

Enhancing Infrared Vision: Progressive Prompt Fusion Network and Benchmark

Jinyuan Liu[†], Zihang Chen[†], Zhu Liu[†], Zhiying Jiang[‡], Long Ma[†], Xin Fan[†], Risheng Liu^{†*}

[†]School of Software Engineering, Dalian University of Technology

[‡]Information Science and Technology College, Dalian Maritime University
atlantis918@hotmail.com, chenzi_hang@mail.dlut.edu.cn

Abstract

We engage in the relatively underexplored task named thermal infrared image enhancement. Existing infrared image enhancement methods primarily focus on tackling individual degradations, such as noise, contrast, and blurring, making it difficult to handle coupled degradations. Meanwhile, all-in-one enhancement methods, commonly applied to RGB sensors, often demonstrate limited effectiveness due to the significant differences in imaging models. In sight of this, we first revisit the imaging mechanism and introduce a Progressive Prompt Fusion Network (PPFN). Specifically, the PPFN initially establishes prompt pairs based on the thermal imaging process. For each type of degradation, we fuse the corresponding prompt pairs to modulate the model’s features, providing adaptive guidance that enables the model to better address specific degradations under single or multiple conditions. In addition, a Selective Progressive Training (SPT) mechanism is introduced to gradually refine the model’s handling of composite cases to align the enhancement process, which not only allows the model to remove camera noise and retain key structural details, but also enhancing the overall contrast of the thermal image. Furthermore, we introduce the most high-quality, multi-scenarios infrared benchmark covering a wide range of scenarios. Extensive experiments substantiate that our approach not only delivers promising visual results under specific degradation but also significantly improves performance on complex degradation scenes, achieving a notable 8.76% improvement. Code is available at <https://github.com/Zihang-Chen/HM-TIR>.

1 Introduction

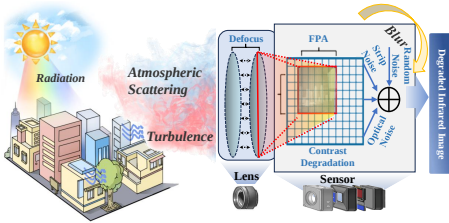


Figure 1: An illustration of the thermal infrared degradation pipeline. Thermal infrared imaging is prone to degradation from external factors such as solar radiation, atmospheric scattering, and turbulence, as well as internal factors like pixel size, internal noise, and jitter.

Thermal Infrared (TIR) imaging captures images by detecting the thermal radiation emitted by objects, typically within the wavelength range of 8 to 14 micrometers. Unlike visible light imaging, TIR does not depend on external light sources, allowing it to function effectively in complete darkness or low-light conditions. Its ability to penetrate smoke, haze, and minor obstructions, coupled with accurate temperature data, makes TIR essential for diverse applications [58, 56], such as object detection [27, 30], semantic segmentation [57], and autonomous driving [26].

Despite its advantages, TIR imaging faces significant challenges that limit its widespread use. The com-

*Corresponding Author

plexity of the imaging process and the reliance on expensive, specialized materials like Mercury Cadmium Telluride (MCT) and Indium Antimonide (InSb) make obtaining high-quality TIR images difficult. Additionally, TIR systems are highly susceptible to external factors such as temperature fluctuations and varying atmospheric conditions, which can degrade image quality. These obstacles underscore the critical need to advance thermal infrared image enhancement techniques.

A considerable number of image enhancement methods have been proposed for TIR or visible images. Techniques such as histogram equalization [42], adaptive filtering [43, 33], and deep learning-based approaches [19, 4, 10, 31] have been utilized to improve image contrast, reduce noise, and enhance overall visual quality. However, these methods exhibit two major limitations. Firstly, enhancement techniques developed for visible images often prove challenging to apply to TIR images due to fundamental differences in imaging modalities, degradation and imaging processes. Secondly, existing enhancement methods only address single degradation, such as denoising or encontrast.

Moreover, a major obstacle in TIR image enhancement is the limited availability of diverse datasets. Although learning-based ways have demonstrated success in various image processing applications, they require large and varied datasets to train effectively and generalize well. However, existing datasets encompass only a narrow range of scenes and conditions, making it challenging to validation.

Incorporating these criteria, this paper presents the Progressive Prompt Fusion Network (PPFN) for enhancing TIR images. PPFN comprises two key components: type and degradation-specific prompts and a prompt fusion module. The degradation-specific prompts guide the model in identifying degradation types, while type-specific prompts differentiate single from composite degradation scenarios. The prompt fusion module integrates prompt pairs to iteratively modulate model features, providing adaptive guidance tailored to specific degradation types in both single and multiple contexts. Additionally, we introduce a Selective Progressive Training (SPT) mechanism for handling composite and single degradations, which iteratively refines each degradation step by using the output from one stage as input for the next in composite scenarios, while applying standard training for single degradations. Consequently, the model effectively eliminates each impairment without interference, resulting in significant performance improvements. Our contributions can be summarized into four key aspects, as follows:

- We propose a PPFN to enhance TIR images, delivering exceptional visual quality in hybrid degradations. To our knowledge, this is the first study addressing TIR enhancement under such multifaceted degradation conditions.
- Addressing intricate degradations in real-world thermal infrared images, we introduce a prompt fusion block that incorporates prior knowledge into the learning process, effectively managing both single and hybrid degradations. Importantly, the prompt fusion block is a plug-and-play module that seamlessly integrates into various existing network architectures, enhancing performance.
- We propose a SPT scheme that optimizes both single and hybrid degradation scenarios, enabling the model to effectively refine complex degradations while ensuring robustness and stability under simpler conditions.
- We establish a high-quality TIR benchmark covering multiple scenarios, named HM-TIR, with all collected images meticulously focused for clarity. This dataset encompasses diverse environments, including urban areas, forests, and oceans, to name a few.

2 Related Work

This section provides a concise overview of existing TIR and visible image enhancement techniques relevant to our study, as well as the necessary benchmarks for learning and empirical evaluation.

2.1 TIR/Visible Image Enhancement

With the growing demands of modern applications, numerous TIR image enhancement methods have been developed, achieving promising results. For TIR denoising, studies [28, 4, 2] have simulated realistic infrared noise by combining various noise types, resulting in significant improvements. Additionally, researchers have addressed specific blur types, including motion blur [49, 13], out-of-focus blur [60], and Gaussian blur that simulates atmospheric effects [53]. These efforts have

substantially enhanced image clarity and detail restoration in infrared imaging. TIR are also vulnerable to other degradations, such as compression artifacts and low resolution. Several studies [1, 12, 24, 25] have tackled these challenges, leading to notable advancements. However, existing methods are typically constrained by specific degradation conditions, which significantly limits their generalization and effectiveness in real-world infrared image processing.

Table 1: Illustration of our benchmark and existing infrared enhancement datasets. The “multiplication” denotes the diverse camera viewpoints, including horizontal, surveillance, driving, etc.

Scene: ①: Road ②: Square ③: City ④: Forest ⑤: Campus ⑥: Coastline ⑦: Residential Zone ⑧: Others							
Corruption: I: Low Contrast		II: Blur		III: Stripe Noise		IV: Optical Noise	
Dataset	Year	Format	# of Images/Videos	Resolution	Camera angle	Scene	Corruption Type
EN [19]	2019	Image	16	256×256	horizontal&surveillance	④⑤⑦	I
Iray [29]	2021	Image	2000	256×192	horizontal	①⑧	III
SBTI [21]	2022	Video	4	640×480	horizontal&surveillance	①③	II
UIRD [16]	2023	Video	17	640×512	horizontal&surveillance	①③	II
TIVID [2]	2024	Video	518	320×256	horizontal	①③④⑦	III IV V
HM-TIR (Ours)	2025	Image	1503	640×512	multiplication	①~⑧	I~V

All-in-One Image Restoration employs a single model to address a range of image degradation issues. PromptIR [40] and ProRes [34] use additional degradation context to introduce task information. IDR [52] explores the model optimization by ingredient-oriented clustering. AutoDIR [14] leverages latent diffusion with degradation-specific text embeddings to automate degradation handling. InstructIR [8] introduces natural language instructions to control restoration. However, most of these methods are only focus visible image enhancement, posing a challenge to apply in TIR images.

2.2 Thermal Image Enhancement Benchmarks

In recent years, several image enhancement benchmarks addressing specific degradations have been introduced, including the Iray Infrared Image Denoising dataset [29] and TIVID [2] for thermal image denoising, EN [19] for contrast enhancement, and SBTI [21] and UIRD [16] for deblurring. The Iray dataset comprises 2,000 pairs of real-world noisy infrared images captured indoors and outdoors alongside their clean counterparts. TIVID includes 518 diverse videos collected with a cooled infrared imaging system to simulate various thermal infrared noises. EN contains 16 internet-sourced images designed to evaluate contrast enhancement. Additionally, SBTI consists of four videos captured on roads and around vehicles, while UIRD includes 17 videos generated through frame interpolation to produce more blurred images.

Table 1 outlines the main attributes of these datasets, including scale, resolution, lighting conditions, and scenario types. Limited resolution, quality, and scenario, degradation types, and overall dataset size variety restrict their applicability for real-world infrared enhancement tasks.

3 Methodology

3.1 Problem Formulation

Infrared imaging systems, especially those using CMOS-based sensors, are prone to additional Fixed-Pattern Noise (FPN) alongside random noise types common in RGB imaging, such as Gaussian and salt-and-pepper noise. Additionally, unlike visible images, which contain detailed visual content and high-quality representation, infrared images capture only thermal distributions, making them particularly vulnerable to atmospheric conditions and temperature differences. As illustrated in Figure 1, this degradation pipeline is affected by several factors, including low contrast due to minimal temperature differences, blurring from environmental radiation effects, and sensor-induced noise, which collectively reduce image clarity and quality. We categorize TIR degradation into three primary types: low contrast, blurring, and noise. The degradation process unfolds in a specific sequence: low contrast occurs first, followed by blurring, and ending with noise.

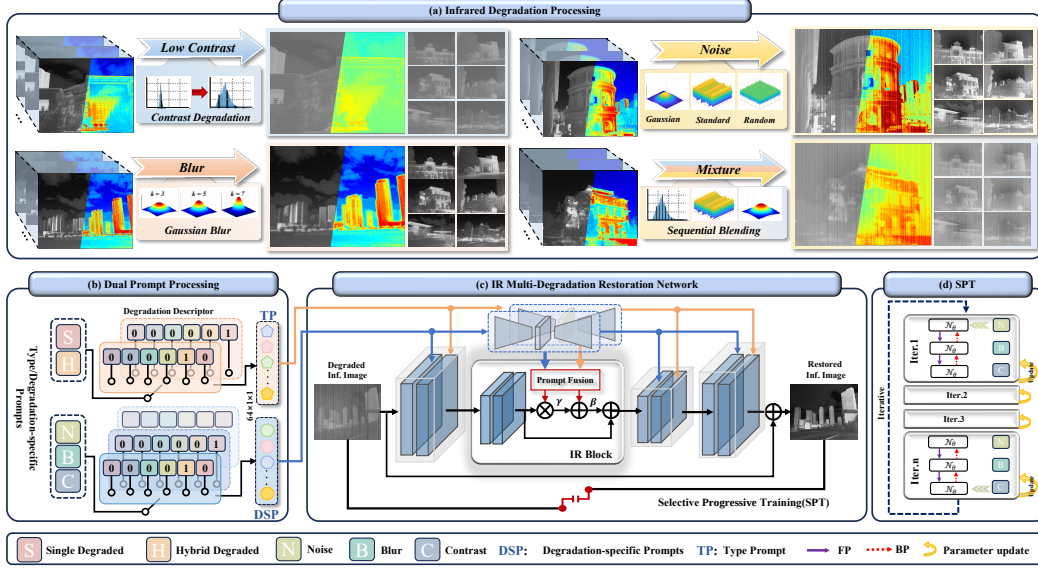


Figure 2: Schematic diagram of the proposed TIR enhancement framework. In subfigure (a), we first illustrate the TIR degradation process, including low contrast, blur, and noise across single and composite degradation scenarios. Subfigures (b) and (c) present details of the PPFN integrated with the image restoration model. Lastly, we depict our SPT in subfigure (d).

Therefore, given an observed clean image $\mathbf{I}_c$, the degraded image $\mathbf{I}_d$ can be formulated as:

$$\mathbf{I}_d = \left(\underbrace{\mathbf{n}_s \circ \mathbf{n}_o}_{\text{FPN}} \circ \underbrace{\mathcal{K}}_{\text{Blur}} \circ \underbrace{\mathcal{C}}_{\text{Low Contrast}} \right) (\mathbf{I}_c) + \mathbf{n}_r, \quad (1)$$

where $\mathcal{C}$, $\mathcal{K}$, $\mathbf{n}_o$, $\mathbf{n}_s$, and $\mathbf{n}_r$ represent the degradation with low contrast, blur kernel, optics, stripe, and additive random noise, respectively. $\circ$ denotes the composition operation.

As shown in Eq. (1), TIR degradation encompasses multiple types that strongly impact TIR images. To enable the base enhancement model to address various degradations in both complex composite and single degradation scenarios, we propose a prompt fusion learning strategy, as described in Sec. 3.2. Furthermore, to improve the model’s stability in addressing composite degradations, we introduce the Selective Progressive Training strategy, as described in Sec. 3.3.

3.2 Prompt Fusion Learning

The primary challenge in infrared image enhancement is the diverse range of degradation types, which single models cannot effectively address. Existing networks typically target specific degradations and struggle with complex composite ones. Although all-in-one restoration frameworks aim to remove multiple degradation types, they often falter with intricate composite degradations. To overcome this, we introduce the Progressive Prompt Fusion Network (PPFN), which enhances image restoration models for more effective infrared enhancement in complex scenarios. As shown in Figure 2(b) and (c), our PPFN comprises a dual-prompt processing module and a prompt fusion module.

In dual prompt processing, we introduce type-specific and degradation-specific prompts. The degradation-specific prompt $\mathbf{P}_{deg} := \{\mathbf{p}_{deg}^n, \mathbf{p}_{deg}^b, \mathbf{p}_{deg}^c\}$ guides the model to adapt degradation types, while the type-specific prompt $\mathbf{P}_{type} := \{\mathbf{p}_{type}^s, \mathbf{p}_{type}^h\}$ is utilized to enable the model to distinguish the difference between single and composite degradation scenarios. Here, n , b , and c represent noise, blurring, and contrast degradation, respectively, while s and h denote single and composite degradation scenarios. The degraded images processed by each step in either single or composite degradation scenarios with specific prompts $\mathbf{p}_{deg}^i \in \mathbf{P}_{deg}$, $i \in \{n, b, c\}$ and $\mathbf{p}_{type}^j \in \mathbf{P}_{type}$, $j \in \{s, h\}$. To extract prompt features, we first obtain the degradation-specific prompt feature $\mathbf{F}_{deg}^p$ and type-specific prompt feature $\mathbf{F}_{type}^p$ using two lightweight prompt encoders,

$\mathbf{E}_{deg}$ and $\mathbf{E}_{type}$, which are expressed as:

$$\begin{aligned}\mathbf{F}_{deg}^p &= \mathbf{E}_{deg}(\mathbf{p}_{deg}^i), i \in \{n, b, c\}, \\ \mathbf{F}_{type}^p &= \mathbf{E}_{type}(\mathbf{p}_{type}^j), j \in \{s, h\}.\end{aligned}\quad (2)$$

To represent the prompt more efficiently and guarantee subsequent injection being a conventional modulation manner [23, 18], a prompt fusion module is introduced. Specifically, we concatenate the two prompt features and then apply a linear layer $\mathcal{W}_{fusion}$, followed by a non-linear activation $\phi(\cdot)$, to obtain the final prompt feature $\mathbf{F}_p$, as expressed below:

$$\mathbf{F}_p = \phi(\mathcal{W}_{fusion}(\text{Cat}(\mathbf{F}_{deg}^p, \mathbf{F}_{type}^p))), \quad (3)$$

where the operator $\text{Cat}(\cdot, \cdot)$ denotes concatenate operation. To integrate the prompt into the model's feature space and enable adaptability across degradation and scenario type, we calculate two channel-wise modulation parameters with suitable dimension, γ and β , by applying a linear layer $\mathcal{W}_p$,

$$\gamma, \beta = \mathcal{W}_p(\mathbf{F}_p). \quad (4)$$

Given the l -th layer feature $\mathbf{F}_l \in \mathbb{R}^{h_l \times w_l \times c_l}$ in restoration model, with calculated modulation parameters $\gamma_l \in \mathbb{R}^{1 \times 1 \times c_l}$ and $\beta_l \in \mathbb{R}^{1 \times 1 \times c_l}$, this adaptation process can be expressed as follows:

$$\tilde{\mathbf{F}}_l = \mathbf{F}_l \otimes (1 + \gamma_l) + \beta_l, \quad (5)$$

where $\tilde{\mathbf{F}}_l$ is the updated model feature that will be passed to the next model block. By integrating PPFN module, the model enables more effective handling of composite degradations.



Figure 3: Example images from our HM-TIR benchmark include: (a) skyscraper, (b) seaside, (c) mountain, (d) cross-sea bridge, (e) pendulum, (f) tower, (g) camping area, (h) commercial street, (i) mansion, (j) square, (k) Ferris wheel, (l) boats, and (m) tourist attraction.

3.3 Selective Progressive Training

To address the distinct challenges of composite and single degradations in TIR enhancement, we introduce the Selective Progressive Training (SPT) mechanism, as described in Figure 2(d). SPT refines the degradation process by progressively enhancing each iteration through feedback loops. For composite degradations, where steps are applied sequentially, each iteration's output feeds into the next, enabling the model to learn and adapt to complex, interdependent degradation patterns. In contrast, for single degradations, where one type of degradation is present, a standard training framework is employed. Given a degradation process with N steps, we generate a sequence degraded images $\mathbf{I}_D := \{\mathbf{I}_d^1, \mathbf{I}_d^2, \dots, \mathbf{I}_d^N\}$ by using clean image $\mathbf{I}_c$ with corresponding to degradation-specific prompts $\mathbf{P}_{deg} := \{\mathbf{p}_{deg}^1, \mathbf{p}_{deg}^2, \dots, \mathbf{p}_{deg}^N\}$. As shown in Figure 2(a), for single degradation scenario, each degraded image is generated by using specific degradation to $\mathbf{I}_c$. While for composite scenario, the k -th degraded image $\mathbf{I}_d^k$ is generated by specific degradation to $k - 1$ -th degraded image $\mathbf{I}_d^{k-1}$. When training network, we set the initial input $\mathbf{I}_{in}^N = \mathbf{I}_d^N$ because the N -th degraded image contains all degradations in composite scenarios. Then network removes each degradation step in reverse order, enhancing the degraded images accordingly. For the k -th iteration of degradation removal training, given the input image $\mathbf{I}_{in}^k$, the restored output image $\mathbf{I}_{rest}^k$ is produced by the restoration model $\mathcal{N}_\theta$. GT for this iteration of the restoration model is defined as follows:

$$\mathbf{I}_{gt}^k = \begin{cases} \mathbf{I}_c, & \text{for single scenario,} \\ \mathbf{I}_d^{k-1}, & \text{for composite scenario.} \end{cases} \quad (6)$$

This setup ensures that only the i -th specific degradation is removed for both single and composite scenarios. Then we calculate the model loss gradient $\nabla_{\theta} \mathcal{L}(\mathbf{I}_{rest}^k, \mathbf{I}_{gt}^k)$ but do not update the network parameters. This approach prevents the model from focusing excessively on any single type of degradation while potentially neglecting others, and ensures that the training sequence does not interfere with single scenario training. For the next iteration’s input, if we use $\mathbf{I}_d^{k-1}$ directly in the composite scenario, the model will be affected by the removal of residual degradation from the previous iteration, leading to a significant drop in performance. To prevent this, we set the input $\mathbf{I}_{in}^{k-1}$ for the enhancement model in the next iteration (if it exists) as:

$$\mathbf{I}_{in}^{k-1} = \begin{cases} \mathbf{I}_d^{k-1}, & \text{for single scenario,} \\ \text{sg}(\mathbf{I}_{rest}^k), & \text{for composite scenario,} \end{cases} \quad (7)$$

where $\text{sg}(\cdot)$ denotes stop gradient operation to reduce training cost. After all iterations are completed, we update the model parameters using the sum of gradients computed across all iterations. In our TIR Enhancement setting, we define a three-step degradation process: noise, blur, and contrast. In the training phase, the degradations are added sequentially for composition scenarios: noise, blurring, and contrast. In the inference phase, we reverse this order to progressively remove the degradation: denoising, deblurring, and decontrast. The procedure is given in Alg. 1.

3.4 High-quality Multi-scenarios TIR Benchmark

Considering that limited diverse data has hindered the development of TIR domain, we establish a high-quality multi-scenario TIR benchmark, HM-TIR. It includes 1,503 TIR images encompassing various object types across different scenarios, as detailed in the last row of Table 1.

Each TIR image has a standard resolution of 640×512 and a wavelength range of 8 to 14 micrometers. To enhance thermal imaging performance by minimizing blur and increasing contrast, we individually adjusted the focus for each scene and secured the settings with mechanical tools before capturing. As shown in Figure 3, the HM-TIR benchmark includes a diverse structured environments, such as skyscrapers and Ferris wheels; unstructured settings like forests; and challenging scenarios like densely populated areas and small targets. We also incorporated various viewing angles, including aerial, eye-level, and low-angle.

Algorithm 1 Selective Progressive Training.

Require: Clean infrared images with $\{\mathbf{I}_c\}$, a restoration Network $\mathcal{N}_{\theta}$ and other necessary hyper-parameters.

- 1: **while** not converged **do**
- 2: Generate $\mathbf{I}_D$ and $\mathbf{P}_{deg}$ by randomly $\mathbf{p}_{type}$;
- 3: $\mathbf{I}_{in}^N = \mathbf{I}_d^N$;
- 4: **for** $k = N, \dots, 1$ **do**
- 5: $\mathbf{I}_{rest}^k = \mathcal{N}_{\theta}(\mathbf{I}_{in}^k, \mathbf{p}_{deg}^k, \mathbf{p}_{type})$;
- 6: Set GT image $\mathbf{I}_{gt}^k$ according to Eq. (6);
- 7: Calculate gradient $\nabla_{\theta} \mathcal{L}(\mathbf{I}_{rest}^k, \mathbf{I}_{gt}^k)$;
- 8: Set next input $\mathbf{I}_{in}^{k-1}$ according to Eq. (7);
- 9: **end for**
- 10: Update parameter θ by gradient descent;
- 11: **end while**
- 12: **return** θ^* .

4 Experimental Results

4.1 Training Details

Training and testing data. In our experiments, we trained the TIR enhancement model on our HM-TIR dataset, which contains 1,503 TIR images encompassing diverse object types across various scenarios. We divided the dataset into 80% for training and 20% for validation, ensuring a balanced evaluation of our model’s performance. For multi-degradation TIR enhancement testing, we created two validation subsets to enable a more detailed assessment: the Normal Set and the Hard Set. The Normal Set comprises images with lower levels of degradation, whereas the Hard Set includes images with more severe degradation. For single-degradation TIR enhancement testing, we applied the same settings as the Hard Set to create three separate single-degradation test subsets.

Training settings. We use Restormer [51] as the baseline model for TIR enhancement to evaluate our proposed module and strategy. All models are implemented in PyTorch on four 4090D GPUs with default settings. For the baseline model, we follow the Gated Degradation pipeline [54] to synthesize degradation, with the probabilities of all gates set to 0.8. During training, we adopt the L1 loss [48] and employ the Adam optimizer with parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Each model is trained

with a batch size of 4, using random cropping and flipping with a patch size of 256×256 . The initial learning rate is set to 8×10^{-5} and decays to 10^{-6} following a cosine annealing schedule. Each model is trained for a total of 300 epochs.

Evaluation metrics. In this work, the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) [48] are employed to assess the quality of the enhancement results under reference-based conditions. The PSNR and SSIM assess the quality of results primarily from the spatial dimension, with larger values indicating better results. For reference-free conditions, three no-reference Image Quality Assessment (IQA) metrics to evaluate image quality: NIMA [44], MUSIQ [15], and NIQE [36]. For NIMA and MUSIQ, higher values indicate better quality, while for NIQE, lower values are preferred.

4.2 Results on Multi-degradation TIR

To evaluate some TIR enhancement models, including WFAF [37], LRSID [3], and TSIRIE [39], as well as visible all-in-one restoration models such as DA-CLIP [32] and DiffUIR [59], we use the Normal Set to compare their TIR enhancement performance with our approach. Quantitative and qualitative comparisons are shown in Figure 4.

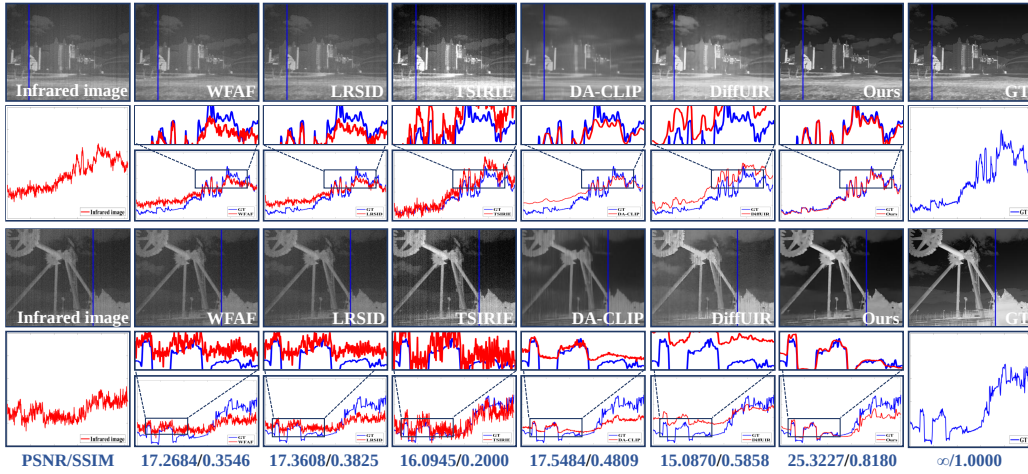


Figure 4: Quantitative and qualitative comparisons of signal performance across competitive image enhancement methods and our proposed approach. The average PSNR and SSIM values in our Normal Set are provided below the comparison figures in **blue**.

TIR enhancement methods such as WFAF, LRSID, and TSIRIE exhibit lower PSNR and SSIM values because they are tailored for single degradations and struggle with complex composite scenarios. In contrast, DA-CLIP and DiffUIR, developed as all-in-one enhancement methods for visible images, perform better; however, differences in imaging models between visible and infrared spectra lead to suboptimal results for infrared images. Our proposed method outperforms these approaches, achieving superior PSNR and SSIM scores and demonstrating enhanced signal restoration across multiple degradation scenarios. Qualitatively, traditional methods like WFAF, LRSID, and TSIRIE produce infrared images with substantial artifacts and background noise, while all-in-one approaches such as DA-CLIP and DiffUIR offer better restoration but still exhibit noticeable blurring and distortion. In contrast, our method excels at preserving critical structural information and fine details, reducing artifacts, and enhancing contrast.

We further evaluate our method alongside competitive approaches on the real-world Iray dataset [29] and adding an additional TIR enhancement method IE-CGAN [19]. Since it only provides the denoised result as ground truth, we use no-reference IQA metrics. Both quantitative and qualitative results are provided in Table 2 and Figure 5. Existing TIR enhancement methods struggle with complex scenarios, typically addressing only one type of degradation. Furthermore, all-in-one enhancement methods designed for visible images are ineffective in handling the specific degradations present in infrared image processing. In contrast, our approach not only outperforms existing methods in IQA scores but also shows superior restoration capabilities in real-world degradation conditions.

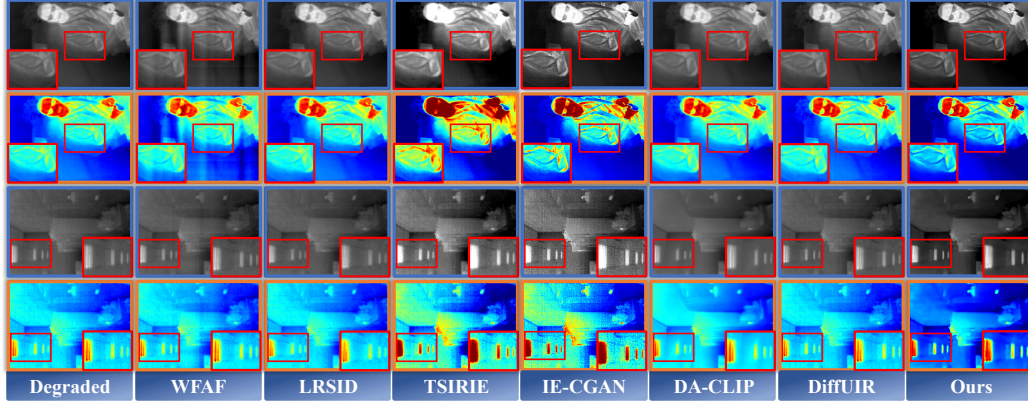


Figure 5: Qualitative comparisons of the competitive enhancement approaches and our method on Iray dataset.

Table 2: Quantitative comparison in Iray dataset. The best is in **red**, and the second-best is in **blue**. “↓” means lower value is better.

Metrics	Degraded	WFAF	LRSID	TSIRE	IE-CGAN	DA-CLIP	DiffUIR	Baseline	Ours
NIMA	3.5326	3.7321	3.5682	3.5359	3.4959	3.7004	3.5935	3.5812	3.8327
MUSIQ	25.2459	25.1264	24.2095	23.7508	29.0350	27.7855	26.8066	27.7829	30.9072
NIQE↓	10.1277	10.3536	8.6838	11.5204	9.4786	9.1896	9.3352	8.7776	8.4693

4.3 Results on Single-degradation TIR

To evaluate three TIR enhancement models with single degradation scenarios, we conduct experiments in test subsets with denoising, deblurring, and contrast enhancement.

For denoising, we compare four state-of-the-art approaches: AP-BSN [22], CycleISP [50], IDR [55], and SDAP [38]. For deblurring, we evaluate leading methods including DeBlurGANv2 [20], MIMO-UNet [7], FFTformer [17], and Stripformer [45]. For contrast enhancement, treated as a combination of haze and low-light enhancement, we include MSBDN [11] and FFA-Net [41] for dehazing, alongside LLFormer [46] and SCI [35] for low-light enhancement. Using three single-degradation subsets, we compared the performance of these methods relative to ours, with qualitative comparisons

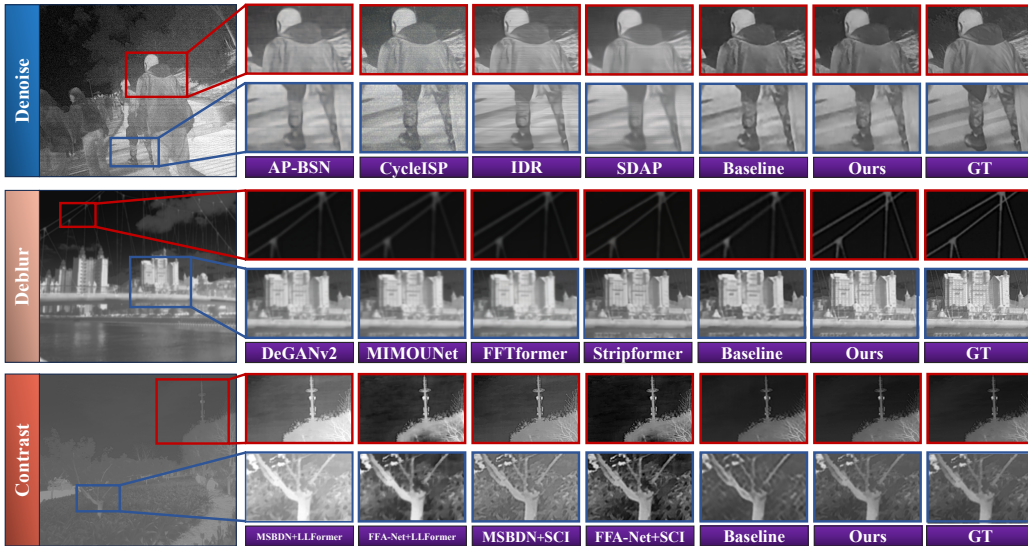


Figure 6: Qualitative comparison of single degradation between visible enhancement methods, baseline and our method on three single-degradation test sets, where “Baseline” refer to Restormer [51].

shown in Figure 6. While existing methods effectively reduce degradation, they still retain artifacts due to modeling differences between TIR and visible images, resulting in lower enhancement quality and fidelity. In contrast, our method delivers superior performance in single-degradation TIR enhancement tasks, effectively reducing noise, recovering fine details, and enhancing contrast while preserving the natural appearance of images.

Table 3: Quantitative results comparing our method with five models in our two test sets, both with and without integration of our PPFN module and SPT strategy. 'Average' refers to the mean value across test sets. † denotes methods using our approaches to train. Baseline and our approaches results are shown in and boxes, respectively. The best result is in **red**, and the second-best in **blue**.

Model	FocalNet PSNR/SSIM	FocalNet [†] PSNR/SSIM	UFormer PSNR/SSIM	UFormer [†] PSNR/SSIM	NAFNet PSNR/SSIM	NAFNet [†] PSNR/SSIM	XRestormer PSNR/SSIM	XRestormer [†] PSNR/SSIM	Restormer PSNR/SSIM	Restormer [†] PSNR/SSIM
Bridge	22.45/0.832 20.80/0.770	24.95/0.865 22.71/0.803	22.75/0.850 20.93/0.780	24.10/0.844 20.95/0.778	24.88/0.859 19.87/0.758	25.98/0.872 22.44/0.786	25.82/0.879 21.52/0.775	25.14/0.890 23.52/0.812	24.28/0.876 21.12/0.783	24.65/0.896 22.27/0.802
Leaning Tower	22.61/0.843 22.89/0.801	26.44/0.861 21.21/0.806	24.50/0.838 18.82/0.723	25.95/0.847 17.72/0.724	26.39/0.849 23.25/0.778	27.24/0.852 23.49/0.806	27.85/0.865 23.29/0.807	29.60/0.877 23.65/0.829	27.80/0.863 23.10/0.825	28.46/0.879 23.23/0.832
Tower	24.89/0.858 25.07/0.836	24.33/0.851 24.86/0.839	23.20/0.838 23.21/0.792	25.71/0.863 25.60/0.839	25.78/0.864 21.16/0.830	26.60/0.867 25.83/0.850	26.46/0.874 22.01/0.854	27.86/0.893 27.40/0.873	29.29/0.903 21.19/0.861	30.40/0.908 28.50/0.882
Two Skyscrapers	20.44/0.719 23.10/0.687	23.91/0.744 24.38/0.700	21.51/0.717 21.96/0.665	21.90/0.721 22.81/0.675	21.14/0.725 23.16/0.671	22.05/0.735 22.57/0.693	24.79/0.740 24.57/0.696	24.13/0.744 26.05/0.709	25.40/0.750 24.71/0.701	25.60/0.755 26.08/0.716
Villa	22.28/0.779 25.37/0.798	26.00/0.855 24.98/0.805	25.17/0.809 24.49/0.782	23.66/0.817 23.23/0.773	24.10/0.815 23.58/0.767	27.70/0.851 25.37/0.797	27.86/0.854 26.13/0.798	29.06/0.865 25.71/0.808	22.46/0.785 25.36/0.806	29.73/0.871 27.26/0.816
Average	21.22/0.778 21.27/0.733	22.63/0.790 21.40/0.740	21.95/0.775 20.31/0.714	21.62/0.768 20.44/0.716	22.29/0.776 21.81/0.731	23.74/0.792 22.30/0.744	23.54/0.801 22.43/0.748	24.75/0.811 23.06/0.758	23.28/0.796 22.87/0.757	25.32/0.818 23.27/0.764

4.4 Ablation studies

Validation of model architectures. In addition to Restormer, we evaluate our PPFN module with four other SOTA image enhancement models: NAFNet [5], UFormer [47], XRestormer [6], and FocalNet [9]. We compare the performance of these five models with and without our PPFN module. Quantitative and qualitative results are presented in Table 3 and Figure 7, respectively.

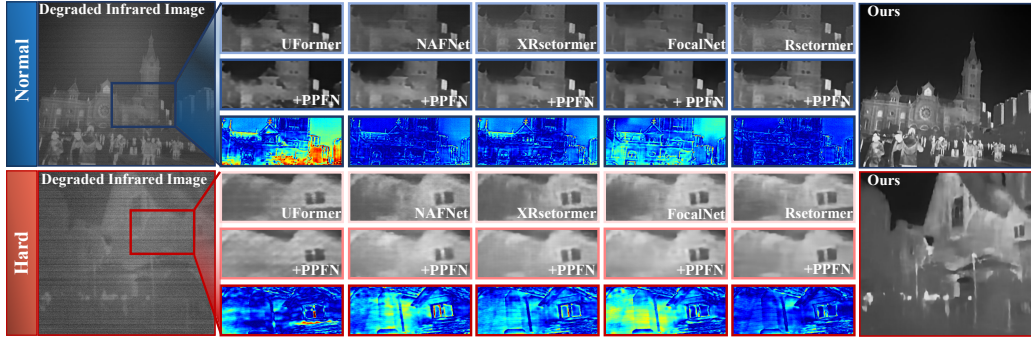


Figure 7: Visual comparison of five advanced methods with and without the integration of our approaches in Normal Set and Hard Set. Our method demonstrates superior visual quality and minimal error.

Quantitative and qualitative results show that all five baseline models exhibit lower PSNR and SSIM values and reduced enhancement quality on both Normal and Hard Sets, indicating their suboptimal performance with complex degradation. In contrast, integrating PPFN with each model consistently improves PSNR, SSIM, and TIR visual quality. Notably, our model achieves the best results, with an improvement of 8.76% on the Normal Set in PSNR.

Study on prompt fusion learning. We train models with the same settings as in previous comparison experiments. Testing is performed on the Hard Set, results are shown in Table 4. In dual prompt processing, applying degradation-specific and type/degradation-specific prompts achieves performance gains of 0.29 dB and 0.40 dB over the baseline, respectively. Regarding the prompt fusion strategy, removing non-linear activation or replacing the concatenation operation with multiplication results in a rapid PSNR decline, with the "Multiply" approach offering only minimal SSIM improvements. SPT reveals that directly applying iterative training to the baseline causes a PSNR drop of 0.23 dB.

Table 4: Ablation studies on the PPFN and SPT strategy. The best is in **red**, and the second-best is in **blue**.

# of Prompt	Prompt Fusion	Iter.	PSNR	SSIM
-	-	-	22.8678	0.7568
-	-	✓	22.6357	0.7524
DSP	-	✓	23.1605	0.7635
TP/DSP	w/o non-linear	✓	23.1487	0.7646
TP/DSP	Multiply	✓	23.1432	0.7627
TP/DSP	PPFN	1	14.5455	0.6125
TP/DSP	PPFN	2	14.6080	0.6261
TP/DSP	PPFN	3	23.2712	0.7643

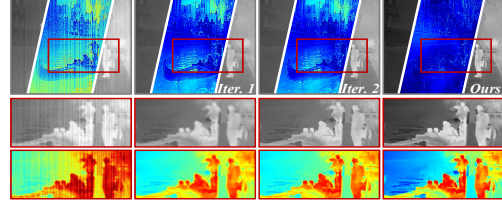


Figure 8: Analyzing the enhanced images and error maps from each iteration. Zoomed and pseudo-color maps for the best view.

Analyzing the enhancement iteration. We demonstrate the enhanced images from each iteration along with corresponding PSNR and SSIM values, as shown in Figure 8 and Table 4, respectively. Note that the iterations progress, specific degradations are incrementally removed, leading to a gradual improvement in both PSNR and SSIM metrics.

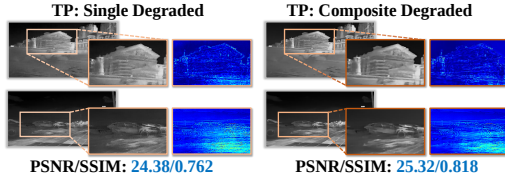


Figure 9: Quantitative and qualitative comparison of TIR enhancement performance between different type-specific prompt setting in Normal Set.

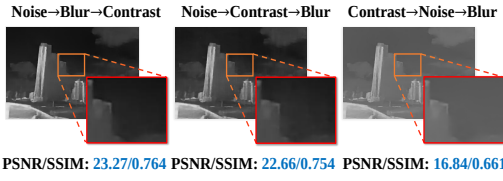


Figure 10: Quantitative and qualitative comparisons of our method with different order in each degradation removal process in Hard Set.

Analyzing the prompt sensitivity. We evaluate the performance of our method with incorrect prompts and order, as shown in Figure 9 and Figure 10. For incorrect prompts, we observe that using a single degradation scenario prompt with compositional degradation results in failure to remove degradation, with artifacts persisting. This indicates that the model struggles to eliminate residual degradation in each iteration under a single scenario. For incorrect order, we demonstrate that the model exhibits lower performance and PSNR when the degradation removal order is incorrect. This supports the hypothesis that optimal artifact removal occurs with a fixed processing order. Our results highlight that the SPT strategy effectively handles fixed-order degradation removal, leading to improved performance.

5 Conclusion

This paper introduced a new way for enhancing TIR images, managing complex degradation through dual-prompt processing and fusion modules. Our training scheme ensures robust performance across various scenarios. We also established a comprehensive TIR benchmark for accurate evaluation. Experiments show that PPFN surpasses existing methods in clarity, detail preservation, and contrast enhancement, advancing TIR image enhancement for broader applications.

Acknowledgment

This work is partially supported by the National Natural Science Foundation of China (Nos.62302078, 62372080, 62450072, U22B2052), the Distinguished Youth Funds of the Liaoning Natural Science Foundation (No.2025JH6/101100001), the Distinguished Young Scholars Funds of Dalian (No.2024RJ002), the China Postdoctoral Science Foundation (No.2023M730741) and the Fundamental Research Funds for the Central Universities.

References

- [1] Neelanjan Bhowmik, Jack W. Barker, Yona Falinie A. Gaus, and Toby P. Breckon. Lost in compression: The impact of lossy image compression on variable size object detection within infrared imagery. In

- Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 369–378, June 2022.
- [2] Lijing Cai, Xiangyu Dong, Kailai Zhou, and Xun Cao. Exploring video denoising in thermal infrared imaging: Physics-inspired noise generator, dataset and model. *IEEE Transactions on Image Processing*, 2024.
 - [3] Yi Chang, Luxin Yan, Tao Wu, and Sheng Zhong. Remote sensing image stripe noise removal: From image decomposition perspective. *IEEE Transactions on Geoscience and Remote Sensing*, 54(12):7018–7031, 2016.
 - [4] Yi Chang, Luxin Yan, Li Liu, Houzhang Fang, and Sheng Zhong. Infrared aerothermal nonuniform correction via deep multiscale residual network. *IEEE Geoscience and Remote Sensing Letters*, 16(7): 1120–1124, 2019.
 - [5] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European Conference on Computer Vision*, pages 17–33. Springer, 2022.
 - [6] Xiangyu Chen, Zheyuan Li, Yuandong Pu, Yihao Liu, Jiantao Zhou, Yu Qiao, and Chao Dong. A comparative study of image restoration networks for general backbone network design. In *European Conference on Computer Vision*, pages 74–91. Springer, 2024.
 - [7] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4641–4650, 2021.
 - [8] Marcos V Conde, Gregor Geigle, and Radu Timofte. Instructir: High-quality image restoration following human instructions. In *European Conference on Computer Vision*, pages 1–21. Springer, 2024.
 - [9] Yuning Cui, Wenqi Ren, Xiaochun Cao, and Alois Knoll. Focal network for image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13001–13011, 2023.
 - [10] Dan Ding, Ye Li, Peng Zhao, Kaitai Li, Sheng Jiang, and Yanxiu Liu. Single infrared image stripe removal via residual attention network. *Sensors*, 22(22):8734, 2022.
 - [11] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2157–2167, 2020.
 - [12] Yongsong Huang, Zetao Jiang, Rushi Lan, Shaoqin Zhang, and Kui Pi. Infrared image super-resolution via transfer learning and psrgan. *IEEE Signal Processing Letters*, 28:982–986, 2021.
 - [13] Haijun Jiang, Fei Chen, Xining Liu, Jesse Chen, Kai Zhang, and Li Chen. Thermal wave image deblurring based on depth residual network. *Infrared Physics & Technology*, 117:103847, 2021.
 - [14] Yitong Jiang, Zhaoyang Zhang, Tianfan Xue, and Jinwei Gu. Autodir: Automatic all-in-one image restoration with latent diffusion. In *European Conference on Computer Vision*, pages 340–359. Springer, 2024.
 - [15] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5148–5157, 2021.
 - [16] Kangwook Ko, Kyujin Shim, Kangil Lee, and Changick Kim. Large-scale benchmark for uncooled infrared image deblurring. *IEEE Sensors Journal*, 23(24):30119–30128, 2023.
 - [17] Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5886–5895, 2023.
 - [18] Xiangtao Kong, Chao Dong, and Lei Zhang. Towards effective multiple-in-one image restoration: A sequential and prompt learning strategy. *arXiv preprint arXiv:2401.03379*, 2024.
 - [19] Xiaodong Kuang, Xiubao Sui, Yuan Liu, Qian Chen, and Guohua Gu. Single infrared image enhancement using a deep convolutional neural network. *Neurocomputing*, 332:119–128, 2019.
 - [20] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8878–8887, 2019.
 - [21] Kangil Lee, Yuseok Ban, and Changick Kim. Motion blur kernel rendering using an inertial sensor: Interpreting the mechanism of a thermal detector. *Sensors*, 22(5):1893, 2022.
 - [22] Wooseok Lee, Sanghyun Son, and Kyoung Mu Lee. Ap-bsn: Self-supervised denoising for real-world images via asymmetric pd and blind-spot network. In *Proceedings of the IEEE/CVF Conference on*

- Computer Vision and Pattern Recognition*, pages 17725–17734, 2022.
- [23] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17452–17462, 2022.
 - [24] Xingyuan Li, Jinyuan Liu, Zhixin Chen, Yang Zou, Long Ma, Xin Fan, and Risheng Liu. Contourlet residual for prompt learning enhanced infrared image super-resolution. In *Proceedings of the European Conference on Computer Vision*, pages 270–288, 2024.
 - [25] Xingyuan Li, Zirui Wang, Yang Zou, Zhixin Chen, Jun Ma, Zhiying Jiang, Long Ma, and Jinyuan Liu. Difisir: A diffusion model with gradient guidance for infrared image super-resolution. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 7534–7544, 2025.
 - [26] Jinyuan Liu, Guanyao Wu, Zhu Liu, Di Wang, Zhiying Jiang, Long Ma, Wei Zhong, Xin Fan, and Risheng Liu. Infrared and visible image fusion: From data compatibility to task adaption. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(4):2349–2369, 2025.
 - [27] Jinyuan Liu, Bowei Zhang, Qingyun Mei, Xingyuan Li, Yang Zou, Zhiying Jiang, Long Ma, Risheng Liu, and Xin Fan. Dcevo: Discriminative cross-dimensional evolutionary learning for infrared and visible image fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2226–2235, June 2025.
 - [28] Li Liu, Luping Xu, and Houzhang Fang. 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, 2019.
 - [29] Qing Liu, Zhaoqi Xu, Jiansheng Wang, and Shuigen Wang. Infrared image denoising database, 2021. URL http://openai.raytrontek.com/apply/E_Image_noise_reduction.html/.
 - [30] Risheng Liu, Zhu Liu, Jinyuan Liu, Xin Fan, and Zhongxuan Luo. A task-guided, implicitly-searched and meta-initialized deep model for image fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(10):6594–6609, 2024.
 - [31] Zhu Liu, Zijun Wang, Jinyuan Liu, Fanqi Meng, Long Ma, and Risheng Liu. Deal: Data-efficient adversarial learning for high-quality infrared imaging. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 28198–28207, 2025.
 - [32] Ziwei Luo, Fredrik K. Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B. Schön. Controlling vision-language models for multi-task image restoration. In *The Twelfth International Conference on Learning Representations*, 2024. URL <https://openreview.net/forum?id=t3vnnLeajU>.
 - [33] Hui Lv, Pengfei Shan, Hongfang Shi, and Li Zhao. An adaptive bilateral filtering method based on improved convolution kernel used for infrared image enhancement. *Signal, Image and Video Processing*, 16(8):2231–2237, 2022.
 - [34] Jiaqi Ma, Tianheng Cheng, Guoli Wang, Qian Zhang, Xinggang Wang, and Lefei Zhang. Prores: Exploring degradation-aware visual prompt for universal image restoration. *arXiv preprint arXiv:2306.13653*, 2023.
 - [35] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5637–5646, 2022.
 - [36] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2012.
 - [37] Beat Münch, Pavel Trtik, Federica Marone, and Marco Stampanoni. Stripe and ring artifact removal with combined wavelet—fourier filtering. *Opt. Express*, 17(10):8567–8591, 2009.
 - [38] Yizhong Pan, Xiao Liu, Xiangyu Liao, Yuanzhouhan Cao, and Chao Ren. Random sub-samples generation for self-supervised real image denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12150–12159, 2023.
 - [39] Zhongxiang Pang, Guihua Liu, Guosheng Li, Jian Gong, Chunmei Chen, and Chao Yao. An infrared image enhancement method via content and detail two-stream deep convolutional neural network. *Infrared Physics & Technology*, 132:104761, 2023.
 - [40] Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. *Advances in Neural Information Processing Systems*, 36:71275–71293, 2023.
 - [41] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 11908–11915, 2020.

- [42] P Shanmugavadivu and K Balasubramanian. Particle swarm optimized multi-objective histogram equalization for image enhancement. *Optics & Laser Technology*, 57:243–251, 2014.
- [43] Yuyi Shao, Yingwei Sun, Mengmeng Zhao, Yankang Chang, Zhouzhou Zheng, Chengliang Tian, and Yan Zhang. Infrared image stripe noise removing using least squares and gradient domain guided filtering. *Infrared Physics & Technology*, 119:103968, 2021.
- [44] Hossein Talebi and Peyman Milanfar. Nima: Neural image assessment. *IEEE Transactions on Image Processing*, 27(8):3998–4011, 2018.
- [45] Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, Chung-Chi Tsai, and Chia-Wen Lin. Stripformer: Strip transformer for fast image deblurring. In *European Conference on Computer Vision*, pages 146–162. Springer, 2022.
- [46] Tao Wang, Kaihao Zhang, Tianrun Shen, Wenhan Luo, Bjorn Stenger, and Tong Lu. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 2654–2662, 2023.
- [47] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17683–17693, 2022.
- [48] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.
- [49] Shi Yi, Li Li, Xi Liu, Junjie Li, and Ling Chen. Hctirdeblur: A hybrid convolution-transformer network for single infrared image deblurring. *Infrared Physics & Technology*, 131:104640, 2023.
- [50] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Cycleisp: Real image restoration via improved data synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2696–2705, 2020.
- [51] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5728–5739, 2022.
- [52] Jinghao Zhang, Jie Huang, Mingde Yao, Zizheng Yang, Hu Yu, Man Zhou, and Feng Zhao. Ingredient-oriented multi-degradation learning for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5825–5835, 2023.
- [53] Jingwen Zhang, Xiaoxuan Zhou, Liyuan Li, Tingliang Hu, and Chen Fansheng. A combined stripe noise removal and deblurring recovering method for thermal infrared remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022.
- [54] Wenlong Zhang, Guangyuan Shi, Yihao Liu, Chao Dong, and Xiao-Ming Wu. A closer look at blind super-resolution: Degradation models, baselines, and performance upper bounds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 527–536, 2022.
- [55] Yi Zhang, Dasong Li, Ka Lung Law, Xiaogang Wang, Hongwei Qin, and Hongsheng Li. Idr: Self-supervised image denoising via iterative data refinement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2098–2107, 2022.
- [56] Zixiang Zhao, Haowen Bai, Yuanzhi Zhu, Jianshe Zhang, Shuang Xu, Yulun Zhang, Kai Zhang, Deyu Meng, Radu Timofte, and Luc Van Gool. Ddfm: denoising diffusion model for multi-modality image fusion. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8082–8093, 2023.
- [57] Zixiang Zhao, Haowen Bai, Jianshe Zhang, Yulun Zhang, Kai Zhang, Shuang Xu, Dongdong Chen, Radu Timofte, and Luc Van Gool. Equivariant multi-modality image fusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 25912–25921, 2024.
- [58] Zixiang Zhao, Lilun Deng, Haowen Bai, Yukun Cui, Zhipeng Zhang, Yulun Zhang, Haotong Qin, Dongdong Chen, Jianshe Zhang, Peng Wang, et al. Image fusion via vision-language model. In *International Conference on Machine Learning*, pages 60749–60765. PMLR, 2024.
- [59] Dian Zheng, Xiao-Ming Wu, Shuzhou Yang, Jian Zhang, Jian-Fang Hu, and Wei-Shi Zheng. Selective hourglass mapping for universal image restoration based on diffusion model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25445–25455, 2024.
- [60] Xiaoxuan Zhou, Jingwen Zhang, Mao Li, Xiaofeng Su, and Fansheng Chen. Thermal infrared spectrometer on-orbit defocus assessment based on blind image blur kernel estimation. *Infrared Physics & Technology*, 130:104538, 2023.