only enhances the representational capacity of the main branch but also reinforces edge features, thereby providing a more robust foundation for the dual-branch mechanism.

Combined Effect of EGI and FADE. When EGI and FADE are integrated simultaneously (A8), the network achieves the best performance, with an average F1 improvement of 31.5%. Notably, on the LAFT dataset, the performance gain reaches 109.0%, indicating that the proposed strategy is well suited for detecting targets with low SNR and vague contours. This further confirms the synergistic effect of the two modules.

Analysis Studies

To further verify whether the performance gains originate from the proposed motivations, we conduct three analysis experiments focusing on the following questions: i) Is the frequency-decoupling strategy in FADE suitable for ISTD? ii) Which level is optimal for embedding EGI? iii) Does the main branch effectively utilize edge-guided information?

Validation of Frequency-Decoupled Strategy. To verify the rationality of the frequency-decoupled design and branch-specific adaptation in FADE, we test different frequency combinations in both branches (Table 4).

B1 serves as DEFANet without FADE. Introducing high-frequency features into the edge branch (B2) yields a +3.4% F1 improvement, while low-frequency input (B3) causes a 10.1% drop, indicating that high-frequency features benefit edge modeling, whereas low-frequency ones introduce interference. Dual-frequency input (B4) performs slightly worse than B2, suggesting that FADE alleviates the adverse effect of low-frequency features on edge modeling.

ID	Layer	MDvsFA	LAFT	SIRST	Mean	Δ Mean
C1	None	0.643	0.271	0.741	0.552	-
C2	1	0.652	0.301	0.760	0.571	+3.4%
C3	2	0.694	0.359	0.779	0.611	+10.6%
C4	3	0.682	0.338	0.772	0.597	+8.2%
C5	4	0.647	0.295	0.749	0.564	+2.1%
C6	5	0.638	0.272	0.737	0.549	−0.5%
C7	all	0.699	0.395	0.811	0.637	+15.4%
C8	2,3	0.703	0.393	0.809	0.635	+15.0%

Table 5: Results of embedding position analysis for EGI.

For the main branch, single-frequency input (B5/B6) yields marginal changes (+1.5%, −0.2%), while combining both frequencies (B7) brings a +5.8% gain, showing that full-frequency cues are crucial for infrared small target perception and FADE can also improve dual-band integration.

Comparing B2, B7, and B8 further confirms that assigning high-frequency features to the edge branch and full-frequency features to the main branch boosts performance by +9.5%, validating the effectiveness of the decoupled and branch-specific frequency design.

In summary, the frequency separation, attention-based fusion, and dual-branch frequency allocation strategies in FADE are well-motivated and empirically effective.

Validation of EGI. To evaluate the layer-wise adaptability of the Edge-Guided Integration (EGI) module, we insert it into different layers. As shown in Table 5, C1 denotes DEFANet without EGI.

Comparing C2–C6, introducing EGI at the 2nd and 3rd (C3, C4) layers yields notable F1 improvements (+10.6%, +8.2%), while the 1st and 4th layers show limited gains (+3.4%, +2.1%). Performance even drops slightly (−0.5%) at the 5th layer (C6), suggesting that mid-level features are more suitable for edge guidance, whereas shallow layers are noise-prone and deep layers suffer from degraded target features.

Further comparing C8 (EGI at only the 2nd/3rd layers) and C7 (EGI at all layers), we observe that C8 already achieves near-optimal performance (+15.0% vs. +15.4%). This indicates that edge-relevant guidance information is primarily concentrated in the shallow-to-middle layers, with limited or even adverse effects from other levels.

Moreover, EGI demonstrates good robustness, showing no significant degradation even when applied to less compatible layers (e.g., −0.5% at C6). This confirms the effectiveness of its group-wise guidance mechanism in selecting features, supporting the design rationale of EGI.

Validation of Dual-Path Group-Wise Guidance Mechanism. To verify whether the main branch truly relies on edge information for modeling, we conduct a pathway effectiveness analysis (Table 6) by controlling the training status and guidance information of the edge branch.

D1 denotes DEFANet without the edge branch. D2 adds a supervised edge branch without employing EGI. D3 keeps the EGI module but removes edge supervision, allowing us to assess whether the structure alone contributes to performance. Furthermore, in the complete DEFANet, D4–D6 re-

ID	EB	S	EGI	EGI Input	MDvsFA	LAFT	SIRST	Mean	Δ Mean
D1					0.643	0.271	0.741	0.552	-
D2	✓	✓			0.665	0.295	0.753	0.571	$+3.4\%$
D3	✓		✓	N	0.659	0.292	0.749	0.567	$+2.7\%$
D4	✓	✓	✓	0	0.597	0.199	0.678	0.491	-11.0%
D5	✓	✓	✓	R	0.438	0.114	0.499	0.350	-36.5%
D6	✓	✓	✓	N	0.703	0.393	0.809	0.635	$+15.0\%$

Table 6: Results of the effectiveness analysis for the dual-path group-wise guidance mechanism. "EB" and "S" denote whether the edge branch is used and supervised, respectively. "EGI Input" indicates the input features for the edge branch in EGI: "N" means unmodified original input, "0" and "R" refer to replacing the input with all-zero and random feature maps, respectively.

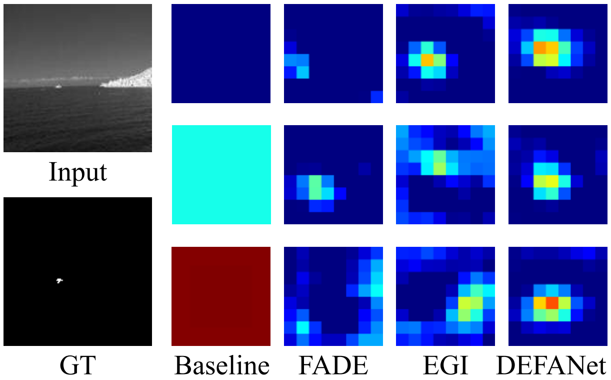

Figure 3: Deep features. From left to right, three representative feature maps are selected from the c-dimension deep semantic features of the baseline model, baseline + FADE, baseline + EGI, and our full model DEFANet, respectively.

place the edge features fed to EGI with zeros, random noise, and original features, respectively, to test whether the main branch effectively utilizes the edge-guided information.

Compared to D1, D2 achieves a 3.4% improvement, suggesting that a supervised edge branch can enhance encoder's features. D3, even without edge supervision, still brings a 2.7% improvement, demonstrating the structural effectiveness of edge branch and EGI. Additionally, D4 and D5 exhibit significant drops (-11.0% and -36.5%), confirming that main branch genuinely relies on edge features in DEFANet. Comparing D6 ($+15.0\%$) with D3 ($+2.7\%$) further confirms that the performance gain stems not from the structure itself, but from the effective guidance of edge branch.

Feature Visualization Analysis Previous studies have highlighted the challenges of infrared small target detection, among which a major issue is the loss of critical features during repeated downsampling in deep networks, due to the small size and low signal-to-noise ratio of infrared small targets (Wang, Zhou, and Wang 2019; Li et al. 2023; Du, Wang, and Cao 2023). We refer to this phenomenon as "deep feature degradation" in the following discussion. To investigate whether the proposed FADE and EGI can alleviate this issue, we visualize the deep features of the network under different configurations.

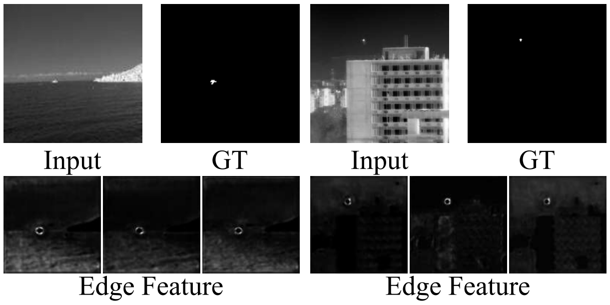

Figure 4: Edge features of DEFANet.

Specifically, Fig. 3 illustrates the deep feature maps under four network settings: Baseline, Baseline+FADE, Baseline+EGI, and the complete DEFANet. In the Baseline, the feature maps' responses are nearly uniform, with limited target-related information, indicating a clear manifestation of deep feature degradation. When the FADE module is applied (without the edge branch), frequency-aware enhancement in the main branch enables the emergence of relevant features, suggesting that frequency-domain modeling contributes to mitigating the degradation phenomenon. With the inclusion of the EGI module, the representation of the target becomes more prominent, demonstrating that the unique edge information of infrared small targets can offer more discriminative cues and thereby strengthen features.

In the complete DEFANet model, which incorporates both FADE and EGI, the deep features exhibit enhanced target awareness and improved localization accuracy, as shown in Fig. 3 (DEFANet). This indicates that the mitigation of the phenomenon of deep feature degradation is a factor contributing to reducing the false alarm rate of DEFANet in practical detection scenarios.

Meanwhile, Fig. 4 presents the edge features extracted by DEFANet, which demonstrate its ability to accurately localize the transition regions between the target and the background. These edge cues are subsequently fused with the holistic target features from the main branch, thereby enhancing the overall feature representation.

Conclusion

To enhance the performance of infrared small targets, this paper proposes DEFANet, a Dual-path Edge-guided Frequency-aware Network that facilitates edge–target collaborative modeling. By explicitly modeling the boundary transitions through a dedicated edge branch and integrating edge cues into the main branch via the proposed EGI, DEFANet effectively strengthens the awareness of infrared small targets. Additionally, FADE leverages frequency-decoupled attention to separately enhance high- and low-frequency features, enhancing both target and edge cues. Extensive experiments on four public datasets validate that DEFANet achieves SOTA performance and exhibits strong generalization capability.

References

Aghaziyarati, S.; Moradi, S.; and Talebi, H. 2019. Small infrared target detection using absolute average difference weighted by cumulative directional derivatives. *Infrared Physics & Technology*, 101: 78–87.

Chen, C. P.; Li, H.; Wei, Y.; Xia, T.; and Tang, Y. Y. 2013. A local contrast method for small infrared target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 52(1): 574–581.

Chen, S.; Ji, L.; Duan, W.; Peng, S.; and Ye, M. 2025. Motion Prior Knowledge Learning with Homogeneous Language Descriptions for Moving Infrared Small Target Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 2186–2194.

Chen, T.; Ye, Z.; Tan, Z.; Gong, T.; Wu, Y.; Chu, Q.; Liu, B.; Yu, N.; and Ye, J. 2024. Mim-istd: Mamba-in-mamba for efficient infrared small target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–13.

Dai, Y.; Wu, Y.; Zhou, F.; and Barnard. 2022. Attentional Local Contrast Networks for Infrared Small Target Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 59(11): 9813–9824.

Dai, Y.; Wu, Y.; Zhou, F.; and Barnard, K. 2021. Asymmetric Contextual Modulation for Infrared Small Target Detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1–10.

Deng, H.; Sun, X.; Liu, M.; Ye, C.; and Zhou, X. 2016. Small infrared target detection based on weighted local difference measure. *IEEE Transactions on Geoscience and Remote Sensing*, 54(7): 4204–4214.

Deshpande, S. D.; Er, M. H.; Venkateswarlu, R.; and Chan, P. 1999. Max-mean and max-median filters for detection of small targets. In *Signal and Data Processing of Small Targets*, volume 3809, 74–83. International Society for Optics and Photonics.

Du, S.; Wang, K.; and Cao, Z. 2022. From Characteristic Response to Target Edge Diffusion: An Approach to Small Infrared Target Detection. *Infrared Physics & Technology*, 124: 104214.

Du, S.; Wang, K.; and Cao, Z. 2023. Bpr-net: Balancing precision and recall for infrared small target detection. *IEEE Transactions on Geoscience and Remote Sensing*.

Duan, W.; Ji, L.; Chen, S.; Zhu, S.; and Ye, M. 2024. Triple-domain feature learning with frequency-aware memory enhancement for moving infrared small target detection. *IEEE Transactions on Geoscience and Remote Sensing*.

Guo, L.; Sun, X.; Zhang, W.; Li, Z.; and Yu, Q. 2023. Small aerial target detection using trajectory hypothesis and verification. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–14.

Guo, T.; Zhou, B.; Luo, F.; Zhang, L.; and Gao, X. 2024. DMFNet: Dual-encoder Multi-stage Feature Fusion Network for Infrared Small Target Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–14.

Hui, B.; Song, Z.; Fan, H.; Zhong, P.; Hu, W.; Zhang, X.; Ling, J.; Su, H.; Jin, W.; Zhang, Y.; and Bai, Y. 2019. A dataset for infrared image dim-small aircraft target detection and tracking under ground/air background. *Science Data Bank*.

Kong, X.; Yang, C.; Cao, S.; Li, C.; and Peng, Z. 2022. Infrared Small Target Detection via Nonconvex Tensor Fibered Rank Approximation. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–21.

Li, B.; Xiao, C.; Wang, L.; Wang, Y.; Lin, Z.; Li, M.; An, W.; and Guo, Y. 2023. Dense nested attention network for infrared small target detection. *IEEE Transactions on Image Processing*, 32: 1745–1758.

Li, Q.; Zhang, M.; Yang, Z.; Yuan, Y.; and Wang, Q. 2024. Edge-guided perceptual network for infrared small target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–10.

Ling, S.; Chen, L.; Wu, Y.; Zhang, Y.; and Gao, Z. 2024. ACANet: attention-based context-aware network for infrared small target detection. *The Journal of Supercomputing*, 1–29.

Liu, C.; Song, X.; Yu, D.; Qiu, L.; Xie, F.; Zi, Y.; and Shi, Z. 2025. Infrared Small Target Detection Based on Prior Guided Dense Nested Network. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 1–15.

Liu, Q.; Liu, R.; Zheng, B.; Wang, H.; and Fu, Y. 2024. Infrared Small Target Detection with Scale and Location Sensitivity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17490–17499.

Liu, S.; Wu, R.; Qu, J.; and Li, Y. 2023. HPN-SOE: Infrared Small Target Detection and Identification Algorithm Based on Heterogeneous Parallel Networks with Similarity Object Enhancement. *IEEE Sensors Journal*.

Nie, J.; Qu, S.; Wei, Y.; Zhang, L.; and Deng, L. 2018. An infrared small target detection method based on multiscale local homogeneity measure. *Infrared Physics & Technology*, 90: 186–194.

Pang, D.; Ma, P.; Shan, T.; Li, W.; Tao, R.; Ma, Y.; and Wang, T. 2022. STTM-SFR: Spatial–temporal tensor modeling with saliency filter regularization for infrared small target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–18.

Qin, Y.; Bruzzone, L.; Gao, C.; and Li, B. 2019. Infrared small target detection based on facet kernel and random walker. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9): 7104–7118.

Sun, X.; Guo, L.; Zhang, W.; Wang, Z.; and Yu, Q. 2021. Small Aerial Target Detection for Airborne Infrared Detection Systems Using LightGBM and Trajectory Constraints. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14: 9959–9973.

Wang, C.; and Wang, L. 2021. Multidirectional Ring Top-Hat Transformation for Infrared Small Target Detection. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14: 8077–8088.

Wang, G.; Zhang, T.; Wei, L.; and Sang, N. 1996. Efficient method for multiscale small target detection from a natural scene. *Optical Engineering*, 35(3): 761–768.

Wang, H.; Zhou, L.; and Wang, L. 2019. 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*, 8509–8518.

Wang, K.; Du, S.; Liu, C.; and Cao, Z. 2022. Interior Attention-Aware Network for Infrared Small Target Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–13.

Wang, W.; Xiao, C.; Dou, H.; Liang, R.; Yuan, H.; Zhao, G.; Chen, Z.; and Huang, Y. 2023. CCRANet: A two-stage local attention network for single-frame low-resolution infrared small target detection. *Remote Sensing*, 15(23): 5539.

Wang, Z.; Wang, C.; Li, X.; Xia, C.; and Xu, J. 2025. MLP-Net: Multi-Layer Perceptron Fusion Network for Infrared Small Target Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 1–13.

Wei, Y.; You, X.; and Li, H. 2016. Multiscale patch-based contrast measure for small infrared target detection. *Pattern Recognition*, 58: 216–226.

Wu, T.; Li, B.; Luo, Y.; Wang, Y.; Xiao, C.; Liu, T.; Yang, J.; An, W.; and Guo, Y. 2023. MTU-Net: Multilevel TransUNet for space-based infrared tiny ship detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–15.

Wu, X.; Hong, D.; and Chanussot, J. 2023. UIU-Net: U-Net in U-Net for infrared small object detection. *IEEE Transactions on Image Processing*, 32: 364–376.

Xu, H.; Zhong, S.; xun Zhang, T.; and Zou, X. 2023a. Multiscale Multilevel Residual Feature Fusion for Real-Time Infrared Small Target Detection. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–16.

Xu, X.; Wang, J.; Zhu, M.; Sun, H.; Wu, Z.; Wang, Y.; Cao, S.; and Liu, S. 2023b. UCDnet: Double U-Shaped Segmentation Network Cascade Centroid Map Prediction for Infrared Weak Small Target Detection. *Remote Sensing*, 15(15): 3736.

Yang, J.; Liu, S.; Wu, J.; Su, X.; Hai, N.; and Huang, X. 2025. Pinwheel-shaped convolution and scale-based dynamic loss for infrared small target detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 9202–9210.

Yuan, S.; Qin, H.; Yan, X.; Akhtar, N.; and Mian, A. 2024. Sctransnet: Spatial-channel cross transformer network for infrared small target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–15.

Zhang, M.; Li, X.; Gao, F.; and Guo, J. 2025a. Irmamba: Pixel difference mamba with layer restoration for infrared small target detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 10003–10011.

Zhang, M.; Shang, W.; Gao, F.; Zhang, Q.; Lu, F.; and Zhang, J. 2025b. Semi-supervised Infrared Small Target Detection with Thermodynamic-Inspired Uneven Perturbation and Confidence Adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 10013–10021.

Zhang, M.; Yue, K.; Li, B.; Guo, J.; Li, Y.; and Gao, X. 2024. Single-Frame Infrared Small Target Detection via Gaussian Curvature Inspired Network. *IEEE Transactions on Geoscience and Remote Sensing*, 62: 1–13.

Zhang, M.; Zhang, R.; Yang, Y.; Bai, H.; Zhang, J.; and Guo, J. 2022. ISNet: Shape matters for infrared small target detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 877–886.

Zhang, Y.; Li, Z.; Siddique, A.; Azeem, A.; Chen, W.; and Cao, D. 2025c. Infrared Small Target Detection Based on Interpretation Weighted Sparse Method. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 1–15.

Zhao, M.; Li, W.; Li, L.; Hu, J.; Ma, P.; and Tao, R. 2022. Single-Frame Infrared Small-Target Detection: A survey. *IEEE Geoscience and Remote Sensing Magazine*, 10: 87–119.

Zhao, M.; Li, W.; Li, L.; Wang, A.; Hu, J.; and Tao, R. 2023. Infrared small UAV target detection via isolation forest. *IEEE Transactions on Geoscience and Remote Sensing*, 61: 1–16.