MDAFNet: Multi-scale Differential Edge and Adaptive Frequency Guided Network for Infrared Small Target Detection

Shuying Li, Qiang Ma, San Zhang, Wuwei Wang, Chuang Yang, *Member, IEEE*

***Abstract*—Infrared small target detection (IRSTD) plays a crucial role in numerous military and civilian applications. However, existing methods often face the gradual degradation of target edge pixels as the number of network layers increases, and traditional convolution struggles to differentiate between frequency components during feature extraction, leading to low-frequency backgrounds interfering with high-frequency targets and high-frequency noise triggering false detections. To address these limitations, we propose MDAFNet (Multi-scale Differential Edge and Adaptive Frequency Guided Network for Infrared Small Target Detection), which integrates the Multi-Scale Differential Edge (MSDE) module and Dual-Domain Adaptive Feature Enhancement (DAFE) module. The MSDE module, through a multi-scale edge extraction and enhancement mechanism, effectively compensates for the cumulative loss of target edge information during downsampling. The DAFE module combines frequency domain processing mechanisms with simulated frequency decomposition and fusion mechanisms in the spatial domain to effectively improve the network's capability to adaptively enhance high-frequency targets and selectively suppress high-frequency noise. Experimental results on multiple datasets demonstrate the superior detection performance of MDAFNet.**

***Index Terms*—IRSTD, gradual degradation, frequency components, multi-scale edge extraction, frequency domain processing.**

I. Introduction

IRSTD is a critical technology in remote sensing and defense, playing an essential role in early warning systems, search and rescue operations, and security surveillance. Infrared imaging sensors can achieve detection under complex conditions such as low visibility, nighttime, and adverse weather by capturing the thermal radiation of targets.However, due to the inherent characteristics of infrared small targets [1], these targets appear as indistinct pixels with small information proportions and minimal texture features. The challenge of accurate detection and positioning is further compounded when they are masked by complex background clutter.

In light of these challenges, researchers have developed a range of traditional IRSTD methods, which are generally categorized into filtering and background suppression strategies [2], human visual system (HVS)-inspired approaches [3], and low-rank sparse representation techniques for background modeling [4]. However, the limited feature representation capability and reliance on manually designed priors result in poor performance when dealing with complex scenes.

In contrast, deep learning approaches have demonstrated superior performance in IRSTD by automatically learning discriminative features from data. Among deep learning-based methods, Convolutional Neural Network (CNN)-based methods [5]–[13] employ encoder-decoder architectures. For example, ACM-Net [5] promoted multi-scale feature interaction across hierarchical layers, while UIU-Net [7] employed a hierarchical U-shaped architecture for dense cross-level feature interaction. Transformer methods [14]–[18] employ self-attention for global context modeling. RKformer [14] combined Transformer and convolution in parallel, leveraging the Runge-Kutta method for feature enhancement. STPSA-Net [15] leveraged semantic tokens and patch-wise spatial attention for accurate localization. However, most models undergo multiple downsampling operations during detection. As network depth increases through repeated downsampling operations, the edge information of small targets undergoes cumulative degradation across network layers. This considerably affects subsequent feature extraction and target localization accuracy. Moreover, traditional convolutions naturally possess smoothing characteristics, lacking explicit frequency processing mechanisms, which causes background clutter and noise interference. In addition, recent frameworks [19], [20] combined CNNs with frequency domain processing. HDNet [19] adopted a dynamic high-pass filter to progressively filter low-frequency background information in the frequency domain, alleviating background interference to some extent. FAA-Net [20] employed DCT to capture high-frequency components and enhance local contrast. However, existing frequency domain methods have several limitations. They only use conventional frequency transforms (e.g., Fourier or DCT) with fixed filtering strategies lacking layer-adaptive modulation. Moreover, these methods lack the ability to differentially model frequency components across directions, and struggle to simultaneously adapt to the distinct frequency requirements of shallow layers retaining high-frequency details and deep layers suppressing high-frequency noise while maintaining semantics. In the edge domain, ISNet [21] designed ODE-inspired edge blocks for shape extraction, but lacks multi-scale differential enhancement capability.

To mitigate the limitations identified above, we develop MDAFNet, which contains two core modules. First, the Multi-Scale Differential Edge (MSDE) module constructs an inde-

This work was supported in part by the National Key Research and Development Program of China under Grant 2022ZD0160404, in part by the National Natural Science Foundation of China under Grant 62575238, 62501511 and 62202376, in part by the Key Research and Development Program of Shaanxi Province under Grant 2025CY-YBXM-048, in part by the Key Scientific Research Program of the Education Department of Shaanxi Province under Grant 24JS048.(*Corresponding author: Chuang Yang.*)

Shuying Li, Qiang Ma, San Zhang and Wuwei Wang are with the School of Artificial Intelligence and School of Automation, Xi'an University of Posts and Telecommunications, Xi'an 710121, China, and also with Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China(e-mail: lishuying@xupt.edu.cn; maqiang123@stu.xupt.edu.cn; zhangsan@xupt.edu.cn; wangwuwei@xupt.edu.cn). Chuang Yang is with the Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong SAR (e-mail: omtcyang@gmail.com).

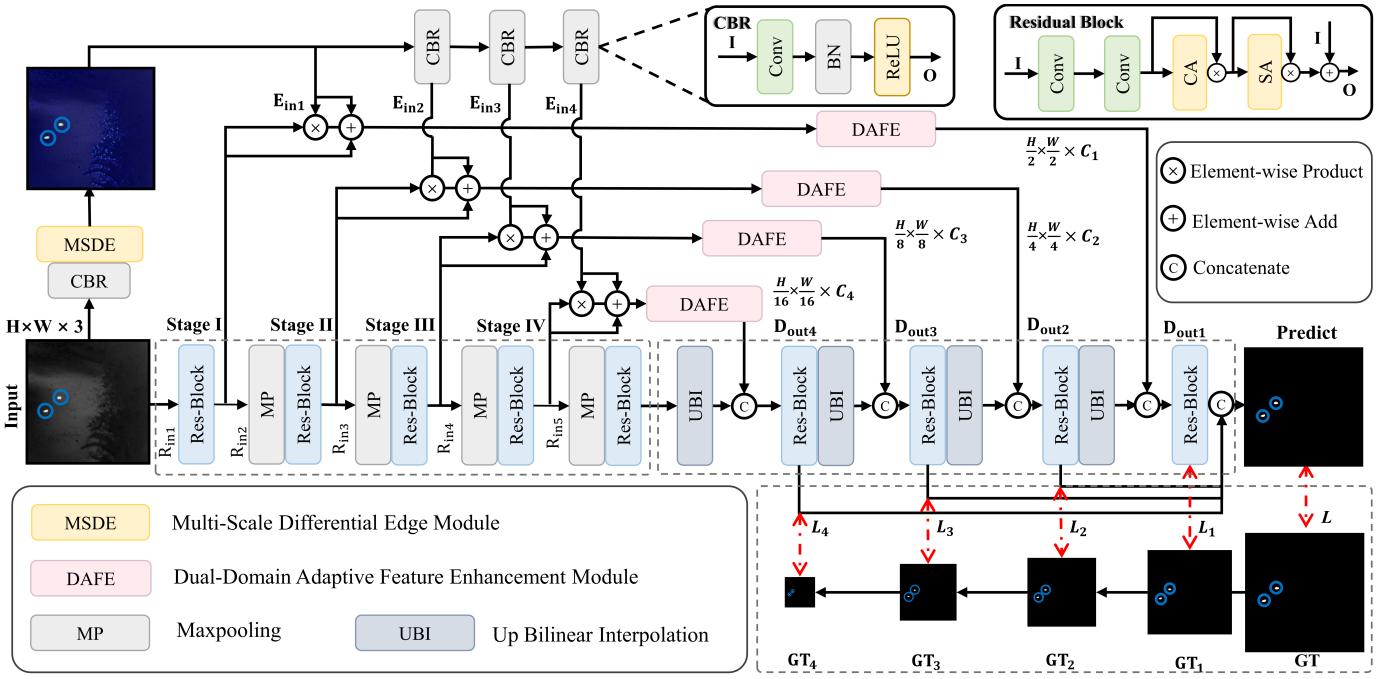

Fig. 1. Schematic diagram of the MDAFNet framework.

pendent edge branch that performs three-way fusion with main branch features at skip connections to offset the cumulative loss of target edge details in the downsampling process of the main branch. Second, the Dual-Domain Adaptive Feature Enhancement (DAFE) module achieves adaptive frequency-guided feature enhancement at skip connections, enabling the network to adaptively enhance high-frequency targets and selectively suppress high-frequency noise. The combination of both thereby enhances the network's detection performance for infrared small targets.

We summarize our key contributions below:

1) We design the MSDE module with multi-scale differential edge enhancement mechanisms to effectively maintain target geometric detail integrity.

2) We develop the DAFE module with adaptive frequency-guided enhancement to effectively discriminate targets from high-frequency noise through dual-domain processing.

3) Extensive experiments across various benchmark datasets show that MDAFNet outperforms state-of-the-art methods, confirming our approach's effectiveness.

II. METHOD

A. Overall architecture

As shown in Fig. 1, MDAFNet employs a U-shaped structure for IRSTD. The encoder contains four stages (Stage I-IV), where each stage employs Residual Blocks (Res-Block) for feature extraction, with feature map resolutions progressively decreasing. To address the cumulative loss of edge information during downsampling and the inadequacy of frequency processing, the network integrates two core modules: MSDE and DAFE. The MSDE module constructs an independent auxiliary edge branch that extracts multi-level edge features ($\mathbf{E}_{\text{in1}}$ to $\mathbf{E}_{\text{in4}}$) and performs adaptive triple-path fusion at skip

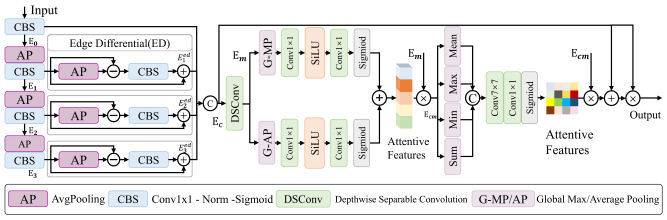

Fig. 2. Architecture of the MSDE module.

connections using main branch features, edge features, and the multiplication of main branch and edge features, effectively compensating for edge information loss. The DAFE module is deployed at each skip connection to enhance the network's capability of adaptively enhancing high-frequency targets and selectively suppressing high-frequency noise through adaptive frequency-guided enhancement. The decoder progressively recovers spatial resolution and fuses multi-scale features through skip connections. The network employs a deep supervision strategy, utilizing SLS Loss [8] to supervise the outputs of each decoder stage, with the output of the final decoder layer serving as the prediction map.

B. Multi-Scale Differential Edge Module

To compensate for edge information loss during downsampling, we propose the MSDE. Specifically, given an input feature $\mathbf{X}$, MSDE first projects it into the edge feature space. Then, it employs hierarchical average pooling to extract edge features at multiple scales ($t = 1, 2, 3$). Notably, at each scale, an Edge Differential (ED) strengthens edge perception through differential operations. Then, the initial feature and edge-enhanced features at all scales are concatenated and fused

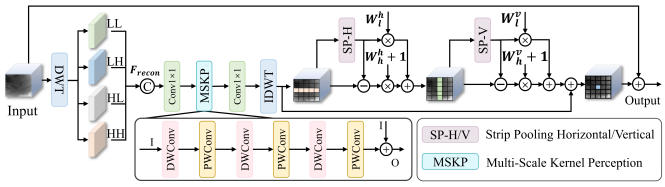

Fig. 3. Architecture of the DAFE module.

through a fusion module to generate a comprehensive edge representation $\mathbf{E}^c$. The operation is defined as:

$$\mathbf{E}_0 = \mathrm{CBS}(\mathbf{X}), \tag{1}$$

$$\mathbf{E}_t = \mathrm{AP}(\mathrm{CBS}(\mathbf{E}_{t-1})), \tag{2}$$

$$\mathbf{E}_t^{\mathrm{ed}} = \mathbf{E}_t + \mathrm{CBS}(\mathbf{E}_t - \mathrm{AP}(\mathbf{E}_t)), \tag{3}$$

$$\mathbf{E}^c = \mathrm{U}([\mathbf{E}_0, \mathbf{E}_1^{\mathrm{ed}}, \mathbf{E}_2^{\mathrm{ed}}, \mathbf{E}_3^{\mathrm{ed}}]), \tag{4}$$

where AP denotes average pooling with kernel size 3×3 and stride 1, CBS represents a 1×1 convolution with batch normalization followed by sigmoid, and U denotes channel concatenation followed by 1×1 convolution.

To enhance multi-scale edge features, we employ channel and spatial attention for refinement. The channel attention captures inter-channel dependencies through dual-path pooling, while the spatial attention leverages multiple statistics to emphasize salient regions. The operation is defined as:

$$\mathbf{E}^m = \mathrm{DSConv}(\mathbf{E}^c), \tag{5}$$

$$\mathbf{E}_{\mathrm{out}}^{\mathrm{ca}} = \mathbf{E}^m \odot \sigma(\phi(\text{G-AP}(\mathbf{E}^m)) + \phi(\text{G-MP}(\mathbf{E}^m))), \tag{6}$$

$$\mathbf{E}_{\mathrm{out}}^{\mathrm{sa}} = \mathbf{E}_{\mathrm{out}}^{\mathrm{ca}} \odot \sigma(\mathrm{Conv}_{1\times1}(\mathrm{SiLU}(\mathrm{Conv}_{7\times7}(\mathcal{M}(\mathbf{E}_{\mathrm{out}}^{\mathrm{ca}}))))), \tag{7}$$

where DSConv is depthwise separable convolution, ϕ denotes two 1×1 convolutions with SiLU, σ is sigmoid, and $\mathcal{M}$ concatenates mean, max, min, and sum pooling.

Finally, a residual connection with element-wise multiplication is applied. The operation is defined as:

$$\mathbf{E}^{\mathrm{out}} = (\mathbf{E}_{\mathrm{out}}^{\mathrm{sa}} + \mathbf{E}^c) \odot \mathbf{E}^c. \tag{8}$$

C. Dual-Domain Adaptive Feature Enhancement

To address the inadequacy of frequency processing in existing methods, we propose the DAFE module for adaptive frequency-guided feature enhancement. For a feature representation $\mathbf{F} \in \mathbb{R}^{C\times H\times W}$, DAFE first decomposes it into four frequency subbands using the Haar wavelet transform. These subbands are then concatenated, reduced in dimension, processed through the Multi-Scale Kernel Perception (MSKP) mechanism, and reconstructed to the spatial domain via inverse wavelet transform. The operation is defined as:

$$\mathbf{F}_{\mathrm{ll}}, \mathbf{F}_{\mathrm{lh}}, \mathbf{F}_{\mathrm{hl}}, \mathbf{F}_{\mathrm{hh}} = \mathrm{DWT}(\mathbf{F}), \tag{9}$$

$$\mathbf{F}_{\mathrm{mskp}} = \mathrm{MSKP}(\mathrm{Conv}_{1\times1}([\mathbf{F}_{\mathrm{ll}}, \mathbf{F}_{\mathrm{lh}}, \mathbf{F}_{\mathrm{hl}}, \mathbf{F}_{\mathrm{hh}}])), \tag{10}$$

$$\mathbf{F}_{\mathrm{recon}} = \mathrm{IDWT}(\mathrm{Split}(\mathrm{Conv}_{1\times1}(\mathbf{F}_{\mathrm{mskp}}))), \tag{11}$$

where $\mathbf{F}_{\mathrm{ll}} \in \mathbb{R}^{C\times\frac{H}{2}\times\frac{W}{2}}$ represents the low-frequency approximation, $\mathbf{F}_{\mathrm{lh}}$, $\mathbf{F}_{\mathrm{hl}}$, $\mathbf{F}_{\mathrm{hh}}$ represent horizontal, vertical, and diagonal high-frequency details, respectively, and MSKP employs stacked depthwise separable convolutions at three scales $(3, 5, 7)$ for multi-scale feature extraction across different receptive fields.

To further refine the frequency components, we apply an Adaptive Frequency Modulation (AFM) mechanism with two-stage strip pooling: horizontal (SP-H) and vertical (SP-V). In each stage, strip-shaped average pooling extracts low-frequency components, while high-frequency components are obtained by subtraction. These frequency components are then modulated using learnable channel-wise parameters. Taking the horizontal stage as an example:

$$\mathbf{F}_h^l = \text{SP-H}(\mathbf{F}_{\mathrm{recon}}), \quad \mathbf{F}_h^h = \mathbf{F}_{\mathrm{recon}} - \mathbf{F}_h^l, \tag{12}$$

$$\mathbf{F}_h^{\mathrm{out}} = \mathbf{w}_l^h \odot \mathbf{F}_h^l + (\mathbf{w}_h^h + 1) \odot \mathbf{F}_h^h, \tag{13}$$

where SP-H denotes horizontal strip pooling with kernel size $(1, K)$, K is the strip length, and $\mathbf{w}_l^h, \mathbf{w}_h^h \in \mathbb{R}^C$ are learnable parameters for modulating low- and high-frequency components, respectively. The term $(\mathbf{w}_h^h + 1)$ ensures that high-frequency information is preserved. The vertical stage follows the same formulation with SP-V using kernel $(K, 1)$ and parameters $\mathbf{w}_l^v$, $\mathbf{w}_h^v$, yielding the final output $\mathbf{F}_{\mathrm{afm}}$.

Finally, the output features are integrated with the input through learnable weighted skip connections. The operation is defined as:

$$\mathbf{F}_{\mathrm{out}} = \boldsymbol{\alpha} \odot \mathbf{F} + \boldsymbol{\beta} \odot \mathbf{F}_{\mathrm{afm}}, \tag{14}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta} \in \mathbb{R}^C$ denote adaptive weights that balance the contributions of the original input and the modulated features.

III. EXPERIMENTS

A. Datasets and Implementation Setup

To thoroughly assess our method's performance, extensive evaluations are performed across three standard IRSTD benchmarks: IRSTD-1K, NUAA-SIRST, and SIRST-Aug. The network is trained using the Adagrad optimizer for 400 epochs with a learning rate initialized at 0.05 and batch size set to 4. For fair comparison, these hyperparameters remain consistent across all datasets. Experiments utilize an NVIDIA RTX 4090 GPU. Model performance is quantified using three metrics: IoU (intersection over union), P_d (probability of detection), and F_a (false alarm rate).

B. Comparsion With SOTA Methods

For performance validation, our approach is compared against 11 SOTA methods, including 8 CNN-based approaches (DNA-Net [6], UIU-Net [7], RDIAN [10], AGPC-Net [22], MSHNet [8], HCF-Net [23], L2SKNet [24], PConv [9]) and 3 Hybrid approaches (ABC [16], MTU-Net [17], SCTransNet [18]). Evaluations are performed across three benchmarks: IRSTD-1K, NUAA-SIRST, and SIRST-Aug, with results analyzed both quantitatively and qualitatively.

1) Quantitative Evaluations Results: Comparative evaluations on three datasets validate our method's effectiveness. Quantitative results in Table I show that MDAFNet achieves superior performance on all datasets, confirming our architecture's effectiveness and robustness. Fig. 4 presents Receiver operating characteristic (ROC) curves across three benchmarks, where Ours consistently attains superior true positive

TABLE I
QUANTITATIVE RESULTS COMPARED AGAINST ADVANCED APPROACHES ON IRSTD-1K, NUAA-SIRST, AND SIRST-AUG. *IoU* (%), *Pd* (%), AND *Fa* ($\times 10^{-6}$) ARE REPORTED, WITH **RED** MARKING THE BEST.

Method	Publication	IRSTD-1K			NUAA-SIRST			SIRST-Aug		
		IoU↑	P_d↑	F_a↓	IoU↑	P_d↑	F_a↓	IoU↑	P_d↑	F_a↓
CNN-based Approaches										
DNA-Net [6]	TIP'22	67.54	92.18	11.77	77.04	99.08	20.05	72.27	**98.35**	51.85
UIU-Net [7]	TIP'22	65.59	88.44	14.73	76.83	98.17	11.71	70.64	96.56	63.47
RDIAN [10]	TGRS'23	63.40	92.86	11.54	74.08	**100.00**	19.87	71.45	98.07	116.58
AGPC-Net [22]	TAES,'23	65.93	91.15	**11.32**	**77.47**	**100.00**	**3.02**	72.73	95.46	128.73
MSHNet [8]	CVPR'24	67.16	**93.88**	15.03	73.65	99.08	19.16	72.64	96.56	91.55
HCF-Net [23]	ICME'24	63.72	89.46	17.38	75.43	98.17	11.18	71.69	97.39	**37.77**
L2SKNet [24]	TGRS'25	65.67	92.18	23.99	75.52	98.17	17.21	73.26	97.94	50.34
PConv [9]	AAAI'25	**67.58**	**93.88**	12.22	75.98	98.17	4.26	**74.04**	93.95	288.71
Hybrid Approaches										
ABC [16]	ICME'23	65.13	88.43	**14.96**	**77.59**	98.17	6.74	72.86	**98.76**	56.64
MTU-Net [17]	TGRS'23	65.44	87.75	16.40	77.46	**100.00**	**5.68**	71.59	96.42	**37.57**
SCTransNet [18]	TGRS'24	**66.07**	**92.52**	15.79	76.18	99.08	26.97	**73.79**	98.21	46.67
MDAFNet (Ours)	–	**70.11**	**95.92**	**8.43**	**79.42**	**100.00**	**3.90**	**75.60**	**99.45**	**15.15**

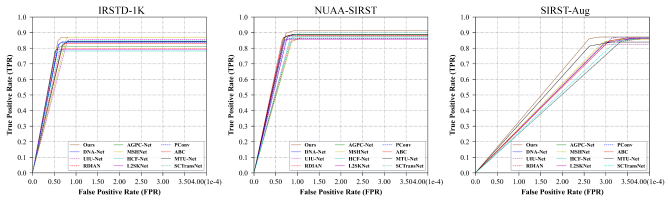

Fig. 4. ROC curves comparison across three datasets.

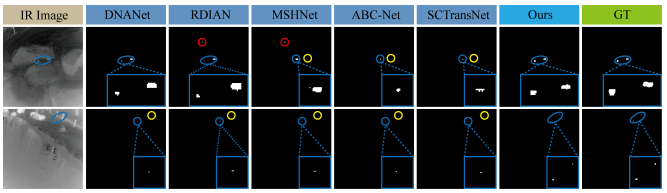

Fig. 5. Visual comparison of detection results for the proposed method against other mainstream approaches. Correctly detected targets are marked in blue, missed targets in yellow, and false alarms in red.

rates at reduced false positive rate levels, further demonstrating enhanced detection reliability.

2) Qualitative Evaluations Results: Fig. 5 presents visualization of two representative examples. Under complex background interference and small target scenarios, existing deep learning methods typically exhibit high false alarm rates and miss detections. Conversely, MDAFNet accurately identifies all genuine targets with more complete boundaries, effectively suppresses background noise, and demonstrates superior detection precision and robustness.

C. Ablation Study

1) Conduct ablation on major components: Comprehensive ablation studies validate each module's contribution. As shown in Table II, individually adding either the MSDE or DAFE module to the baseline brings performance improvements, while the complete MDAFNet integrating both modules achieves optimal performance, significantly outperforming any single-module configuration, fully validating the effectiveness and necessity of the MSDE and DAFE modules.

TABLE II
ABLATION STUDY ON IRSTD-1K WITH *IoU* (%), *Pd* (%), AND *Fa* ($\times 10^{-6}$), PARA. (M), AND FLOPS (G).

Module	IoU	P_d	F_a	Para.	FLOPs
Baseline	67.16	93.88	15.03	4.07	6.11
Baseline+MSDE	68.26	94.90	11.99	4.19	6.59
Baseline+DAFE	68.86	94.56	9.41	4.33	6.36
Baseline+MSDE+DAFE	**70.11**	**95.92**	**8.43**	4.45	6.83

2) Selection of multi-scale branch numbers of MSDE: To examine how multi-scale branch quantity in MSDE affects performance, comparative evaluations are conducted using the IRSTD-1K dataset. Table III demonstrates that optimal results are obtained with Width=4. Too few differential edge branches cannot sufficiently extract edge information, while too many branches lead to feature redundancy. Balancing detection accuracy and computational cost, the width parameter of MSDE is set to 4.

3) Triple-path fusion effectiveness validation: Comparative experiments involving four distinct fusion approaches are conducted to evaluate our triple-path strategy. As shown in Table IV, basic fusion methods such as simple addition and element-wise multiplication fail to effectively improve performance. In contrast, the proposed adaptive triple-path fusion strategy achieves optimal performance by integrating features through three adaptive pathways, effectively compensating for edge information loss.

4) Ablation study on removing individual components from DAFE: Ablation experiments evaluate DAFE's component contributions by sequentially removing each module. As shown in Table V, removing any component leads to performance degradation. Specifically, removing DWT/IDWT significantly reduces detection rate, removing MSKP increases F_a,

TABLE III
ABLATION STUDY ON THE WIDTH PARAMETER OF MSDE.

Width	IoU	P_d	F_a	Para.	FLOPs
3	67.85	93.54	11.39	4.187	6.565
4	**68.26**	**94.90**	11.99	4.188	6.587
5	67.60	93.88	15.03	4.188	6.609
6	67.94	91.84	**10.93**	4.188	6.631

TABLE IV
TRIPLE-PATH FUSION EFFECTIVENESS VALIDATION.

No.	Module	IoU	P_d	F_a
1	None	67.16	93.88	15.03
2	Simple addition	66.97	92.18	14.27
3	Element-wise multiply	67.20	93.54	16.40
4	Adaptive Dual-path	67.17	93.20	14.20
5	**Adaptive Triple-path (Ours)**	**68.26**	**94.90**	**11.99**

TABLE V
ABLATION STUDY OF DAFE COMPONENTS.

Module	IoU	P_d	F_a	Para.	FLOPs
DAFE	**68.86**	**94.56**	**9.41**	4.331	6.357
w/o DWT/IDWT	67.81	91.50	10.02	4.153	6.504
w/o MSKP	67.28	**94.56**	12.98	4.440	6.413
w/o AFM	67.64	93.20	14.42	4.331	6.353
w/o Adaptive Residual	68.69	94.22	9.717	4.331	6.357

and removing AFM causes the most significant performance drop, validating its critical role in suppressing high-frequency noise. Each DAFE component positively impacts overall detection performance.

IV. CONCLUSION

We present MDAFNet for IRSTD. The MSDE module effectively preserves target geometric detail integrity through a multi-scale differential edge enhancement mechanism. Additionally, the DAFE module achieves effective separation of high-frequency targets and noise through adaptive frequency guidance. Experiments conducted on several benchmark datasets reveal that MDAFNet achieves satisfactory results, confirming our approach's effectiveness. Future work intends to explore lightweight network architectures to adapt to real-time detection scenarios.

REFERENCES

[1] M. Zhao, W. Li, L. Li, J. Hu, P. Ma, and R. Tao, "Single-frame infrared small-target detection: A survey," *IEEE Geoscience and Remote Sensing Magazine*, vol. 10, no. 2, pp. 87–119, June 2022.

[2] S. D. Deshpande, M. H. Er, R. Venkateswarlu, and P. Chan, "Max-mean and max-median filters for detection of small targets," in *Signal and Data Processing of Small Targets 1999*, vol. 3809. SPIE, 1999, pp. 74–83.

[3] C. L. P. Chen, H. Li, Y. Wei, T. Xia, and Y. Y. Tang, "A local contrast method for small infrared target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 574–581, Jan 2014.

[4] C. Gao, D. Meng, Y. Yang, Y. Wang, X. Zhou, and A. G. Hauptmann, "Infrared patch-image model for small target detection in a single image," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4996–5009, Dec 2013.

[5] Y. Dai, Y. Wu, F. Zhou, and K. Barnard, "Asymmetric contextual modulation for infrared small target detection," in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Jan 2021, pp. 949–958.

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

[7] X. Wu, D. Hong, and J. Chanussot, "Uiu-net: U-net in u-net for infrared small object detection," *IEEE Transactions on Image Processing*, vol. 32, pp. 364–376, 2023.

[8] Q. Liu, R. Liu, B. Zheng, H. Wang, and Y. FU, "Infrared small target detection with scale and location sensitivity," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2024, pp. 17 490–17 499.

[9] J. Yang, S. Liu, J. Wu, X. Su, N. Hai, and X. Huang, "Pinwheel-shaped convolution and scale-based dynamic loss for infrared small target detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 9, 2025, pp. 9202–9210.

[10] H. Sun, J. Bai, F. Yang, and X. Bai, "Receptive-field and direction induced attention network for infrared dim small target detection with a large-scale dataset irdst," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–13, 2023.

[11] C. Yang, X. Han, T. Han, H. Han, B. Zhao, and Q. Wang, "Edge approximation text detector," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 35, no. 9, pp. 9234–9245, Sep. 2025.

[12] C. Yang, X. Han, T. Han, Y. Su, J. Gao, H. Zhang, Y. Wang, and L.-P. Chau, "Signeye: Traffic sign interpretation from vehicle first-person view," *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 11, pp. 19 413–19 425, Nov 2025.

[13] C. Yang, H. Ma, and Q. Wang, "Instance mask growing on leaf." in *BMVC*, 2023, pp. 4–6.

[14] M. Zhang, H. Bai, J. Zhang, R. Zhang, C. Wang, J. Guo, and X. Gao, "Rkformer: Runge-kutta transformer with random-connection attention for infrared small target detection," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, pp. 1730–1738.

[15] S. Liu, B. Qiao, S. Li, Y. Wang, and L. Dang, "Patch spatial attention networks for semantic token transformer in infrared small target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1–14, 2025.

[16] P. Pan, H. Wang, C. Wang, and C. Nie, "Abc: Attention with bilinear correlation for infrared small target detection," in *2023 IEEE International Conference on Multimedia and Expo (ICME)*, July 2023, pp. 2381–2386.

[17] T. Wu, B. Li, Y. Luo, Y. Wang, C. Xiao, T. Liu, J. Yang, W. An, and Y. Guo, "Mtu-net: Multilevel transunet for space-based infrared tiny ship detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–15, 2023.

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

[19] M. Xu, C. Yu, Z. Li, H. Tang, Y. Hu, and L. Nie, "Hdnet: A hybrid domain network with multiscale high-frequency information enhancement for infrared small-target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1–15, 2025.

[20] S. Zhuang, Y. Hou, M. Qi, and D. Wang, "Faa-net: A frequency-aware attention network for single-frame infrared small target detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 74, pp. 1–16, 2025.

[21] M. Zhang, R. Zhang, Y. Yang, H. Bai, J. Zhang, and J. Guo, "Isnet: Shape matters for infrared small target detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 877–886.

[22] T. Zhang, L. Li, S. Cao, T. Pu, and Z. Peng, "Attention-guided pyramid context networks for detecting infrared small target under complex background," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, no. 4, pp. 4250–4261, Aug 2023.

[23] S. Xu, S. Zheng, W. Xu, R. Xu, C. Wang, J. Zhang, X. Teng, A. Li, and L. Guo, "Hcf-net: Hierarchical context fusion network for infrared small object detection," in *2024 IEEE International Conference on Multimedia and Expo (ICME)*, July 2024, pp. 1–6.

[24] F. Wu, A. Liu, T. Zhang, L. Zhang, J. Luo, and Z. Peng, "Saliency at the helm: Steering infrared small target detection with learnable kernels," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1–14, 2025.