PTFUSION: PHOTOMETRIC AND TEXTURE PRIOR LEARNING FOR ILLUMINATION-ADAPTIVE INFRARED AND VISIBLE IMAGE FUSION

Anonymous authors
Paper under double-blind review

ABSTRACT

The fusion of infrared and visible images in low-light environments faces the problem of degraded photometric and texture information. The existing methods usually couple photometric restoration with texture compensation, which limits their adaptability to different lighting conditions. To address these issues, we propose PTFusion, a multi-illumination image fusion method based on photometric and texture priors. Specifically, we use images under multiple lighting conditions to learn photometric priors, mapping different lighting states to a unified photometric representation space to enhance the model's adaptability to different lighting conditions. At the same time, we construct a texture prior network to inject texture features into the fusion feature space to compensate for the lack of texture information and insufficient local structural details caused by lighting degradation. In the fusion process, we use photometric priors to adapt visible features and inject texture priors to compensate for lighting degradation and texture deficiencies. Then, prior of photometric and texture is used to supervise the fusion results, improving the photometric fidelity and texture details of the fused image. The experimental results show that our method has excellent multi-light adaptation ability and exhibits competitive comprehensive performance in fusion tasks and downstream visual tasks. The code will be released publicly.

1 INTRODUCTION

The fusion of infrared and visible images aims to integrate complementary information from different modal images into fused image, obtaining a more comprehensive scene representation Tang et al. (2023c); Lv et al. (2024); Bi et al. (2026). Infrared images can perceive the thermal radiation information and have good perception ability under low-light and complex weather conditions. Visible images can provide rich texture, color, and detail information. The fused images can effectively enhance scene perception in complex environments, and has therefore been widely applied in fields such as night surveillance, autonomous driving, and object detection Tang et al. (2022a).

Most existing methods typically assume that visible images can provide stable and reliable texture and structural information Ma et al. (2021); Tang et al. (2023b); Zhao et al. (2023b). However, in real-world scenarios, complex lighting changes can significantly affect the brightness, color, and texture information of visible images. In response to the performance degradation caused by low-light environments, researchers have proposed a series of methods for low-light scenes. DIVFusion Tang et al. (2023a) uses Retinex-based joint enhancement and fusion. SLFusion Lv et al. (2025) combines illumination enhancement with structural perception. DDBF Zhang et al. (2024) uses infrared guidance for low-light visible enhancement. SDSFusion Chen et al. (2025) introduces semantic perception into enhancement and fusion. LENFusion Chen et al. (2024) adopts brightness feedback for adaptive enhancement. LEFuse Cheng et al. (2025b) jointly performs low-light enhancement and infrared-visible fusion. DAFusion Wang et al. (2025a) learns degradation-aware representations for feature fusion. RDMFuse Wang et al. (2025b) employs Retinex decomposition for hierarchical illumination-adaptive fusion. However, existing methods are mainly tailored to low-light conditions, limiting their ability to maintain stable photometric and texture representations under diverse illumination.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

To address the above issue, we propose PTFusion, an illumination-adaptive fusion method that learns complementary photometric and texture priors for unified representation and effective fusion under diverse illumination conditions. Specifically, we use images under low light, normal light, and high light conditions to learn photometric and texture priors, in order to respectively characterize photometric changes and texture degradation under different lighting conditions. For photometric priors, we learn a unified photometric representation through multi illumination image reconstruction, mapping low light, normal light, and high light images to a consistent normal light feature space. After training, freeze the photometric encoder and transfer it to the visible branch of the fusion network. Simultaneously utilizing the normal light results generated by the prior decoder to perform photometric supervision on the fused image, improving the photometric fidelity and multi illumination adaptability of the fused result. For texture priors, we learn stable high-frequency texture representations from images under different lighting conditions to provide texture compensation for the fusion process. At the same time, using the texture prior network generated repair results as a reference, texture supervision is provided for the fusion results, further enhancing their texture details and structural expression ability. In addition, in response to the stable target perception ability of infrared images under complex lighting conditions, we introduce target perception constraints to enhance infrared information preservation and suppress texture compensation interference in key target areas, thereby improving the structural fidelity of the target area. Through the above design, PTFusion synergistically utilizes photometric recovery, texture compensation, and target information preservation to provide a unified representation and supervision for infrared and visible fusion under multiple lighting conditions. The experimental results show that PTFusion can improve the fusion quality under complex lighting conditions and effectively enhance the performance of downstream advanced visual tasks.

In summary, our main contributions are as follows:

- We propose PTFusion, a fusion method for infrared and visible images under multiple lighting conditions, which can adapt to different lighting environments such as low light, normal light, and high light, and achieve stable and effective image fusion under multiple lighting conditions.
- We decouple photometric and texture priors, construct a fusion and supervision mechanism guided by dual priors, adapt visible features through photometric priors, and use texture priors to compensate for texture and local structural information during the fusion process, thereby improving the fusion effect.
- The experimental results show that PTFusion exhibits excellent fusion performance under multi-light conditions, demonstrating strong multi-light adaptability and effectively improving the performance of subsequent advanced visual tasks.

2 Related Work

Conventional Image Fusion Methods. Traditional fusion methods usually assume that the input image can provide reliable information, so they mainly focus on complementary information extraction and feature reconstruction between different modalities. With the development of deep learning technology, data-driven image fusion methods have made significant progress. U2Fusion Xu et al. (2022) proposes a universal image fusion framework that can be applied to various fusion scenarios. FusionGAN Ma et al. (2019) introduces generative adversarial networks into image fusion tasks to improve the visual quality of fused images. SwinFusion Ma et al. (2022) utilizes Swin Transformer to model long-range dependency relationships, in order to obtain fusion results with more consistent structure and semantic expression. With the continuous deepening of image fusion research, researchers have further developed task driven fusion methods. SeAFusion Tang et al. (2022b) combines image fusion and semantic segmentation networks for joint training to improve the performance of fused images in downstream tasks. TarDal Liu et al. (2022b) combines object detection and semantic segmentation networks to enhance the semantic preservation ability of fusion results through task related features. MetaFusion Zhao et al. (2023a) introduces detection networks based on meta learning frameworks to achieve task oriented adaptive fusion learning. FILM Zhao et al. (2024), TextFusion Cheng et al. (2025a), and TextFuse Lv et al. (2026) further utilize the semantic features to guide the fusion process.

Low-Light Image Fusion Methods. In real-world scenarios, input images may not always provide reliable information. Under low light conditions, visible images are prone to severe brightness degradation, color distortion, and missing texture details, making it difficult for traditional fusion methods to achieve ideal fusion results. DIVFusion Tang et al. (2023a) combines Retinex theory to alleviate low light, texture loss, and color distortion issues. SLFusion Lv et al. (2025) combines illumination enhancement and structural perception modeling to preserve key structural information in low light areas. DDBF Zhang et al. (2024) utilizes the high contrast information of infrared images to guide low light visible image enhancement. SDSFusion Chen et al. (2025) introduces semantic perception mechanism to improve the fusion quality in degraded scenes. LENFusion Chen et al. (2024) achieves adaptive brightness enhancement through brightness feedback mechanism, improving fusion performance in dark environments. LEFuse Cheng et al. (2025b) combines low light enhancement and fusion optimization, and is designed for dark texture restoration and local overexposure suppression. DAFusion Wang et al. (2025a) guides feature fusion by learning degraded perceptual representations to reduce the propagation of degraded information during the fusion process. RDMFuse Wang et al. (2025b) uses Retinex decomposition for hierarchical feature fusion, combined with adaptive lighting optimization to alleviate low light degradation. Although these methods improve low-light fusion, their scene-specific enhancement strategies limit the stable photometric and textural representation of fused images under diverse illumination.

3 METHOD

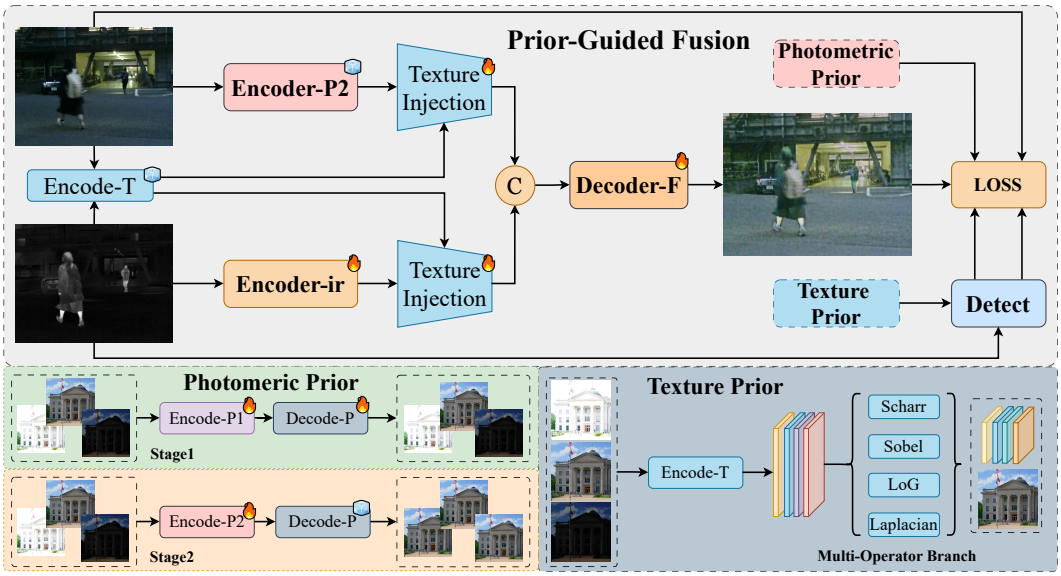


Figure 1: Overall framework of PTFusion, including photometric prior learning, texture prior learning, and prior-guided fusion.

3.1 OVERALL FRAMEWORK

As shown in Fig. 1, we propose PTFusion, which learns transferable photometric and texture priors to alleviate photometric degradation and texture loss under diverse illumination conditions. The entire framework consists of three stages: photometric prior learning, texture prior learning, and prior guided fusion. For photometric priors, we use images under low light, normal light, and high light conditions for training, allowing the network to learn photometric changes under different lighting conditions and map them to a unified normal photometric representation space. For the input visible image I_{vi}, a pre trained photometric prior encoder $E_P(\cdot)$ extracts photometric features:

$$F_{vi}^{P} = E_P(I_{vi}), \tag{1}$$

where F_{vi}^{P} represents the visible image photometric features obtained by pretraining the photometric encoder. For texture priors, we construct a texture prior network to learn high-frequency textures

and local structural representations under different lighting conditions. For the input visible and infrared images, the texture prior network extracts corresponding texture prior features:

$$F_{vi}^{T} = E_T(I_{vi}), \qquad F_{ir}^{T} = E_T(I_{ir}), \tag{2}$$

where $E_T(\cdot)$ represents the texture prior network, and F_{vi}^{T} and F_{ir}^{T} represent the visible and infrared texture prior features, respectively. Then, injecting texture features as compensation information into the corresponding fusion feature space to enhance the high-frequency details and local structural expression ability of the fusion features. To preserve infrared target fidelity, we introduce target-aware constraints that suppress texture compensation in key target regions. After photometric feature adaptation and texture prior injection, visible and infrared features are represented as F_{vi}^{PT} and F_{ir}^{T}, respectively. Subsequently, the two modal features are input into the fusion module:

$$I_{fuse} = F_{fuse}\left(F_{vi}^{PT}, F_{ir}^{T}\right), \tag{3}$$

where $F_{fuse}(\cdot)$ represents the fusion module, and I_{fuse} represents the fused image. In addition, PTFusion further uses the outputs of the prior networks to provide reliable photometric and texture supervision for fusion learning. Ultimately, the fusion process achieves adaptive complementary fusion of infrared and visible information under different lighting conditions, under the combined action of photometric prior and texture prior.

3.2 Photometric Prior Learning

We construct a photometric prior network and learn transferable photometric priors through multi-lighting image restoration tasks. The photometric prior network adopts an encoder decoder structure and is trained in two stages. In the first stage, we train a complete encoder and decoder to establish a stable image reconstruction capability for the network.

$$\hat{I}_{low}, \hat{I}_{normal}, \hat{I}_{high} = D_P\left(E_P(I_{\text{low}}, I_{\text{normal}}, I_{\text{high}})\right). \tag{4}$$

In the second stage, we freeze the decoder and update only the encoder, mapping low-, normal-, and high-illumination inputs to their corresponding normal-light targets:

$$\hat{I}_{normal} = D_P\left(E_P(I_{\text{low}}, I_{\text{normal}}, I_{\text{high}})\right), \tag{5}$$

where $\hat{I}$ represents the restored image. Since the decoder remains unchanged in this stage, the encoder needs to map the input under different lighting conditions to the unified photometric feature space established by the decoder, in order to learn photometric representations with cross lighting adaptability. After training, we freeze the photometric prior encoder and apply it to the visible branch in the fusion stage:

$$F_{vi}^{P} = E_P(I_{vi}), \tag{6}$$

where F_{vi}^{P} represents the photometric-adaptive visible feature, providing stable photometric information for fusion under diverse illumination conditions. Through the above training, the photometric prior encoder can extract more consistent photometric representations under different lighting conditions, and transferring them to the fusion network can reduce the dependence of the fusion model on specific lighting states.

3.3 Texture Prior Learning

We construct a texture prior network to provide texture priors for subsequent fusion. The texture prior network adopts a shared encoder and a multi branch structure. A shared texture encoder first extracts high-dimensional texture features from the input. The features are then processed by four branches using Scharr, Sobel, LoG, and Laplacian operators, with fixed gradient constraints to capture complementary local texture responses. The resulting representations are concatenated to form structured texture prior features:

$$F^{T} = E_T(I), \tag{7}$$

where F^{T} represents the structured texture prior features corresponding to the input image.

$$\mathcal{L}_T = \sum_{k \in \mathcal{K}} \lambda_k \mathcal{C}\left(\hat{G}_k - G_k^{gt}\right), \tag{8}$$

where $\mathcal{K}$ represents different fixed texture operators, $\hat{G}k$ and Gk^{gt} represent predicted texture responses and corresponding real texture responses, λ_k is the weight of each branch, and $\mathcal{C}(\cdot)$ represents the Charbonnier loss function. After training, we freeze the texture prior network and apply it separately to visible images and infrared images:

$$F_{vi}^{T} = E_T(I_{vi}), \qquad F_{ir}^{T} = E_T(I_{ir}). \tag{9}$$

where F_{vi}^{T} and F_{ir}^{T} respectively represent the texture prior features of visible and infrared images. The texture prior emphasizes local edges and structural details, providing stable texture information for fusion under diverse illumination conditions.

3.4 Prior-Guided Fusion Network

After obtaining pretrained photometric and texture priors, we construct a prior guided fusion network that incorporates both types of prior knowledge into fusion process. For the input visible image, we use the frozen photometric prior encoder to extract visible features with photometric adaptability:

$$F_{vi}^{P} = E_P(I_{vi}), \tag{10}$$

where F_{vi}^{P} represents visible features with photometric adaptability, providing stable photometric information for fusion under diverse illumination conditions. At the same time, we use frozen texture prior networks to extract texture priors for visible and infrared images respectively:

$$F_{vi}^{T} = E_T(I_{vi}), \qquad F_{ir}^{T} = E_T(I_{ir}). \tag{11}$$

Subsequently, we introduced texture prior features into the fusion network to compensate for the insufficient texture and structural information in visible and infrared images under different lighting conditions. Considering that texture compensation may interfere with the original structure of infrared targets, we further introduce target masks to protect reliable infrared information in critical target areas. Finally, the prior enhanced features of the two modalities are fused and reconstructed to obtain a fused image.

$$I_{fuse} = F_{fuse}\left(F_{vi}^{PT}, F_{ir}^{T}\right), \tag{12}$$

In order to fully utilize the reliable information provided by the pretrained prior network, we further introduce prior supervision in the fusion stage. We construct a joint loss function consisting of strength consistency, gradient consistency, structural similarity, and color consistency during the fusion phase. Among them, the enhanced visible image is provided by a pretrained photometric prior network to constrain the photometric performance of the fusion result. The gradient response generated by the texture prior network is used to provide texture structure supervision. Firstly, we use intensity consistency loss to constrain the overall photometric performance of the fusion result, and additionally enhance the information preservation of the infrared target area:

$$\mathcal{L}_{int} = \mathcal{C}(I_{vi}^{P}, I_{fuse}) + \mathcal{C}(I_{ir}, I_{fuse}) + 2\mathcal{C}(I_{ir}^{tar}, I_{fuse}^{tar}), \tag{13}$$

where $\mathcal{C}(\cdot,\cdot)$ represents the Charbonnier loss. In order to preserve the texture and local structural information in the source image, we further utilize the Sobel response generated by the texture prior network for gradient supervision:

$$\mathcal{L}_{grad} = \mathcal{C}(\nabla I_{fuse}, I_{vi}^{T}) + \mathcal{C}(\nabla I_{fuse}, I_{ir}^{T}) + 2\mathcal{C}\left(\nabla I_{fuse}^{tar}, \nabla I_{ir}^{tar}\right). \tag{14}$$

where $\nabla(\cdot)$ represents the gradient operator, I_{vi}^{T} and I_{ir}^{T} represent the Sobel texture response obtained by the texture prior network from visible and infrared images, respectively. In addition, we use structural similarity loss to further constrain the structural consistency between the fusion result and the source image:

$$\mathcal{L}_{ssim} = \mathcal{SSIM}(I_{vi}^{P}, I_{fuse}) + \mathcal{SSIM}(I_{ir}, I_{fuse}), \tag{15}$$

where $\mathcal{SSIM}(\cdot,\cdot)$ represents SSIM loss. In order to reduce the distortion of color information during the fusion process, we further constrain the color composition relationship between the fused image and the enhanced visible image in the normalized RGB space. For any RGB image, the normalized color representation is defined as:

$$\widetilde{I}^{c}(p) = \frac{I^{c}(p)}{\sum_{j\in\{R,G,B\}} I^{j}(p) + \epsilon}, \qquad c \in \{R, G, B\}. \tag{16}$$

On this basis, the color consistency loss is defined as:

$$\mathcal{L}_{color} = \mathcal{C}\left(\widetilde{I}_F, \widetilde{I}_V^E\right). \tag{17}$$

By constraining in the normalized RGB space, the impact of overall brightness differences on color errors can be reduced, thereby better maintaining the color composition relationship of visible images.

$$\mathcal{L}_{fuse} = \mathcal{L}_{int} + \mathcal{L}_{grad} + \mathcal{L}_{ssim} + 2\mathcal{L}_{color}. \tag{18}$$

Through the above design, PTFusion utilizes both photometric and textural priors during the fusion process, thus balancing the photometric and color information of visible, scene texture details, and reliable infrared target structure under different lighting conditions.

4 EXPERIMENT

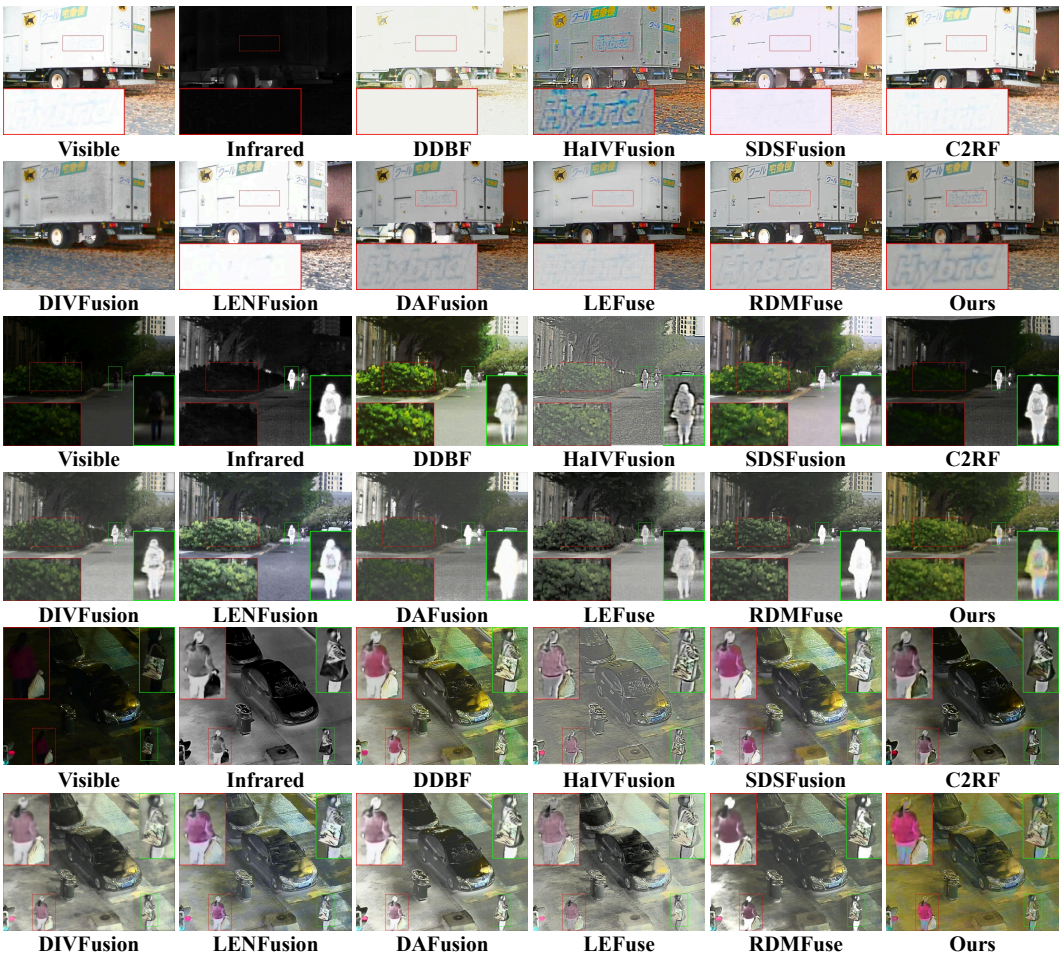


Figure 2: The performance of PTFusion and other methods in fusion task.

4.1 EXPERIMENTAL CONFIGURATIONS

During the training phase, we trained the photometric prior network and texture prior network using the MEF Cai et al. (2018) dataset, and trained the fusion network on the MSRS dataset. To comprehensively evaluate the effectiveness of PTFusion, we conducted qualitative and quantitative experiments on the MSRS Tang et al. (2022c), LLVIP Jia et al. (2021), M3FD Liu et al. (2022a), and DDL Tang et al. (2025a) datasets. To verify the effectiveness of the proposed method, we selected

9 fusion methods for comparison, including DDBF Zhang et al. (2024), HaIVFusion Gao et al. (2025), SDSFusion Chen et al. (2025), C2RF Tang et al. (2025b), DIVFusion Tang et al. (2023a), LENFusion Chen et al. (2024), DAFusion Wang et al. (2025a), LEFuse Cheng et al. (2025b) and RDMFuse Wang et al. (2025b). All comparison methods adopted their official implementation to ensure the fairness of experimental comparison. In terms of fusion quality evaluation, we use PSNR, CC, $Q^{ab/f}$, SSIM, MS-SSIM, and FMI to comprehensively evaluate the fusion results from multiple aspects such as pixel consistency, structural preservation ability, and information fidelity. We used the AdamW optimizer to optimize the network. The batch size for both the photometric prior network and the texture prior network is set to 4, and the batch size for the fusion network is set to 6. The initial learning rate of all networks is set to 0.001 and trained for 100 epochs. All experiments were conducted on the PyTorch framework and NVIDIA A40 GPU.

Table 1: Quantitative comparison on MSRS, M3FD and LLVIP datasets. **Red**: best. **Blue**: second. **Black**: third. ↓: lower is better.

MSRS	DDBF	HaIVFusion	SDSFusion	C2RF	DIVFusion	LENFusion	DAFusion	LEFuse	RDMFuse	Ours
PSNR	55.756	56.425	55.830	**62.430**	56.555	56.945	**57.675**	56.925	56.724	**61.276**
CC	0.4597	0.4649	0.4805	0.4612	**0.5312**	0.4873	0.4925	0.4638	**0.5164**	**0.5386**
$N^{ab/f}$ ↓	**0.3802**	0.8113	0.4325	**0.1956**	0.4240	0.6021	0.4170	0.4948	0.5643	**0.2986**
SSIM	0.8185	0.5020	0.8089	0.6748	**0.8477**	0.7455	**0.8607**	0.7794	0.7565	**0.8958**
MS-SSIM	0.7795	0.5932	0.8088	0.6068	**0.8134**	0.7314	**0.8434**	0.7780	0.7698	**0.8895**
FMI_{pixel}	0.8746	0.8752	**0.8811**	0.8764	0.8810	0.8685	**0.8962**	0.8597	0.8688	**0.8983**
M3FD	DDBF	HaIVFusion	SDSFusion	C2RF	DIVFusion	LENFusion	DAFusion	LEFuse	RDMFuse	Ours
PSNR	54.509	**60.954**	56.739	**61.450**	60.344	58.581	60.605	59.148	58.928	**62.955**
CC	0.3853	0.4006	0.4384	0.4212	**0.5286**	0.4585	**0.5512**	0.4771	0.5006	**0.5653**
$N^{ab/f}$ ↓	**0.1547**	0.7089	0.4148	**0.2245**	0.3921	0.5467	0.3965	0.4880	0.6520	**0.0975**
SSIM	0.8685	0.7404	0.8272	0.6591	**0.8806**	0.7961	**0.8944**	0.8484	0.7262	**0.8951**
MS-SSIM	0.8005	0.7656	0.8280	0.6765	**0.8490**	0.7968	**0.8830**	0.8515	0.7406	**0.8938**
FMI_{pixel}	**0.8643**	0.8597	0.8605	0.8433	0.8638	0.8459	**0.8760**	0.8619	0.8441	**0.8774**
LLVIP	DDBF	HaIVFusion	SDSFusion	C2RF	DIVFusion	LENFusion	DAFusion	LEFuse	RDMFuse	Ours
PSNR	57.167	57.988	57.368	**62.526**	**58.387**	57.948	58.456	57.731	58.270	**62.411**
CC	0.5816	0.5085	0.5801	0.5744	**0.6389**	0.5714	**0.6812**	0.5815	0.5956	**0.6678**
$N^{ab/f}$ ↓	0.4943	0.5616	0.4456	**0.2096**	0.3648	0.4951	**0.2349**	0.5038	0.5319	**0.2272**
SSIM	0.7933	0.7988	0.7944	0.6667	**0.8405**	0.7759	**0.9119**	0.7565	0.7662	**0.8845**
MS-SSIM	0.7444	0.7726	0.7453	0.6829	**0.7734**	0.7314	**0.8966**	0.7262	0.7053	**0.8483**
FMI_{pixel}	0.8427	**0.8694**	0.8414	0.8579	0.8679	0.8389	**0.8919**	0.8293	0.8445	**0.8731**
DDL Lowlight	DDBF	HaIVFusion	SDSFusion	C2RF	DIVFusion	LENFusion	DAFusion	LEFuse	RDMFuse	Ours
PSNR	57.620	57.510	58.074	**61.761**	57.810	58.199	58.657	57.618	**58.697**	**61.233**
CC	0.5057	0.4437	0.5194	0.4956	**0.5697**	0.4949	**0.6343**	0.4966	0.5379	**0.5489**
$N^{ab/f}$ ↓	0.3362	0.7199	0.2777	**0.1085**	**0.3214**	0.4553	0.3262	0.4025	0.4376	**0.1919**
SSIM	**0.8681**	0.5251	**0.8705**	0.6234	0.8512	0.7951	0.7663	0.8160	0.7962	**0.8855**
MS-SSIM	**0.8963**	0.5951	**0.9108**	0.5455	0.8619	0.8251	0.7353	0.8552	0.8571	**0.8945**
FMI_{pixel}	0.8706	0.8577	**0.8759**	0.8715	0.8664	0.8563	**0.8729**	0.8540	0.8585	**0.8781**
DDL Overexposure	DDBF	HaIVFusion	SDSFusion	C2RF	DIVFusion	LENFusion	DAFusion	LEFuse	RDMFuse	Ours
PSNR	51.673	57.443	53.154	55.223	57.624	54.855	**58.012**	**58.621**	56.283	**60.029**
CC	0.2927	0.4585	0.3859	0.3997	**0.5494**	0.4320	**0.6307**	**0.5146**	0.4937	0.5121
$N^{ab/f}$ ↓	**0.1954**	0.7485	0.4131	0.3269	**0.3139**	0.5169	0.3688	0.3410	0.5642	**0.2128**
SSIM	0.7172	0.5550	0.7050	0.5528	**0.8623**	0.7478	0.8459	**0.8753**	0.7156	**0.9066**
MS-SSIM	0.6512	0.6481	0.7517	0.5301	**0.8525**	0.7597	0.8501	**0.8920**	0.7688	**0.9197**
FMI_{pixel}	0.8405	0.8481	0.8399	0.8340	**0.8656**	0.8435	**0.8665**	0.8555	0.8356	**0.8679**

4.2 Comparative Experiment

Qualitative Results. The fusion results of different fusion methods under various lighting conditions are shown in Fig. 2. It can be observed that PTFusion exhibits more stable and consistent visual quality under different lighting conditions such as overexposure, low light at night, and low light during the day. Further observation reveals that PTFusion exhibits good comprehensive abilities in color preservation and texture restoration, and can balance visible texture information with infrared target saliency under different lighting conditions. In high light scenarios, the textual information on the surface of the vehicle can still be clearly preserved. In low light scenarios, the fusion result can effectively restore the color of the scene and maintain the significant structural information of the infrared target. In addition, the fused results exhibit good consistency in brightness, color, and structure across different illumination conditions, demonstrating the stable compensation provided by the photometric and texture priors for multi-illumination fusion. Overall, the qualitative results

show that PTFusion can maintain stable fusion performance under different lighting conditions and effectively coordinate scene texture, color information, and infrared target information.

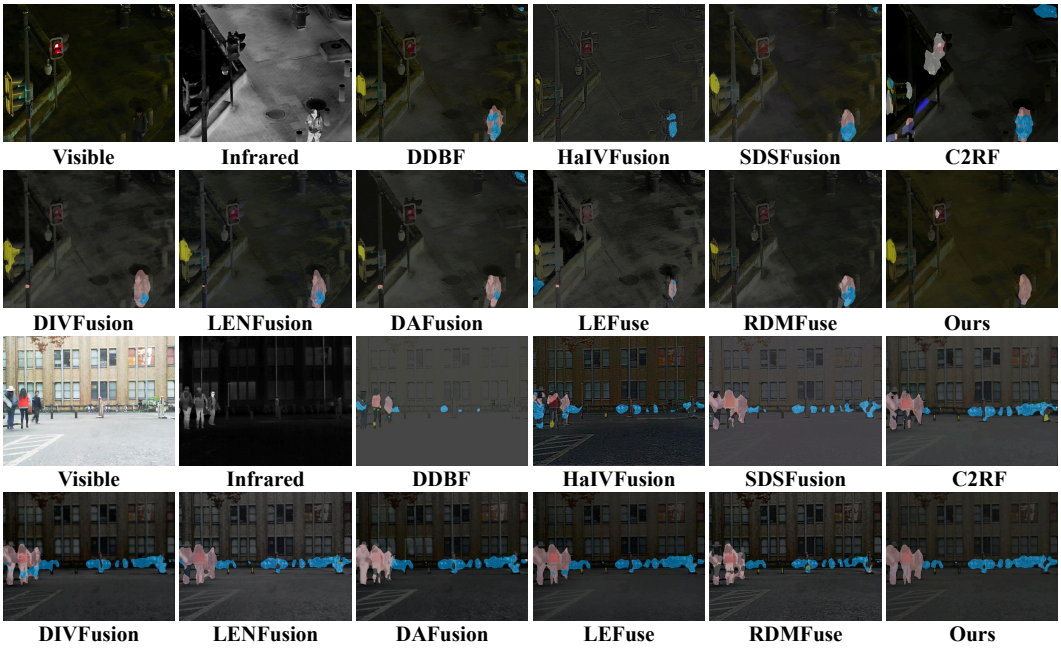


Figure 3: The performance of PTFusion and other methods in segmentation task.

Quantitative Results. In order to comprehensively evaluate the fusion performance of PTFusion in different degraded environments, we conducted quantitative comparisons on real datasets MSRS, LLVIP, and M3FD, as well as degraded datasets DDL. As shown in Table 1, PTFusion achieved competitive comprehensive performance on real datasets, indicating that the method can effectively coordinate infrared and visible information under real lighting conditions and achieve stable fusion. PTFusion also achieved excellent results on degraded datasets such as DDL, indicating that the learned photometric and texture priors can effectively cope with different types of image degradation and have cross scene adaptability. Based on the comprehensive analysis of different lighting conditions, degradation types, and various evaluation indicators, it can be seen that PTFusion can maintain relatively stable fusion performance in different degradation environments. The above experimental results have verified the effectiveness of PTFusion in terms of fusion quality and multi-light adaptability. In particular, when the degradation conditions vary, the pre-trained priors provide relatively stable photometric and texture information for the fusion network, thereby mitigating the impact of degraded inputs on the fused results.

Segmentation Performance. To further evaluate the support capability of the fusion results for downstream semantic segmentation tasks, we conducted semantic segmentation experiments on the fused images using DeepLabV3+ Chen et al. (2018). As shown in Fig. 3, PTFusion can obtain relatively accurate segmentation results in both low light and strong light environments. At the same time, under complex lighting conditions, PTFusion can effectively maintain the structural integrity of the target area, reduce the impact of lighting degradation on the expression of the target area, and provide more comprehensive visual information for subsequent segmentation. As shown in Table 2, PTFusion has achieved excellent performance in semantic segmentation evaluation metrics, further verifying the effective support capability of our method for advanced visual tasks. Overall, PTFusion not only improves the fusion quality in complex lighting environments, but also enhances th of key target areas, providing more reliable visual representations for high-level tasks.

4.3 Ablation Study

To verify the effectiveness of photometric and texture priors, we removed both types of priors for ablation experiments and qualitatively compared the fusion results. As shown in Fig. 3, after re-

Table 2: Segmentation mIoU comparison on the road obstacle dataset. **Red**: best. **Blue**: second.

Segment	bump	cone	guardrail	car_stop	curve	bike	person	car	mIoU
DDBF	16.85	61.70	74.08	57.82	18.46	62.37	55.42	**84.48**	53.90
HaIVFusion	42.06	41.46	30.11	60.10	3.62	34.78	31.16	66.56	38.73
SDSFusion	36.61	**63.10**	**78.81**	59.92	24.60	62.67	58.71	83.84	58.53
C2RF	**58.26**	55.69	57.71	62.32	31.43	54.29	**62.93**	78.29	57.62
DIVFusion	48.44	**64.75**	76.29	60.92	30.87	54.88	55.55	84.32	**59.50**
LENFusion	43.31	62.07	70.58	60.73	30.76	58.16	54.69	84.35	58.08
DAFusion	40.62	61.92	64.59	59.26	26.07	59.93	61.57	82.94	57.11
LEFuse	27.56	62.63	76.35	**62.29**	**36.04**	**62.73**	49.82	84.05	57.68
RDMFuse	37.58	63.07	**82.30**	57.37	25.31	**62.74**	57.56	84.21	58.77
Ours	**61.11**	62.61	77.05	**64.39**	**37.22**	62.34	**62.06**	**85.50**	**64.03**

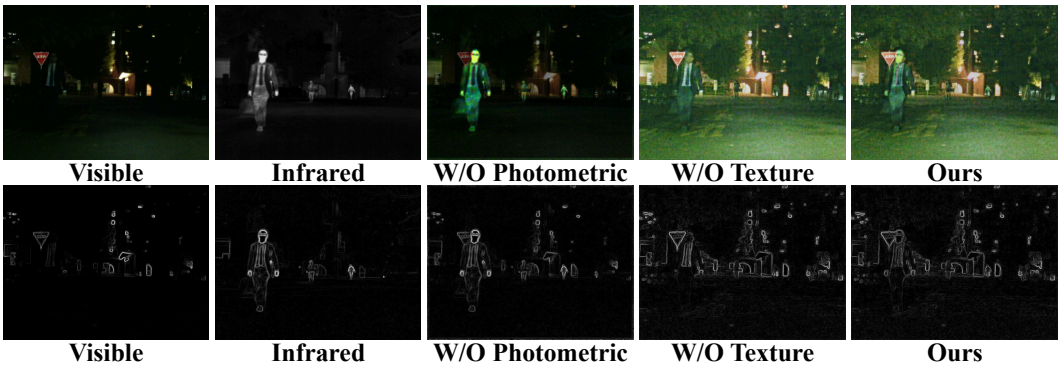


Figure 4: The performance of PTFusion and other methods in segmentation task.

moving the photometric prior, the adaptability of the fusion result to different lighting conditions significantly decreases, making it difficult to maintain stable photometric performance. This indicates that photometric prior can effectively enhance the multi light adaptability of the fusion process. After removing texture priors, the texture details and local structure of the fused image decreased, and some areas showed missing details, indicating that texture priors can effectively compensate for the loss of texture information caused by lighting degradation. Moreover, the two priors exhibit complementary effects across different illumination conditions, with the photometric prior stabilizing appearance and the texture prior preserving fine details. The above results indicate that the photometric prior and texture prior play a role in both photometric characterization and texture details, respectively, jointly improving the fusion performance of PTFusion under different lighting conditions.

5 Conclusion

In this paper, we proposed PTFusion, a photometric and texture prior learning framework for infrared and visible image fusion under diverse illumination conditions. To address the degradation of photometric and textural information under diverse illumination conditions, PTFusion introduces photometric and texture priors into the fusion process. During fusion, the photometric prior is used to adapt visible features, while texture priors are injected into the fusion features to compensate for missing texture details and local structures. In addition, prior-generated results are employed as reference supervision to provide photometric and texture guidance for fusion learning, while target-aware constraints help preserve reliable infrared information in important target regions. Extensive experiments on multiple real and degraded datasets demonstrate that PTFusion achieves competitive fusion performance under diverse illumination conditions. Moreover, improved downstream semantic segmentation performance further demonstrates the effectiveness of PTFusion in high-level task. Overall, PTFusion enables illumination-adaptive infrared and visible image fusion through photometric and structured texture priors.

AI USE STATEMENT

In this work, we used ChatGPT for code modification, technical discussions, and LaTeX formatting. All research ideas, experimental designs, results, analyses, and conclusions were developed and verified by the authors. All AI-assisted content was reviewed by the authors, who take full responsibility for the final manuscript.

REFERENCES

Chongke Bi, Chaoqun Sima, and Lu Yang. RegFuse: Registration Meets Infrared and Visible Image Fusion. In *ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 10417–10421, 2026. doi: 10.1109/ICASSP55912.2026.11461749.

Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018.

Jun Chen, Liling Yang, Wei Liu, Xin Tian, and Jiayi Ma. LENFusion: A Joint Low-Light Enhancement and Fusion Network for Nighttime Infrared and Visible Image Fusion. *IEEE Transactions on Instrumentation and Measurement*, 73:1–15, 2024.

Jun Chen, Liling Yang, Wei Yu, Wenping Gong, Zhanchuan Cai, and Jiayi Ma. SDSFusion: A Semantic-Aware Infrared and Visible Image Fusion Network for Degraded Scenes. *IEEE Transactions on Image Processing*, 34:3139–3153, 2025.

Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 833–851, 2018.

Chunyang Cheng, Tianyang Xu, Xiao-Jun Wu, Hui Li, Xi Li, Zhangyong Tang, and Josef Kittler. TextFusion: Unveiling the power of textual semantics for controllable image fusion. *Information Fusion*, 117:102790, 2025a. ISSN 1566-2535. doi: https://doi.org/10.1016/j.inffus.2024.102790.

Muhang Cheng, Haiyan Huang, Xiangyu Liu, Hongwei Mo, Xiongbo Zhao, and Songling Wu. LEFuse: Joint low-light enhancement and image fusion for nighttime infrared and visible images. *Neurocomputing*, pp. 129592, 2025b.

Xiang Gao, Yongbiao Gao, Aimei Dong, Jinyong Cheng, and Guohua Lv. HaIVFusion: Haze-free infrared and visible image fusion. *IEEE/CAA Journal of Automatica Sinica*, 12(10):2040–2055, 2025.

Xinyu Jia, Chuang Zhu, Minzhen Li, Wenqi Tang, and Wenli Zhou. LLVIP: A visible-infrared paired dataset for low-light vision. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3496–3504, 2021.

Jinyuan Liu, Xin Fan, Zhanbo Huang, Guanyao Wu, Risheng Liu, Wei Zhong, and Zhongxuan Luo. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5802–5811, 2022a.

Jinyuan Liu, Xin Fan, Zhanbo Huang, Guanyao Wu, Risheng Liu, Wei Zhong, and Zhongxuan Luo. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5802–5811, 2022b.

Guohua Lv, Chaoqun Sima, Yongbiao Gao, Aimei Dong, Guangxiao Ma, and Jinyong Cheng. SIGFusion: Semantic information-guided infrared and visible image fusion. *IEEE Transactions on Instrumentation and Measurement*, 73:1–18, 2024. doi: 10.1109/TIM.2024.3457951.

Guohua Lv, Xiang Gao, Aimei Dong, Zhonghe Wei, and Jinyong Cheng. SLFusion: A Structure-Aware Infrared and Visible Image Fusion Network for Low-Light Scenes. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(12):11877–11892, 2025.

Guohua Lv, Chaoqun Sima, Juncan Wang, Yongbiao Gao, Zhonghe Wei, and Jinyong Cheng. TextFuse: Text Information Guided Infrared and Visible Image Fusion. *IEEE Transactions on Multimedia*, pp. 1–15, 2026. doi: 10.1109/TMM.2026.3673589.

Jiayi Ma, Wei Yu, Pengwei Liang, Chang Li, and Junjun Jiang. FusionGAN: A generative adversarial network for infrared and visible image fusion. *Information Fusion*, 48:11–26, 2019.

Jiayi Ma, Linfeng Tang, Meilong Xu, Hao Zhang, and Guobao Xiao. STDFusionNet: An infrared and visible image fusion network based on salient target detection. *IEEE Transactions on Instrumentation and Measurement*, 70:1–13, 2021.

Jiayi Ma, Linfeng Tang, Fan Fan, Jun Huang, Xiaoguang Mei, and Yong Ma. SwinFusion: Cross-domain long-range learning for general image fusion via swin transformer. *IEEE/CAA Journal of Automatica Sinica*, 9(7):1200–1217, 2022.

Linfeng Tang, Yuxin Deng, Yong Ma, Jun Huang, and Jiayi Ma. SuperFusion: A versatile image registration and fusion network with semantic awareness. *IEEE/CAA Journal of Automatica Sinica*, 9(12):2121–2137, 2022a.

Linfeng Tang, Jiteng Yuan, and Jiayi Ma. Image fusion in the loop of high-level vision tasks: A semantic-aware real-time infrared and visible image fusion network. *Information Fusion*, 82: 28–42, 2022b.

Linfeng Tang, Jiteng Yuan, Hao Zhang, Xingyu Jiang, and Jiayi Ma. PIAFusion: A progressive infrared and visible image fusion network based on illumination aware. *Information Fusion*, 83-84:79–92, 2022c. ISSN 1566-2535. doi: https://doi.org/10.1016/j.inffus.2022.03.007.

Linfeng Tang, Xinyu Xiang, Hao Zhang, Meiqi Gong, and Jiayi Ma. DIVFusion: Darkness-free infrared and visible image fusion. *Information Fusion*, 91:477–493, 2023a.

Linfeng Tang, Hao Zhang, Han Xu, and Jiayi Ma. Rethinking the necessity of image fusion in high-level vision tasks: A practical infrared and visible image fusion network based on progressive semantic injection and scene fidelity. *Information Fusion*, 99:101870, 2023b. ISSN 1566-2535. doi: https://doi.org/10.1016/j.inffus.2023.101870.

Linfeng Tang, Hao Zhang, Han Xu, and Jiayi Ma. Deep learning-based image fusion: A survey. *Journal of Image and Graphics*, 28(1):3–36, 2023c.

Linfeng Tang, Yeda Wang, Zhanchuan Cai, Junjun Jiang, and Jiayi Ma. ControlFusion: A Controllable Image Fusion Network with Language-Vision Degradation Prompts. In *Advances in Neural Information Processing Systems*, volume 38, pp. 118304–118327, 2025a.

Linfeng Tang, Qinglong Yan, Xinyu Xiang, Leyuan Fang, and Jiayi Ma. C2RF: Bridging Multi-modal Image Registration and Fusion via Commonality Mining and Contrastive Learning. *International Journal of Computer Vision*, 133(8):5262–5280, 2025b.

Xue Wang, Zheng Guan, Wenhua Qian, Jinde Cao, Runzhuo Ma, and Cong Bi. A degradation-aware guided fusion network for infrared and visible image. *Information Fusion*, pp. 102931, 2025a. ISSN 1566-2535.

Xue Wang, Wenhua Qian, Zheng Guan, Jinde Cao, RunZhuo Ma, and Chengchao Wang. A retinex decomposition model-based deep framework for infrared and visible image fusion. *IEEE Journal of Selected Topics in Signal Processing*, 19(1):154–168, 2025b.

Han Xu, Jiayi Ma, Junjun Jiang, Xiaojie Guo, and Haibin Ling. U2Fusion: A unified unsupervised image fusion network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1): 502–518, 2022.

Hao Zhang, Linfeng Tang, Xinyu Xiang, Xuhui Zuo, and Jiayi Ma. Dispel Darkness for Better Fusion: A Controllable Visual Enhancer Based on Cross-Modal Conditional Adversarial Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 26477–26486, 2024.

Wenda Zhao, Shigeng Xie, Fan Zhao, You He, and Huchuan Lu. Metafusion: Infrared and visible image fusion via meta-feature embedding from object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13955–13965, 2023a.

Zixiang Zhao, Haowen Bai, Jiangshe Zhang, Yulun Zhang, Shuang Xu, Zudi Lin, Radu Timofte, and Luc Van Gool. CDDFuse: Correlation-Driven Dual-Branch Feature Decomposition for Multi-Modality Image Fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5906–5916, June 2023b.

Zixiang Zhao, Lilun Deng, Haowen Bai, Yukun Cui, Zhipeng Zhang, Yulun Zhang, Haotong Qin, Dongdong Chen, Jiangshe Zhang, Peng Wang, and Luc Van Gool. Image fusion via vision-language model. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2024.