MIPI 2023 Challenge on Nighttime Flare Removal: Methods and Results

Yuekun Dai   Chongyi Li   Shangchen Zhou   Ruicheng Feng   Qingpeng Zhu
Qianhui Sun   Wenxiu Sun   Chen Change Loy   Jinwei Gu   Shuai Liu   Hao Wang
Chaoyu Feng   Luyang Wang   Guangqi Shao   Chenguang Zhang   Xiaotao Wang
Lei Lei   Dafeng Zhang   Xiangyu Kong   Guanqun Liu   Mengmeng Bai
Jia Ouyang   Xiaobing Wang   Jiahui Yuan   Xinpeng Li   Chengzhi Jiang
Ting Jiang   Wenjie Lin   QiWu   Mingyan Han   Jinting Luo   Lei Yu
Haoqiang Fan   Shuaicheng Liu   Bo Yan   Zhuang Li   Yadong Li   Hongbin Wang
Soonyong Song   Minghan Fu   Rayyan Azam Khan   Fangxiang Wu   Zhao Zhang
Suiyi Zhao   Huan Zheng   Yangcheng Gao   Yanyan Wei   Jiahuan Ren   Bo Wang
Yan Luo   Shuaibo Gao   Wenhui Wu   Sicong Kang   Nikhil Akalwadi
Ankit Raichur   Vinod Patil   Allabakash G   Swaroop A   Amogh Joshi
Chaitra Desai   Ramesh Ashok Tabib   Ujwala Patil   Uma Mudenagudi Sicheng Li
Ruoxi Zhu   Jiazheng Lian   Shusong Xu   Zihao Liu   Sabari Nathan   Priya Kansal

Abstract

Developing and integrating advanced image sensors with novel algorithms in camera systems are prevalent with the increasing demand for computational photography and imaging on mobile platforms. However, the lack of high-quality data for research and the rare opportunity for in-depth exchange of views from industry and academia constrain the development of mobile intelligent photography and imaging (MIPI). With the success of the 1st MIPI Workshop@ECCV 2022, we introduce the second MIPI challenge including four tracks focusing on novel image sensors and imaging algorithms. In this paper, we summarize and review the Nighttime Flare Removal track on MIPI 2023. In total, 120 participants were successfully registered, and 11 teams submitted results in the final testing phase. The developed solutions in this challenge achieved state-of-the-art performance on Nighttime Flare Removal. A detailed description of all models developed in this challenge is provided in this paper. More details of this challenge and the link to the dataset can be found at https://mipi-challenge.org/MIPI2023/.

1. Introduction

Lens flare is a common optical phenomenon that occurs when intense light is scattered or reflected within a lens system, resulting in a distinctive radial-shaped bright area and light spots in captured photos. In mobile platforms such as monitor lenses, smartphone cameras, UAVs, and autonomous driving cameras, daily wear and tear, fingerprints, and dust can function as a grating, exacerbating lens flare and making it particularly noticeable at night. Thus, flare removal algorithms are highly desired

Flares can be categorized into three main types: scattering flares, reflective flares, and lens orbs. In this competition, we mainly focus on removing the scattering flares, as they are the most prevalent type of nighttime image degradation. Early attempts at scattering flare removal were made by Wu et al. [24], who proposed a dataset with physically-based synthetic flares and flare photos taken in a darkroom. However, this approach did not perform well in nighttime situation. To address this issue, Dai et al. [3, 4] created a new synthetic dataset specifically designed for nighttime scenes. This challenge is based on the subset of [3] and aims to restore the flares-corrupted images with different complicated degradations. Further details will be discussed in the following sections.

We hold this challenge in conjunction with the second MIPI Challenge which will be held on CVPR 2023. Similar to the first MIPI challenge [6, 18, 26, 27, 25], we are seeking an efficient and high-performance image restoration algorithm to be used for recovering flare corrupted images. MIPI 2023 consists of four competition tracks:

- **RGB+ToF Depth Completion** uses sparse and noisy ToF depth measurements with RGB images to obtain a

Yuekun Dai[1] (ykdai005@ntu.edu.sg), Chongyi Li[1] (chongyi.li@ntu.edu.sg), Shangchen Zhou[1], Ruicheng Feng[1], Qingpeng Zhu[3], Qianhui Sun[2], Wenxiu Sun[3], Chen Change Loy[1], Jinwei Gu[2,3] are the MIPI 2023 challenge organizers ([1] Nanyang Technological University, [2] SenseBrain, [3] SenseTime Research and Tetras.AI). The other authors participated in the challenge. Please refer to Appendix for details.
MIPI 2023 challenge website: https://mipi-challenge.org/MIPI2023/

complete depth map.

- **RGBW Sensor Fusion** fuses Bayer data and a monochrome channel data into Bayer format to increase SNR and spatial resolution.

- **RGBW Sensor Re-mosaic** converts RGBW RAW data into Bayer format so that it can be processed by standard ISPs.

- **Nighttime Flare Removal** is to improve nighttime image quality by removing lens flare effects.

2. MIPI 2023 Nighttime Flare Removal

To facilitate the development of efficient and high-performance flare removal solutions, we provide a high-quality dataset to be used for training and testing and a set of evaluation metrics that can measure the performance of developed solutions. This challenge aims to advance research on nighttime flare removal.

2.1. Datasets

In this competition, we will provide a synthetic flare dataset with 5,000 scattering flare images [3] in $1440{\times}1440{\times}3$. These images can be added to the provided flare-free background to synthesize paired data for training. We will provide detailed annotations for each component of the synthetic flare, including streak, glare, and light source. To help the participants evaluate the performance of their models, we provide 100 real captured flare-corrupted/flare-free validation image pairs. Finally, models will be evaluated on 100 extra captured test image pairs.

2.2. Evaluation

In this competition, we use the standard Peak Signal To Noise Ratio (PSNR) as our evaluation metrics. The restoration results in the glare and streak region will be evaluated by using glare region PSNR (G-PSNR) and streak region PSNR (S-PSNR). Since the ground truth images can not offer perfect light source image without any flares, we only compute the PSNR for the region without the light source in this challange. Finally, global PSNR is calculated in the region without light source. These three metrics will be averaged to get the final score. Participants can view their score, G-PSNR, S-PSNR of their submission to optimize the model's performance on scattering flares' different components.

2.3. Challenge Phase

The challenge consisted of the following phases:

1. Development: The registered participants get access to the data and baseline code, and are able to train the models and evaluate their running time locally.

2. Validation: The participants can upload their models to the remote server to check the fidelity scores on the validation dataset, and to compare their results on the validation leaderboard.

3. Testing: The participants submit their final results, code, models, and factsheets.

3. Challenge Results

Among 120 registered participants, 11 teams successfully submitted their results, code, and factsheets in the final test phase. Table 1 reports the final test results and rankings of the teams. Only three teams train their models with extra data, and several top-ranked teams apply ensemble strategies (self-ensemble, model ensemble, or both). The methods evaluated in Table 1 are briefly described in Section 4 and the team members are listed in Appendix.

Finally, the MiAlgo team is the first place winner of this challenge, while Samsung Research China - Beijing (SRC-B) team win the second place and MegFR team is the third place, respectively. Because the quantitative metrics of the competition can not perfectly reflect everyone's results, we also conducted an additional user study with 20 people. In this user study, we choose all test results for the teams in top 5 and complete the light source for these results. We ask the user to pick the best flare-removed images among top 5 results. Based on the result of user study, we present the **Best Visualization Award** to the team of Samsung Research China - Beijing.

4. Methods

MiAlgo. For the nighttime flare removal task, this team proposes a two-stage network(see Figure 1) that consists of an erasing module and an inpainting module. In the first stage, borrowing from [3], they use the Uformer[20] as the erasing module, aiming to remove the flares as much as possible. In the second stage, they use the AOT-GAN[29] as the inpainting module to restore the details on the flare area, the input is the original flare image concatenated with the output of erasing module. The output of the inpainting module has excellent visualization effects, but its PSNR is relatively low. They figure out that the ground-truth image is not entirely flare-free, so they blend the output of the inpainting module with the original flare input to get the final output, which improves the PSNR by 0.4 dB. Regarding data generation, they collect several nighttime images as base images and randomly adjust the colors of the flares to blue, yellow, and white, which are common on the ground. They also add light and local haze to the base image to simulate the more realistic flare-corrupted images. The data augmentation uses very common methods such as random affine, color jitter, etc.

Table 1. Results of MIPI 2023 challenge on nighttime flare removal. 'Runtime' for per image is tested and averaged across the validation datasets, and the image size is 512×512. 'Params' denotes the total number of learnable parameters.

| Team Name	User Name	Metric			Params (M)	Runtime (s)	Platform	Extra data	Ensemble
		Score	S-PSNR	G-PSNR					
MiAlgo	mialgo_ls	$29.44_{(1)}$	$28.59_{(1)}$	$28.89_{(1)}$	65.56	2.2	Nvidia RTX 3090	Yes	self-ensemble + model
SRC-B	xiaozhazha	$29.16_{(2)}$	$28.21_{(2)}$	$28.59_{(2)}$	23.56	2.0	Nvidia A100	Yes	self-ensemble + model
MegFR	jiangchengzhi, lxp0_0	$27.66_{(3)}$	$26.56_{(3)}$	$27.05_{(3)}$	20.4	0.168	Nvidia 2080Ti	-	self-ensemble
AntIns	Yjingyu	$26.03_{(4)}$	$24.04_{(6)}$	$25.68_{(5)}$	50.88	2.83	Nvidia Tesla V100	-	-
ActionBrain-ETRI	soony	$26.03_{(5)}$	$24.49_{(4)}$	$25.36_{(8)}$	41	0.24	Nvidia RTX A6000	-	-
USask-Flare	eason	$25.86_{(6)}$	$24.19_{(5)}$	$25.38_{(6)}$	5.3	0.64	Nvidia RTX 3090	-	-
LVGroup_HFUT	HuanZheng	$25.82_{(7)}$	$23.36_{(9)}$	$25.92_{(4)}$	29.16	0.03	Nvidia RTX 3090	-	-
szzzzz01	gsb	$25.72_{(8)}$	$24.00_{(7)}$	$25.26_{(9)}$	/	0.47	Nvidia A100	-	-
CEVI_Explorers	Lowlight_Hypnotise,Niksx	$25.52_{(9)}$	$23.40_{(8)}$	$25.26_{(10)}$	82.2	2.07	Nvidia RTX 3090	-	-
AI ISP	zrx	$25.46_{(10)}$	$23.17_{(10)}$	$25.38_{(7)}$	21.71	0.29	Nvidia A100	Yes	-
Couger AI	SabariNathan	$20.76_{(11)}$	$15.88_{(11)}$	$22.71_{(11)}$	1.6428	0.21	Nvidia Tesla V100	-	-
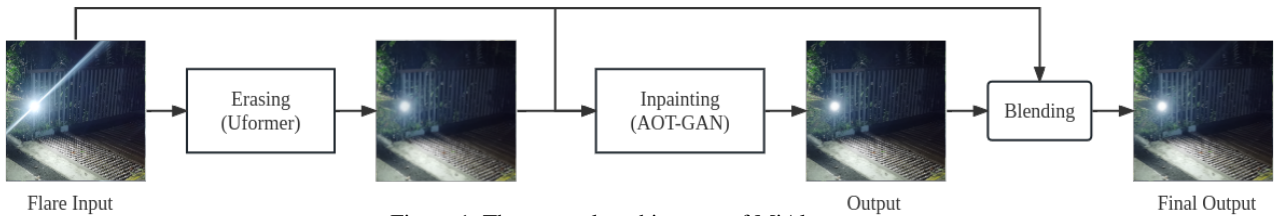
Figure 1. The network architecture of MiAlgo team.

When training, they first train the erasing module using synthetic data generated online for about 100,000 iterations. Then fix the weight of the erasing module and train the inpainting module for about 100,000 iterations. The initial learning rate is 1e-3 and they use cosine annealing to reduce the learning rate. The batch size is 16 and the patch size is 512. They train on 8 Tesla V100 GPUs for about 2 days. During testing, they use the self-ensemble(X8) and the model-ensemble(X4) strategy to get the highest PSNR.

SRC-B. In the process of removing nighttime flare, it is crucial to have a large receptive field because flare can occupy a substantial portion of an image, even potentially the entire image. However, the conventional window-based Transformer approaches restrict the receptive field within the window, limiting its ability to capture global features. And the flare can cause the dark regions to become brighter and result in a loss of contrast and alteration of the frequency characteristics of the image. To address these challenges, SRC-B introduced FF-Former (see Figure 2), which is based on Fast Fourier Convolution (FFC) and is designed to extract global frequency features for enhancing nighttime flare removal. To achieve this, SRC-B incorporates a Spatial Frequency Block (SFB) [30] after the Swin Transformer, which forms the Swin Fourier Transformer Block (SFTB). This configuration enables the establishment of long dependencies and the extraction of global features. Unlike the traditional Transformer, which relies on global self-attention, the SFB module only performs convolution computation, making it both effective and efficient.

SFTB is the basic module for our FF-Former, and the channel number C is 32 in the first SFTB of Encoder. We
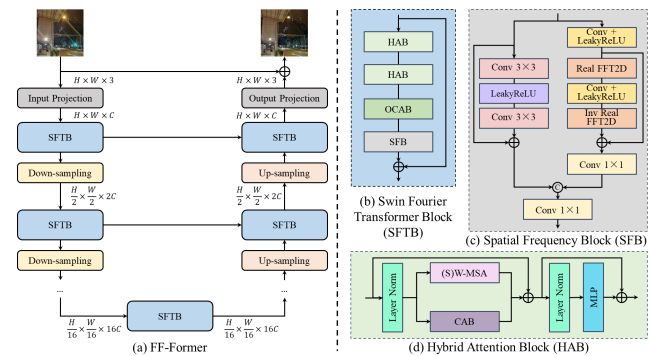

Figure 2. The network architecture of SRC-B team.

set 4 levels in the Encoder and Decoder module for extracting multi-scale features. Following Dai et al. [3], we crop the input flare-free and flare-corrupted images into 512×512 with batch size of 2 to train our FF-Former. We use the Adam with $\beta_1 = 0.9$ and $\beta_2 = 0.99$ to optimize the Charbonnier L1 loss function [9] for nighttime flare removal. The initial learning rate is 1e-4 and we use CosineAnnealingLR with 600,000 maximum iterations and 1e-7 minimum learning rate to adjust the learning rate. we also use horizontal and vertical flips for data enhancement.

MegFR. The team MegFR proposes a new data synthesis method that can synthesize flare images with distributions that are closer to real data and propose a new learning method based on signal separation to remove flare more finely.

Since it is difficult to obtain real Flare-corrupted images and flare-free image pairs, this team uses the officially provided flare images and background images to synthesize a

large number of image pairs for training. Given a flare-free image I_b as background, they first randomly select t from 5000 flare images and denote it as $F = \{f_1, \ldots, f_t\}$, in this paper t is set to 4. Firstly, they perform random inverse gamma correction on I_b and f_i:

$$
\begin{aligned}
I_b^g &= \gamma^{-1}(I_b; \theta^y), \\
f_i^g &= \gamma^{-1}(f_i; \theta_i^f),
\end{aligned}
\tag{1}
$$

where θ^y and θ_i^f are random numbers in $[1.8, 2.2]$, which represents the correction strength. Then in order to simulate more morphological flares, they do a Gaussian blur with random intensity on each flare:

$$
f_i^b = g(f_i^g; \theta_i^b),
\tag{2}
$$

where g represents Gaussian blur, and θ_i^b represents the size of the Gaussian kernel, which is a random number in $[5, 21]$. At the same time, in order to simulate the different brightness of the main light source and other light sources in the same scene, this team multiplies different gain values for different flares. Then after randomly offsetting the positions of all the flares, finally these flares are added to get the combination of multiple flares.

$$
f_c = \sum_{i=1}^{t} gain_i * \Phi(f_i^b),
\tag{3}
$$

where Φ means randomly offsetting the center of the flares, and $gain_i$ is a random number in $[0.8, 1.0]$. Finally, this team gets the flare-corrupted image I_x through:

$$
I_x = \gamma(I_b^g + f_c; \theta),
\tag{4}
$$

where γ is gamma correction, θ is a random number in $[1.8, 2.2]$. The corresponding flare-free image I_y and flare image I_f could be obtained in the following:

$$
\begin{aligned}
I_y &= \gamma(I_b^g; \theta), \\
I_f &= \gamma(f_c; \theta).
\end{aligned}
\tag{5}
$$

As shown in Fig. 3,after generating the data, this team uses a Uformer [19] ψ for signal separation:

$$
Y = \psi(I_x),
\tag{6}
$$

where $Y \in \mathbb{R}^{w \times h \times 6}$ is the result of signal separation, the first three channels Y_c are image content signals, and the last three channels Y_f are flare signals. This team uses L1 loss to supervise and train the ψ:

$$
loss_{L1} = |Y_c - I_y| + |Y_f - I_f|.
\tag{7}
$$

Perceptual loss can facilitate regularizing the network to produce images having more structural similarity with the
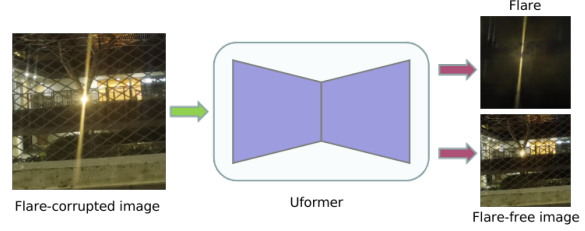

Figure 3. The pipeline of MegFR team.

ground truth [8, 10]. So this team further introduces perceptual loss into our training:

$$
loss_p = |f_p - f_g|^2,
\tag{8}
$$

where f_p and f_g respectively represent the features of Y_c and I_y. In particular, the features here refer to the features of the last convolution of an alexNet pre-trained on ImageNet.

The final loss is a combination of the two losses:

$$
loss = loss_{L1} + loss_p.
\tag{9}
$$

AntIns. The team AntIns proposes an effective nighttime flare removal pipeline for MIPI challenge. Firstly, they employed a strong image restoration model, Uformer[20], as a base model. Then, some improvements were made to the loss function, and an effective data augmentation method was proposed for the problem of the nighttime flare removal. Finally, through model fusion and test-time augmentation, the performance of the model was further improved and ultimately achieved competitive results in the challenge.

The team uses Uformer-B(Base) as a base model, which is trained from scratch. They synthesize training images using Flickr24K as background images and 5k scattering flare images from Flare7K. The team also designs a night data augmentation strategy (Night Data Aug) for the background images. As shown in Figure 4, the Night Data Aug has four modes, with one randomly selected for each image during training. The loss function includes both L1 loss and perceptual loss. Additionally, the team assign different weights to the L1 loss for areas inside and outside the flare, and the weight for areas inside the flare is 5, while outside the flare it is 1. The weight of perceptual loss is 1. The training samples use a patch size of 512, and the optimizer is AdamW [12] with an initial learning rate of 0.0002. The team trains for 85 epochs and find that longer training might improve the results. Ultimately, the proposed pipeline achieves a score of 26.03 on the test set of the MIPI challenge.

ActionBrain-ETRI. This team proposes a cascaded neural network to reduce distortions by flares in dark images. This network is designed to concatenate the identical Uformer models serially as shown in Figure 5. The

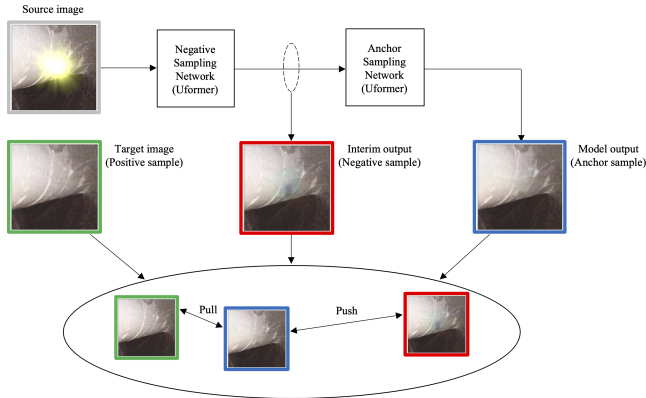Figure 4. Augmented image examples of Night Data Aug strategy.

Figure 5. The network architecture of ActionBrain-ETRI team.

proposed network assumes that the second part of the model is able to compensate for errors induced from the first one [23]. The Uformer model is provided as one of the baseline codes. The input shape of the model is $batchsize \times 3 \times 512 \times 512$, and the output shape is $batchsize \times 6 \times 512 \times 512$. The earlier 3 channels in the output shape are image predictions, and the rest 3 channels are mask predictions. This team uses image predictions only. Weights of the first model are configured by copying provided pre-trained checkpoint. Their gradients are fixed to get stable negative samples. The second model is set to be trainable, and it produced anchor samples. This team assumes that target images were dealt with positive samples. The proposed network is optimized by MSE L_{MSE}, triplet [16] $L_{triplet}$ and perceptual loss [10] L_{PL}:

$$
\begin{aligned}
&L_{ActionBrain-ETRI}(x_i^a, x_i^p, x_i^n) \\
&= \lambda \times L_{MSE}(x_i^a, x_i^p) + \delta \times L_{PL}(x_i^a, x_i^p) \quad (10)\\
&+ (1 - \lambda) \times L_{triplet}(x_i^a, x_i^p, x_i^n),
\end{aligned}
$$

where x_i^n, x_i^p, x_i^a are negative, positive, anchor samples of the i-th triplet, $\lambda = 0.6$, and $\delta = 0.001$.

To train the proposed network, dual GPUs (NVIDIA Quadro RTX A6000) are used. Then the models are trained using provided images and pre-trained checkpoints without extra resources. The proposed network has 40,892,876 parameters (20,446,438 trainable and 20,446,438 non-trainable parameters). Here, hyperparameters for learning rate, batch size, number of workers, and epochs are set to 0.0002, 6, 6, and 50 respectively. At each training and testing step, time is spent around 2900 and 24 seconds per epoch. Based on calculations, the inference time per image is estimated to be 0.24 seconds, given that the number of target images is 100.

USask-Flare. This team proposes a novel Light source guided Spatial transformer Generative Adversarial Network, namely LS-GAN. The overall network pipeline is shown in Figure 6. LS-GAN follows a U-Shaped structure, which has an encoder, a decoder and massive skip connections. The Locally-enhanced Window (LeWin) block is adopting the design in Uformer [20] and SRGAN [10] is adopted as the discriminator to form the adversarial loss L_{adv}. LS-GAN has two key components to achieve better flare-removal performance, which contain the designed Mask-guided Hard-Attention module (MHA module) and the hybrid objective function (including smooth L1 loss, perceptual loss, gradient loss, and adversarial loss).

They observe that the conditions of the light source in the image play an important role in the resulting flares. Thus, in the designed MHA module, this team uses light source masks as strong prior knowledge to lead the network training. In detail, they use the image saturation threshold and use image opening operation to preprocess the network input flare images and localize the light source area. By using the sigmoid function as the activation function, the light source areas get higher initial weights and non-light source areas get initial lower weights, which allows the network to focus more on the light source zones in the flare reflection removal process. Their proposed hybrid loss can be defined as:

$$
\mathcal{L}_{total} = \mathcal{L}_{smooth} + \alpha \mathcal{L}_{per} + \beta \mathcal{L}_{mge} + \gamma \mathcal{L}_{adv}, \quad (11)
$$

where $\alpha = 0.01, \beta = 0.01$ and $\gamma = 0.005$ are the hyperparameters weighting for each loss function. They use ImageNet [5] pre-trained VGG16 [17] as the loss network to measure perceptual similarity to form the perceptual loss. They use the Sobel operator for gradient calculation on both X-axis and Y-axis to form the mean gradient error (mge) loss. The smooth-L1 loss and adversarial loss are the same as [7, 28].

To preserve the original light source, this team blends the input light source back into the network prediction images. In detail, they compute a binary mask according to the saturated threshold. This binary mask indicates the light source zone of the input image. The masked area in the net-
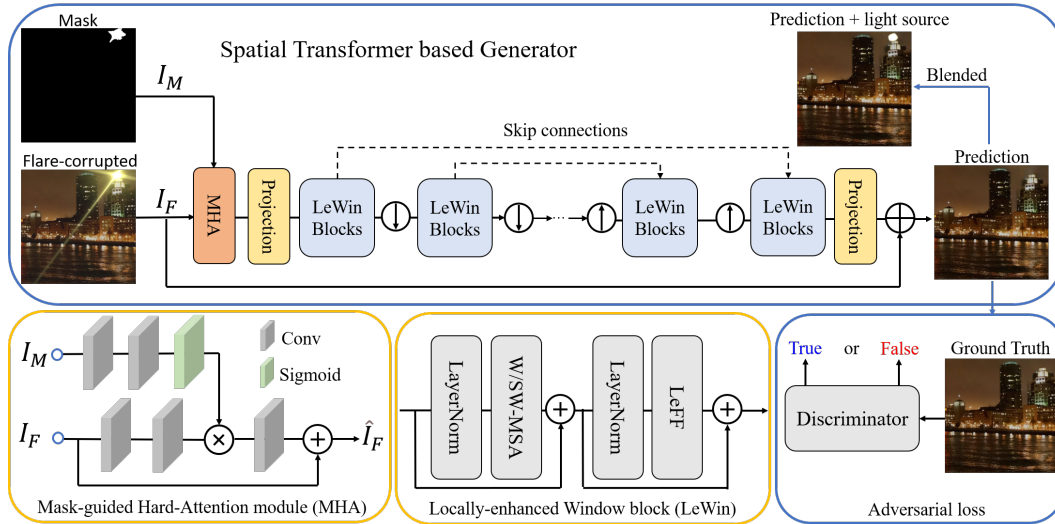
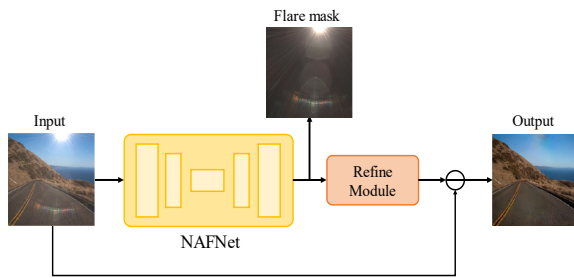Figure 6. The network architecture of USask-Flare team.

Figure 7. The network architecture of LVGroup_HFUT team.

work output is then replaced by the input pixels, producing a more realistic final result.

For the network training details, this team resized raw training samples to 512×512 before feeding them into the network. They randomly rotate each image by 90,180,270 degrees and also utilize horizontal and vertical flips as our data augmentation strategies. The batch size is set to 2 due to memory limitation. They use Adam with β_1=0.9, β_2=0.999 as our optimizer, and a specific decay strategy is adopted, where the initial learning rate is set to 1e-4 and decays 0.5 times at 100, and 150 epochs for a total of 500 epochs. The discriminator uses the same optimizer and weight decay strategies.

LVGroup_HFUT. This team try to build an effective model with flare guidance for nighttime flare removal (see Figure 7). To be specific, the author first employ NAFNet[2] as the backbone to generate a flare since NAFNet is a very strong model for image restoration. According to the obtained flare and the synthesis mechanism of flare-interrupted image, the author try to get a flare-free image by subtracting the original image and the flare images. Please note that the authors use the designed refine module to allow for flexible subtraction rather than strict subtraction.

The authors use L_1 loss and frequency reconstruction loss (weight:0.1) to constrain the predicted flare mask and final flare-free output. The proposed solution is implemented based on PyTorch version 1.10 and NVIDIA RTX 3090 with 24G memory. During training, the authors first perform a series of data augment operations sequentially as follows: 1):random crop to 384x384; 2): vertical flip with probability 0.5; 3): horizontal flip with probability 0.5. The authors train the model for 90 epochs on provided training dataset with initial learning rate 1e-4 and batch size 6. For testing (inferring), The authors feed original resolution (512x512) images into the pretrained model and directly generate the final flare-free images.

szzzzz01. This team proposes a flare-region attention recurrent network for flare removal, and the pipeline is shown in Figure. 8. Fundamentally, flare removal is achieved by utilizing surrounding pixels to restore the center corrupted pixel. Therefore, different surrounding corrupted pixels should make a different contribution to the restoration. Based on that, this team designs a spatial attention mask M for a corrupted image before feeding it into the restoration module. To make full use of the attention mask, rather than one-step restoration, they gradually remove the flare within multi-steps. Concretely, a flare detected network (FDN) with a Unet [15] architecture is first pretrained to generate the rough flare region F, whose loss is:

$$\mathcal{L}_{FDN} = \|F - F_{ref}\|_2^2 + \|\nabla F - \nabla F_{ref}\|_2^2, \quad (12)$$

where F_{ref} denotes the corresponding flare reference and ∇ is the gradient operation.

During each restoration step, the mask generation network (MGN) with an architecture of two convolutional layers takes $1 - F_k$ as input and outputs the spatial attention

mask M_k. Then the restoration network with a Uformer [20] architecture takes $M_k \odot I_k$ as input and produces recovered result $\hat{I}_{k+1}$, where $\odot$ is element-wise multiplication. The reconstruction loss is as follows:

$$\mathcal{L}_{re} = \sum_{k=1}^{K} \gamma_k(\|\hat{I}_k - I_{ref}\|_F^2 + 1 - SSIM(\hat{I}_k - I_{ref}))$$
$$+ \|\phi(\hat{I}_k) - \phi(I_{ref})\|_1, \tag{13}$$

where K is the total number of recurrent times set to be 3, $\phi(\cdot)$ represents a pre-trained VGG19 feature extractor, $SSIM(\cdot)$ denotes the structural similarity loss, and γ_k is hyper-parameter set to 1/32, 1/8, and 1 in different steps. Besides, those severely polluted pixels that are nearer the brightest light source should be assigned small weights, such that the generated attention mask is similar to $1 - F_k$, which leads to a mask loss:

$$\mathcal{L}_m = \lambda \sum_{k=1}^{K} \|\hat{M}_k - (1 - F_k)\|_F^2, \tag{14}$$

where λ is a hyper-parameter set to be 0.1. The total loss function is $\mathcal{L}_{total} = \mathcal{L}_{re} + \mathcal{L}_m$.

The authors train FDN for 300 epochs alone, and train the MGN and Uformer for 200 epochs together with batch size and learning rate set to be 4 and 0.0001 respectively. All the experiments are implemented on Nvidia A100 GPU with pytorch framework.
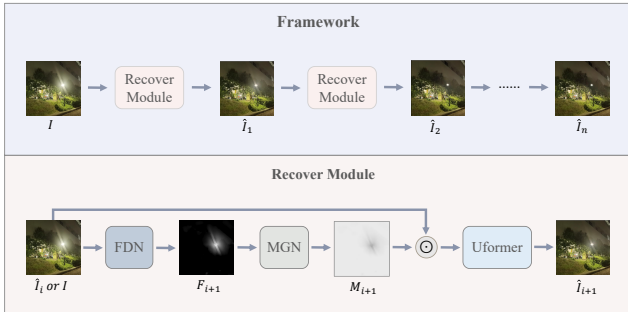

Figure 8. The framework of szzzzz01 team.

CEVI_Explorers. This team proposes a pipeline DeFlare-Net for the removal of flare artifacts from images. The proposed architecture (DeFlare-Net) is shown in Figure 9. DeFlare-Net comprises two main modules, i.e., Flare Removal Network (FRN), and Light Source Detection Network (LSDN) as shown in Figure 9. FRN and LSDN are built upon U-Net-based architecture [21] with skip connection across scales. Given a flared image input I_{flare}, the aim of the proposed DeFlare-Net is to learn flare artifacts, and light source and generate a flare-free image. FRN detects and removes flare by masking the flare region

X_{flare}. FRN implicitly learns to segment light source along with flare region, as shown in Figure 9. Removal of the light source creates additional artifacts resulting in unpleasant observation. Towards this, the authors introduce LSDN to retain the light source in accordance with the ground truth image/true scene. LSDN learns to segment the light source from the scene for retaining the light source in the flare-free image. The removal of flare from the input flare image I_{flare} is guided by the flare mask, resulting in the implicit removal of a light source. To retain the light source in the image, the light source mask is blended with the flare-removed scene, to obtain the final flare-free image.

Loss Functions. To optimize the task of flare removal using DeFlare-Net the authors propose $L_{DeFlare}$ as a weighted combination of Flare loss L_{flare}, Lightsource loss L_{ls}, and Reconstruction loss L_{recon}. To learn the segmentation of flares in the images, we incorporate L_{flare} [14]. L_{flare} is computed between the input flare image and the flare mask. To learn the segmentation of light source in the true scene/ground truth image, the authors incorporate light source loss L_{ls}. L_{ls} is computed between the light source masks of the ground truth image and the predicted mask. To facilitate the overall restoration of the flare-free image the authors use L_{recon} computed between ground truth image and the flare-free image. The proposed $L_{DeFlare}$ loss is given as,

$$\mathbb{L}_{DeFlare} = \alpha * \mathbb{L}_{flare} + \beta * \mathbb{L}_{ls} + \gamma * \mathbb{L}_{recon}, \tag{15}$$

where α, β, and γ are weights and we heuristically set $\alpha = \beta = \gamma = 0.33$.

The DeFlare-Net architecture was written using Python (v3.8) and PyTorch framework. All experiments were performed on NVIDIA RTX 3090 GPU with 24GB memory and AMD Ryzen Threadripper processor.

AI ISP. This team adopts Uformer [20] architecture and makes several improvements in loss function and data synthesis to boost the network's performance.

Using L1 loss and focal frequency loss, the flare regions should be given a higher weight when calculating the loss function to get better restoration results in these regions. Thus, both global loss like [24] and regional loss are used. The predicted image is replaced by $\hat{I}_g$ and $\hat{I}_r$ when calculating the global loss and regional loss respectively:

$$\hat{I}_g = I_0 \odot M_l + f(I_f, \Theta) \odot (1 - M_l),$$
$$\hat{I}_r = I_0 \odot (M_l \vee M_{nf}) + f(I_f, \Theta) \odot (1 - (M_l \vee M_{nf})), \tag{16}$$

where M_l and M_{nf} denote binary masks indicating the light source and non-flare regions respectively, I_0 and I_f denote the clean image and flare-polluted image respectively, $f(I_f, \Theta)$ denotes the flare removal network, $\odot$ denotes Hadamard product, and $\vee$ denotes logic OR. The
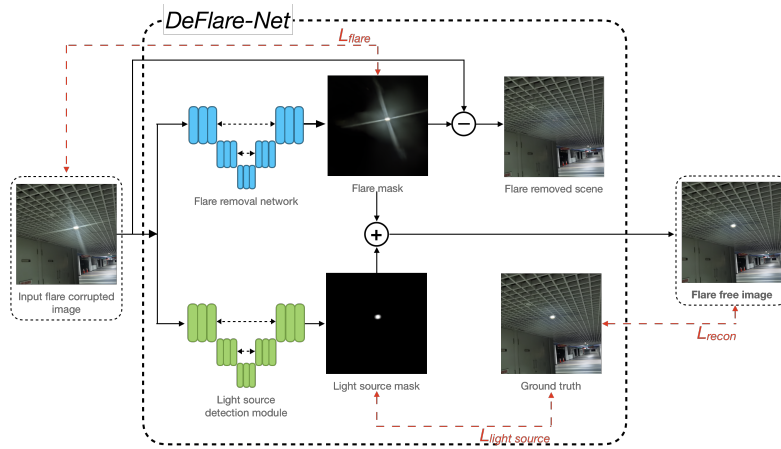
Figure 9. The network architecture of DeFlare-Net for removal of flare artefacts proposed by CEVI_Explorers team.
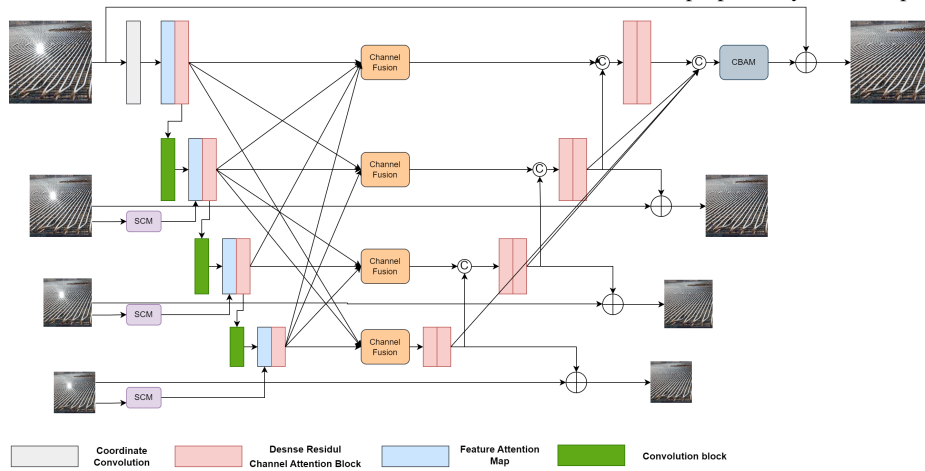

Figure 10. The network architecture of Couger AI team.

masks are generated using threshold segmentation when loading the training samples. The total loss function is the weighted sum of the global loss and regional loss.

To get better performance on nighttime scenes, this team uses extra 260 RAW2RGB images from SID [1], which are manually cropped to remove flare and light source regions. The model is trained for 80 epochs on the Flickr24K [31] and SID dataset, and for another 60 epochs on Flickr24K only. Besides, in view of the fact that there may exist more than one light source in real images, this team synthesizes images with multi light sources of different scales in addition to images with only a single light source.

The proposed solution is implemented based on PyTorch on Nvidia A100 GPUs. Each batch contains 16 samples, among which 11 images have one light source, four images have two light sources, and one image has three light sources. Adam optimizer is applied with learning rate 10^{-4}.

Couger AI. The network is inspired by [6, 13]. The proposed network is a multi-level U-net-based system (see Figure 10). It includes four encoder and four decoder blocks.

The input image is passed to the Coordinate[11] convolution layer to enhance the spatial quality. The output of the Coordinate convolution layer passed to the Encoder block. The encoder block contains the convolution layer with a 3x3 filter, Feature Attention map block, and Dense Residual Channel Attention Block[32] [DRCA], followed by a downsampling layer. The output of the encoder block is passed to the channel fusion block and concatenated to the corresponding decoder block. The decoder block contains the upsampling layer with two DRCA blocks. Each decoder block is fused with the input image and supervised by the loss function. Each decoder level's output is concatenated along the channel axis, followed by the CBAM[22] attention block before the final supervision.

5. Acknowledgements

References

[1] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to See in the Dark. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 8

[2] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European Conference on Computer Vision*, 2022. 6

[3] Yuekun Dai, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Flare7k: A phenomenological nighttime flare removal dataset. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2022. 1, 2, 3

[4] Yuekun Dai, Yihang Luo, Shangchen Zhou, Chongyi Li, and Chen Change Loy. Nighttime smartphone reflective flare removal using optical center symmetry prior. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2023. 1

[5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009. 5

[6] Ruicheng Feng, Chongyi Li, Shangchen Zhou, Wenxiu Sun, Qingpeng Zhu, Jun Jiang, Qingyu Yang, Chen Change Loy, Jinwei Gu, Yurui Zhu, et al. Mipi 2022 challenge on underdisplay camera image restoration: Methods and results. In *European Conference on Computer Vision Workshops*, 2022. 1, 8

[7] Minghan Fu, Huan Liu, Yankun Yu, Jun Chen, and Keyan Wang. Dw-gan: A discrete wavelet transform gan for non-homogeneous dehazing. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 5

[8] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*. Springer, 2016. 4

[9] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(11):2599–2613, 2018. 3

[10] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 4, 5

[11] Rosanne Liu, Joel Lehman, Piero Molino, Felipe Petroski Such, Eric Frank, Alex Sergeev, and Jason Yosinski. An intriguing failing of convolutional neural networks and the coordconv solution. *Advances in Neural Information Processing Systems*, 31, 2018. 8

[12] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2018. 4

[13] Sabari Nathan and Priya Kansal. Skeletonnetv2: A dense channel attention blocks for skeleton extraction. In *IEEE International Conference on Computer Vision*, 2021. 8

[14] Xiaotian Qiao, Gerhard P Hancke, and Rynson WH Lau. Light source guided single-image flare removal from unpaired data. In *IEEE International Conference on Computer Vision*, 2021. 7

[15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015. 6

[16] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 815–823, 2015. 5

[17] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2014. 5

[18] Wenxiu Sun, Qingpeng Zhu, Chongyi Li, Ruicheng Feng, Shangchen Zhou, Jun Jiang, Qingyu Yang, Chen Change Loy, Jinwei Gu, Dewang Hou, et al. Mipi 2022 challenge on rgb+ tof depth completion: Dataset and report. In *European Conference on Computer Vision Workshops*, 2022. 1

[19] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2022. 4

[20] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restorationn. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2022. 2, 4, 5, 7

[21] Jia Hao Wong and Satish T S Bukkapatnam. A review of deep learning approaches for removing flare artifacts in optical microscopy. *IEEE Reviews in Biomedical Engineering*, 12:20–34, 2019. 7

[22] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *European Conference on Computer Vision*, 2018. 8

[23] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 5

[24] Yicheng Wu, Qiurui He, Tianfan Xue, Rahul Garg, Jiawen Chen, Ashok Veeraraghavan, and Jonathan T. Barron. How to train neural networks for flare removal. In *IEEE International Conference on Computer Vision*, 2021. 1, 7

[25] Qingyu Yang, Guang Yang, Jun Jiang, Chongyi Li, Ruicheng Feng, Shangchen Zhou, Wenxiu Sun, Qingpeng Zhu, Chen Change Loy, Jinwei Gu, et al. Mipi 2022 challenge on quad-bayer re-mosaic: Dataset and report. In *European Conference on Computer Vision Workshops*, 2022. 1

[26] Qingyu Yang, Guang Yang, Jun Jiang, Chongyi Li, Ruicheng Feng, Shangchen Zhou, Wenxiu Sun, Qingpeng Zhu, Chen Change Loy, Jinwei Gu, et al. Mipi 2022 challenge on rgbw sensor fusion: Dataset and report. In *European Conference on Computer Vision Workshops*, 2022. 1

[27] Qingyu Yang, Guang Yang, Jun Jiang, Chongyi Li, Ruicheng Feng, Shangchen Zhou, Wenxiu Sun, Qingpeng Zhu,

Chen Change Loy, Jinwei Gu, et al. Mipi 2022 challenge on rgbw sensor re-mosaic: Dataset and report. In *European Conference on Computer Vision Workshops*, 2022. 1

[28] Yankun Yu, Huan Liu, Minghan Fu, Jun Chen, Xiyao Wang, and Keyan Wang. A two-branch neural network for non-homogeneous dehazing via ensemble learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 5

[29] Yanhong Zeng, Jianlong Fu, Hongyang Chao, and Baining Guo. Aggregated contextual transformations for high-resolution image inpainting. *IEEE Transactions on Visualization and Computer Graphics*, 2022. 2

[30] Dafeng Zhang, Feiyu Huang, Shizhuo Liu, Xiaobing Wang, and Zhezhu Jin. Swinfir: Revisiting the swinir with fast fourier convolution and improved training for image super-resolution. *arXiv preprint arXiv:2208.11247*, 2022. 3

[31] Xuaner Zhang, Ren Ng, and Qifeng Chen. Single image reflection separation with perceptual losses. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 8

[32] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 8

A. Teams and Affiliations

MiAlgo.

Title: Erasing and Inpainting network for nighttime flare removal
Members:
Shuai Liu[1] (liushuai21@xiaomi.com)
Hao Wang[1] Chaoyu Feng[1] Luyang Wang[1] Guangqi Shao[1] Chenguang Zhang[1] Xiaotao Wang[1] Lei Lei[1]
Affiliations:
[1] Xiaomi Inc., China

SRC-B.

Members:
Dafeng Zhang[1] (dfeng.zhang@samsung.com)
Xiangyu Kong[1] Guanqun Liu[1] Mengmeng Bai[1] Jia Ouyang[1] Xiaobing Wang[1] Jiahui Yuan[1]
Affiliations:
[1] Samsung Research China - Beijing (SRC-B), China

MegFR.

Title: Nighttime Flare Removal through Signal Separation
Members:
Xinpeng Li[1] (lixinpeng@megvii.com)
Chengzhi Jiang[1], Ting Jiang[1], Wenjie Lin[1], Qi Wu[1], Mingyan Han[1], Jinting Luo[1], Lei Yu[1], Haoqiang Fan[1] and Shuaicheng Liu[2,1*]
Affiliations:

[1] Megvii Technology
[2] University of Electronic Science and Technology of China (UESTC)

AntIns.

Title: An Effective Nighttime Flare Removal Pipeline for MIPI Challenge
Members:
Bo Yan (lengyu.yb@antgroup.com)
Zhuang Li Yadong Li Hongbin Wang
Affiliations:
Ant Group, China

ActionBrain-ETRI.

Title: Suppressing Flares in the Dark Images with Cascaded U-Former Architecture
Members:
Soonyong Song (soony@etri.re.kr)
Affiliations:
Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea

USask-Flare.

Title: A Light source guided Spatial transformer Generative Adversarial Network for single Image Flare Removal
Members:
Minghan Fu (fig072@mail.usask.ca)
Rayyan Azam Khan Fangxiang Wu
Affiliations:
College of Engineering, University of Saskatchewan, Canada

LVGroup_HFUT.

Title: Flare-guide nonlinear activation free network for nighttime flare removal
Members:
Zhao Zhang[1] (cszzhang@gmail.com)
Suiyi Zhao[1] Huan Zheng[1] Yangcheng Gao[1] Yanyan Wei[1] Jiahuan Ren[1] Bo Wang[1] Yan Luo[1]
Affiliations:
[1] Laboratory of Multimedia Computing, Hefei University of Technology, China

szzzzz01.

Title: A Flare-region Attention Recurrent Network for Flare Removal
Members:
Shuaibo Gao[1] (gaosb201@gmail.com)

Wenhui Wu[1]   Sicong Kang[1]
Affiliations:
College of Electronics and Information Engineering, Shenzhen University, China

CEVI Explorers.

Title: DeFlare-Net: Flare Detection and Removal Network
Members:
Nikhil Akalwadi (nikhil.akalwadi@kletech.ac.in)
Ankit Raichur   Vinod Patil   Allabakash G
Swaroop A   Amogh Joshi   Chaitra Desai
Ramesh Ashok Tabib   Ujwala Patil   Uma Mudenagudi
Affiliations:
Center of Excellence in Visual Intelligence (CEVI)- KLE Technological University, Hubballi, Karnataka, India

AI ISP.

Title: Improvements of loss function and data synthesis for nighttime flare removal
Members:
Ruoxi Zhu[1]   Jiazheng Lian[2](jzlian20@fudan.edu.cn)
Sicheng Li   Shusong Xu   Zihao Liu
Affiliations:
[1]School of Microelectronics, Fudan University, China
[2]Academy for Engineering and Technology, Fudan University, China

Couger AI.

Title: Light Weight Dense Residual Channel attention model for flare removal
Members:
Sabari Nathan[1] (sabari@couger.co.jp)
Priya Kansal[1]
Affiliations:
[1] Couger Inc, Tokyo, Japan