Journal of Electronic Research and Application, 2026, Volume 10, Issue 4
http://ojs.bbwpublisher.com/index.php/JERA
Online ISSN: 2208-3510
Print ISSN: 2208-3502

Infrared Image Feature Enhancement Under Complex Backgrounds Using an Improved UNetFPN

Qiuyu Wang*, Quanli Wang

Tianjin University of Technology and Education, Tianjin 300222, China

****Corresponding author:*** Qiuyu Wang, 1448308659@qq.com

Abstract: Infrared images acquired under complex background conditions are often affected by background clutter, local high-response interference, and non-uniform fluctuations, which may reduce target saliency and local discriminability. To address this issue, this paper proposes an improved UNetFPN-based feature enhancement network, termed CBAM-UNetFPN. Built on an encoder-decoder architecture, the proposed method introduces a feature pyramid fusion mechanism to combine shallow spatial details with deep semantic information, and incorporates an attention enhancement strategy to enhance target-related responses while suppressing redundant background activations. Experiments were conducted on three public infrared image datasets, namely NUDT-SIRST, IRSTD-1k, and WideIRSTD-Weak, and the enhancement performance was evaluated using the signal-to-clutter ratio, background suppression factor, and contrast gain. The results show that the proposed method achieves stable enhancement performance across scenes with different levels of complexity. Comparative experiments further indicate that CBAM-UNetFPN can better balance target response enhancement and background suppression under complex background conditions, thereby improving the local discriminability between target regions and the surrounding background.

Keywords: Infrared image feature enhancement; Complex backgrounds; UNetFPN; Background suppression

Online publication: May 14, 2026

1. Introduction

Infrared imaging technology features passive sensing, all-day operation, and strong environmental adaptability, and has been widely used in target surveillance, early warning, maritime search and rescue, and complex scene perception [1]. However, in practical scenarios involving cloud occlusion, sea-surface fluctuations, complex terrain backgrounds, and non-uniform noise, useful target information is often affected by background clutter and locally high-response interference [2]. As a result, infrared images usually suffer from insufficient saliency, reduced contrast, and degraded local discriminability, which increases the difficulty of subsequent image analysis [3–5].

Existing infrared image feature enhancement methods mainly include spatial-domain filtering, local contrast analysis, frequency-domain saliency detection, and deep learning-based approaches [6]. Traditional methods can improve target responses in relatively simple scenes, but under complex background conditions they are easily disturbed by edge clutter, texture interference, and bright background responses, resulting in limited background suppression capability and weak adaptability [7,8]. Deep learning-based methods have shown advantages in multi-scale feature extraction and complex scene representation [9]. Nevertheless, when target-related responses are weak and highly similar to background interference in local statistical characteristics and spatial distribution, methods relying on single-scale features or simple fusion strategies may still produce insufficient target enhancement and redundant background activation [10,11].

To address these issues, this paper proposes an improved UNetFPN-based infrared image feature enhancement network, termed CBAM-UNetFPN. Built on an encoder-decoder architecture, the proposed method introduces a feature pyramid fusion mechanism to combine shallow spatial details with deep semantic information, and incorporates an attention enhancement strategy to strengthen target-related responses while suppressing non-target high-response interference. Experiments conducted on three public datasets, namely NUDT-SIRST, IRSTD-1k, and WideIRSTD-Weak, demonstrate that the proposed method can improve local feature representation quality under complex background conditions and maintain a good balance between target enhancement and background suppression [12–14].

2. Method design

2.1. Local feature characteristics of infrared images under complex backgrounds

Infrared image feature enhancement under complex background conditions is mainly challenged by weak target responses, strong background clutter interference, and low local discriminability. In practical scenes, target regions usually appear as compact local bright spots with limited spatial extent, whereas background interference is more likely to present as continuous textures, structural edges, or large-scale gray-level fluctuations. Therefore, effective enhancement methods should preserve and strengthen target-related responses while suppressing non-target high-response regions in the background, so as to improve local discriminability between target regions and the surrounding background.

2.2. Infrared image feature enhancement method based on an improved UNetFPN

To address weak target responses and severe background clutter interference in infrared images under complex background conditions, this paper proposes an improved UNetFPN-based feature enhancement network, termed CBAM-UNetFPN. Built on an encoder-decoder architecture, the proposed method introduces a feature pyramid fusion mechanism to combine shallow spatial details with deep semantic information, thereby enhancing the representation of target-related regions while retaining complex background context. In addition, an attention enhancement strategy is incorporated to adaptively emphasize target responses and suppress redundant background activations. The overall architecture of the proposed network is shown in **Figure 1**.

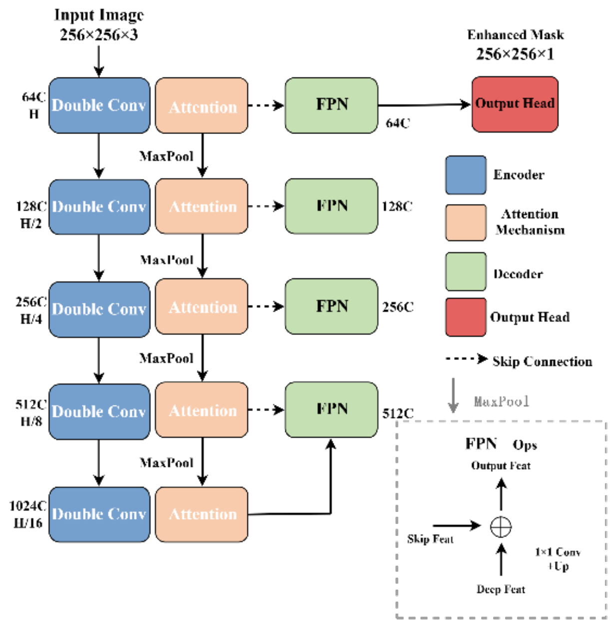

Figure 1. Overall architecture of the proposed CBAM-UNetFPN.

As shown in **Figure 1**, CBAM-UNetFPN mainly consists of three components: the encoder, the feature pyramid fusion module, and the decoder. The encoder extracts hierarchical features through successive convolution and downsampling operations, where shallow layers preserve local spatial details and deep layers encode richer semantic and contextual information. To alleviate the loss of weak target features during deep feature extraction, a feature pyramid network is introduced between the encoder and decoder to perform cross-layer fusion of multi-scale features. During feature reconstruction, the decoder progressively upsamples the fused features to generate the enhanced output, while the embedded attention mechanism further strengthens target-related responses and suppresses non-target interference. As a result, the enhanced image presents clearer and more concentrated target responses with a smoother background, thereby improving local feature representation quality under complex background conditions.

3. Experimental results and analysis

3.1. Experimental setup

Experiments were conducted on three public infrared image datasets, namely NUDT-SIRST, IRSTD-1k, and WideIRSTD-Weak, to evaluate the enhancement performance of the proposed method under different levels of scene complexity [12–14]. The results were quantitatively assessed using the signal-to-clutter ratio (SCR), background suppression factor (BSF), and contrast gain (CG). All input images were resized to 256×256 and normalized to [0,1]. Data augmentation, including random flipping and rotation, was applied during training. The model was implemented in PyTorch and trained on an NVIDIA GeForce RTX 3090 GPU using the Adam optimizer with an initial learning rate of 1×10^{-4}, a batch size of 16, and 100 epochs. The loss function was defined as a weighted combination of Dice Loss and Binary Cross-Entropy Loss:

$$L = 0.7L_{Dice} + 0.3L_{BCE} \tag{1}$$

where L_{Dice} and L_{BCE} denote Dice Loss and Binary Cross-Entropy Loss, respectively.

3.2. Results and discussion

To provide an intuitive demonstration of the enhancement performance of the proposed method under complex background conditions, typical samples from the IRSTD-1k dataset were selected for visual analysis, as shown in **Figure 2**.

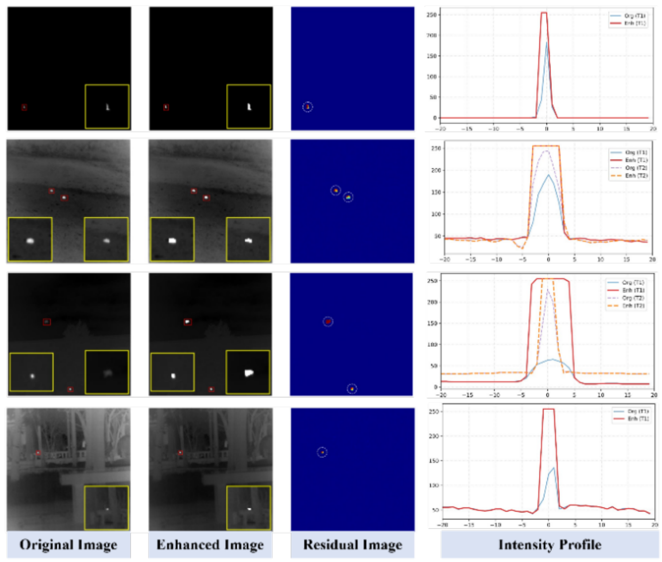

Figure 2. Typical feature enhancement results on the IRSTD-1k dataset.

On the IRSTD-1k dataset, complex interference such as ground textures, bright edges, and cloud streaks is more prominent, placing higher demands on feature enhancement robustness. As shown in **Figure 2**, the proposed method can effectively suppress non-target responses while preserving stable enhancement at target locations. Compared with the original images, the enhanced results present more concentrated target responses and a smoother background, thereby improving the contrast between target regions and the surrounding background. In adjacent-target and multi-target scenarios, good target separability is still maintained, indicating stable feature representation under complex real-scene conditions.

To further evaluate the effectiveness of the proposed method, comparative experiments were conducted against four representative traditional enhancement methods, including Top-Hat, LCM, MPCM, and SR. Quantitative results on the IRSTD-1k dataset are presented in **Table 1**.

Table 1. Quantitative comparison of different methods on the three datasets

Model	NUDT-SIRST			IRSTD-1k			WideIRSTD-Weak		
	SCR	BSF	CG	SCR	BSF	CG	SCR	BSF	CG
Top-Hat	17.96	6.48	6.01	19.04	4.07	3.02	-28.91	3.05	-38.07
LCM	7.03	-9.25	-4.78	11.36	-9.10	-4.64	-35.14	-6.92	-44.24
MPCM	18.26	3.08	6.30	26.55	11.33	10.05	-28.08	5.88	-37.23
SR	13.60	-2.32	1.65	12.10	-5.17	-3.93	-31.48	-3.36	-40.65
Ours	99.87	90.47	97.25	72.52	64.53	72.10	11.35	64.64	8.38

As shown in **Table 1**, the proposed CBAM-UNetFPN achieves the best overall performance across the three datasets. On the NUDT-SIRST dataset, the proposed method attains 99.87 dB, 90.47 dB, and 97.25

dB in SCR, BSF, and CG, respectively, indicating strong enhancement capability in scenes with relatively clear targets and moderate background complexity. On the IRSTD-1k dataset, the corresponding values are 72.52 dB, 64.53 dB, and 72.10 dB, which remain significantly higher than those of the comparative methods, demonstrating effective target enhancement and background suppression under complex real-scene conditions. On the more challenging WideIRSTD-Weak dataset, although the SCR and CG values decrease to 11.35 dB and 8.38 dB, respectively, the BSF still reaches 64.64 dB, indicating that the proposed method can still suppress background clutter effectively under extremely low signal-to-clutter ratio conditions. Overall, the results demonstrate that the combination of multi-scale feature fusion and attention enhancement is beneficial for improving local feature representation quality in infrared images under complex background conditions.

4. Conclusion

To address the problems of insufficient target saliency, strong background clutter interference, and poor local discriminability in infrared images under complex background conditions, this paper proposes an improved UNetFPN-based feature enhancement network, termed CBAM-UNetFPN. Built on an encoder-decoder architecture, the proposed method introduces a feature pyramid fusion mechanism to combine shallow spatial details with deep semantic information, and incorporates an attention enhancement strategy to adaptively strengthen target responses while suppressing redundant activations in complex backgrounds, thereby improving the local feature representation quality of infrared images. Experiments conducted on three public datasets, namely NUDT-SIRST, IRSTD-1k, and WideIRSTD-Weak, show that the proposed method achieves stable enhancement performance across scenes with different levels of complexity. In scenes with relatively salient targets and moderate background complexity, the proposed method can effectively balance target enhancement and background suppression. Under complex real-background and low signal-to-clutter ratio conditions, although the enhancement effect is somewhat limited, the model can still improve local discriminability by suppressing background fluctuations. Comparative results further demonstrate that the proposed method maintains a favorable balance between target response preservation and background clutter suppression under complex background conditions.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Wang Q, Jin P, Wu Y, et al., 2025, Infrared Image Enhancement: A Review. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2025(18): 3281–3299.

[2] Zhang R, Liu G, Zhang Q, et al., 2025, Detail-Aware Network for Infrared Image Enhancement. IEEE Transactions on Geoscience and Remote Sensing, 2025(63): 5000314.

[3] Liu L, Song R, Li J, et al., 2025, ASMF: A Self-Supervised Atmospheric Scatter Model-Based Fusion Network for Infrared Image Enhancement. IEEE Transactions on Geoscience and Remote Sensing, 2025(63): 1–13.

[4] Wu Z, Li Y, Zhang R, et al., 2025, TEFormer: Thermal Infrared Image Enhancement by Preserving Spatial

Consistency and Details. IEEE Transactions on Geoscience and Remote Sensing, 2025(63): 1–14.

[5] Cai L, Dong X, Zhou K, et al., 2024, Exploring Video Denoising in Thermal Infrared Imaging: Physics-Inspired Noise Generator, Dataset, and Model. IEEE Transactions on Image Processing, 2024(33): 3839–3854.

[6] Jiang N, Zhang Y, Li Q, et al., 2025, An Infrared Image Denoising Method Based on Wavelet Dual-Branch Diffusion Model. IEEE Transactions on Geoscience and Remote Sensing, 2025(63): 1–13.

[7] Liu L, Xu L, Fang H, 2020, Simultaneous Intensity Bias Estimation and Stripe Noise Removal in Infrared Images Using the Global and Local Sparsity Constraints. IEEE Transactions on Geoscience and Remote Sensing, 58(3): 1777–1789.

[8] Zhang T, Li L, Cao S, et al., 2023, Attention-Guided Pyramid Context Networks for Detecting Infrared Small Target Under Complex Background. IEEE Transactions on Aerospace and Electronic Systems, 59(4): 4250–4261.

[9] Guo T, Zhou B, Luo F, et al., 2024, DMFNet: Dual-Encoder Multistage Feature Fusion Network for Infrared Small Target Detection. IEEE Transactions on Geoscience and Remote Sensing, 2024(62): 1–14.

[10] Li B, Li X, Li S, et al., 2024, Cross-Layer Feature Guided Multiscale Infrared Small Target Detection. IEEE Geoscience and Remote Sensing Letters, 2024(21): 1–5.

[11] Sun H, Bai J, Yang F, et al., 2023, 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, 2023(61): 5000513.

[12] Li B, Xiao C, Wang L, et al., 2023, Dense Nested Attention Network for Infrared Small Target Detection. IEEE Transactions on Image Processing, 2023(32): 1745–1758.

[13] Zhang M, Zhang R, Yang Y, et al., 2022, ISNet: Shape Matters for Infrared Small Target Detection, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 877–886.

[14] Li B, Ying X, Li R, et al., 2025, ICPR 2024 Competition on Resource-Limited Infrared Small Target Detection Challenge: Methods and Results, Pattern Recognition. Competitions: 27th International Conference, ICPR 2024, 62–77.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.