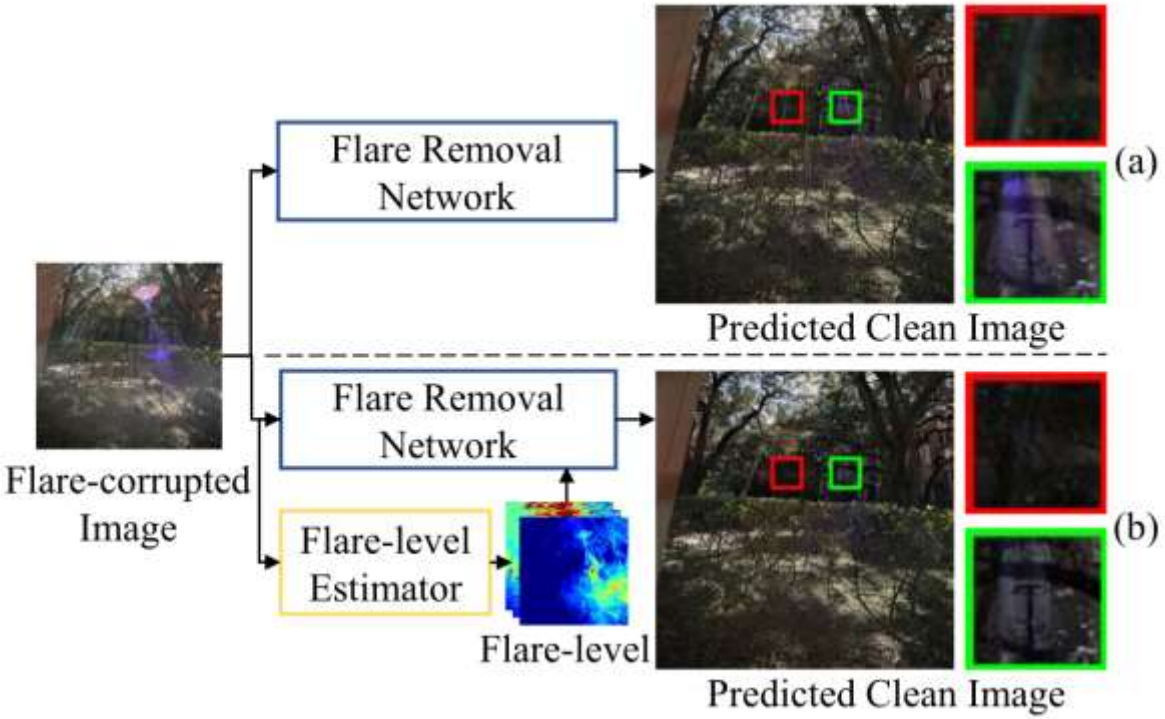
Toward Blind Flare Removal Using Knowledge-Driven Flare-Level Estimator

Haoyou Deng , Lida Li , Feng Zhang , Zhiqiang Li, Bin Xu,
Qingbo Lu, Changxin Gao, and Nong Sang

IEEE Transactions on Image Processing , 2024

Motivation

Due to the variety of flares, the local flare corruption level in real flare images differs from here to there. Obviously, an ideal network is required to remove flare artifacts in accordance with the local corruption level. However, no prior work explicitly models flare corruption levels for effective flare removal. In our view, there is still room to enhance the quality of flare removed images in this direction.



Contributions

• We focus on the blind property of flare artifacts in a corrupted image and propose a **flare-level estimator** to predict its corruption level to facilitate networks.

• We develop a flare-level modulator to integrate flare-level into networks and devise a **flare-aware block** for more accurate flare recognition and reconstruction.

• We construct a new flare dataset named **WiderFlare** with diverse flare-corrupted scenarios to provide a more comprehensive evaluation for flare removal method.

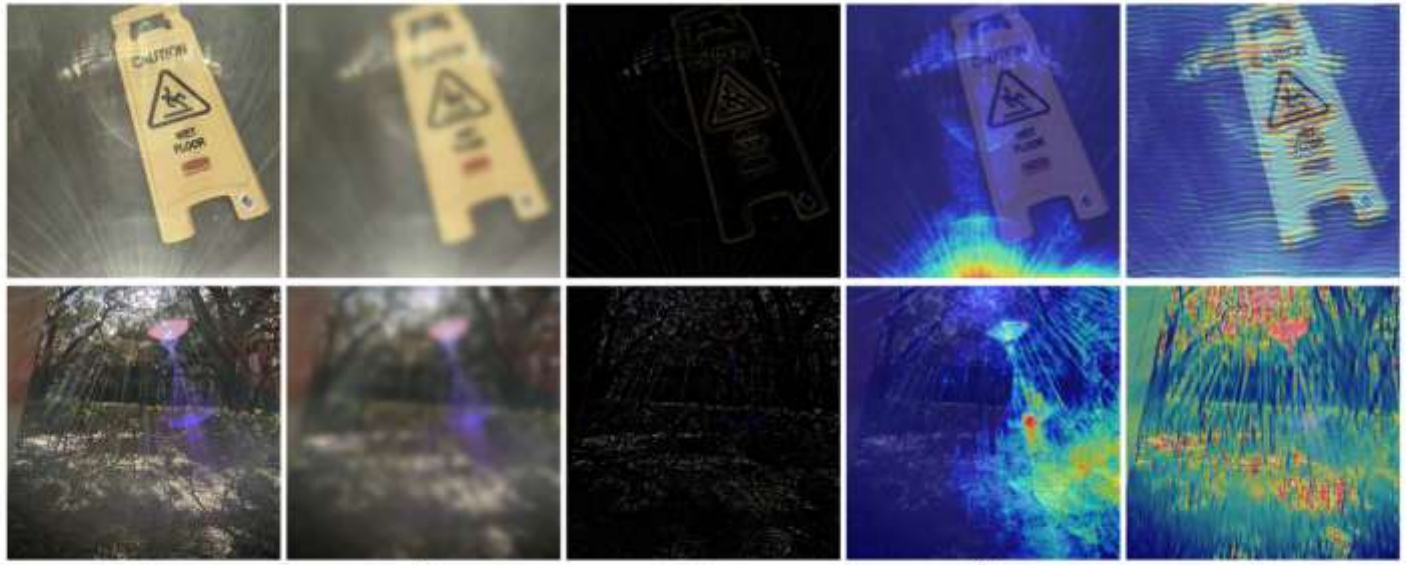
• We achieve state-of-the-art results for flare removal on two existing challenging benchmark datasets, as well as our newly constructed WiderFlare dataset.

Flare-Level Estimator--Flare-Level of Low-Frequency Component (FLFC)

By isolating the flare-corrupted images shown in Fig. 2(a) in the frequency domain, we extract the low-frequency component and high-frequency component of images in Fig. 2(b) and Fig. 2(c), respectively. Notably, Fig. 2(b) illustrates that most flares are encapsulated within the low-frequency component, characterized by small gradients. Furthermore, Fig. 2(c) presents that the high-frequency component contains residual flares that exhibit elevated saturation levels, such as streak, shimmer, ghost etc., displaying distinct shapes in contrast to the background.

We apply the gradient-stop function $\delta\tau(\cdot)$ with threshold τ to separate the part with a significant value of a gradient map $G \in R^{H\times W}$.

$$\delta_\tau(G_{i,j}) = \begin{cases} G_{i,j}, & if \ ||G_{i,j}||_2 \geq \tau \\ 0, & otherwise. \end{cases}$$



(a) (b) (c) (d) (e)

Fig. 2. Visualization of estimated flare-levels. Given (a) flare-corrupted images composed of (b) low-frequency component and (c) high-frequency component, both (d) the flare-level of low-frequency component and (e) the flare-level of high-frequency component can be predicted in our flare-level estimator.

Flare-Level Estimator--Flare-Level of High-Frequency Component (FHFC)

We estimate FHFC based on visual representation. We observe that the majority of high-frequency components are roughly in an elliptical shape. In this paper, we leverage some handcrafted filters with different kernels for template matching, including Gabor filters and anisotropic Gaussian filters.

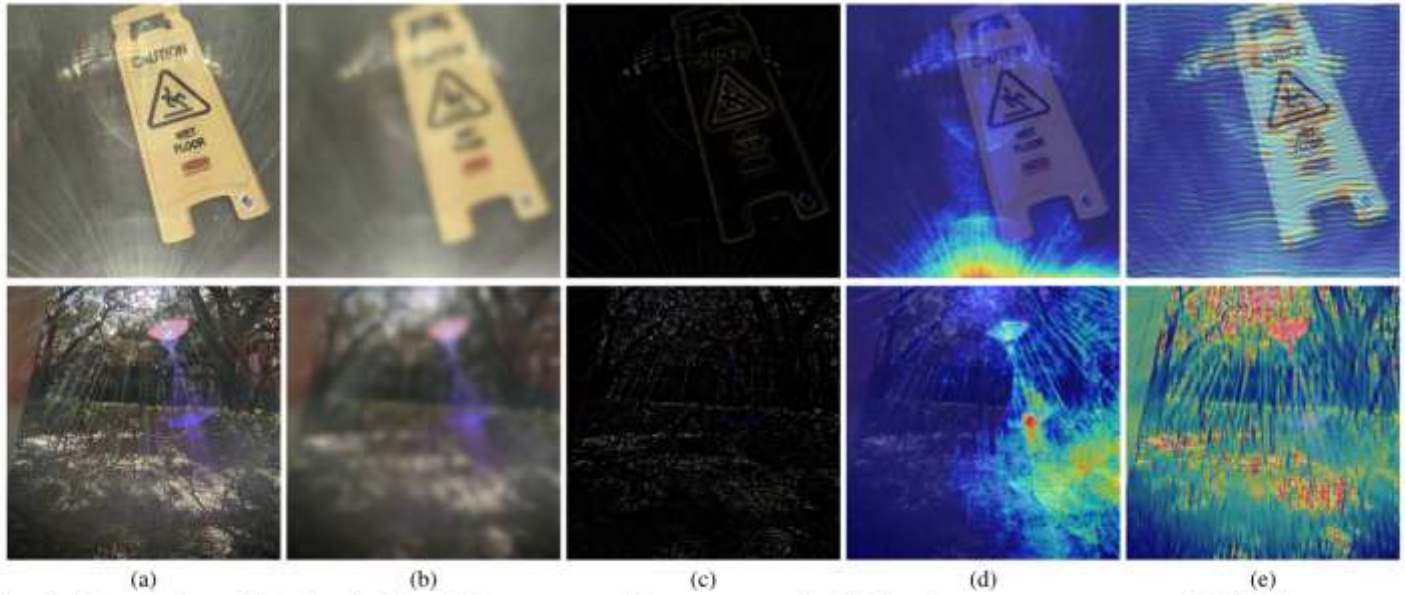


Fig. 2. Visualization of estimated flare-levels. Given (a) flare-corrupted images composed of (b) low-frequency component and (c) high-frequency component, both (d) the flare-level of low-frequency component and (e) the flare-level of high-frequency component can be predicted in our flare-level estimator.

Network Architecture--Flare-Level Modulator

Inspired by [1], we design a flare-level modulator to modulate the skip features using dynamic convolutions. The kernels of the dynamic convolutions are generated based on the flare-level estimations.

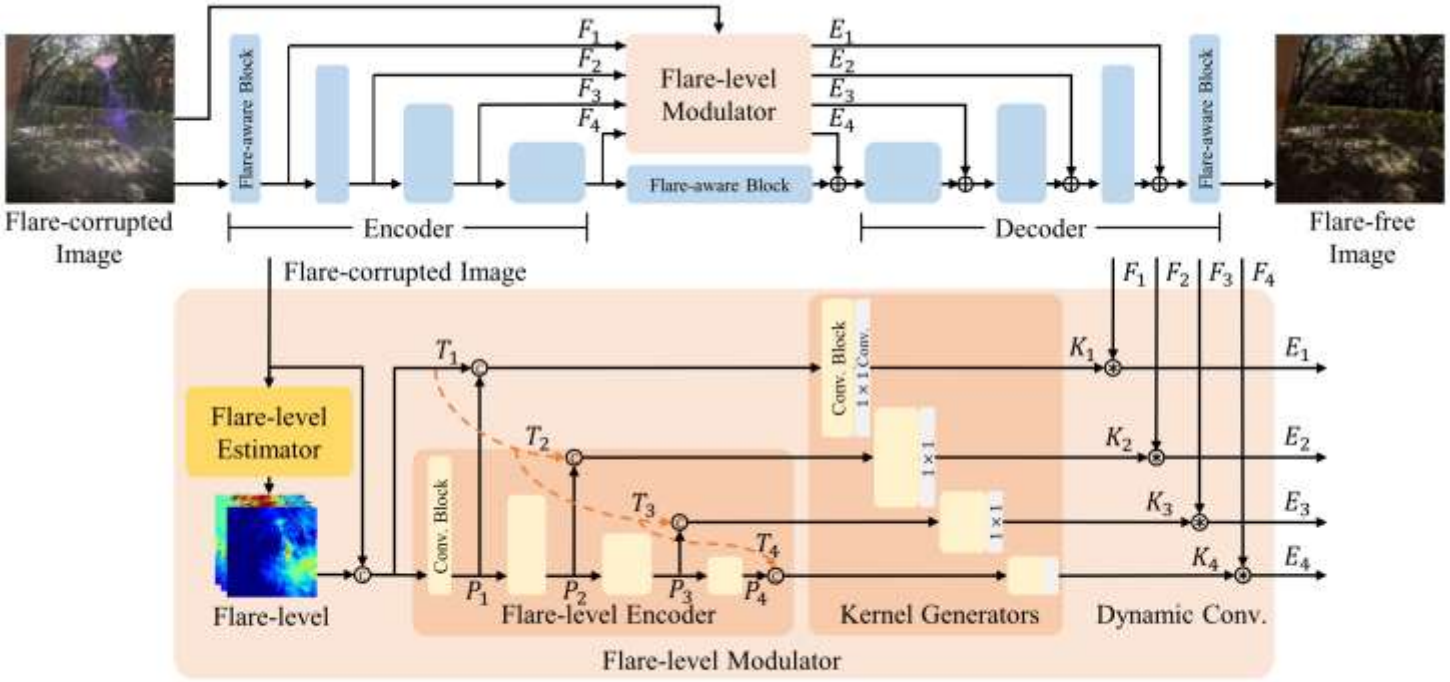


Fig. 3. The overview of our network architecture. The U-shaped restoration backbone consists of several flare-aware blocks stacked at each stage with skip feature maps modulated by the flare-level modulator.

[1] R. Feng, C. Li, H. Chen, S. Li, C. C. Loy, and J. Gu, "Removing diffraction image artifacts in under-display camera via dynamic skip connection network," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2021, pp. 662–671.

Network Architecture--Flare-Level Modulator

First, the flare-level estimator predicts the flare corruption level from the flare-corrupted image. Then, given the flare-level concatenated with the corrupted image as input T1, the flare-level encoder extracts scale-specific feature maps P1, P2, P3, and P4. These feature maps are concatenated with downsampled T1, T2, T3, and T4 in each scale and then fed into the corresponding kernel generators. Each kernel generator comprises a convolutional block similar to the flare-level encoder followed by a 1 × 1 convolution layer.

For each pixel of the features, the dynamic convolution adopts a position-aware s × s kernel weight for convolution calculation. These modulated features are then cast to the decoder of the restoration backbone at their corresponding stages.

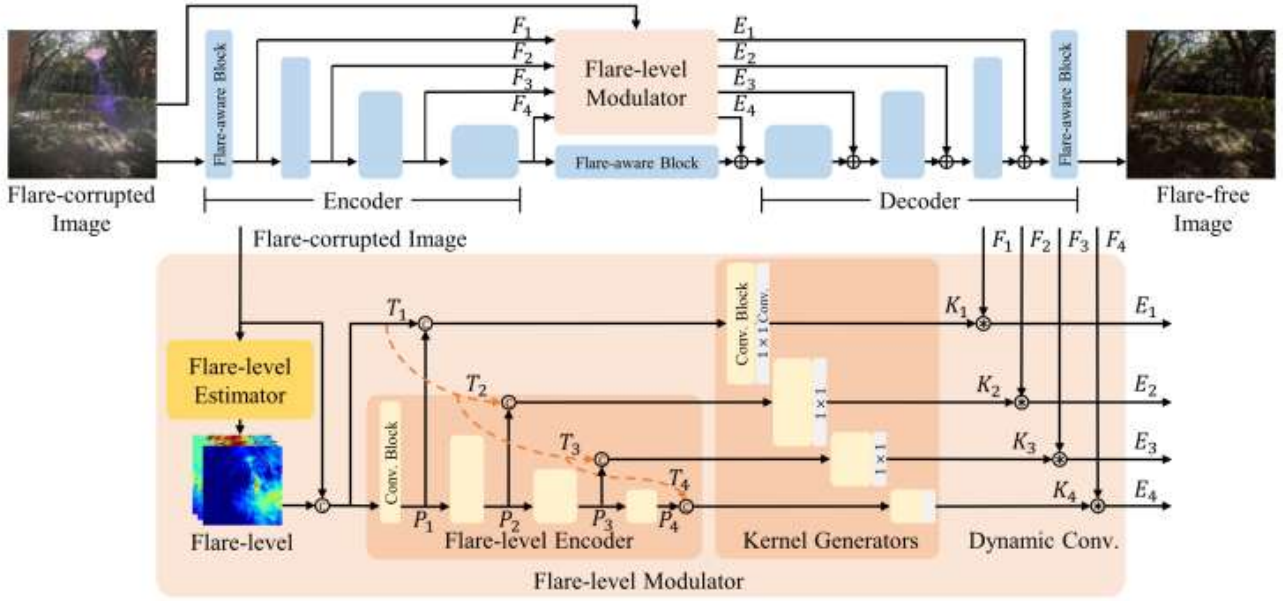


Fig. 3. The overview of our network architecture. The U-shaped restoration backbone consists of several flare-aware blocks stacked at each stage with skip feature maps modulated by the flare-level modulator.

Network Architecture--Flare-Aware Block

We insert SimAM between a Simple Gate and a convolution layer, where the Simple Gate is implemented as an activation function like ReLU. Adding a convolution layer and a depthwise convolution layer at the beginning produces the first residual block. Cascading another residual block formed by two convolution layers and a Simple Gate yields our flare-aware block. For the U-Net architecture, we replace all of its basic blocks with our flareaware blocks.

SimAM[1] is a 3D weights attention module based on the theory of neuroscience that an active neuron suppresses the activities of surrounding neurons.

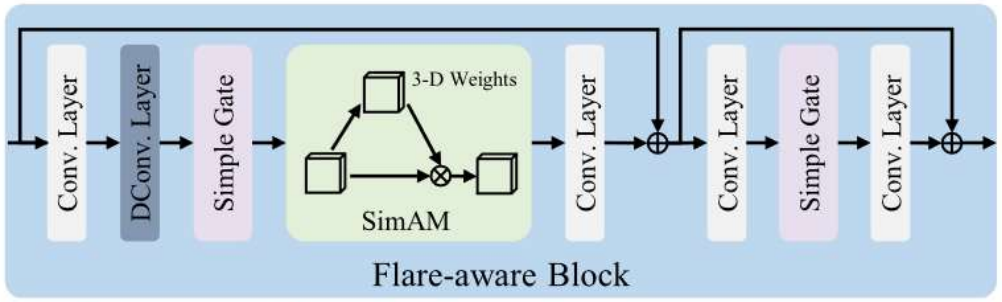


Fig. 4. Illustration of flare-aware block. The flare-aware block is composed of two residual blocks cascaded. One consists of a convolution layer, a depthwise convolution layer, a Simple Gate, a SimAM and a convolution layer in line. The other is formed by two convolution layers and a Simple Gate.

Loss Function

$$\mathcal{L}_{Total} = \mathcal{L}(\hat{I}, I_0) + \mathcal{L}(\hat{F}, F), \quad \mathcal{L} = \mathcal{L}_1 + \alpha \mathcal{L}_{LPIPS},$$

[1] L. Yang, R. Zhang, L. Li, and X. Xie, “SimAM: A simple, parameter free attention module for convolutional neural networks,” in Proc. 38th Int. Conf. Mach. Learn., vol. 139, 2021, pp. 11863–11874.

WIDERFLARE DATASET

To construct a comprehensive flare removal benchmark, we collect a new dataset in the wild using action cameras as well as camera drones in both consumer and enterprise lines for subjective evaluation, namely WiderFlare. The equivalent focal length of lenses used for data collection ranges from 24mm to 405mm, and the apertures range from f/2.8 to f/11, satisfying various settings of daily photography. In total, we obtain 500 real-world flarecorrupted images.

we capture pure flare images in a dark laboratory where we set up a professional sun lamp as the light source and adjust its luminance to mimic as many light conditions as possible. Different lenses of action cameras and drones are used to collect pure flare images, including wide lenses, zoom lenses, etc. Finally, we obtain 800 pure-flare images.

(a) Example of flare-corrupted images

(b) Example of pure-flare images

Fig. 6. Example images in our WiderFlare dataset. The WiderFlare dataset contains diverse real-world scenarios with diverse flare artifacts in (a). Additionally, (b) presents some pure flares in WiderFlare dataset.

TABLE I

THE COMPARISON OF REAL-WORLD TEST SUBSET BETWEEN DAYTIME FLARE DATASET [21], FLARE7K [22], FLARE7K++ [23] AND WIDERFLARE. NOTABLY, SINCE FLARE7K++ [23] SHARES THE SAME REAL TEST SUBSET AS FLARE7K [22], WE COMBINE THEM FOR DISCUSSION

Dataset	Statistics		Corruption level		Capture time			Flare artifact diversity		
	Number	Device types	Low	High	Daytime	Dusk	Nighttime	Light source	Green spot	Colorful ghost
Daytime flare dataset [21]	20	1	✗	✓	✓	✗	✗	✗	✓	✓
Flare7K [22] (Flare7K++ [23])	100	1	✗	✓	✗	✗	✓	✓	✗	✗
WiderFlare (ours)	**500**	4	✓	✓	✓	✓	✓	✓	✓	✓

Experiments--Results on Daytime Flare Dataset

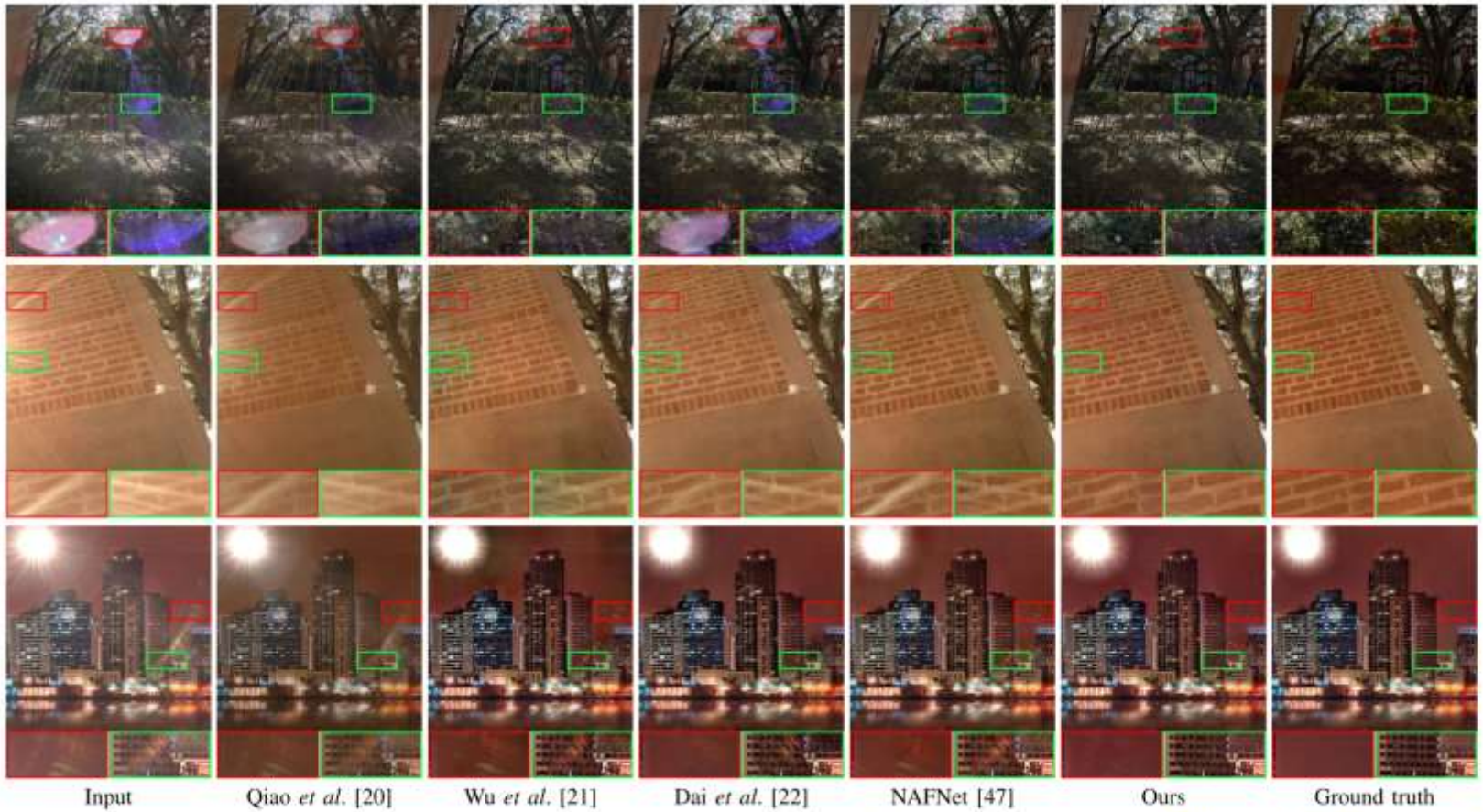


Fig. 7. Visual comparison on daytime flare dataset. From left to right are flare-corrupted inputs, the results of Qiao et al. [20], Wu et al. [21], Dai et al. [22], NAFNet [47], our method, and ground truth. From the zoomed boxes, our method can effectively remove diverse flare artifacts, such as ghosts and shimmers.

Experiments--Results on Daytime Flare Dataset

TABLE II

QUANTITATIVE COMPARISON ON DAYTIME FLARE DATASET.THE BEST RESULTS ARE SHOWN IN BOLD.† INDICATES OFFICIAL RESULTS FROM ORIGINAL PAPERS AND ‡ INDICATES OUR IMPLEMENTATION. - INDICATES THAT THE RESULTS ARE NOT PROVIDED IN THE ORIGINAL PAPERS

Method	GFLOPs	Params (M)	Real			Synthetic		
			PSNR(↑)	SSIM(↑)	LPIPS(↓)	PSNR(↑)	SSIM(↑)	LPIPS(↓)
Qiao *et al.* [20]	280.91	12.17	21.07	0.749	0.174	20.67	0.762	0.182
Wu *et al.* [21] †	219.17	31.03	25.55	**0.850**	-	30.37	**0.944**	-
Wu *et al.* [21] ‡	219.17	31.03	24.89	0.827	0.060	32.66	0.925	0.018
Dai *et al.* [22]	164.36	20.60	25.82	0.836	0.055	34.55	0.928	0.014
NAFNet [47]	63.88	67.06	25.90	0.837	0.056	34.24	0.929	0.014
NAFNet ×3 [47]	189.29	79.55	25.94	0.838	0.052	34.26	0.931	0.013
HINet [50]	171.12	22.18	25.23	0.832	0.059	33.46	0.922	0.015
MIRNet [51]	204.68	2.11	24.71	0.833	0.064	33.51	0.927	0.015
Restormer [4]	146.36	4.87	25.51	0.841	0.058	35.49	0.936	0.010
Ours	188.20	24.63	**26.63**	0.842	**0.048**	**36.64**	0.941	**0.009**

Experiments--Results on Flare7K Dataset

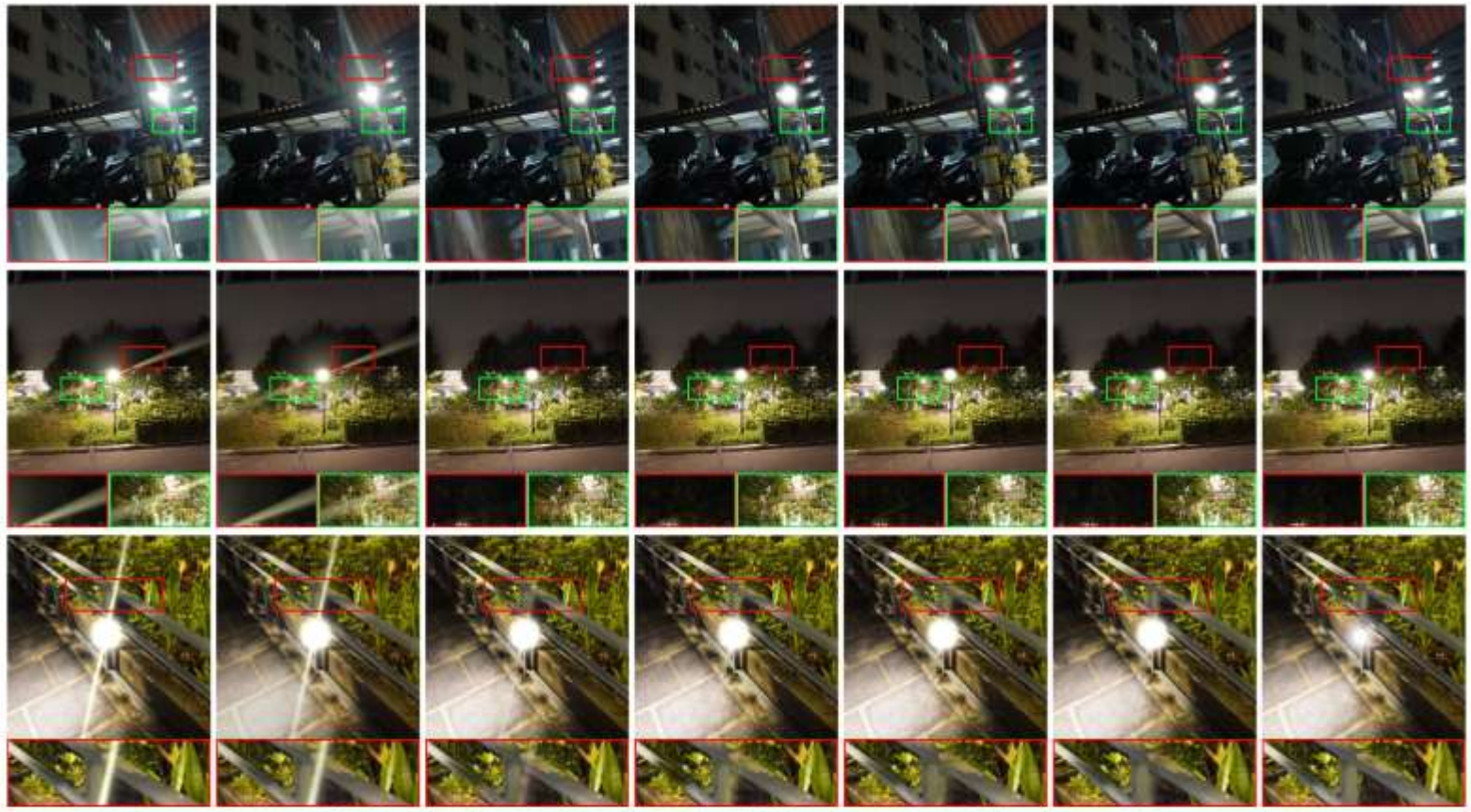


Fig. 8. Visual comparison on Flare7K dataset. From left to right are flare-corrupted inputs, the results of Qiao et al. [20], Wu et al. [21], Dai et al. [22], NAFNet [47], our method, and ground truth. Guided by our FHFC flare-level estimator, our method produces clean and natural outputs, whereas other methods produce artifacts at the boundary of shimmers.

Experiments--Results on Flare7K Dataset

TABLE III

QUANTITATIVE COMPARISON ON FLARE7K DATASET.THE BEST RESULTS ARE SHOWN IN BOLD. † INDICATES OFFICIAL RESULTS FROM ORIGINAL PAPERS AND ‡ INDICATES OUR IMPLEMENTATION

Method	GFLOPs	Params (M)	Real			Synthetic		
			PSNR(↑)	SSIM(↑)	LPIPS(↓)	PSNR(↑)	SSIM(↑)	LPIPS(↓)
Qiao *et al.* [20]	280.91	12.17	22.61	0.848	0.078	22.89	0.896	0.060
Wu *et al.* [21]	219.17	31.03	25.58	0.886	0.058	29.12	0.957	0.022
Dai *et al.* [22] †	164.36	20.60	26.98	0.890	0.047	30.47	0.965	0.017
Dai *et al.* [22] ‡	164.36	20.60	26.39	0.889	0.049	30.36	0.965	0.018
NAFNet [47]	63.88	67.06	25.64	0.889	0.060	29.32	0.960	0.025
NAFNet ×3 [47]	189.29	79.55	25.88	0.890	0.057	30.13	0.964	0.022
HINet [50]	171.12	22.18	26.13	0.889	0.055	30.24	0.961	0.020
MIRNet [51]	204.68	2.11	25.26	0.887	0.053	29.64	0.961	0.024
Restormer [4]	146.36	4.87	26.30	0.895	0.048	30.06	0.963	0.020
Ours	188.20	24.63	**27.05**	**0.901**	**0.045**	**30.52**	**0.966**	**0.016**

Experiments-- Results on WiderFlare Dataset

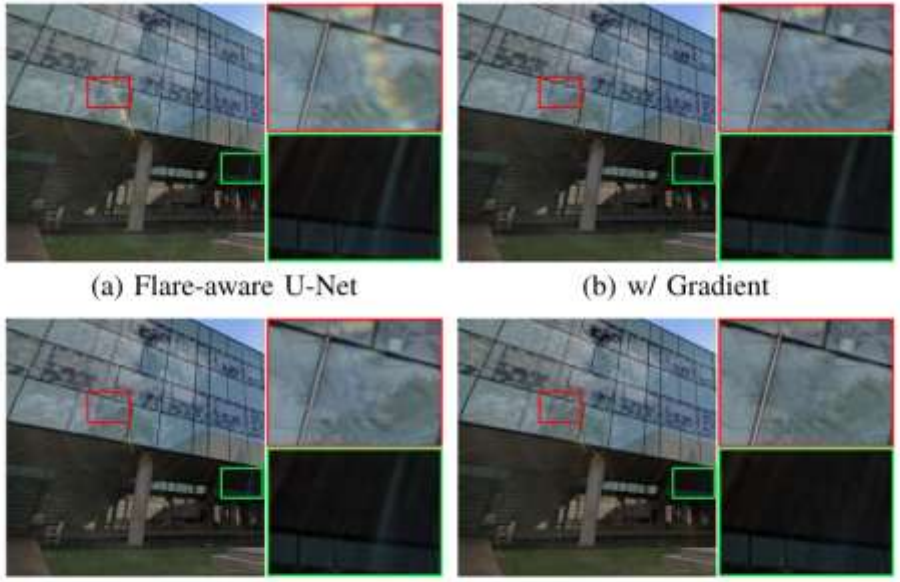
Fig. 9. Visual comparison on WiderFlare dataset. From left to right are flare-corrupted inputs, the results of Qiao et al. [20], Wu et al. [21], Dai et al. [22], NAFNet [47], and our method. With the knowledge-driven flare-level estimator, our method can effectively address a wide range of real-world flares and restore scene details.

Ablation Study -- Effect of Flare-Level Estimator

TABLE IV

EFFECT OF FLARE-LEVEL ESTIMATOR AND FLARE-AWARE BLOCK ON DAYTIME FLARE DATASET. THE STUDY IS PERFORMED IN PLAIN U-NET AND FLARE-AWARE U-NET. THE BEST RESULTS ARE SHOWN IN BOLD

Method	Real		Synthetic	
	PSNR(↑)	SSIM(↑)	PSNR(↑)	SSIM(↑)
Plain U-Net	24.89	0.827	32.66	0.925
w/ Gradient	25.23	0.829	33.42	0.923
w/ Shape	25.24	0.830	33.47	0.925
w/ Gradient + Shape	25.32	0.831	33.95	0.926
Flare-aware U-Net	26.11	0.833	36.06	0.937
w/ Gradient	26.24	0.837	36.42	0.939
w/ Shape	26.26	0.839	36.20	0.938
w/ Gradient + Shape	**26.63**	**0.842**	**36.64**	**0.941**



(a) Flare-aware U-Net (b) w/ Gradient

(c) w/ Shape (d) w/ Gradient + Shape

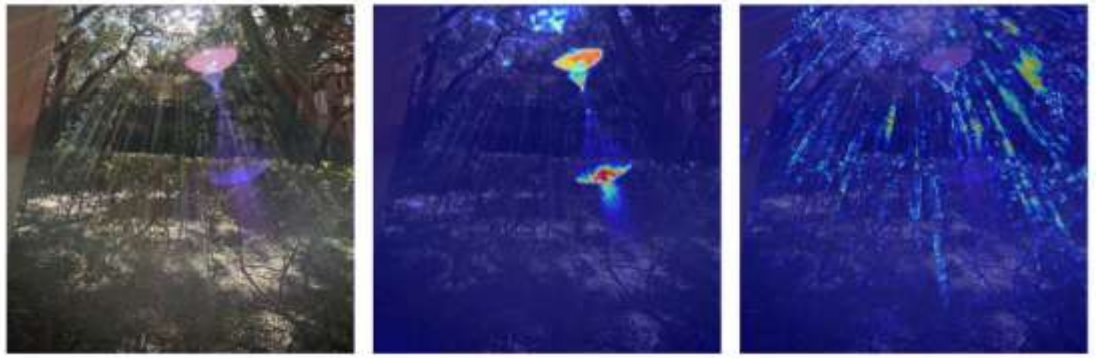
Fig. 10. Visualization of effect of flare-level estimator. With the gradient and shape flare-levels, the rainbow-like flare and shimmer are removed clearly.

Experiments-- Effect of Flare-Aware Block

TABLE IV

EFFECT OF FLARE-LEVEL ESTIMATOR AND FLARE-AWARE BLOCK ON DAYTIME FLARE DATASET. THE STUDY IS PERFORMED IN PLAIN U-NET AND FLARE-AWARE U-NET. THE BEST RESULTS ARE SHOWN IN BOLD

Method	Real		Synthetic	
	PSNR(↑)	SSIM(↑)	PSNR(↑)	SSIM(↑)
Plain U-Net	24.89	0.827	32.66	0.925
w/ Gradient	25.23	0.829	33.42	0.923
w/ Shape	25.24	0.830	33.47	0.925
w/ Gradient + Shape	25.32	0.831	33.95	0.926
Flare-aware U-Net	26.11	0.833	36.06	0.937
w/ Gradient	26.24	0.837	36.42	0.939
w/ Shape	26.26	0.839	36.20	0.938
w/ Gradient + Shape	**26.63**	**0.842**	**36.64**	**0.941**



(a) Network Input (b) Weight 1 (c) Weight 2

Fig. 11. Visualization of attention weights in flare-aware block. Given the flare-corrupted input as (a), activation maps of the 3D attention are visualized in (b) and (c). The color ranging from blue to red represents the value from small to large.