Nighttime Flare Removal via Wavelet-Guided and Gated-Enhanced Spatial-Frequency Fusion Network

Yun Liu[1], Guang Yang[1]*, Tao Li[1], Weisi Lin[2]

[1]College of Artificial Intelligence, Southwest University
[2]College of Computing and Data Science, Nanyang Technological University
yunliu@swu.edu.cn, learnforai@gmail.com, lt3088919588@email.swu.edu.cn, wslin@ntu.edu.sg

Abstract

Nighttime flares, caused by complex scattering and reflections from artificial light sources, significantly degrade image quality and hinder downstream visual tasks. Existing deflare networks usually struggle to jointly capture and fuse latent spatial and frequency features. In this paper, we propose a novel Wavelet-guided and Gated-enhanced Spatial-frequency Fusion Network (WGSF-Net) for nighttime flare removal. WGSF-Net is primarily composed of two key modules: Wavelet-guided Fusion Block (WFB) and Local-Global Block (LGB). Specifically, WFB integrates a Multi-level Wavelet Enhancement Block (MWEB) and a Spatial-Frequency Fusion Network (SFFN) to effectively extract hierarchical spatial and frequency features through a coarse-to-fine strategy based on multi-level wavelet decomposition. To better suppress flare artifacts, LGB is designed to jointly capture local and global information: a Gated-Enhanced Attention Block (GEAB) selectively amplifies critical local features through a gated network and a difference network, and the subsequent SFFN performs global spatial-frequency fusion via depthwise separable convolution and partial Fourier convolution. This design enables LGB to effectively disentangle flare-corrupted regions and restore fine-grained details, making it particularly suited for challenging real-world flare scenarios. Extensive experiments on both synthetic and real datasets show that WGSF-Net achieves state-of-the-art performance in nighttime flare removal, outperforming existing methods across five evaluation metrics.

Code — https://github.com/gyang666/WGSF-Net

Introduction

Nighttime imaging often suffers from lens flare caused by light scattering and internal reflections from intense artificial light sources such as vehicle headlights and street lamps. These flare artifacts typically appear as streaks, halos, or glows that obscure critical scene details, severely degrading perceptual clarity and impairing the performance of downstream vision applications, such as detection, recognition, navigation, and visual tracking tasks. Therefore, effective flare removal is essential for improving perceptual quality and enhancing the overall reliability of vision-based systems in nighttime conditions.

---

*Corresponding author

Figure 1: Flare removal results on a real-world nighttime flare-corrupted image.

Early physics-based methods leverage specially designed optical components, such as anti-reflective coatings (Chen et al. 2010) and lens hoods (Ramesh 2008) to suppress the flares. However, these methods may degrade image quality and exhibit inherent limitations in flare mitigation. On the other hand, computation-based approaches typically employ Bayesian optimization (Wu and Tang 2005), layer decomposition (Zhang et al. 2018c), and deconvolution (Faulkner, Kotre, and Louka 1989) to remove flares, whereas their effectiveness is usually constrained to a specific type of flare and may incorrectly identify artificial light sources as flares.

Recent deep learning-based approaches have achieved promising results for flare removal. Flare7K (Dai et al. 2022) and Flare7K++ (Dai et al. 2024) datasets are constructed for flare removal, and several classic network architectures, including U-Net (Ronneberger, Fischer, and Brox 2015), HINet (Chen et al. 2021), MPRNet (Zamir et al. 2021), Restormer (Zamir et al. 2022), and Uformer (Wang et al. 2022), are trained on them as baselines. (Dai et al. 2023) constructs the BracketFlare dataset and proposes an optical center symmetry prior to remove the reflective flares. (Kotp and Torki 2024) introduces a two-stage flare removal framework consisting of a depth map estimation Transformer and a general Uformer. FBNet (Lian et al. 2025) leverages the Jetmap mask-guided cross-attention mechanism and subspace projection to improve flare localization and separation. (Zhou et al. 2025) proposes an ISP-guided flare removal framework that models automatic exposure and tone mapping to synthesize more realistic training data, integrates ad-

versarial curve learning for cross-device generalization, and introduces a convex illumination-based strategy for accurate multi-light source recovery. SGLFR-Net (He et al. 2025) designs a generation-based lens flare removal network trained in a self-supervised manner. Unfortunately, these methods overlook frequency-domain cues, which are critical for capturing long-range dependencies and periodic flare patterns, often resulting in residual artifacts or incomplete flare suppression, as shown in Fig. 1.

Recently, several learning-based networks have been proposed to leverage frequency-domain features for flare removal. FF-Former (Zhang et al. 2023) employs Fast Fourier Convolution (FFC) (Chi, Jiang, and Mu 2020) to construct a U-shaped network architecture for capturing global frequency features. SFSNiD (Cong et al. 2024) integrates spatial-frequency interaction with local brightness constraints to suppress nighttime glows. SGSFT (Ma et al. 2025) constructs a self-prior guided spatial and Fourier Transformer to infer flare-free regions. LPFSformer (Chen et al. 2025) designs a location prior guidance module to focus on flare-corrupted regions and distinguish flare features through spatial and frequency interacting learning. Although recent flare removal methods use frequency-domain features, they still fail to fully exploit hierarchical structures and effectively emphasize flare-related components.

To address these issues, we propose WGSF-Net, a novel Wavelet-guided and Gated-enhanced Spatial-frequency Fusion Network for flare removal, as depicted in Fig. 2. WGSF-Net effectively fuses multi-scale high- and low-frequency features that are highly relevant to flare patterns, primarily owing to two key components: Wavelet-guided Fusion Block (WFB) and Local-Global Block (LGB). More concretely, WFB consists of two modules: a Multi-level Wavelet Enhancement Block (MWEB) and a Spatial-Frequency Fusion Network (SFFN). MWEB employs multi-level wavelet decomposition and residual learning to progressively refine low- and high-frequency features, capturing long-range dependencies in low-frequency components while preserving fine-grained textures in high-frequency ones. To further alleviate flare artifacts, LGB integrates a Gated-Enhanced Attention Block (GEAB) and an SFFN. GEAB employs a gating mechanism to selectively amplify flare-related patterns by controlling the network's attention toward flare regions. SFFN, utilized in both WFB and LGB, employs depthwise separable convolution and partial Fourier convolution to efficiently integrate spatial information with global frequency features while reducing redundancy. The main contributions are summarized as follows:

- We propose WGSF-Net, a novel flare removal framework that integrates wavelet-guided multi-scale feature extraction and gated-enhanced spatial-frequency fusion, enabling hierarchical cross-domain representation and effective localization of flare-corrupted regions.

- We propose three key modules: MWEB for progressive multi-level frequency refinement, GEAB for gated attention on flare regions, and SFFN for efficient spatial-frequency fusion. Embedded within WFB and LGB, these modules enable WGSF-Net to effectively extract

multi-scale features and suppress flare artifacts.

- Experiments verify that WGSF-Net outperforms state-of-the-art methods on both real and synthetic datasets, achieving a 7.93% reduction in LPIPS on real data and a 2.80% improvement in S-PSNR on synthetic data.

Related Work

Traditional Flare Removal Methods

Lens flares are common optical artifacts around strong light sources that arise from internal refraction and scattering within lens elements during nighttime photography. Existing hardware-based methods employ physical optical principles to mitigate flare artifacts. Anti-reflective coatings (Chen et al. 2010) and lens hoods (Ramesh 2008) are commonly used physical solutions to suppress surface light reflections and block extreme-angle incident light, thereby reducing lens flare. On the other hand, computation-based methods adopt post-processing strategies to remove flares. For instance, a Bayesian optimization framework (Wu and Tang 2005) has been proposed for flare removal by treating it as a shadow removal problem. Deconvolution-based techniques (Faulkner, Kotre, and Louka 1989; Seibert, Nalcioglu, and Roeck 1985) mitigate flare artifacts by modeling optical degradation. A layer decomposition approach (Zhang et al. 2018c) has been developed to separate and eliminate the flare component. These traditional methods are typically limited to specific flare types and struggle to distinguish flares from natural light sources.

Learning-based Flare Removal Methods

Early work (Wu et al. 2021) constructs a semi-synthetic dataset and designs a pix2pix model based on U-Net (Ronneberger, Fischer, and Brox 2015) to remove flares. (Qiao, Hancke, and Lau 2021) takes the relationships between light sources and flare regions into account and develops a CycleGAN-based network (Zhu et al. 2017) for flare removal. However, due to the limited diversity of flare types in the training data, these methods struggle to handle complex and realistic flare patterns.

To increase the diversity of flare datasets, Flare7K (Dai et al. 2022) and Flare7K++ (Dai et al. 2024), which incorporate scattering and reflective flare images, are proposed for flare removal. Based on Flare7K and Flare7K++, Restormer (Zamir et al. 2022) and Uformer (Wang et al. 2022) are commonly used baselines trained on these datasets. Subsequently, FF-Former (Zhang et al. 2023) enhances flare removal by extracting global frequency features using FFC. LPFSformer (Chen et al. 2025) incorporates location priors into a Transformer-based network to improve region-specific flare suppression. SGSFT (Ma et al. 2025) learns the location and intensity of flares through a self-prior extraction module and extracts both spatial contextual and frequency-domain features for effective flare removal. In addition, several nighttime dehazing methods (Liu et al. 2023b,a; Cong et al. 2024; Jin et al. 2023; Liu et al. 2025; Lin et al. 2025) have been proposed to eliminate glow artifacts caused by artificial light sources in nighttime scenes.
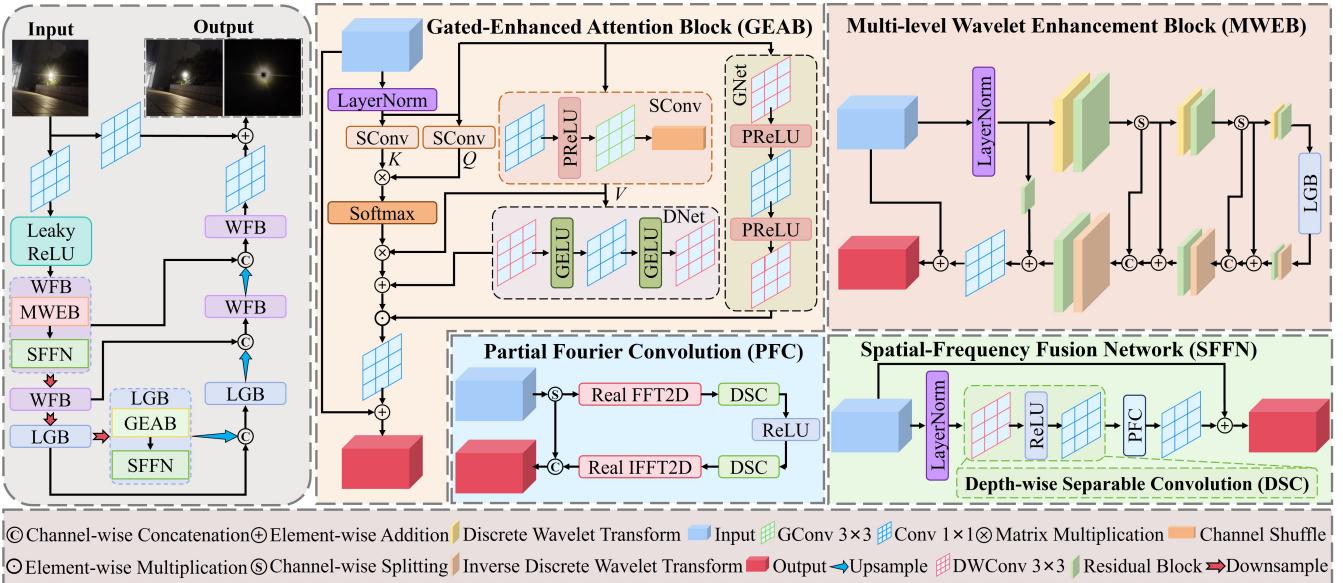
Figure 2: Overall architecture of our WGSF-Net. WGSF-Net is composed of two key modules: Wavelet-guided Fusion Block (**WFB**) and Local-Global Block (**LGB**). WFB incorporates MWEB for progressive frequency feature extraction, combined with SFFN for effective spatial-frequency features fusion. LGB includes GEAB to highlight flare-relevant regions and leverages SFFN to further refine and fuse local-global features for enhanced flare removal. Here, **Downsample/Upsample** denotes a 1×1 convolution layer combined with a PixelUnshuffle/PixelShuffle operation (Shi et al. 2016), respectively.

Existing methods insufficiently exploit frequency-domain information and lack effective region-selective mechanisms, limiting their ability to capture hierarchical features and distinguish flare-corrupted regions from unaffected areas.

Methodology

Framework Overview

As illustrated in Fig. 2, WGSF-Net adopts a U-shaped end-to-end architecture consisting of two key modules: WFB and LGB. WFB comprises MWEB and SFFN to extract hierarchical representations, where MWEB captures low-frequency structures and high-frequency details through progressive wavelet decomposition, and SFFN fuses spatial and frequency information. LGB integrates GEAB with SFFN, in which a gated attention mechanism selectively identifies flare-related regions, and the subsequent spatial–frequency fusion incorporates global frequency cues.

Multi-level Wavelet Enhancement Block (MWEB)

Previous frequency-domain flare removal networks (Zhang et al. 2023; Chen et al. 2025; Ma et al. 2025) typically extract features at a single scale, limiting their ability to capture both coarse and fine-grained frequency cues and thus resulting in incomplete suppression of flare artifacts, especially under complex lighting variations. To address this, we design MWEB to progressively perform multi-level wavelet decomposition, enabling hierarchical extraction of frequency-aware features that more accurately isolate flare patterns across scales and substantially enhance deflare performance.

Given the input feature maps $F_{in} \in \mathbb{R}^{H \times W \times C}$ and a pre-defined wavelet decomposition level $L = 3$, MWEB first performs layer normalization (Ba, Kiros, and Hinton 2016) on F_{in} and then a multi-level Discrete Wavelet Transform (DWT) (Finder et al. 2024) is utilized to progressively decompose the feature maps into four frequency sub-bands at each scale. For each level $i = 1, 2, 3$, the feature map F_{i-1} is separated into one low-frequency part F_i^{LL} and three high-frequency parts F_i^{LH}, F_i^{HL} and F_i^{HH}:

$$\left\{ F_i^{LL}, F_i^{LH}, F_i^{HL}, F_i^{HH} \right\} = \text{DWT}(F_{i-1}),$$
$$\text{where } F_0 = \text{LN}(F_{in}) \tag{1}$$

After each wavelet decomposition, a lightweight residual block (ResBlock), comprising a 3×3 depthwise convolution (DWConv), ReLU, a 1×1 pointwise convolution, ReLU, and a 3×3 DWConv with a residual connection, is employed to enhance feature representations and compensate for potential information loss caused by the downsampling inherent in the wavelet transform. Moreover, to fully leverage the complementary information across different frequency bands, a mixed-subband strategy is adopted for residual feature learning. Specifically, instead of processing each wavelet subband ($LL, LH, HL,$ and HH) independently, each subband is split into four parts, which are then recombined to form four new subbands, with each new subband containing one quarter of the features from each original subband. These fused subbands, enriched with complementary information, are fed into the residual module to facilitate feature learning. Subsequently, only the low-frequency component F_i^{LL} is propagated to the next decomposition level, which facilitates a coarse-to-fine structure extraction while suppressing

high-frequency flare artifacts in a progressive manner.

Following the final decomposition level ($i = 3$), the LGB module integrates a gated attention mechanism individually performed on each wavelet subband (LL, LH, HL, and HH), and a spatial-frequency fusion block to effectively distinguish between flare and non-flare regions and facilitate the joint utilization of spatial and frequency features, thereby enhancing flare suppression while preserving essential image details.

During the reconstruction stage, the Inverse Discrete Wavelet Transform (IDWT) is progressively applied in a bottom-up fashion to recover spatial resolution at each level. Specifically, the reconstructed feature $\tilde{F}_{i-1}$ is obtained as:

$$\begin{cases} \tilde{F}_{i-1} = \text{IDWT}\left(\tilde{F}_i + F_i^{LL}, F_i^{LH}, F_i^{HL}, F_i^{HH}\right) \\ \tilde{F}_2 = \text{IDWT}\left(\text{LGB}\left(F_3^{LL}, F_3^{LH}, F_3^{HL}, F_3^{HH}\right)\right) \end{cases} \quad (2)$$

Here, $\tilde{F}_i$ denotes the reconstructed feature maps from the previous level, and $F_i^{LL}, F_i^{LH}, F_i^{HL}$, and F_i^{HH} are the low- and high-frequency components obtained during the wavelet decomposition. Similarly, a residual block is also appended after each reconstruction level to refine structural details in the progressively recovered features.

Finally, the reconstructed features $\tilde{F}_0$ are fused with F_0^{res} via a 1×1 convolution, and the result is then combined with F_{in} through a residual connection to produce F_{out}:

$$F_{out} = F_{in} + \text{Conv}(F_0^{\text{res}} + \tilde{F}_0) \quad (3)$$

where F_0^{res} is the output of a residual block applied to F_0.

Gated-Enhanced Attention Block (GEAB)

Most attention-based flare removal networks (Zhang et al. 2023; Ma et al. 2025; Lian et al. 2025) rely on architectures for contextual modeling. However, these designs often struggle to explicitly differentiate between flare-affected and flare-free regions, resulting in insufficient suppression of flare artifacts. To address this limitation, we propose a novel GEAB module, which integrates a learnable gating mechanism into the attention module, enabling the network to adaptively highlight flare-relevant features.

Given the input feature maps $F_{in} \in \mathbb{R}^{H \times W \times C}$, we first apply layer normalization (Ba, Kiros, and Hinton 2016) to F_{in}, followed by three Shuffle Convolutions (SConv), each consisting of a 1×1 convolution, a PReLU activation (He et al. 2015), a 3×3 group convolution (GConv), and a channel shuffle operation (Zhang et al. 2018b), to generate channel-expanded queries and keys ($Q, K \in \mathbb{R}^{H \times W \times C \cdot h}$), and values ($V \in \mathbb{R}^{H \times W \times C}$) as representations. These are then split into h subspaces ($i = 1, \ldots, h$) along channel dimension, denoted as $Q_i, K_i \in \mathbb{R}^{H \times W \times C}$ and $V_i \in \mathbb{R}^{H \times W \times \frac{C}{h}}$, where $i = 1, \ldots, h$ indicates the i-th subspace. The multi-head attention is calculated as:

$$\text{Attn}(Q, K, V) = \text{Cat}_{i=1}^{h}\left(\text{Softmax}\left(\frac{Q_i K_i^\top}{\sqrt{d}}\right) V_i\right) \quad (4)$$

where $d = C/h$ denotes the head dimension and Cat operator represents channel-wise concatenation. Compared to the classic multi-head mechanism, our channel-expanded Q and K assign C-dimensional queries and keys to each head, effectively alleviating the subspace bottleneck and facilitating richer contextual representations.

Furthermore, to enhance attention with spatial details and flare-aware guidance, we introduce two auxiliary branches: DNet, which compensates for spatial detail loss in the Value (V) branch; and GNet, which serves as an adaptive spatial gate to highlight flare-relevant regions and guide attention accordingly. Specifically, DNet employs two 3×3 depthwise convolutions for local texture extraction and a 1×1 convolution for channel integration, followed by GELU (Hendrycks and Gimpel 2016) to preserve contrast between flare and non-flare regions, enhancing detail recovery. GNet shares the same structure as DNet but replaces GELU with PReLU (He et al. 2015), enabling stronger activation of flare-related features and guiding attention toward degraded regions. The gated-enhanced attention output O is formulated as follows:

$$O = (\text{Attn}(Q, K, V) + \text{DNet}(V)) \cdot \text{GNet}(\text{LN}(F_{in})) \quad (5)$$

Finally, the residual connection forms the enhanced feature output F_{out}:

$$F_{out} = F_{in} + \text{Conv}(O) \quad (6)$$

Spatial-Frequency Fusion Network (SFFN)

To effectively integrate multi-level wavelet decomposition features from MWEB with gated-enhanced attention features from GEAB, SFFN is designed to jointly fuse spatial and frequency domain information to enrich feature representation, enhancing the network's ability to distinguish and remove flare artifacts.

Given the input feature maps $F_{in} \in \mathbb{R}^{H \times W \times C}$, we first apply layer normalization to normalize the features. Then, a lightweight Depthwise Separable Convolution (DSC), consisting of a 3×3 depthwise convolution, a ReLU activation, and a 1×1 pointwise convolution, is utilized to expand the channel dimension to $\frac{3C}{2}$, enriching the feature representation and facilitating effective separation into spatial and frequency branches:

$$F_{DSC} = \text{DSC}\left(\text{LN}\left(F_{in}\right)\right) \quad (7)$$

Next, motivated by (Chen et al. 2023), the Partial Fourier Convolution (PFC) module splits channels into a spatial branch F_{spatial} with C channels to preserve spatial details, and a frequency branch F_{freq} with $\frac{C}{2}$ channels for frequency-domain processing. The frequency branch is transformed via a 2D Real Fast Fourier Transform (FFT), with real and imaginary parts concatenated along the channel dimension. To further enhance these frequency features, two DSC modules with a ReLU activation in between are applied. An inverse Real FFT then reconstructs the refined features back to the spatial domain. Finally, F_{spatial} and the reconstructed frequency features are concatenated channel-wise to produce the PFC output F_{PFC}.

Finally, a 1×1 convolution is applied to fuse the spatial-frequency features, and the fused result is combined with F_{in} through a residual connection to produce F_{out}:

$$F_{out} = F_{in} + \text{Conv}(\text{PFC}(F_{DSC})) \quad (8)$$

Dataset	Method	Publication	PSNR↑	SSIM↑	LPIPS↓	G-PSNR↑	S-PSNR↑
	Input	–	22.56	0.857	0.0777	19.556	13.105
Previous Data Synthesis Pipelines	Zhang (Zhang et al. 2020)	MM'20	21.02	0.784	0.1738	19.868	13.062
	Sharma (Sharma and Tan 2021)	CVPR'21	20.49	0.826	0.1115	17.790	12.648
	Wu (Wu et al. 2021)	ICCV'21	24.61	0.871	0.0598	21.772	16.728
Flare7K Data Synthesis Pipeline	U-Net (Ronneberger, Fischer, and Brox 2015)	MICCAI'15	26.11	0.879	0.055	–	–
	HINet (Chen et al. 2021)	CVPRW'21	26.74	0.882	0.048	–	–
	MPRNet* (Zamir et al. 2021)	CVPR'21	26.14	0.878	0.050	–	–
	Restormer* (Zamir et al. 2022)	CVPR'22	26.28	0.883	0.054	–	–
	Dai (Dai et al. 2022)	NeurIPS'22	26.98	0.890	0.047	23.507	21.563
	FF-Former (Zhang et al. 2023)	CVPRW'23	27.35	0.901	0.044	–	–
	Zhou (Zhou et al. 2023)	ICCV'23	25.18	0.872	0.055	22.112	20.543
	SGSFT (Ma et al. 2025)	TASE'25	27.57	0.897	0.045	23.845	22.636
	FBNet (Lian et al. 2025)	TCE'25	27.35	0.895	0.043	–	–
Flare7K++ Data Synthesis Pipeline	U-Net (Ronneberger, Fischer, and Brox 2015)	MICCAI'15	27.19	0.894	0.0452	23.527	22.647
	HINet (Chen et al. 2021)	CVPRW'21	27.55	0.892	0.0464	24.081	22.907
	MPRNet* (Zamir et al. 2021)	CVPR'21	27.04	0.893	0.0481	23.490	22.267
	Restormer* (Zamir et al. 2022)	CVPR'22	27.60	0.897	0.0447	23.828	22.452
	Dai++ (Dai et al. 2024)	TPAMI'24	27.63	0.894	0.0428	23.949	22.603
	Kotp and Torki (Kotp and Torki 2024)	ICASSP'24	27.66	0.897	0.0422	23.987	22.847
	SGSFT (Ma et al. 2025)	TASE'25	28.08	0.904	0.0416	24.477	23.306
	LPFSformer (Chen et al. 2025)	TCSVT'25	28.24	0.905	0.0422	24.793	23.876
	FBNet (Lian et al. 2025)	TCE'25	27.71	0.903	0.0423	24.226	22.775
	WGSF-Net (Ours)	AAAI'26	**28.52**	**0.907**	**0.0383**	**24.984**	**24.084**

Table 1: Quantitative results on Flare7K real test set. "*" denotes models with reduced parameters due to the limited GPU memory. The best result is highlighted in **red** and the second best result is highlighted in blue.

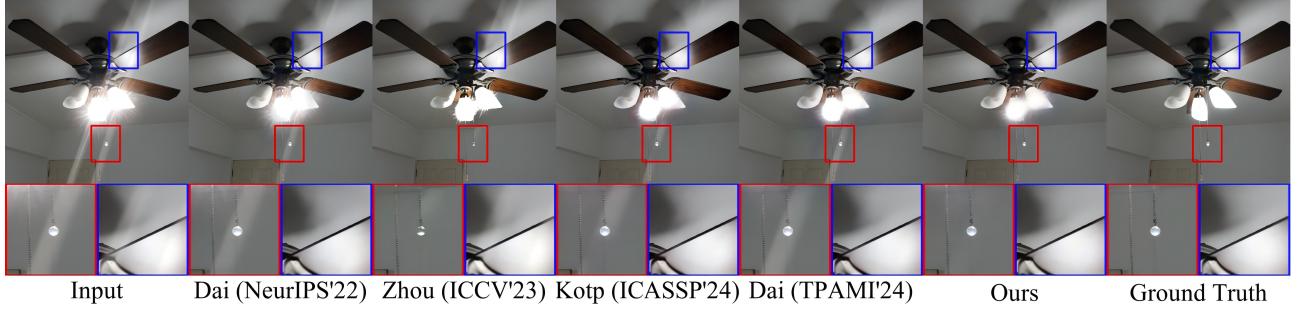Figure 3: Visual comparisons on a real-world flare-corrupted image from Flare7K real test set.

Input | Dai (NeurIPS'22) | Zhou (ICCV'23) | Kotp (ICASSP'24) | Dai (TPAMI'24) | Ours | Ground Truth

Figure 4: Visual comparisons on a synthetic flare-corrupted image from Flare7K synthetic test set.

Input | Dai (NeurIPS'22) | Zhou (ICCV'23) | Kotp (ICASSP'24) | Dai (TPAMI'24) | Ours | Ground Truth

Loss Function

Following (Dai et al. 2024), we employ the pixel-wise loss $\mathcal{L}_1$, the perceptual loss $\mathcal{L}_{vgg}$, and the reconstruction loss $\mathcal{L}_{rec}$ for supervised training. In addition, the Fourier loss $\mathcal{L}_{fft}$ (Qi, Wang, and Liu 2025) constrains the clean image reconstruction in the frequency domain. Overall, the total

Table 2 content:

Dataset	Method	Publication	PSNR↑	SSIM↑	LPIPS↓	G-PSNR↑	S-PSNR↑
	Input	-	22.77	0.921	0.060	18.804	13.927
Previous Data Synthesis Pipeline	Zhang (Zhang et al. 2020)	MM'20	21.04	0.841	0.136	19.625	14.330
	Sharma (Sharma and Tan 2021)	CVPR'21	20.01	0.865	0.111	–	–
	Wu (Wu et al. 2021)	ICCV'21	27.88	0.952	0.031	–	–
Flare7K Data Synthesis Pipeline	U-Net (Ronneberger, Fischer, and Brox 2015)	MICCAI'15	29.07	0.958	0.022	–	–
	HINet (Chen et al. 2021)	CVPRW'21	29.97	0.959	0.021	–	–
	MPRNet* (Zamir et al. 2021)	CVPR'21	29.87	0.959	0.020	–	–
	Restormer* (Zamir et al. 2022)	CVPR'22	29.45	0.950	0.025	–	–
	Dai (Dai et al. 2022)	NeurIPS'22	30.47	0.965	**0.017**	25.735	25.279
	FF-Former (Zhang et al. 2023)	CVPRW'23	30.88	**0.969**	0.019	–	–
	Zhou (Zhou et al. 2023)	ICCV'23	28.78	0.939	0.029	23.779	22.237
Flare7K++ Data Synthesis Pipeline	Dai++ (Dai et al. 2024)	TPAMI'24	29.50	0.962	0.021	24.685	24.155
	Kotp and Torki (Kotp and Torki 2024)	ICASSP'24	29.57	0.961	0.021	24.879	24.458
	WGSF-Net (Ours)	AAAI'26	**30.91**	**0.969**	0.018	**26.264**	**25.988**

Table 2: Quantitative results on Flare7K synthetic test set. "*" denotes models with reduced parameters due to limited GPU memory. The best result is highlighted in **red** and the second best result is highlighted in blue.
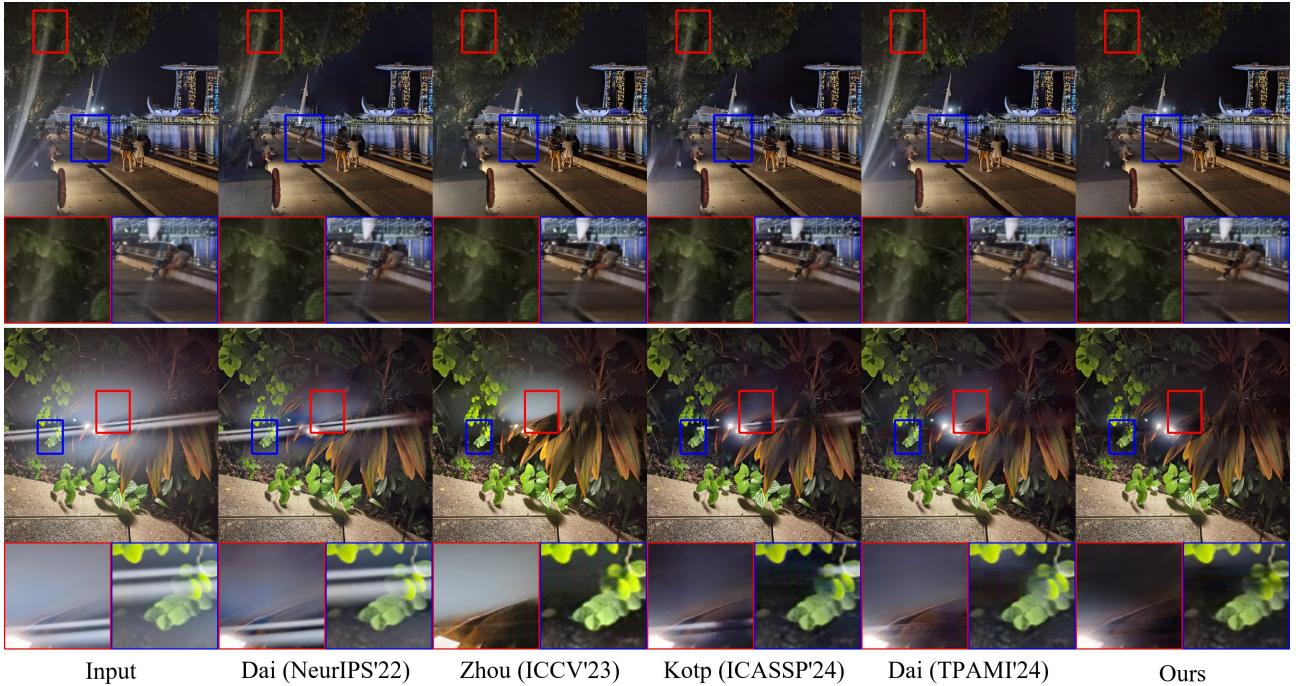

Figure 5: Visual comparisons on real flare-corrupted images from Unpaired Official Flare Corrupted (UOFC) (Dai et al. 2022).

loss $\mathcal{L}$ is defined as:

$$\mathcal{L} = \lambda_1 \mathcal{L}_1 + \lambda_{vgg} \mathcal{L}_{vgg} + \lambda_{rec} \mathcal{L}_{rec} + \lambda_{fft} \mathcal{L}_{fft} \quad (9)$$

where $\lambda_1 = 1.0$, $\lambda_{vgg} = 1.0$, $\lambda_{rec} = 2.0$, and $\lambda_{fft} = 0.1$.

Experiments

Experiment Setup

Datasets. For network training, we follow the data synthesis pipeline and preprocessing strategy proposed by (Dai et al. 2024), which dynamically generates paired flare-corrupted and flare-free images by using 23,949 background images randomly sampled from 24K Flickr dataset (Zhang, Ng, and Chen 2018) and flare patterns from Flare7K++ dataset. For testing, we perform both quantitative and qualitative comparisons on 100 pairs of real nighttime flare images and 100 pairs of synthetic nighttime flare images from the Flare7K (Dai et al. 2022) test set.

Implementation Details. Our network is trained in an end-to-end fashion for 600,000 iterations using the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.99$, and a fixed learning rate of 1×10^{-4}. During training, input images are randomly cropped into patches of size $512 \times 512 \times 3$, and a batch size

Table 3 (Ablation studies):

Ablation study					Flare7K real test set					Flare7K synthetic test set				
Loss function	$\mathcal{L}_1$	$\mathcal{L}_{vgg}$	$\mathcal{L}_{fft}$	$\mathcal{L}_{rec}$	PSNR↑	SSIM↑	LPIPS↓	G-PSNR↑	S-PSNR↑	PSNR↑	SSIM↑	LPIPS↓	G-PSNR↑	S-PSNR↑
	✓	✗	✗	✗	28.18	0.899	0.0491	24.906	22.778	29.29	0.958	0.0307	24.534	22.992
	✓	✓	✗	✗	28.35	0.903	0.0415	24.905	23.770	30.22	0.966	0.0205	25.276	25.213
	✓	✓	✓	✗	28.46	**0.907**	0.0408	24.956	23.994	30.01	0.963	0.0236	25.180	24.573
Network components	SFFN	GEAB	MWEB	ResBlock	PSNR↑	SSIM↑	LPIPS↓	G-PSNR↑	S-PSNR↑	PSNR↑	SSIM↑	LPIPS↓	G-PSNR↑	S-PSNR↑
	✗	✓	✓	✓	28.15	0.905	0.0416	24.660	23.852	29.26	0.962	0.0231	24.694	24.363
	✓	✗	✓	✓	28.33	0.905	0.0414	24.733	23.508	29.51	0.962	0.0240	24.834	24.178
	✓	✓	✗	✓	28.42	0.904	0.0419	**25.030**	23.824	30.11	0.966	0.0217	25.306	24.823
	✓	✓	✓	✗	28.44	0.905	0.0397	25.023	23.985	30.33	**0.969**	0.0203	25.331	25.611
Modules					PSNR↑	SSIM↑	LPIPS↓	G-PSNR↑	S-PSNR↑	PSNR↑	SSIM↑	LPIPS↓	G-PSNR↑	S-PSNR↑
	w/o WFB				28.25	0.905	0.0405	24.655	23.787	29.81	0.967	0.0212	24.849	24.667
	w/o LGB				28.09	0.904	0.0438	24.597	23.008	29.95	0.964	0.0235	24.950	24.442
Our full WGSF-Net framework					**28.52**	**0.907**	**0.0383**	24.984	**24.084**	**30.91**	**0.969**	**0.0182**	**26.264**	**25.988**

Table 3: Ablation studies on various configurations. Best results are highlighted in **red**.

of 2 is adopted. The channel dimension C is set to 32. All our experiments are conducted on a single NVIDIA RTX 4090 GPU (24GB) using the PyTorch framework.

Evaluation Metrics. PSNR, SSIM (Wang et al. 2004), and LPIPS (Zhang et al. 2018a) are used to evaluate image restoration quality, while G-PSNR and S-PSNR (Dai et al. 2024) measure the performance of glare and streak removal.

Comparisons with State-of-the-art Methods

Quantitative Comparisons. Tables 1 and 2 present the quantitative results on the Flare7K real and synthetic test sets, respectively. On the real test set, our WGSF-Net consistently outperforms all competing methods across five metrics. Specifically, our WGSF-Net achieves a 7.93% reduction in LPIPS, indicating better perceptual quality, and improves G-PSNR and S-PSNR by 0.77% (0.191dB) and 0.87% (0.208dB), respectively, reflecting superior restoration performance in flare-corrupted regions. On the synthetic test set, WGSF-Net achieves the best performance on four metrics and a comparable LPIPS score. In particular, G-PSNR and S-PSNR are further improved by 2.06% (0.529dB) and 2.80% (0.709dB), respectively. These results demonstrate the superior restoration performance and flare removal capability of our proposed network on both real and synthetic datasets.

Qualitative Comparisons. Figs. 3 and 4 illustrate the visual comparisons of flare removal results on real and synthetic flare-corrupted images, respectively. As highlighted in the red and blue zoomed-in regions of Figs. 3-4, our WGSF-Net not only suppresses flare artifacts more effectively but also better preserves scene details, producing more natural and visually consistent results with the ground truth. Furthermore, to demonstrate generalization in real-world scenarios, Fig. 5 shows visual comparisons on unpaired real flare-corrupted images from Flare7K (Dai et al. 2022). Our WGSF-Net produces significantly clearer and more visually compelling results than state-of-the-art methods, highlighting its superior perceptual quality under practical conditions.

Ablation Study

Effectiveness of Loss Function. Table 3 shows that each loss function contributes to overall performance, with WGSF-Net achieving the best results across five objective metrics when all four losses are combined.

Effectiveness of Network Modules. From Table 3, incorporating **SFFN** improves G-PSNR by 1.31%/6.36% (0.324/1.570dB) and reduces LPIPS by 7.93%/21.21% on the real/synthetic test sets, validating its effectiveness in flare removal. **GEAB** enhances S-PSNR by 2.45%/7.49% (0.576/1.810dB) and provides a 7.49%/24.17% reduction in LPIPS on the real/synthetic test sets. With LPIPS reduced by 8.59%/16.13% on the real/synthetic test sets, **MWEB** verifies the importance of multi-scale wavelet feature extraction for detail restoration. **ResBlock** leads to an LPIPS drop of 3.53% and 10.34% on real and synthetic test sets, respectively, highlighting its contribution to texture refinement. Overall, all four modules contribute positively to our WGSF-Net framework.

Effectiveness of WFB and LGB. Both WFB and LGB modules contribute to consistent improvements for all five evaluation metrics on real and synthetic data, validating their effectiveness in flare removal and scene restoration.

Conclusion

In this paper, we have proposed a novel end-to-end flare removal network, called WGSF-Net, which integrates multi-level wavelet-guided features with a gated-enhanced attention mechanism to enable effective spatial and frequency-domain fusion. To fully and selectively leverage frequency-domain features, we design three key modules: MWEB for coarse-to-fine multi-scale frequency feature extraction, GEAB for adaptive attention to flare-related regions, and SFFN for fusing spatial and frequency features. Quantitative and qualitative comparisons on both real and synthetic datasets demonstrate that our WGSF-Net consistently outperforms state-of-the-art methods in removing flare artifacts.

Acknowledgments

References

Ba, J. L.; Kiros, J. R.; and Hinton, G. E. 2016. Layer Normalization. *arXiv preprint arXiv:1607.06450*.

Chen, G.-Y.; Dong, W.; Fan, G.; Su, J.-N.; Gan, M.; and Philip Chen, C. 2025. LPFSformer: Location Prior Guided Frequency and Spatial Interactive Learning for Nighttime Flare Removal. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(4): 3706–3718.

Chen, H.-T.; Zhou, J.; O'Hara, J. F.; Chen, F.; Azad, A. K.; and Taylor, A. J. 2010. Antireflection Coating Using Metamaterials and Identification of Its Mechanism. *Physical Review Letters*, 105(7): 073901.

Chen, J.; Kao, S.-h.; He, H.; Zhuo, W.; Wen, S.; Lee, C.-H.; and Chan, S.-H. G. 2023. Run, Don't Walk: Chasing Higher FLOPS for Faster Neural Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12021–12031.

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 (CVPR) Workshops*, 182–192.

Chi, L.; Jiang, B.; and Mu, Y. 2020. Fast Fourier Convolution. *Advances in Neural Information Processing Systems (NeurIPS)*, 33: 4479–4488.

Cong, X.; Gui, J.; Zhang, J.; Hou, J.; and Shen, H. 2024. A Semi-supervised Nighttime Dehazing Baseline with Spatial-Frequency Aware and Realistic Brightness Constraint. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2631–2640.

Dai, Y.; Li, C.; Zhou, S.; Feng, R.; and Loy, C. C. 2022. Flare7K: A Phenomenological Nighttime Flare Removal Dataset. *Advances in Neural Information Processing Systems (NeurIPS)*, 35: 3926–3937.

Dai, Y.; Li, C.; Zhou, S.; Feng, R.; Luo, Y.; and Loy, C. C. 2024. Flare7K++: Mixing Synthetic and Real Datasets for Nighttime Flare Removal and Beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(11): 7041–7055.

Dai, Y.; Luo, Y.; Zhou, S.; Li, C.; and Loy, C. C. 2023. Nighttime Smartphone Reflective Flare Removal Using Optical Center Symmetry Prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 20783–20791.

Faulkner, K.; Kotre, C.; and Louka, M. 1989. Veiling Glare Deconvolution of Images Produced by X-ray Image Intensifiers. In *Third International Conference on Image Processing and its Applications*, 669–673.

Finder, S. E.; Amoyal, R.; Treister, E.; and Freifeld, O. 2024. Wavelet Convolutions for Large Receptive Fields. In *European Conference on Computer Vision (ECCV)*, 363–380.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 1026–1034.

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 (AAAI)*, volume 39, 3464–3472.

Hendrycks, D.; and Gimpel, K. 2016. Gaussian Error Linear Units (Gelus). *arXiv preprint arXiv:1606.08415*.

Jin, Y.; Lin, B.; Yan, W.; Yuan, Y.; Ye, W.; and Tan, R. T. 2023. Enhancing Visibility in Nighttime Haze Images Using Guided APSF and Gradient Adaptive Convolution. In *Proceedings of the 31st ACM International Conference on Multimedia (ACM MM)*, 2446–2457.

Kotp, Y.; and Torki, M. 2024. Flare-Free Vision: Empowering Uformer with Depth Insights. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2565–2569.

Lian, J.; Liu, J.; Jing, M.; Zeng, X.; Liu, Z.; Zhou, J.; and Fan, Y. 2025. Nighttime Glare Removal for Consumer Electronics via Latent Space Transformation and Feature-Enhanced Attention Mechanism. *IEEE Transactions on Consumer Electronics*, 71(2): 6719–6733.

Lin, B.; Jin, Y.; Wending, Y.; Ye, W.; Yuan, Y.; and Tan, R. T. 2025. NightHaze: Nighttime Image Dehazing via Self-Prior Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 39, 5209–5217.

Liu, Y.; Wang, X.; Hu, E.; Wang, A.; Shiri, B.; and Lin, W. 2025. VNDHR: Variational Single Nighttime Image Dehazing for Enhancing Visibility in Intelligent Transportation Systems via Hybrid Regularization. *IEEE Transactions on Intelligent Transportation Systems*, 26(7): 10189–10203.

Liu, Y.; Yan, Z.; Chen, S.; Ye, T.; Ren, W.; and Chen, E. 2023a. NightHazeFormer: Single Nighttime Haze Removal Using Prior Query Transformer. In *Proceedings of the 31st ACM International Conference on Multimedia (ACM MM)*, 4119–4128.

Liu, Y.; Yan, Z.; Tan, J.; and Li, Y. 2023b. Multi-Purpose Oriented Single Nighttime Image Haze Removal Based on Unified Variational Retinex Model. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(4): 1643–1657.

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*, 22: 11996 – 12011.

Qi, K.; Wang, B.; and Liu, Y. 2025. A Self-prompt Based Dual-domain Network for Nighttime Flare Removal. *Engineering Applications of Artificial Intelligence*, 144: 110103.

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 (ICCV)*, 4177–4185.

Ramesh, R. 2008. Glare Aware Photography: 4D Ray Sampling for Reducing Glare Effects of Camera Lenses. *ACM Transactions on Graphics*, 27(3): 54.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 234–241.

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.

Sharma, A.; and Tan, R. T. 2021. Nighttime Visibility Enhancement by Increasing the Dynamic Range and Suppression of Light Effects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11977–11986.

Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A. P.; Bishop, R.; Rueckert, D.; and Wang, Z. 2016. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1874–1883.

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 (CVPR)*, 17683–17693.

Wu, T.-P.; and Tang, C.-K. 2005. A Bayesian Approach for Shadow Extraction from a Single Image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 480–487.

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 (ICCV)*, 2239–2247.

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 (CVPR)*, 5728–5739.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; Yang, M.-H.; and Shao, L. 2021. Multi-Stage Progressive Image Restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 14821–14831.

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 (CVPR) Workshops*, 2824–2832.

Zhang, J.; Cao, Y.; Zha, Z.-J.; and Tao, D. 2020. Nighttime Dehazing with a Synthetic Benchmark. In *Proceedings of the 28th ACM International Conference on Multimedia (ACM MM)*, 2355–2363.

Zhang, R.; Isola, P.; Efros, A. A.; Shechtman, E.; and Wang, O. 2018a. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 586–595.

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 (CVPR)*, 4786–4794.

Zhang, X.; Zhou, X.; Lin, M.; and Sun, J. 2018b. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6848–6856.

Zhang, Z.; Feng, H.; Xu, Z.; Li, Q.; and Chen, Y. 2018c. Single Image Veiling Glare Removal. *Journal of Modern Optics*, 65(19): 2220–2230.

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*, 47(9): 7396–7409.

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 (ICCV)*, 12969–12979.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2223–2232.