SLCFormer: Spectral-Local Context Transformer with Physics-Grounded Flare Synthesis for Nighttime Flare Removal

Xiyu Zhu[1,2], Wei Wang[1,2*], Xin Yuan[2], Xiao Wang[1]

[1]School of Computer Science and Technology, Wuhan University of Science and Technology
[2]Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan University of Science and Technology
xiyu_zhu@wust.edu.cn, wangwei8@wust.edu.cn, xinyuan@wust.edu.cn, wangxiao2021@wust.edu.cn

Abstract

Lens flare is a common nighttime artifact caused by strong light sources scattering within camera lenses, leading to hazy streaks, halos, and glare that degrade visual quality. However, existing methods usually fail to effectively address nonuniform scattered flares, which severely reduces their applicability to complex real-world scenarios with diverse lighting conditions. To address this issue, we propose SLCFormer, a novel spectral-local context transformer framework for effective nighttime lens flare removal. SLCFormer integrates two key modules: the Frequency Fourier and Excitation Module (FFEM), which captures efficient global contextual representations in the frequency domain to model flare characteristics, and the Directionally-Enhanced Spatial Module (DESM) for local structural enhancement and directional features in the spatial domain for precise flare removal. Furthermore, we introduce a ZernikeVAE-based scatter flare generation pipeline to synthesize physically realistic scatter flares with spatially varying PSFs, bridging optical physics and data-driven training. Extensive experiments on the Flare7K++ dataset demonstrate that our method achieves state-of-the-art performance, outperforming existing approaches in both quantitative metrics and perceptual visual quality, and generalizing robustly to real nighttime scenes with complex flare artifacts.

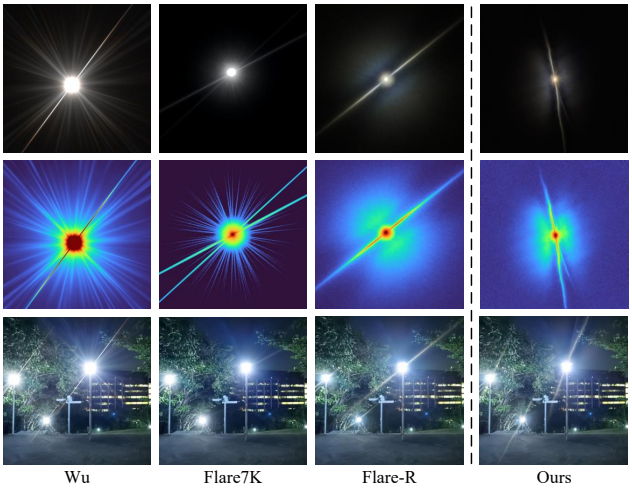

Figure 1: The comparison of scattering flare. We improve on the original dataset by embedding optical aberrations in the PSF for spatially varying scattering flares, as visualized in the intensity distribution heatmap.

Introduction

Lens flare is a common optical phenomenon that is generated primarily from strong light entering a camera lens and scattering through several internal components such as lens surfaces, coatings, or iris mechanisms (Ernst, Akenine-Möller, and Jensen 2005; Hullin et al. 2011). These artifacts are especially prominent at night, where bright light sources like streetlights or headlights contrast sharply with dark backgrounds, severely degrading image quality and hindering downstream vision tasks (Shao et al. 2025; Li et al. 2025) such as object detection (Girshick et al. 2014), autonomous driving (Bojarski et al. 2016) and depth estimation.

To address the lens flare problem, researchers have used traditional hardware and software solutions. Hardware designs modify optical systems (lens coatings or aperture adjustments) to suppress flare (Raut et al. 2011), but they are costly and inflexible. Software-based techniques rely on brightness thresholding (Vitoria and Ballester 2019), edge cues (Xie and Tu 2015), or image decomposition, yet they only handle simple flare patterns and struggle with diverse real-world distributions.

With the rise of deep learning in computer vision (Vaswani et al. 2017; Zamir et al. 2022), flare removal techniques have been driven from experiential rules to end-to-end learning frameworks. An important benchmark in this transition is the work of (Wu et al. 2021) who debuted the first large-scale synthetic dataset for lens flare removal, enabling supervised learning across diverse flare types and lighting conditions. They used a semi-synthetic approach to synthesize flare images with clean scenes, which greatly improved the feasibility of deep learning-based methods for flare removal. Building on this foundation, the Flare7K (Dai et al. 2022) and enhanced Flare7K++ (Dai et al. 2023) datasets incorporate more different flare categories (including glare, light sources, and reflection artifacts), enriching the diversity and realism of the training data.

Recent methods have explored increasingly sophisticated

*Corresponding author: wangwei8@wust.edu.cn

architectures to address the multifaceted nature of flare artifacts. Despite some progress in recent years, most existing flare removal models still face significant problems in representational and practical applicability. Many methods treat flare removal as a traditional image restoration problem without explicitly distinguishing flare artifacts from scene content. Additionally, many networks use attention-based modules (Liu et al. 2021), which are computationally expensive, though they are powerful. Some recent attempts have employed multi-branch designs to preserve better detail, but the cost of architecture and training instability increased. These challenges highlight the need for more efficient flare removal frameworks that are adaptable to a wide range of flare types while preserving image fidelity and remaining computationally tractable.

To solve this problem, we first construct a ZernikeVAE-based scatter flare generation pipeline for enhancing the complexity of the dataset with the realism of synthetic flares (shown in Fig. 1). The method generates the diffracted flare patterns by modeling the spatially varying point spread function (PSF) to break the limitation of the simple superposition of traditional synthetic flare. Unlike previous works that assume radially symmetric or uniform flare distributions, this is the first to capture non-uniform flares lacking central symmetry in both intensity and direction. On this basis, we propose SLCFormer, a hybrid frequency-space transformer for nighttime flare removal. It integrates two core modules: FFEM, which leverages Fourier transform and channel attention to efficiently capture global frequency features of flare, and DESM, which enhances local structural and directional modeling via directionally-aware convolutions. In summary, our contributions are as follows:

- A ZernikeVAE-based scatter flare generation pipeline that synthesizes more realistic and complex scatter flares by modeling physically spatially varying PSFs.
- We design SLCFormer, a novel spectral-local context transformer architecture that integrates frequency-domain global context modeling with spatial-domain directional feature enhancement for effective flare removal.
- Our method achieves superior performance compared to most existing models on both synthetic and real-world datasets.

Related Work

Datasets for Flare Removal

Earlier studies tended to use synthetic datasets, which limit the generalizability of the models. (Wu et al. 2021) proposed a semi-synthetic dataset, but the diversity and realism of the captured flares were limited because they were acquired under the same conditions. To address this, the Flare7K (Dai et al. 2022) dataset introduced a more complex flare pattern using optical flares that simulate real-world scattering and reflective flares. Dai's follow-up result, Flare7K++ (Dai et al. 2023), integrated real flares from contaminated lenses and enhanced the dataset's ability to model both scattered and reflected flares, resulting in better training for flare removal.

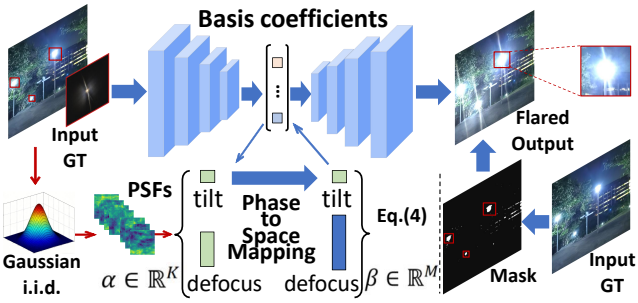

Figure 2: The ZernikeVAE-based flare synthesis pipeline. The simulator generates physically realistic flare images by modeling multi-dimensional PSFs with Zernike polynomials and Fourier optics.

Other datasets have explored different aspects of the problem (Jin et al. 2023; Qu et al. 2024; He et al. 2025). (Qiao, Hancke, and Lau 2021) collected unpaired flare-damaged images and pairs of flare-free images, which enabled an unpaired learning strategy but made it difficult to train pixel-level models. (Deng et al. 2024) proposed the WiderFlare dataset, a real-world dataset customized for flare removal benchmarking in a blind environment, which contains flare severity annotations for each image. However, many existing datasets rely only on gamma-corrected images, ignoring non-linear FSP operations, all of which significantly affect the perceived shape and visibility of flares.

Model for Flare Removal

On the basis of these datasets, various models have been proposed to solve the flare removal problem. Early studies focused on removing masking flares (Debevec et al. 2004; Seibert, Nalcioglu, and Roeck 1985). These methods, which usually rely on physical priors and deconvolution techniques, work well for smooth artifacts but fail for complex flare patterns.

The advent of deep learning has led to a new wave of methods (Krizhevsky, Sutskever, and Hinton 2012). (Wu et al. 2021) used their semi-synthetic dataset to train the network with U-Net (Ronneberger, Fischer, and Brox 2015) as the backbone. However, its reliance on traditional thresholding makes the model susceptible to unsaturated light sources. For dazzle diversity and model generalization, the latest methods propose end-to-end frameworks. These methods achieve better light source protection and flare suppression (Dai et al. 2023).

Recent methods integrate stronger architectural priors and perceptual constraints (Qiao, Hancke, and Lau 2021; Matta et al. 2024; Zhang et al. 2025; Zhou et al. 2025; Wu et al. 2024b). For example, LPFSformer (Chen et al. 2024) introduced a location prior to guide learning, improving positional dependency modeling. FF-Former (Zhang et al. 2023) combined Swin Transformer and Fourier processing for better global and high-frequency flare removal. MFDNet (Jiang et al. 2024) reduced computational cost by multi-band processing instead of direct image correction, while SparseUFormer (Wu et al. 2024a) used a sparse transformer to cap-

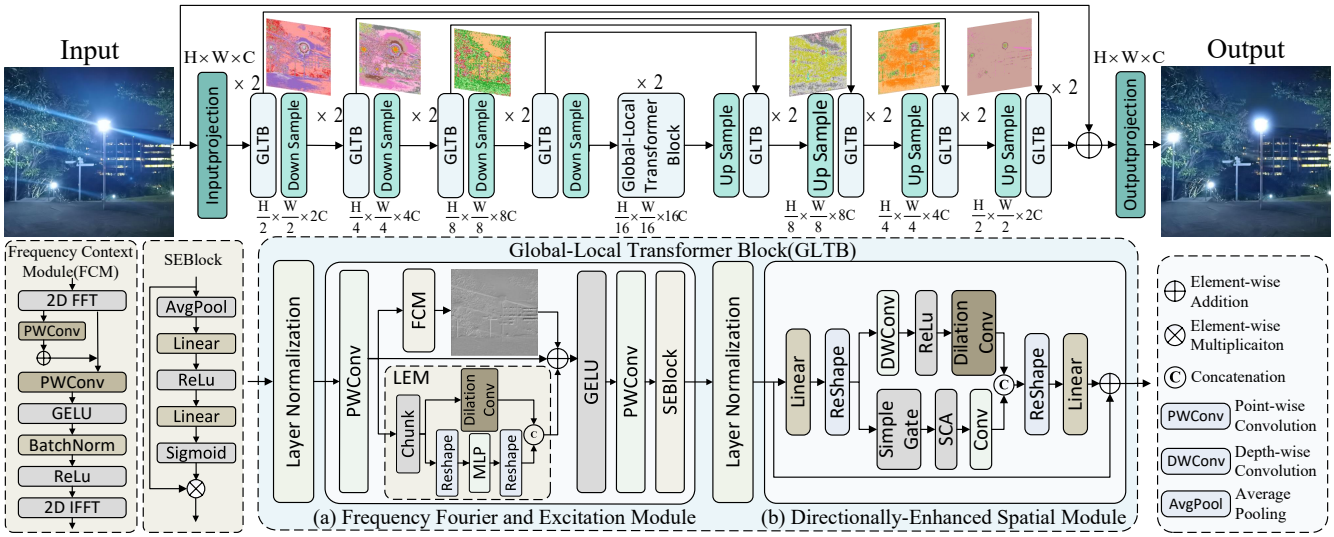

Figure 3: Overview of the proposed network architecture. The network adopts a U-shaped encoder-decoder structure, where each stage consists of Global-Local Transformer Blocks (GLTBs). (a) The detailed architecture of Frequency Fourier and Excitation Module FFEM, which integrates frequency representations with channel-wise excitation for efficient global modeling. (b) Directionally-Enhanced Spatial Module (DESM) enhances the feed-forward network by integrating directional and spatial structural information through convolutional design.

ture multi-scale features and focus on flare-relevant regions. However, there are still difficulties in generating more realistic data and handling complex flares, and we improve the synthesis method and design a stronger network module to enhance the flare removal effect.

Dataset

Atmospheric Turbulence Modeling

Atmospheric turbulence modeling is fundamental for simulating realistic optical distortions in computer vision and computational photography. Specifically, the turbulence-induced phase distortion $\phi(x, y)$ is modeled as a stochastic process with spatial frequency characteristics described by the refractive index structure constant C_n^2, leading to random perturbations in the optical wavefront.

Recent simulation frameworks adopt Zernike polynomial decomposition (Mao, Chimitt, and Chan 2021; Chimitt et al. 2022) to efficiently approximate these distortions. The resulting point spread function (PSF) is then computed by taking the squared magnitude of the Fourier transform of the complex exponential of the phase. This approach effectively captures diffraction and spatially varying blur characteristics without requiring expensive full-wave simulations.

ZernikeVAE-Based Scatter Flare Generation

Although the Scattering flare in Flare7K dataset enables efficient flare synthesis for supervised training, it lacks physical realism because real-world flares often exhibit spatially varying distortions caused by lens aberrations, diffraction, and micro-scale scattering effects. Consequently, conventional synthetic flares tend to appear overly sharp, lacking the natural smoothness gradients and localized blur complexity of real nighttime scatter flares.

A widely used point spread function (PSF) $h(\mathbf{x})$ for flare synthesis pipeline is defined with the pupil function at output plane of aperture:

$$h(\mathbf{x}) = |\mathcal{F}\{P(\mathbf{x})\}|^2 \tag{1}$$

$$P(\mathbf{x}) = A(\mathbf{x}) \exp(j\phi(\mathbf{x})) \tag{2}$$

where $\mathcal{F}$ denotes the Fourier transform, and $A(\mathbf{x})$ is the aperture function. Since PSFs adopted in Flare 7K **cite** is mainly used for stimulation of single light resources without interactions, the four types of aperture function $A(\mathbf{x})$ are approximated with PSFs as the sum of their Fourier transformations, which is observed with four separate symmetry PSF patterns. This assumption is not always held, since flare lies between two light sources is affected and exaggerated with both light sources with possible local glow blurring. To better describe this interacted flare phenomena, we revisit the locally flare degradation image I from flare-free input I_{clear} with a set of spatially varying PSFs:

$$I(\mathbf{x}_i) = \sum_{j=1}^{N} h_{\mathbf{x}_i}(\mathbf{u}_j) I_{clear}(\mathbf{u}_j), \quad i = 1, \ldots, N. \tag{3}$$

$$h_{\mathbf{x}}(\mathbf{u}) = \left| \mathcal{F}\left(A(\mathbf{x}) e^{-j2\pi\phi_{\mathbf{x}}(\boldsymbol{\rho})} \right) \right|^2 \tag{4}$$

Here, the phase function $\phi_{\mathbf{x}}(\boldsymbol{\rho})$ is defined per pixel at each coordinate $\mathbf{x}$. The variable $\boldsymbol{\rho} \in R^2$ is the phase coordinate in the Fourier space.

$$\phi_{\mathbf{x}}(\boldsymbol{\rho}) = \sum_{i=1}^{N} a_i Z_i(\rho) \tag{5}$$

where a_i denotes the number of PSF basis. $Z_i(\rho)$ are the orthogonal Zernike polynomials representing various aberration modes, including tilt, defocus, astigmatism, coma, and higher-order distortions. To simplify the equation (5) in numerical formation, the PSFs in a spatial decomposition are introduced in cite:

$$h_{\mathbf{x}}(\mathbf{u}) = \sum_{i=1}^{N} \beta_i \boldsymbol{\varphi}_i(\mathbf{u}) \tag{6}$$

where $\boldsymbol{\varphi}_i$ spatial basis functions for the PSFs with locally pixel-based coefficients β_i. This equation enables the local blurring and uneven flare described in the degraded images.

Unlike traditional scalar PSF models, our simulator constructs a multi-dimensional representation, embedding each sampled PSF into a high-dimensional learned kernel space using precomputed basis dictionaries derived from empirical optical data.

After simulation, our pipeline adopts an encoder-decoder architecture to refine the generated flare. The encoder processes the concatenated Zernike coefficients and kernel size map to estimate the latent distribution's mean μ and log-variance $\log \sigma^2$. A latent code z is then sampled via the reparameterization trick to model stochastic optical distortions. The decoder reconstructs the final ZernikeVAE flare, improving SLCFormer's generalization for nighttime flare removal.

Flare Removal Framework

Overall Architecture

As shown in Fig. 3, our proposed network follows a hierarchical U-shaped encoder-decoder architecture, designed to effectively remove lens flare while preserving structural integrity. Given a corrupted-flare image $I_{\text{flare}} \in R^{H\times W\times 3}$, our goal is to reconstruct a clean image $I_{\text{deflare}} \in R^{H\times W\times 3}$. To achieve this, we first apply a convolutional feature extractor to obtain an initial feature representation $F_s \in R^{H\times W\times C}$, which is fed into the encoder path for hierarchical processing.

Each stage of the encoder and decoder is composed of a Global-Local Transformer Block (GLTB), which integrates two key modules: the Frequency Fourier and Excitation Module (FFEM) and the Directionally-Enhanced Spatial Module (DESM). Specifically, FFEM is designed to enhance global context modeling by leveraging spatial and frequency representations. In contrast, DESM complements this by focusing on local structural enhancement and spatially adaptive attention.

To preserve multiscale information and ensure spatial consistency, we adopt pixel unshuffle and pixel shuffle for downsampling and upsampling. Moreover, long-range skip connections are employed to bridge corresponding encoder and decoder stages.

Finally, the decoder reconstructs the output, and a residual connection between the input and output is applied before the final sigmoid activation to produce the deflared image.

Frequency Fourier and Excitation Module(FFEM)

FFEM is designed to efficiently capture global contextual representations through frequency-domain analysis, complemented by local enhancement to preserve structural precision. The proposed module comprises three core components: a Local Enhancement Module (LEM), a Frequency Context Module (FCM), and a Squeeze-and-Excitation (SE) Block, which the architecture is shown in Fig. 3(a).

We first apply a point-wise convolution to the input feature $X \in R^{H\times W\times C}$, generating an intermediate representation F, which is then split into two parts along the channel dimension:

$$F_{\text{local}}, F_{\text{global}} = \text{Split}(\text{PWConv}(X)) \tag{7}$$

the two branches are then processed separately: F_{local} is fed into LEM to extract spatially localized features, while F_{global} is passed to FCM to capture long-range dependencies in the frequency domain. The processed features are then concatenated and passed through a GELU activation and another pointwise convolution to fuse them:

$$X = \text{PWConv}\Big(\text{GELU}\big(\text{LEM}(F_{\text{local}}) \| \text{FCM}(F_{\text{global}})\big)\Big) \tag{8}$$

finally, the features X are further refined by a Squeeze-and-Excitation (SE) Block, resulting in the final output. Next, we elaborate on the internal structure and functionality of the three core components: FCM, LEM and the SEBlock.

Frequency Context Module (FCM). The Fourier Transform provides an efficient way to capture global structures by converting spatial features into frequency components. The Fast Fourier Transform (FFT) enables holistic frequency-domain analysis with significantly lower computational cost ($\mathcal{O}(N \log N)$) than spatial-domain methods such as multi-head self-attention (MSA)($\mathcal{O}(N^2)$). MSA suffers from redundant computations in modeling global relationships. In contrast, FFT compactly represents global structures through orthogonal frequency bases, avoiding redundancy while maintaining scalability. Thus, we adopt FFT as an efficient substitute for MSA.

The FCM first performs the 2D Fast Fourier Transform to project the input $X \in R^{H\times W\times C}$ into the frequency domain. After frequency transformation, the real and imaginary parts are concatenated along the channel dimension and processed by two lightweight 1×1 convolutions and a non-linear activation. The final frequency representation is then projected back to the spatial domain using the inverse FFT. The entire transformation process can be summarized as follows:

$$F = \mathcal{F}(X) \tag{9}$$

$$F = \text{Conv}_{1\times 1}(F) + F \tag{10}$$

$$F = \text{GELU}(\text{Conv}_{1\times 1}(F)) \tag{11}$$

$$X = X + \mathcal{F}^{-1}\big(\text{BN}(\text{ReLU}(F))\big) \tag{12}$$

where $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$ denote a fast Fourier transform (FFT) and an inverse transform (IFFT), respectively; $\text{Conv}_{1\times 1}(\cdot)$ denotes a 1×1 convolution operation in the frequency domain; GELU and ReLU are nonlinear activation functions; and BN denotes a batch normalization operation.

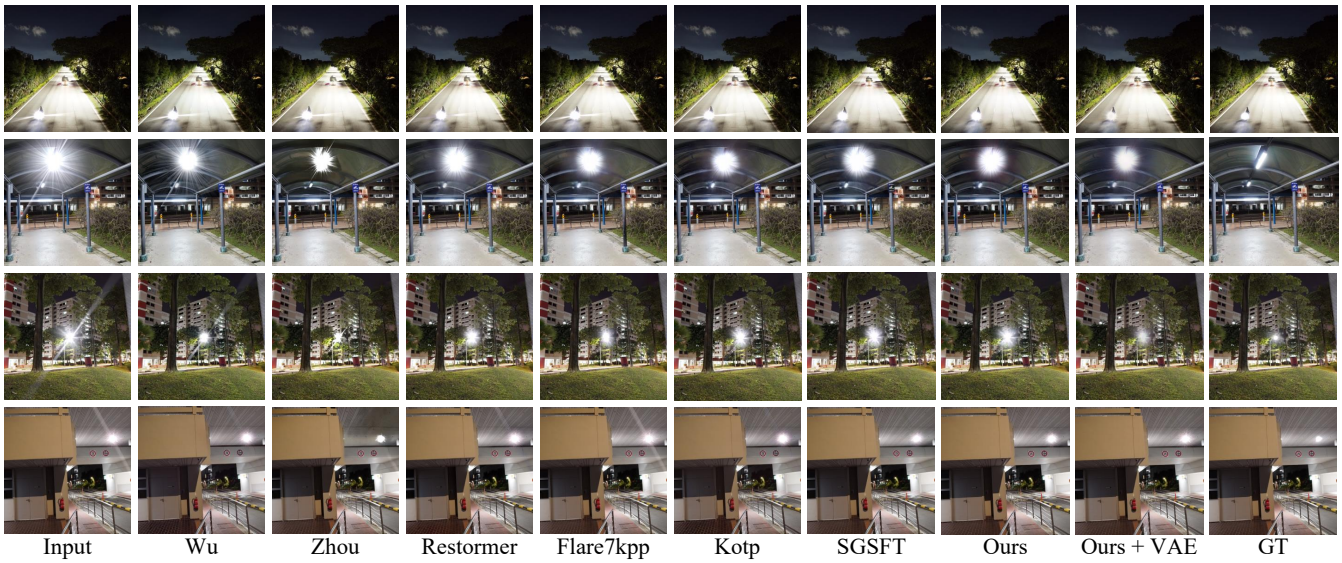

Figure 4: Visual comparison of flare removal on synthetic nighttime flare images.

In addition, the final output is added back to the original input through residual concatenation X to enhance global modeling while preserving structure information.

Local Enhancement Module (LEM). While FFT excels at capturing long-range dependencies, its global receptive field tends to overlook certain local details. To address this limitation, we introduce LEM to compensate for local feature extraction. LEM operates on the input feature $F_{\text{local}} \in R^{B\times C\times H\times W}$, and first splits it equally into two parts: $F_1 \in R^{B\times \frac{C}{2}\times H\times W}$ and $F_2 \in R^{B\times \frac{C}{2}\times H\times W}$. The first half is passed through a spatially independent token-wise MLP. F_2 is processed using a dilated convolution to capture spatial dependencies. The two features are then concatenated along the channel axis to form the locally enhanced output.

Squeeze-and-Excitation (SE) Block. To improve network representation, we integrate a Squeeze-and-Excitation (SE) (Hu, Shen, and Sun 2018) block after the module.

Given an input feature $F \in R^{B\times H\times W\times C}$, the SE block first applies a global average pooling to generate a channel descriptor. This descriptor passes through two fully connected layers with ReLU and sigmoid activations, obtaining channel weights $S \in R^{B\times C}$, which are then multiplied F to produce the refined output. The entire process can be summarized as follows:

$$F = F \odot \sigma(w_2(\delta(w_1(\text{AvgPool}(F))))) \tag{13}$$

here, δ and σ are ReLU and sigmoid, and $\odot$ denotes channel-wise multiplication.

Directionally-Enhanced Spatial Module (DESM)

Traditional FFNs often lack the ability to capture spatial structure and directional cues, which are essential for accurately restoring flare-degraded regions. DESM aims to introduce spatially-aware interactions within the FFN by injecting directional and local structural information through convolutional design.

As shown in Fig. 3(b), input tokens are projected to a higher-dimensional space and reshaped into 2D feature maps, then split into two branches. One applies depth-wise and dilated convolutions (X_{dir}) to encode spatial and directional features, while the other uses a SimpleGate with channel attention (X_{gate}) to enhance structural representations. The outputs are concatenated and linearly projected:

$$X = \text{Linear}(\text{Concat}(X_{\text{dir}}, X_{\text{gate}})) \tag{14}$$

Loss Function

In order to efficiently supervise the flare removal, we employ an integrated loss function with the following expression:

$$L = \lambda_1 L_1 + \lambda_2 L_{\text{vgg}} + \lambda_3 L_{\text{rec}} + \lambda_4 L_{\text{hf}} \tag{15}$$

where L_1, L_{vgg}, L_{rec} denote the standard $L1$ loss, the perceptual loss (Simonyan and Zisserman 2014), and the reconstruction loss (Dai et al. 2023), respectively.

To enhance the recovery of high-frequency details, we further introduce a high-frequency loss L_{hf} combining Laplacian and Sobel gradients to capture fine flare structures:

$$L_{\text{hf}} = \tfrac{1}{2}\|\nabla_{\text{lap}}(\hat{I}) - \nabla_{\text{lap}}(I)\| + \tfrac{1}{2}\|\nabla_{\text{sob}}(\hat{I}) - \nabla_{\text{sob}}(I)\| \tag{16}$$

where $\hat{I}$ and I are the predicted and ground-truth images, and $\nabla_{\text{lap}}, \nabla_{\text{sob}}$ denote Laplacian and Sobel operators. The weighting coefficients, λ_1, λ_2, λ_3, λ_4 are empirically set to 0.5, 0.5, 1.0, and 1.0, respectively.

Experiments

Datasets

To train our flare removal model, we construct a supervised training set based on the Flare7K++ dataset (Dai et al. 2023),

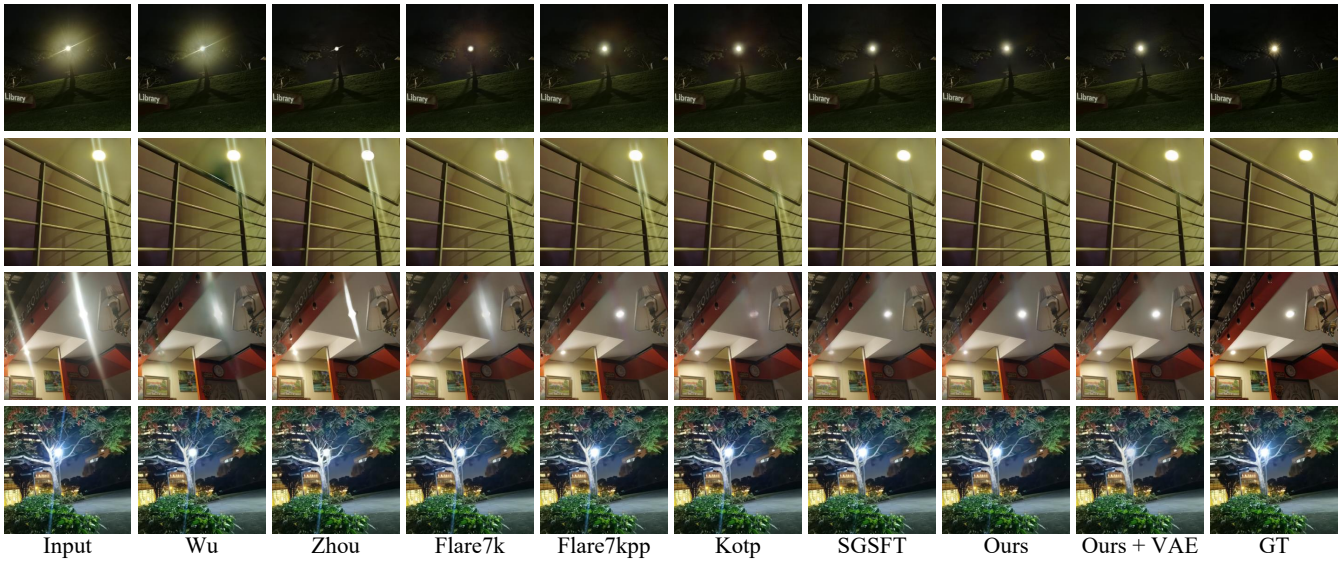

Figure 5: Visual comparison of flare removal on real-world nighttime flare images.

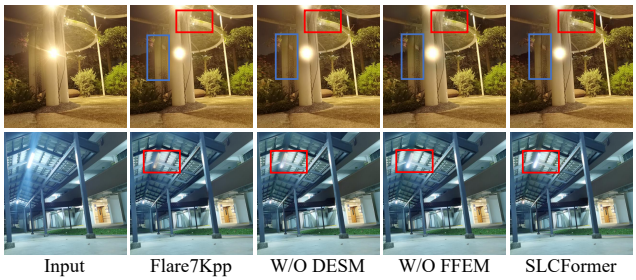

Figure 6: Ablation studies on the proposed method.

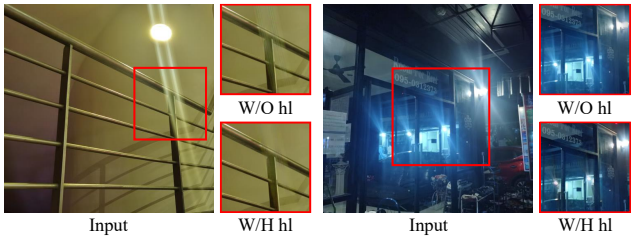

Figure 7: Ablation studies on the proposed loss function.

excluding 5,000 scattering flares from Flare7K and 964 from Flare-R. Background images are sampled from Flickr-24K (Zhang, Ng, and Chen 2018). To ensure that the network learns to restore both the underlying scene and the light source, we use the light source.

To enhance realism and diversity, we apply a series of photometric and geometric augmentations to flare and background images during synthesis. These operations improve the robustness of the model by simulating a wide range of flare appearances. Detailed augmentation parameters are provided in the supplementary material.

Implementation Details

We implement our model using the PyTorch framework and train it on an NVIDIA RTX 3090 GPU. During training, both flare-corrupted and flare-free input images are cropped to a fixed resolution of 512×512. We use a batch size of 2 and adopt the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.99$. The initial learning rate is set to 1e-4 and is scheduled using a MultiStepLR strategy, with the learning rate reduced by a factor of 0.5 after 200K iterations. The total number of training iterations is 400K.

Evaluation Metrics

To comprehensively evaluate the performance of our model, we adopt several restoration metrics, including PSNR, SSIM(Wang et al. 2004), and LPIPS(Zhang et al. 2018). In addition, we incorporate two specialized metrics proposed by (Dai et al. 2023), namely S-PSNR and G-PSNR, which are designed to assess the effectiveness of flare removal in localized streak and glare regions, respectively.

Comparison with Previous Methods

To validate the advantages of our proposed SLCFormer, we conduct both qualitative and quantitative evaluations against a range of flare removal and image restoration methods on both the Flare7K real and synthetic test dataset, including U-Net (Ronneberger, Fischer, and Brox 2015), HINet (Chen et al. 2021), Restormer (Zamir et al. 2022), Uformer (Wang et al. 2022), etc.

Qualitative Evaluation. We compare visual results on both real-world nighttime flare and synthetic flare removal in Fig. 4 and Fig. 5. The recovered images produced by SLCFormer exhibit noticeably cleaner flare removal. Moreover, SLCFormer trained with VAE-augmented scatter flares produces more realistic outputs.

Test	Real_Images					Synthetic_Images				
Metrics	PSNR↑	SSIM↑	LPIPS↓	G-PSNR↑	S-PSNR↑	PSNR↑	SSIM↑	LPIPS↓	G-PSNR↑	S-PSNR↑
Wu (Wu et al. 2021)	24.613	0.871	0.0598	21.772	16.728	28.260	0.954	0.0331	24.757	21.108
Zhou (Zhou et al. 2023)	25.149	0.883	0.0576	22.053	17.865	28.779	0.939	0.0286	23.779	22.237
U-Net (Ronneberger et al. 2015)	27.189	0.894	0.0452	23.527	22.647	29.389	0.933	0.0276	24.271	23.338
HINet (Chen et al. 2021)	27.548	0.892	0.0464	*24.081*	*22.907*	29.462	0.938	0.0248	24.454	23.534
Restormer* (Zamir et al. 2022)	27.597	*0.897*	0.0447	23.828	22.452	29.489	0.958	0.0244	24.654	23.078
Flare7K (Dai et al. 2022)	26.978	0.890	0.0466	23.507	21.563	-	-	-	-	-
Flare7Kpp (Dai et al. 2023)	27.633	0.894	0.0428	23.949	22.603	29.498	*0.962*	0.0210	24.685	24.155
Flare-Free (Kotp and Torki 2024)	*27.662*	*0.897*	*0.0422*	*23.987*	*22.847*	*29.573*	0.961	*0.0205*	**24.879**	*24.458*
SGSFT (Ma et al. 2025)	<u>28.077</u>	<u>0.904</u>	<u>0.0416</u>	<u>24.477</u>	**23.305**	<u>29.576</u>	<u>0.966</u>	<u>0.0200</u>	*24.745*	**24.914**
SLCFormer (ours)	**28.092**	**0.905**	**0.0400**	**24.497**	<u>23.287</u>	**29.798**	**0.968**	**0.0195**	<u>24.792</u>	<u>24.578</u>

Table 1: Quantitative comparison on both the real and synthetic tests from Flare7K++. Bold values indicate the best performance among all methods, while underlined values indicate the second-best and italicised values indicate the third-best. Models marked with "*" have reduced parameters due to GPU memory limitations. Note that since all models are trained on Flare7K++, and the architecture of Flare7K is exactly the same as Uformer but trained on Flare7K instead of Flare7K++, the results of Flare7K on the synthetic test set are omitted.

Method	PSNR↑	SSIM↑	LPIPS↓	G-PSNR↑	S-PSNR↑
W/O Both	27.633	0.894	0.0428	23.949	22.603
W/O FFEM	27.888	0.902	0.0408	24.473	23.137
W/O DESM	27.804	0.899	0.0430	24.372	22.926
Full Model	**28.092**	**0.905**	**0.0400**	**24.497**	**23.287**

Table 2: Ablation studies on the proposed method.

Method	PSNR↑	SSIM↑	LPIPS↓	G-PSNR↑	S-PSNR↑
W/O L_{hf}	27.824	0.898	0.0431	24.302	22.986
W/H L_{hf}	**28.092**	**0.905**	**0.0400**	**24.497**	**23.287**

Table 3: Ablation study on the proposed loss function.

Quantitative Evaluation. Table 1 summarizes the full-reference metrics on the Flare7K real and synthetic tests. SLCFormer achieves a PSNR of 28.092 dB in the real test images. In SSIM, our model reaches 0.905. For perceptual fidelity, SLCFormer attains the lowest LPIPS score of 0.0400, indicating closer alignment with ground-truth perceptual quality.

Ablation Study

Network Structure. We further illustrate qualitative comparisons in Figure 6. In the first group of images, results produced by the full model appear cleaner and less hazy, while other configurations show varying degrees of turbidity. Although w/o FFEM generates relatively clearer results compared to other ablated models, it still fails to match the clarity of the full model. Additionally, the shimmer in the scene is effectively removed only by the full model and w/o DESM, whereas other models fail to eliminate it.

Additionally, only the full model preserves the light source region completely intact in the second group. Other models remove parts of the light source, leading to unnatural results. These results demonstrate that both modules are essential for flare removal.

Loss Function. We further evaluate the impact of the proposed high-frequency loss by conducting ablation experiments. As shown in Table 3, incorporating the high-frequency loss leads to consistent improvements across all evaluation metrics. Qualitative results in Figure 7 show that incorporating the high-frequency loss yields sharper details and clearer structures, with shimmer flares more effectively suppressed and overall flare intensity visibly reduced.

Conclusion

In this paper, we propose a ZernikeVAE-based scatter flare generation pipeline that synthesizes physically realistic flare patterns by modeling spatially varying point spread functions (PSFs) with interpretable Zernike polynomials. This approach enriches the diversity and complexity of flare distributions, thereby improving data realism and enhancing the generalization capability of learning-based flare removal models. Based on this, we introduced SLCFormer, integrating the Frequency Fourier and Excitation Module (FFEM) and Directionally-Enhanced Spatial Module (DESM) to jointly model global frequency context and local structural features for effective nighttime flare removal. Experiments demonstrate that our model outperforms most current methods, though it introduces additional computational overhead and the ZernikeVAE-based synthesis may not fully capture real flare complexity. Future work will explore lightweight architectures and more realistic flare synthesis to extend applicability to broader optical degradation restoration tasks.

Acknowledgments

This work was supported financially by the Natural Science Foundation of China (62202347).

References

Bojarski, M.; Del Testa, D.; Dworakowski, D.; Firner, B.; Flepp, B.; Goyal, P.; Jackel, L. D.; Monfort, M.; Muller, U.; Zhang, J.; et al. 2016. End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*.

Chen, G.-Y.; Dong, W.; Fan, G.; Su, J.-N.; Gan, M.; and Chen, C. P. 2024. LPFSformer: Location Prior Guided Frequency and Spatial Interactive Learning for Nighttime Flare Removal. *IEEE Transactions on Circuits and Systems for Video Technology*.

Chen, L.; Lu, X.; Zhang, J.; Chu, X.; and Chen, C. 2021. Hinet: Half instance normalization network for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 182–192.

Chimitt, N.; Zhang, X.; Mao, Z.; and Chan, S. H. 2022. Real-time dense field phase-to-space simulation of imaging through atmospheric turbulence. *IEEE Transactions on Computational Imaging*, 8: 1159–1169.

Dai, Y.; Li, C.; Zhou, S.; Feng, R.; and Loy, C. C. 2022. Flare7K: A Phenomenological Nighttime Flare Removal Dataset.

Dai, Y.; Li, C.; Zhou, S.; Feng, R.; and Loy, C. C. 2023. Flare7K++: Mixing Synthetic and Real Datasets for Nighttime Flare Removal and Beyond.

Debevec, P.; Reinhard, E.; Ward, G.; and Pattanaik, S. 2004. High dynamic range imaging. In *ACM SIGGRAPH 2004 Course Notes*, 14–es.

Deng, H.; Li, L.; Zhang, F.; Li, Z.; Xu, B.; Lu, Q.; Gao, C.; and Sang, N. 2024. Towards Blind Flare Removal Using Knowledge-driven Flare-level Estimator. *IEEE Transactions on Image Processing*.

Ernst, M.; Akenine-Möller, T.; and Jensen, H. W. 2005. Interactive rendering of caustics using interpolated warped volumes. In *Proceedings of Graphics Interface 2005*, 87–96.

Girshick, R.; Donahue, J.; Darrell, T.; and Malik, J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 580–587.

He, Y.; Wang, W.; Wu, W.; and Jiang, K. 2025. Disentangle nighttime lens flares: self-supervised generation-based lens flare removal. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 3464–3472.

Hu, J.; Shen, L.; and Sun, G. 2018. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7132–7141.

Hullin, M.; Eisemann, E.; Seidel, H.-P.; and Lee, S. 2011. Physically-based real-time lens flare rendering. In *ACM SIGGRAPH 2011 papers*, 1–10.

Jiang, Y.; Chen, X.; Pun, C.-M.; Wang, S.; and Feng, W. 2024. Mfdnet: Multi-frequency deflare network for efficient nighttime flare removal. *The Visual Computer*, 1–14.

Jin, Z.; Chen, S.; Feng, H.; Xu, Z.; and Chen, Y. 2023. Toward real flare removal: A comprehensive pipeline and a new benchmark. *arXiv preprint arXiv:2306.15884*.

Kotp, Y.; and Torki, M. 2024. Flare-free vision: Empowering uformer with depth insights. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2565–2569. IEEE.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

Li, Z.; Zheng, C.; Chen, B.; and Wu, S. 2025. Neural-Augmented HDR Imaging via Two Aligned Large-Exposure-Ratio Images. *IEEE Transactions on Instrumentation and Measurement*.

Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022.

Ma, T.; Kai, Z.; Miao, X.; Liang, J.; Peng, J.; Wang, Y.; Wang, H.; and Liu, X. 2025. Self-Prior Guided Spatial and Fourier Transformer for Nighttime Flare Removal. *IEEE Transactions on Automation Science and Engineering*.

Mao, Z.; Chimitt, N.; and Chan, S. H. 2021. Accelerating atmospheric turbulence simulation via learned phase-to-space transform. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 14759–14768.

Matta, G. R.; Siddartha, R.; Girish, R. S. V.; Sharma, S.; and Mitra, K. 2024. GN-FR: Generalizable Neural Radiance Fields for Flare Removal. *arXiv preprint arXiv:2412.08200*.

Qiao, X.; Hancke, G. P.; and Lau, R. W. 2021. Light source guided single-image flare removal from unpaired data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4177–4185.

Qu, L.; Zhou, S.; Pan, J.; Shi, J.; Chen, D.; and Yang, J. 2024. Harmonizing Light and Darkness: A Symphony of Prior-guided Data Synthesis and Adaptive Focus for Nighttime Flare Removal. *arXiv preprint arXiv:2404.00313*.

Raut, H. K.; Ganesh, V. A.; Nair, A. S.; and Ramakrishna, S. 2011. Anti-reflective coatings: A critical, in-depth review. *Energy & Environmental Science*, 4(10): 3779–3804.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, 234–241. Springer.

Seibert, J. A.; Nalcioglu, O.; and Roeck, W. 1985. Removal of image intensifier veiling glare by mathematical deconvolution techniques. *Medical physics*, 12(3): 281–288.

Shao, S.; Pei, Z.; Chen, W.; Chen, P. C.; and Li, Z. 2025. Iebins: Iterative elastic bins for monocular depth estimation and completion. *International Journal of Computer Vision*, 133(5): 2463–2486.

Simonyan, K.; and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Vitoria, P.; and Ballester, C. 2019. Automatic flare spot artifact detection and removal in photographs. *Journal of Mathematical Imaging and Vision*, 61(4): 515–533.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.

Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; and Li, H. 2022. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 17683–17693.

Wu, S.; Liu, F.; Bai, Y.; Han, H.; Wang, J.; and Zhang, N. 2024a. Flare Removal Model Based on Sparse-UFormer Networks. *Entropy*, 26(8): 627.

Wu, W.; Wang, W.; Wang, Z.; Jiang, K.; and Li, Z. 2024b. For Overall Nighttime Visibility: Integrate Irregular Glow Removal With Glow-Aware Enhancement. *IEEE Transactions on Circuits and Systems for Video Technology*.

Wu, Y.; He, Q.; Xue, T.; Garg, R.; Chen, J.; Veeraraghavan, A.; and Barron, J. T. 2021. How to train neural networks for flare removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2239–2247.

Xie, S.; and Tu, Z. 2015. Holistically-nested edge detection. In *Proceedings of the IEEE international conference on computer vision*, 1395–1403.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M.-H. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5728–5739.

Zhang, D.; Ouyang, J.; Liu, G.; Wang, X.; Kong, X.; and Jin, Z. 2023. Ff-former: Swin fourier transformer for nighttime flare removal. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2824–2832.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 586–595.

Zhang, W.; Shang, W.; Ren, D.; and Zuo, W. 2025. Flare-Aware RWKV for Flare Removal. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.

Zhang, X.; Ng, R.; and Chen, Q. 2018. Single image reflection separation with perceptual losses. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4786–4794.

Zhou, Y.; Liang, D.; Chen, S.; and Huang, S.-J. 2025. Image Lens Flare Removal Using Adversarial Curve Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zhou, Y.; Liang, D.; Chen, S.; Huang, S.-J.; Yang, S.; and Li, C. 2023. Improving lens flare removal with general-purpose pipeline and multiple light sources recovery. In *Proceedings of the IEEE/CVF international conference on computer vision*, 12969–12979.