

MFDNet: Multi-Frequency Deflare Network for Efficient Nighttime Flare Removal

Yiguo Jiang · Xuhang Chen · Chi-Man Pun · Shuqiang Wang · Wei Feng

Abstract When light is scattered or reflected accidentally in the lens, flare artifacts may appear in the captured photos, affecting the photos' visual quality. The main challenge in flare removal is to eliminate various flare artifacts while preserving the original content of the image. To address this challenge, we propose a lightweight Multi-Frequency Deflare Network (MFDNet) based on the Laplacian Pyramid. Our network decomposes the flare-corrupted image into low and high-frequency bands, effectively separating the illumination and content information in the image. The low-frequency part typically contains illumination information, while the high-frequency part contains detailed content information. So our MFDNet consists of two main modules: the Low-Frequency Flare Perception Module (LFFPM) to remove flare in the low-frequency part and the Hierarchical Fusion Reconstruction Module (HFRM) to reconstruct the flare-free image. Specifically, to perceive flare from a global perspective while retaining detailed information for image restoration, LFFPM utilizes Transformer to extract global information while utilizing a convolutional neural network to capture detailed local features. Then HFRM gradually fuses the outputs of LFFPM with the high-frequency component of the image through feature aggregation. Moreover, our MFDNet can reduce the computational cost by processing in multiple frequency bands instead of directly removing the flare on the input image. Experimental results demonstrate that our approach outperforms state-of-the-art methods in removing nighttime flare on real-world and synthetic images from the

Flare7K dataset. Furthermore, the computational complexity of our model is remarkably low.

Keywords Flare removal · Multi-frequency · CNN · Transformer · Efficient

1 Introduction

Photographs taken in nighttime scenes with bright light sources often exhibit flare artifacts, which occur as a result of undesired scattering and reflection of intense light within the camera lens. Light scattering and reflection are common in real camera lenses, particularly in consumer-grade mobile phone cameras. Daily wear and tear, along with the presence of fingerprints and dust, can unintentionally cause light scattering or reflections in the lens. These flare artifacts not only impact the aesthetics of the photograph but also degrade the detailed visual information, hindering image comprehension. As a result, there is a strong demand for a reliable and effective nighttime flare removal algorithm.

The flare patterns in photographs are affected by the lens properties and shooting environment, which encompass factors such as the design of the optics lens, manufacturing imperfections, lens smudges, and the light source's position and angle relative to the lens. The diversity of these factors results in flares with different shapes, positions, and colors. Typical flare artifacts include glare, streaks, bright colored lines, shimmer, saturated blobs, and many others. The variety in appearance of flare artifacts makes it challenging to remove them entirely from a photograph while preserving other content information, especially when multiple flare patterns exist within a single image.

Traditional flare removal methods include hardware-based methods [3, 25] and software-based methods [1,

Corresponding author: Chi-Man Pun
E-mail: cmpun@umac.mo

This work was supported in part by the Science and Technology Development Fund, Macau SAR, under Grants 0141/2023/RIA2 and 0193/2023/RIA3.

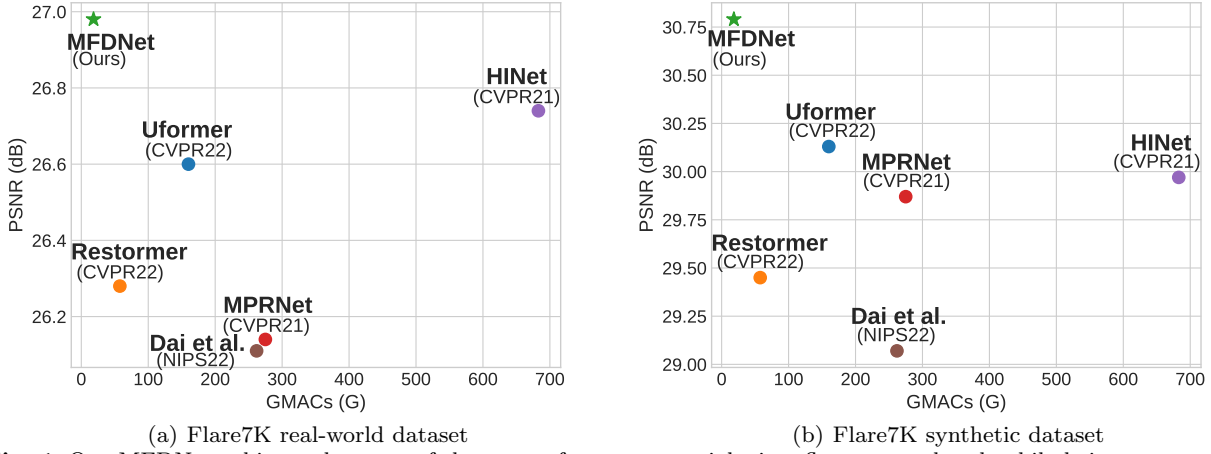


Fig. 1 Our MFDNet achieves the state-of-the-art performance on nighttime flare removal task while being computationally efficient.

5, 14, 29, 32, 45]. Advanced materials and refined optical designs can contribute to more specialized lenses that reduce flare artifacts. Applying an anti-reflective coating is another widely used way of reducing flare impact. Nevertheless, these hardware methods can alleviate some flare effects but cannot wholly remove various flares. Furthermore, these hardware-based methods are not useful in images containing flares and are relatively expensive. To address these above problems, some software-based algorithms have emerged for removing flares. These methods usually follow two steps: detecting flares based on their characteristics and then removing them. However, software-based methods struggle to handle a broad range of flare artifacts.

Recently, Deep learning-based methods [8, 13, 35–38] for removing flare have emerged. Nevertheless, most of these existing methods view flare removal as a general image restoration task and do not consider effectively decoupling the image’s illumination and content information. As a result, these methods might not fully eliminate the flare artifacts or cause degradation of the image content during the removal process. Moreover, these methods directly utilize deep networks to globally manipulate the flare-corrupted image, which results in high computational costs. Consequently, their computational complexity exponentially increases with image resolution, making it unfeasible to apply to high-resolution images, reducing the algorithm’s applicability.

In this paper, we propose a lightweight Multi-Frequency Deflare Network (MFDNet) based on the Laplacian Pyramid [4] for nighttime flare removal. Our proposed method aims to effectively eliminate various flare artifacts while preserving the integrity of the original image. Inspired by the reversible frequency-band decomposition frame-

work of a Laplacian Pyramid [4], our MFDNet decouples illumination and content information by decomposing the image into low and high-frequency bands. It performs flare removal in the low-frequency part of the image, followed by gradual fusion with the high-frequency part to reconstruct the flare-free image. At the same time, because our method performs flare removal in the low-frequency part, where the resolution is low, it can effectively reduce the computational complexity. As shown in Figure 1, our method of first decoupling the flare-corrupted image and then removing the flare can effectively eliminate the flare artifacts while minimizing the computational complexity.

Specifically, we propose the Low-Frequency Flare Perception Module (LFFPM) for flare removal in the low-frequency part. In the task of flare removal, a large receptive field is essential due to the extensive coverage of the flare. Thus, global information is crucial in accurately identifying the flare. Considering the Transformer’s proficiency in capturing long-range pixels, the Low-Frequency Flare Perception Module (LFFPM) utilizes the Transformer for global feature extraction and refinement. In order to alleviate the limitation of Transformers in capturing local dependencies and reduce the model’s computational complexity, LFFPM uses a convolution-based encoder-decoder structure to enhance local detailed feature representation. In addition, we propose the Hierarchical Fusion Reconstruction Module (HFRM) for an efficient fusion of high-frequency information. In HFRM, features from the high-frequency component and the results of the Low-Frequency Flare Perception Module (LFFPM) are aggregated at each layer to construct the Laplacian Pyramid for the final reconstruction.

In summary, the contributions of this paper are as follows:

1. We propose a lightweight and effective Multi-Frequency Deflare Network (MFDNet) that removes nighttime flare artifacts by decoupling the image’s illumination and content information into different frequency bands.
2. We design the Low-Frequency Flare Perception Module (LFFPM) to remove flares in the low-frequency part, which utilizes convolution to capture local features and self-attention to model long-range dependencies.
3. We design the Hierarchical Fusion Reconstruction Module (HFRM), which gradually aggregates features from the high-frequency bands and LFFPM’s results to reconstruct the final flare-free image.
4. Extensive experiments demonstrate that our method achieves state-of-the-art performance on nighttime flare removal task while maintaining low computational complexity.

The subsequent sections are structured as follows. Section 2 discusses related works. Section 3 is devoted to the details of our proposed MFDNet elaborately. Section 4 presents our extensive experiments and analysis, and the conclusion is summarized in Section 5.

2 Related Work

2.1 Image Restoration

During the process of capturing photographs, various factors, including unfavorable weather conditions, optics-induced diffraction, and relative motion between the camera and object, can lead to the deterioration of image quality. This degradation results in the loss of important information in the captured image, necessitating the restoration of the image to its original quality. Typical image restoration tasks include but are not limited to image deblurring [9], image denoising [42], image dehazing [10], rain removal [26], reflection removal [19], shadow removal [28], and more. Recently, some state-of-the-art image restoration methods [6, 8, 18, 22, 24, 31, 35, 37–39, 43] have appeared to handle different image restoration tasks.

2.2 Flare Removal

2.2.1 Traditional methods

Traditional flare removal methods include hardware-based methods and software-based methods. Most hardware-based methods aim to reduce flare artifacts by improv-

ing optical designs and camera lens materials. Boynton et al. [3] propose a fluid-filled camera lens to alleviate flare artifacts caused by light reflections. Macleod et al. [25] employ a neutral density filter to minimize reflective flare artifacts. Another commonly used hardware-based approach is to apply an anti-reflective coating to the camera lens. However, this coating may interfere with other coatings like anti-scratch and anti-fingerprint, and it is usually only designed to work for specific light wavelengths and angles of incidence. While these specific hardware approaches can eliminate some lens flare artifacts, they can often not address unforeseeable flares such as those generated by fingerprints or dust on the lens. In addition, these hardware-based methods are usually expensive, and none of them can deal with photographs that already exhibit flare artifacts.

In response to the above problems, some algorithms have been proposed to remove flare artifacts. Seibert et al. [29] and Faulkner et al. [14] propose to use deconvolution to remove flare artifacts. Zhang et al. [45] propose a method to remove flare by separating the image into a flared part and a scene part. Other methods [1, 5, 32] employ a two-step approach, where they detect flare based on their features and subsequently eliminate flare artifacts while reconstructing affected areas through inpainting [11]. However, these approaches might wrongly identify bright regions as flare artifacts. Additionally, they may not be effective in dealing with various patterns of flare artifacts in complex scenarios.

2.2.2 Deep learning-based methods

Recently, deep learning-based methods have achieved good results in some low-level image restoration tasks [6, 8, 18, 22, 31, 35, 37–40, 43]. The success of these deep learning methods typically depends on domain-specific image training datasets. However, collecting large-scale pairs of flare-corrupted and flare-free images from real-world scenes can be a labor-intensive and time-consuming process. Consequently, the development of deep learning-based flare removal algorithms has been sluggish, with only a limited amount of related work emerging in recent years.

Wu et al. [36] proposed the first deep learning-based method for daytime flare removal utilizing a U-Net [27] model to reconstruct flare-free images. Their approach involved a post-processing step to reintegrate the light source into the restored image. Many subsequent methods, including ours, adopted a similar pipeline by first removing the flare and then blending the light source back into the image. Dai et al. [13] proposed a method for synthesizing flare by simulating the optical principles of nighttime flare generation. This enables them to

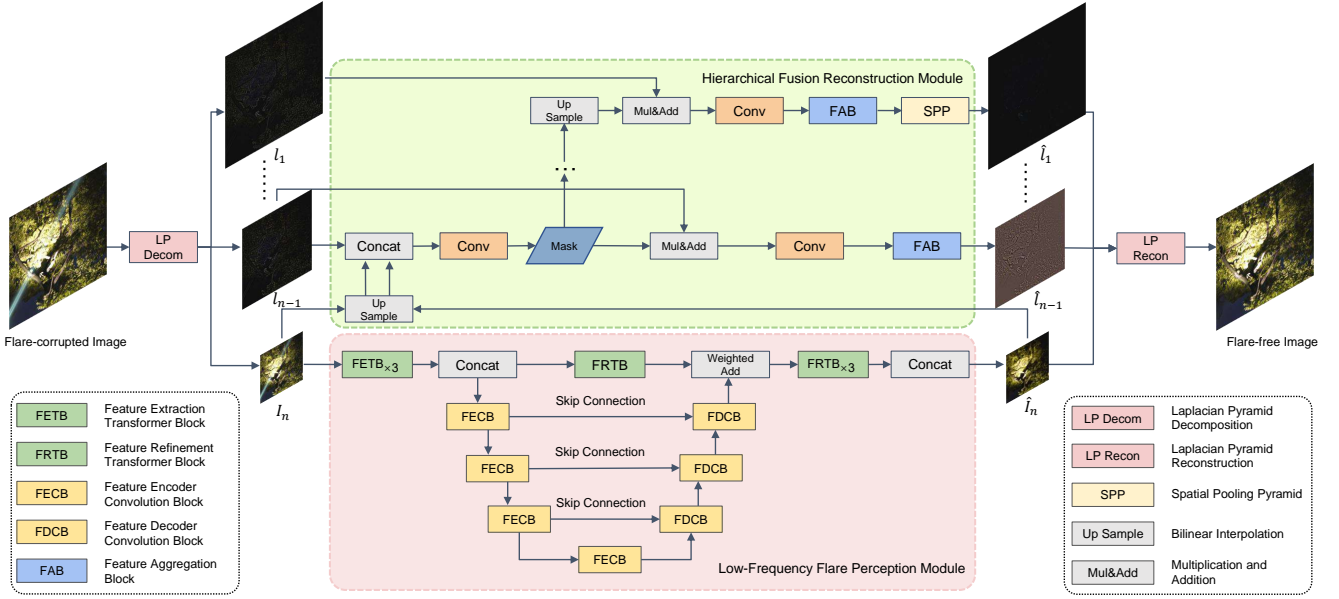


Fig. 2 The structure of our proposed MFDNet. Given a nighttime flare-corrupted image, our MFDNet first decouples its content and illumination information by decomposing this image into high and low-frequency bands. Then MFDNet performs flare removal in the low-frequency part of the image, followed by gradual fusion with the detailed high-frequency part to reconstruct the final flare-free image.

construct datasets consisting of paired flare-corrupted and flare-free images. They created Flare7K, the first benchmark dataset for nighttime flare removal, which serves as a valuable resource for tackling this challenging task. With the proposed Flare7K dataset, they follow Wu et al. [36] and train a U-Net [27] as a baseline network. Meanwhile, they train some state-of-the-art image restoration methods in the Flare7K dataset, including HINet [8], MPRNet [38], Restormer [37], and Uformer [35] to build the flare removal benchmark. FF-Former [40] proposes a U-shape network based Fast Fourier Convolution (FFC) for nighttime flare removal, which addresses the issue of limited receptive field in traditional window-based Transformer approaches.

HINet [8] integrates Instance Normalization (IN) into the basic module to build the HIN module, which improves the performance of the image restoration network. MPRNet [38] proposes a multi-stage architecture that incrementally learns recovery functions for degraded inputs, thereby dividing the entire restoration process into more manageable steps. Restormer [37] proposes an encoder-decoder Transformer model to learn multi-scale representations of high-resolution images without decomposing them into local windows. Uformer [35] proposes a universal U-shaped Transformer for various image restoration tasks, which is built on the basic locally-enhanced window Transformer module and is efficient and effective. In Section 4, we compare our MFDNet with these state-of-the-art methods in Flare7K benchmark.

3 Methodology

3.1 Overview

For the nighttime flare removal task, we propose a lightweight Multi-Frequency Deflare Network (MFDNet) based on the Laplacian Pyramid. As shown in Figure 2, our MFDNet consists of two primary modules: the Low-Frequency Flare Perception Module (LFFPM) and the Hierarchical Fusion Reconstruction Module (HFRM). Next, we first describe the overall pipeline of MFDNet, and then we detail the LFFPM in Section 3.2 and the HFRM in Section 3.3.

The Laplacian Pyramid (LP) [4] is a frequency-band image decomposition technique derived from the Gaussian Pyramid (GP). The main idea of the LP method is to decompose an image linearly into high-frequency and low-frequency bands. And based on the LP, the image reconstruction process can be implemented with precision and reversibility. Specifically, the LP is obtained by calculating the difference between adjacent layers in the Gaussian Pyramid. According to [4, 23], the illumination information of the image is more related to the low-frequency band and the high-frequency component contains more detailed content information such as textures. Inspired by the above properties of LP, we decouple the image’s illumination and content information, remove the flare artifacts in the low-frequency part of the image, and subsequently fuse the low-frequency flare-free image with the detailed high-frequency infor-

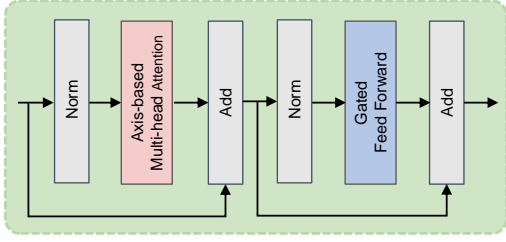


Fig. 3 The structure of Feature Extraction Transformer Block (FETB) and Feature Refinement Transformer Block (FRTB).

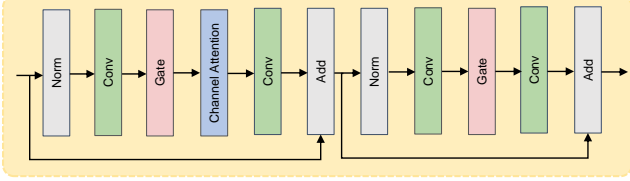


Fig. 4 The structure of Feature Encoder Convolution Block (FECB) and Feature Decoder Convolution Block (FDCB).

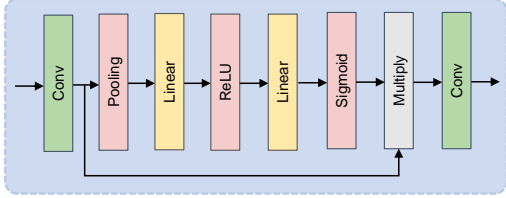


Fig. 5 The structure of Feature Aggregation Block (FAB).

mation to restore the final flare-free image. The entire process is depicted in Figure 2.

Specifically, given a nighttime flare-corrupted image $I \in R^{H \times W \times 3}$, MFDNet first decomposes it into a Laplacian Pyramid, generating a set of different frequency parts $L = [l_1, l_2, \dots, l_{n-1}]$ and the lowest frequency image I_n , where $H \times W$ denotes the spatial dimension, and n is the number of LP's decomposed levels. The components of L have progressively reducing resolutions from $H \times W$ to $\frac{H}{2^{n-1}} \times \frac{W}{2^{n-1}}$, and I_n has $\frac{H}{2^n} \times \frac{W}{2^n}$ pixels. After getting I_n , we input it into the Low-Frequency Flare Perception Module (LFFPM) for flare removal and get a low-frequency flare-free image \hat{I}_n . I_n and \hat{I}_n are then provided to the Hierarchical Fusion Reconstruction Module (HFRM), which fuses each layer of L with \hat{I}_n incrementally to create $\hat{L} = [\hat{l}_1, \hat{l}_2, \dots, \hat{l}_{n-1}]$ for reconstruction. L and \hat{L} share a one-to-one mirror relationship, so the final flare-free image \hat{I} can be reconstructed by \hat{I}_n and \hat{L} . Because our method only needs to remove flares in the low-frequency components, it significantly reduces computational complexity.

3.2 Low-Frequency Flare Perception Module

The Low-Frequency Flare Perception Module (LFFPM) mainly includes the following parts, the Feature Ex-

traction Transformer Block (FETB) and the Feature Refinement Transformer Block (FRTB) for extracting and refining features, and a U-shaped network with skip connections, composed of the Feature Encoder Convolution Block (FECB) and the Feature Decoder Convolution Block (FDCB). Transformers have been demonstrated to be more effective than CNNs in modeling long-range dependencies. As depicted in Figure 2, we use the Feature Extraction Transformer Block (FETB) in the Low-Frequency Flare Perception Module (LFFPM) to extract global features from the input image I_n . Moreover, at the end of LFFPM, we use the Feature Refinement Transformer Block (FRTB) to generate the enhanced features. After multiple FRTBs and FRTBs, we concatenate the output of each block to obtain longer-distance features. Inspired by [33], FETB and FRTB use axis-based self-attention (ASA) to reduce computational complexity and gated feed-forward network (GFFN) to capture more critical features. Traditional self-attention's computational complexity is quadratic with the input resolution. ASA computes self-attention sequentially on the height and width axes across the channel dimension, resulting in linear complexity. GFFN applies GELU and elementwise product to eliminate less relevant features in two parallel paths, then combines the relevant features through element-wise summation. As shown in Figure 3, the computation of FETB and FRTB are represented as:

$$\begin{aligned} \mathbf{F}' &= \text{ASA}(\text{LN}(\mathbf{F})) + \mathbf{F}, \\ \hat{\mathbf{F}} &= \text{GFFN}(\text{LN}(\mathbf{F}')) + \mathbf{F}', \end{aligned} \quad (1)$$

where \mathbf{F} is the input of FETB and FRTB. \mathbf{F}' and $\hat{\mathbf{F}}$ are the outputs of ASA and GFFN, respectively. LN represents the layer normalization [2].

In order to alleviate the limitation of Transformers in capturing local dependencies, the Low-Frequency Flare Perception Module (LFFPM) uses a convolution-based U-shape encoder-decoder structure consisting of Feature Encoder Convolution Block (FECB) and Feature Decoder Convolution Block (FDCB) to enhance detailed feature representation, as shown in the Figure 2. Additionally, to ensure that both the global and local features are adequately fused, the Low-Frequency Flare Perception Module (LFFPM) automatically learns the weighted summation of the global long-distance features extracted by FETB and the output of FDCB. Inspired by [7], FECB and FDCB are designed as simple and efficient nonlinear activation networks. As shown in Figure 4, FECB and FDCB mainly include the following parts: layer normalization [2], convolution, Gate, and Channel Attention [16]. Gate divides the feature map into two parts in the channel dimension and mul-

tiplies them. The formula of Gate is:

$$\text{Gate}(\mathbf{X}, \mathbf{Y}) = \mathbf{X} \odot \mathbf{Y}, \quad (2)$$

where \mathbf{X} and \mathbf{Y} are feature maps.

The Channel Attention(CA) mechanism can capture global information efficiently. The formula of Channel Attention is:

$$\text{CA}(\mathbf{X}) = \mathbf{X} * \text{WPool}(\mathbf{X}), \quad (3)$$

where \mathbf{X} denotes the feature map, W is fully-connected layers, and Pool represents the global average pooling. $*$ is channel-wise product operation.

3.3 Hierarchical Fusion Reconstruction Module

As shown in Figure 2, after obtaining the low-frequency flare-free image \hat{I}_n through the Low-Frequency Flare Perception Module (LFFPM), the Hierarchical Fusion Reconstruction Module (HFRM) fuses \hat{I}_n with the high-frequency parts $L = [l_1, l_2, \dots, l_{n-1}]$ layer by layer to obtain $\hat{L} = [\hat{l}_1, \hat{l}_2, \dots, \hat{l}_{n-1}]$ for reconstruction.

We concatenate the upsampled I_n and \hat{I}_n with l_{n-1} , and then input them into a simple network to learn a mask M_{n-1} to guide the fusion process. The specific fusion process is as follows:

$$l'_{n-1} = l_{n-1} \otimes \mathbf{M}_{n-1} + l_{n-1}, \quad (4)$$

where \otimes represents the pixel-wise multiplication. To refine the fused features further, we perform a dilated convolution operation on l'_{n-1} , then use the Feature Aggregation Block (FAB) to re-weight the significance of features.

Through the above steps, the low-frequency flare-free image \hat{I}_n achieves the feature fusion operation with one layer in the high-frequency parts $L = [l_1, l_2, \dots, l_{n-1}]$. By analogy, we need to upsample \mathbf{M}_{n-1} through linear interpolation and perform the same fusion operation with the next layer in the high-frequency parts $L = [l_1, l_2, \dots, l_{n-1}]$. Behind the Feature Aggregation Block (FAB) in the final layer, we use the Spatial Pooling Pyramid (SPP) [15] to facilitate remixing multi-context features. This way, we get $\hat{L} = [\hat{l}_1, \hat{l}_2, \dots, \hat{l}_{n-1}]$ to reconstruct the final flare-free image.

As shown in Figure 5, the Feature Aggregation Block (FAB) structure is simple. Inspired by [12], we use a squeeze-and-excitation block [16] in FAB to learn the weights of different channel features through some linear layers and pooling layers, which can automatically preserve essential features. Additionally, convolution is used to squeeze the features and match the original channels.

3.4 Loss Function

We train our model with different losses, including the mean square error (MSE) loss L_M , structural similarity loss L_S [34], and perceptual loss [44].

Given the final output image I_{out} and ground truth image I_{gt} , the perceptual loss L_P is defined as:

$$L_P = \sum_l \|F_l(I_{out}) - F_l(I_{gt})\|_2^2, \quad (5)$$

where F is a pre-trained AlexNet [17] feature extractor. We compute the l_2 distance between $F_l(I_{out})$ and $F_l(I_{gt})$ for layer l .

In summary, our total loss function can be expressed as:

$$L_{total} = \lambda_m L_M + \lambda_s L_S + \lambda_p L_P, \quad (6)$$

where we empirically set $\lambda_m = 1$, $\lambda_s = 0.3$ and $\lambda_p = 0.7$ respectively.

4 Experiments and Results

4.1 Setup

4.1.1 Dataset

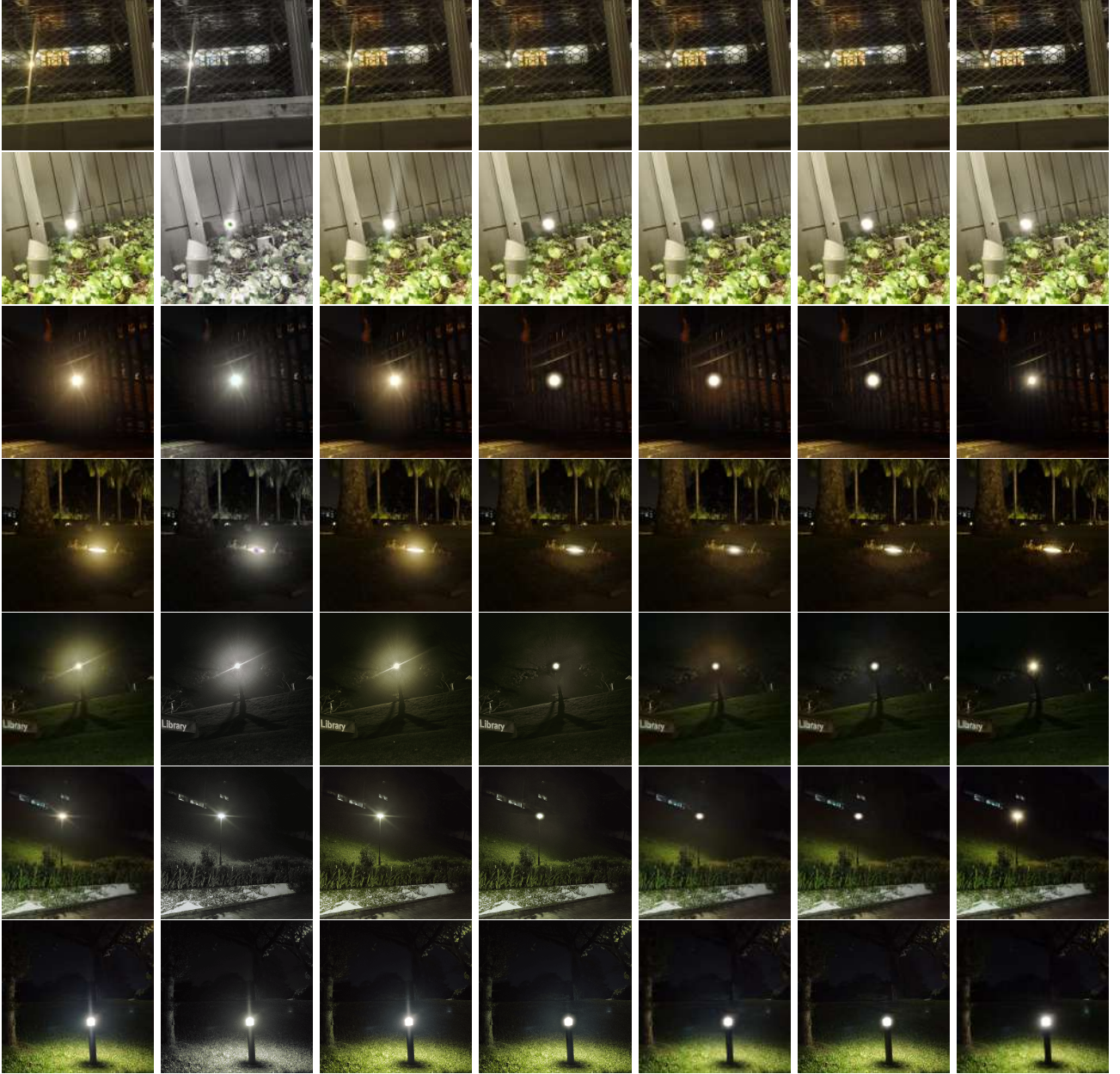
We train our MFDNet based on the Flare7K dataset [13]. Flare7K is currently the largest publicly available nighttime flare dataset, consisting of 5,000 scattering and 2,000 reflective flare images. The dataset comprises 25 types of scattering flares and 10 types of reflective flares. We combine the Flare7K flare image with 23949 flare-free images sampled from the 24K Flickr images [45] to create paired flare-corrupted and flare-free images for training. Furthermore, Flare7K provides 100 real-world flare-corrupted images and 100 synthetic flare-corrupted images, along with their corresponding ground truth, which serve as the test dataset.

4.1.2 Implementation Details

Our implementation is based on PyTorch. We use the Adam optimizer with a learning rate of $1e - 4$. For a fair comparison, we use the same data augmentation strategy and post-processing step as [13]. To augment the training samples, we use random rotation, translation, shear, scale, and flip transformations. During the post-processing stage, we extract the saturated regions of the input image and superimpose it back onto the deflared image to recover the light source. In addition, our MFDNet is designed as a scalable model and the number of LP's decomposed layers n is related to the size

Table 1 Quantitative comparison on Flare7K real-world and synthetic datasets in terms of PSNR, SSIM, LPIPS. The best results are highlighted in bold.

Dataset\Method		Input	Zhang [41]	Sharma [30]	Wu [36]	Dai [13]	HINet [8]	MPRNet [38]	Restormer [37]	Uformer [35]	MFDNet (Ours)
Real-world	PSNR \uparrow	22.56	21.02	20.49	24.61	26.11	26.74	26.14	26.28	26.60	26.98
	SSIM \uparrow	0.857	0.784	0.826	0.871	0.879	0.882	0.878	0.883	0.892	0.895
	LPIPS \downarrow	0.078	0.174	0.112	0.060	0.055	0.048	0.050	0.054	0.051	0.051
Synthetic	PSNR \uparrow	22.77	21.04	20.01	27.88	29.07	29.97	29.87	29.45	30.13	30.79
	SSIM \uparrow	0.921	0.841	0.865	0.952	0.958	0.959	0.959	0.950	0.965	0.966
	LPIPS \downarrow	0.060	0.136	0.111	0.031	0.022	0.021	0.020	0.025	0.020	0.019

**Fig. 6** Visual comparison on Flare7K real-world nighttime flare-corrupted images.

of the input image. According to [20, 21, 23], increasing n reduces the computational burden but degrades the

performance. As the number of LP's decomposed layers n increases, the image size of the lowest frequency part



Fig. 7 Visual comparison on Flare7K real-world nighttime flare-corrupted images.

decreases, which can reduce the computational cost. However, as n increases, the amount of effective information in the lowest-frequency components diminishes, which impacts the low-frequency deflaring and the quality of the reconstructed image. Consequently, the model’s performance deteriorates with increasing depth. We set up $n = 3$ to get a trade-off between the computational cost and the performance.

4.1.3 Evaluation Metrics

We utilize various metrics to evaluate our MFDNet’s performance. Specifically, we measure the quality of nighttime flare removal using Structural Similarity (SSIM) [34], Peak Signal-to-Noise Ratio (PSNR), and Learned Perceptual Image Patch Similarity (LPIPS) [44]. Additionally, we measure the method’s practicality using GMACs, parameters, and inference time.

4.2 Comparison with the State-of-the-art Methods

We compare the performance of MFDNet with several state-of-the-art methods for nighttime flare removal in Flare7K real-world and synthetic datasets. These methods include a nighttime dehazing method [41], a nighttime visual enhancement method [30], two U-Net-based flare removal methods [13, 36], and HINet [8], MPRNet [38], Restormer [37], Uformer [35].

4.2.1 Qualitative Evaluation

We first compare the visual results of real-world nighttime flare removal in Figure 6 and Figure 7. The visualization results show that the flare-free images re-

covered by our MFDNet are closer to the ground-truth images. Subsequently, we provide a visual comparison of removing synthetic flares in Figure 8. Our MFDNet still restores cleaner flare-free images than those of the other algorithms. As shown in Figure 6, Figure 7, and Figure 8, the previous methods could not completely remove the flare, or alter the original color and texture details of the input image while removing the flare. Our method can remove the flare while preserving the original image information.

Specifically, to further prove the effectiveness and robustness of our method, we compare the flare removal results of our MFDNet and other state-of-the-art methods for different flare patterns. In the fourth row of Figure 6, the input image contains the glare artifact, in the sixth row of Figure 6, the input image contains the shimmer flare, in the second row of Figure 7, the input image contains the streak flare, in the fifth row of Figure 8, the input image contains colored lines. It can be seen that for these different flare patterns, the deflared image recovered through our method is closer to the ground truth. At the same time, our MFDNet can remove multiple flare patterns that exist within a single image, such as the third row of Figure 6, which contains both shimmer and glare artifacts, the first row of Figure 7, and the last row of Figure 8, which contain both streak and glare artifacts.

4.2.2 Quantitative Evaluation

Here, we perform a thorough quantitative evaluation of our MFDNet. Following the benchmark provided by [13], we use full-reference metrics PSNR, SSIM [34], and LPIPS [44] to measure the performance of different

Table 2 Comparison of the computational cost and time consumption on images from 512×512 resolution to 4K resolution. 'N.A.' denotes that the method cannot handle the input image of this size, and 'OOM' means that the method causes the out-of-memory issue for this specific resolution. The best results are highlighted in bold.

Method	512×512			1024×1024			1080p			2K			4K		
	GMACs (G)	Params (M)	Time (s)	GMACs (G)	Params (M)	Time (s)	GMACs (G)	Params (M)	Time (s)	GMACs (G)	Params (M)	Time (s)	GMACs (G)	Params (M)	Time (s)
Dai [13]	261.90	34.53	0.052	1047.61	34.53	0.195	N.A.	N.A.	N.A.	3682.99	34.53	0.686	OOM	OOM	OOM
HINet [8]	682.86	88.67	0.119	2731.43	88.67	0.448	N.A.	N.A.	N.A.	OOM	OOM	OOM	OOM	OOM	OOM
MPRNet [38]	274.97	1.93	0.101	1099.87	1.93	0.377	2175.04	1.93	0.760	OOM	OOM	OOM	OOM	OOM	OOM
Restormer [37]	57.61	2.98	0.106	230.46	2.98	0.394	455.74	2.98	0.771	OOM	OOM	OOM	OOM	OOM	OOM
Uformer [35]	160.09	20.43	0.137	640.35	20.43	0.536	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
MFDNet (ours)	18.25	6.32	0.034	73.01	6.32	0.075	144.53	6.32	0.153	256.66	6.32	0.301	578.12	6.32	0.783

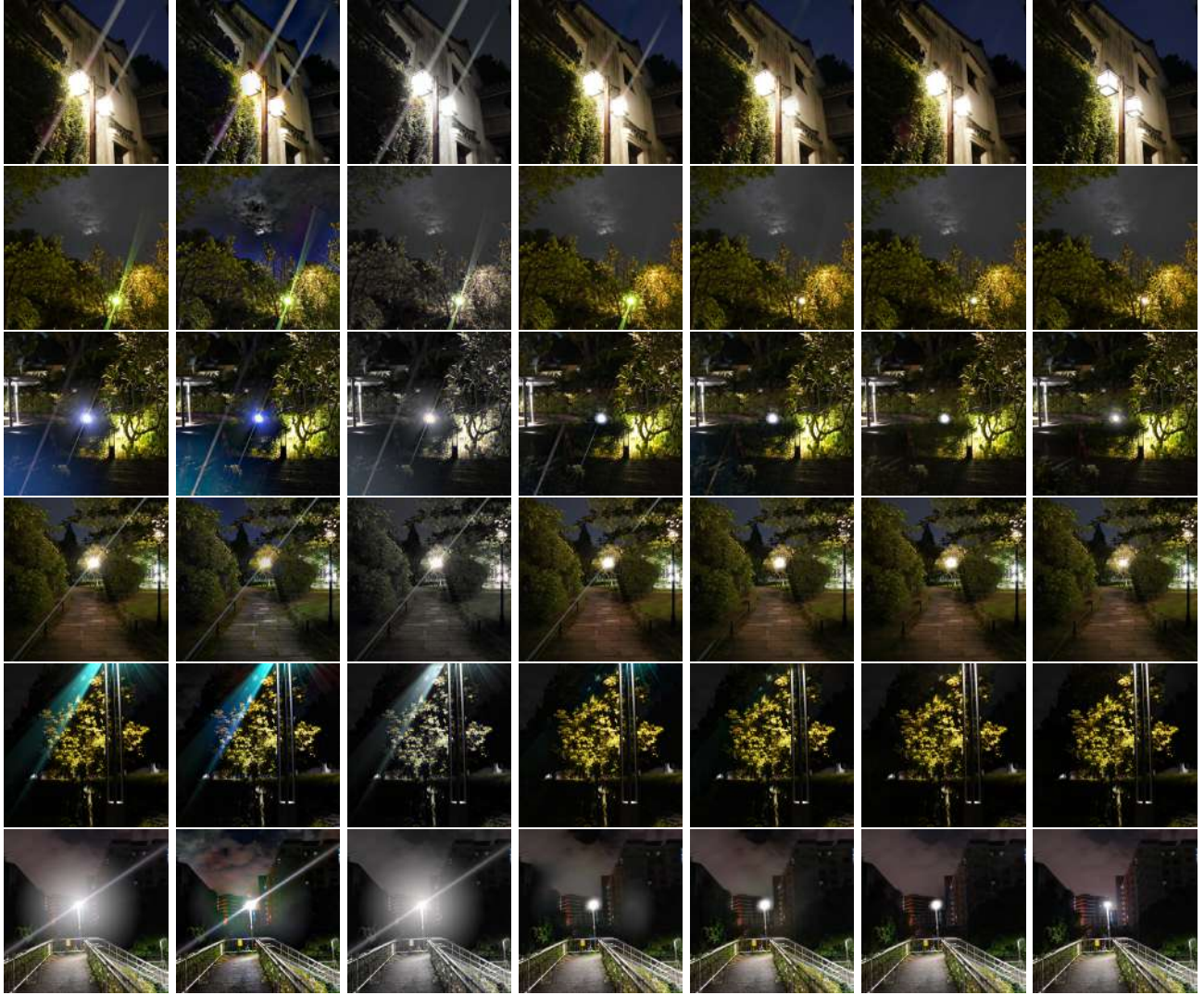


Fig. 8 Visual comparison on Flare7K synthetic nighttime flare-corrupted images.

methods. In order to better compare the comprehensiveness of the algorithm for removing various flares, we re-evaluate Uformer to make it capable of removing both reflected flares and scattered flares, to replace the original method in [13], which can only remove scattered flares. The results are reported in Table 1. Our MFDNet achieves the highest average PSNR and

SSIM than the previous state-of-the-art methods for real-world nighttime flare removal. For flare removal in synthetic nighttime flare-corrupted images, Our MFDNet performs better in terms of PSNR, SSIM, and LPISP than the previous state-of-the-art methods. Specifically, our MFDNet achieves 30.79 dB on PSNR, which is at least 0.66 dB higher than all other methods.



Fig. 9 Flare removal results of our MFDNet on 4K nighttime flare-corrupted images.

4.2.3 Efficiency Analysis

In Table 2, we compare the GMACs, parameters, and inference time of different methods on images from 512×512 resolution to 4K resolution. Each result in Table 2 is an average of 100 tests on an NVIDIA GTX 2080Ti GPU with 11G RAM, where the N.A. denotes that the method cannot handle the input image of this size, and OOM means that the method causes the out-of-memory issue for this specific resolution. As shown in Figure 1 and Table 2, our proposed MFDNet outperforms other methods by a significant margin in terms of inference time and GMACs, while also achieving superior deflare performance than others. Specifically, for 512×512 resolution images, the GMACs of our method are 18.25G, which is much smaller than the 57.61G of the second-best Restormer [37]. Our inference time is 0.034s, which is much smaller than the 0.052s of the second-best Dai [13]. Furthermore, our MFDNet can process images of any size from 512×512 to 4K resolution. MPRNet [38] and Uformer [35] can only handle 512×512 and 1024×1024 resolution images, and none of those state-of-the-art methods can handle 4K resolution images.

Currently, there are no high-resolution nighttime flare image datasets, and the Flare7K dataset that we used for training and testing is at 512×512 resolution. To evaluate our MFDNet’s performance on 4K

resolution images, we followed the pipeline of [13] by combining flare artifacts into real 4K nighttime images to generate 4K flare-corrupted images. The results of our method on 4K images are illustrated in Figure 9. Despite the use of 512×512 resolution images for our training, our MFDNet can still remove flare on 4K images. This demonstrates the effectiveness and efficiency of our method in removing flare by decoupling the illumination information and content information of the image.

Table 3 Ablation experiments of individual components in our MFDNet on Flare7K dataset.

Dataset\Network		w/o FETB+FRTB	w/o FECB+FDCB	w/o FAB	MFDNet
Real-world	PSNR \uparrow	26.36	24.94	26.70	26.98
	SSIM \uparrow	0.890	0.872	0.894	0.895
	LPIPS \downarrow	0.054	0.066	0.053	0.051
Synthetic	PSNR \uparrow	30.32	25.95	30.61	30.79
	SSIM \uparrow	0.963	0.925	0.964	0.966
	LPIPS \downarrow	0.021	0.046	0.021	0.019

4.3 Ablation Studies

We conduct ablation studies to evaluate the contributions of the different components of our proposed method. Specifically, we measure the contributions of the Feature Extraction Transformer Block (FETB) and the Feature Refinement Transformer Block (FRTB), the Feature Encoder Convolution Block (FECB) and



Fig. 10 Visual results for the ablation studies.

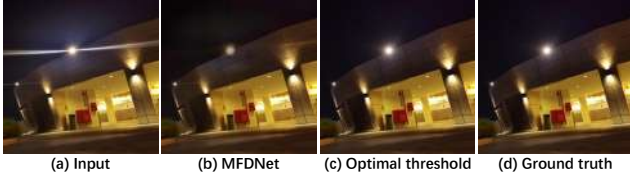


Fig. 11 Visual results for the limitation case. From left to right, the sequence includes the input flare image, the result of our MFDNet, our enhanced result after applying the optimal luminance threshold, and the ground truth image.

the Feature Decoder Convolution Block (FDCB), and the Feature Aggregation Block (FAB). The results from Table 3 and Figure 10 show that these components of our MFDNet are all effective for flare removal.

The Transformer block for global feature extraction and refinement can improve the deflare performance, because global information is crucial in accurately identifying the flare. Additionally, the convolution-based encoder-decoder structure provides substantial capabilities to remove flare artifacts and restore more image details. As can be seen from Table 3 and Figure 10, after removing the FECB and FEDB, the deflare effect of the model drops significantly, indicating that in the task of flare removal, it is very important to capture the detailed information of the input flare-corrupted image. Moreover, the FAB helps to integrate information from different frequency bands to reconstruct the final flare-free image.

4.4 Limitations

However, the proposed method exhibits some limitations that need to be addressed. Firstly, if the light source is diminutive and faint, accompanied by glare flare, the contrast between the light source and adjacent flare areas may not be distinctly pronounced. Under such circumstances, our approach might not fully restore the light source. To address this issue, we might manually determine an optimal luminance threshold for saturation during restoration as shown in Figure 11. In future work, we will endeavor to resolve this challenge by employing adaptive techniques to recover the light source. Secondly, our method may fail to completely

remove flare shimmers with very high brightness and extensive shapes, probably because they have obvious texture features in the high-frequency band. We intend to resolve these shortcomings in future studies.

5 Conclusion

We propose a highly efficient Multi-Frequency Deflare Network (MFDNet) for the nighttime flare removal task, which significantly reduces the computational burden when handling high-resolution images while simultaneously removing flare artifacts successfully. We disentangle illumination information from content information by decomposing the input image with the Laplacian Pyramid. We leverage Transformer’s ability to capture long-range pixel dependencies and convolutional neural networks’ capability to capture local features to remove flare comprehensively. We developed a hierarchical fusion reconstruction step to adaptively refine high-frequency components, to generate a final flare-free image. Dividing frequency bands to process flare-corrupted images can substantially decrease computational complexity. We demonstrate through extensive experiments that our method is superior and more efficient than other state-of-the-art nighttime flare removal methods.

References

1. Asha, C., Bhat, S.K., Nayak, D., Bhat, C.: Auto removal of bright spot from images captured against flashing light source. In: IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics, pp. 1–6. IEEE (2019)
2. Ba, J.L., Kiros, J.R., Hinton, G.E.: Layer normalization. ArXiv preprint [abs/1607.06450](https://arxiv.org/abs/1607.06450) (2016)
3. Boynton, P.A., Kelley, E.F.: Liquid-filled camera for the measurement of high-contrast images. In: Cockpit Displays X, vol. 5080, pp. 370–378 (2003)
4. Burt, P.J., Adelson, E.H.: The laplacian pyramid as a compact image code. In: Readings in computer vision, pp. 671–679. Elsevier (1987)
5. Chabert, F.: Automated lens flare removal (2015)
6. Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., Ma, S., Xu, C., Xu, C., Gao, W.: Pre-trained image processing transformer. In: CVPR, pp. 12,299–12,310 (2021)

7. Chen, L., Chu, X., Zhang, X., Sun, J.: Simple baselines for image restoration. In: ECCV, pp. 17–33 (2022)
8. Chen, L., Lu, X., Zhang, J., Chu, X., Chen, C.: Hinet: Half instance normalization network for image restoration. In: CVPR, pp. 182–192 (2021)
9. Chen, L., Zhang, J., Li, Z., Wei, Y., Fang, F., Ren, J., Pan, J.: Deep richardson–lucy deconvolution for low-light image deblurring. *International Journal of Computer Vision* pp. 1–18 (2023)
10. Chougule, A., Bhardwaj, A., Chamola, V., Narang, P.: Agd-net: Attention-guided dense inception u-net for single-image dehazing. *Cognitive Computation* pp. 1–14 (2023)
11. Criminisi, A., Pérez, P., Toyama, K.: Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on image processing* **13**(9), 1200–1212 (2004)
12. Cun, X., Pun, C., Shi, C.: Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan. In: AAAI, pp. 10,680–10,687 (2020)
13. Dai, Y., Li, C., Zhou, S., Feng, R., Loy, C.C.: Flare7k: A phenomenological nighttime flare removal dataset. In: NeurIPS (2022)
14. Faulkner, K., Kotre, C., Louka, M.: Veiling glare deconvolution of images produced by x-ray image intensifiers. In: Third International Conference on Image Processing and its Applications, 1989., pp. 669–673 (1989)
15. He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence* **37**(9), 1904–1916 (2015)
16. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: CVPR, pp. 7132–7141 (2018)
17. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NeurIPS, pp. 1106–1114 (2012)
18. Li, W., Lu, X., Lu, J., Zhang, X., Jia, J.: On efficient transformer-based image pre-training for low-level vision. In: IJCAI (2021)
19. Li, Y., Yan, Q., Zhang, K., Xu, H.: Image reflection removal via contextual feature fusion pyramid and task-driven regularization. *IEEE Transactions on Circuits and Systems for Video Technology* **32**(2), 553–565 (2022)
20. Li, Z., Chen, X., Pun, C.M., Cun, X.: High-resolution document shadow removal via a large-scale real-world dataset and a frequency-aware shadow erasing net. In: ICCV, pp. 12,415–12,424 (2023). DOI 10.1109/ICCV51070.2023.01144
21. Li, Z., Chen, X., Wang, S., Pun, C.M.: A large-scale film style dataset for learning multi-frequency driven film enhancement. In: IJCAI (2023)
22. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: ICCV, pp. 1833–1844 (2021)
23. Liang, J., Zeng, H., Zhang, L.: High-resolution photorealistic image translation in real-time: A laplacian pyramid translation network. In: CVPR, pp. 9392–9400 (2021)
24. Luo, Q., Liao, Y., Jing, B., Gao, X., Chen, W., Tan, K.: Hir-net: a simple and effective heterogeneous image restoration network. *Signal, Image and Video Processing* pp. 1–12 (2023)
25. Macleod, H.A., Macleod, H.A.: Thin-film optical filters. CRC press (2010)
26. Ragini, T., Prakash, K., Cheruku, R.: Detformer: A novel efficient transformer framework for image deraining. *Circuits, Systems, and Signal Processing* pp. 1–23 (2023)
27. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI, pp. 234–241. Springer (2015)
28. Sahoo, S., Nanda, P.K.: Adaptive feature fusion and spatio-temporal background modeling in kde framework for object detection and shadow removal. *IEEE Transactions on Circuits and Systems for Video Technology* **32**(3), 1103–1118 (2022)
29. Seibert, J.A., Nalcioğlu, O., Roeck, W.: Removal of image intensifier veiling glare by mathematical deconvolution techniques. *Medical physics* **12**(3), 281–288 (1985)
30. Sharma, A., Tan, R.T.: Nighttime visibility enhancement by increasing the dynamic range and suppression of light effects. In: CVPR, pp. 11,977–11,986 (2021)
31. Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., Li, Y.: Maxim: Multi-axis mlp for image processing. In: CVPR, pp. 5769–5780 (2022)
32. Vitoria, P., Ballester, C.: Automatic flare spot artifact detection and removal in photographs. *Journal of Mathematical Imaging and Vision* **61**(4), 515–533 (2019)
33. Wang, T., Zhang, K., Shen, T., Luo, W., Stenger, B., Lu, T.: Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In: AAAI, vol. 37, pp. 2654–2662 (2023)
34. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **13**(4), 600–612 (2004)
35. Wang, Z., Cun, X., Bao, J., Zhou, W., Liu, J., Li, H.: Uformer: A general u-shaped transformer for image restoration. In: CVPR, pp. 17,683–17,693 (2022)
36. Wu, Y., He, Q., Xue, T., Garg, R., Chen, J., Veeraraghavan, A., Barron, J.T.: How to train neural networks for flare removal. In: ICCV, pp. 2239–2247 (2021)
37. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H.: Restormer: Efficient transformer for high-resolution image restoration. In: CVPR, pp. 5728–5739 (2022)
38. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: CVPR, pp. 14,821–14,831 (2021)
39. Zha, Z., Yuan, X., Zhou, J., Zhu, C., Wen, B.: Image restoration via simultaneous nonlocal self-similarity priors. *IEEE Transactions on Image Processing* **29**, 8561–8576 (2020)
40. Zhang, D., Ouyang, J., Liu, G., Wang, X., Kong, X., Jin, Z.: Ff-former: Swin fourier transformer for nighttime flare removal. In: CVPR Workshops, pp. 2824–2832 (2023)
41. Zhang, J., Cao, Y., Zha, Z., Tao, D.: Nighttime dehazing with a synthetic benchmark. In: ACM MM, pp. 2355–2363 (2020)
42. Zhang, J., Wang, F., Zhang, H., Shi, X.: Compressive sensing spatially adaptive total variation method for high-noise astronomical image denoising. *The Visual Computer* pp. 1–13 (2023)
43. Zhang, M., Desrosiers, C.: High-quality image restoration using low-rank patch regularization and global structure sparsity. *IEEE Transactions on Image Processing* **28**(2), 868–879 (2019)
44. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: CVPR, pp. 586–595 (2018)
45. Zhang, Z., Feng, H., Xu, Z., Li, Q., Chen, Y.: Single image veiling glare removal. *Journal of Modern Optics* **65**(19), 2220–2230 (2018)