

Reflective Flare Removal Using Image Bursts

Ke Xu, Zheng Dong, Weiwei Xu, Gerhard Petrus Hancke, *Fellow, IEEE*, Rynson W.H. Lau *Senior Member, IEEE*

Abstract—When the camera is facing the light source, light scattering and reflections within the lens system may produce both scattering and reflective flares in the captured image. While scattering flares are typically radial and have been largely addressed by existing flare removal methods via modeling in the image spatial domain, reflective flares (which appear as bright spots or ghostings of the light source) are more challenging to remove as they may have similar intensity values and shapes to the light source in the camera-finished images. We observe that in an image burst, reflective flares tend to present larger spatial-temporal disturbances than the light source, as their disturbances tend to be amplified internally by the lens system during the camera shake. This phenomenon can be used to help separate reflective flares from light sources for detection. Inspired by this, we propose a novel neural approach for removing reflective flares based on image bursts. We first formulate an image formation model to model the “moving” reflective flares in a burst, which produces realistic data for training deep models. We then propose a novel neural network with a novel reflective flare localization (REFO) module for locating the reflective flares, a novel guided burst-feature alignment (GBFA) module for handling the spatial and color misalignments caused by camera shakes, and an adaptive fusion (ADFU) module for removing reflective flares. Extensive experiments show that our approach outperforms state-of-the-art methods and generalizes well on image bursts captured in the wild.

Index Terms—Reflective Flare Removal, Image Enhancement, Image Dynamics Modeling, Burst sRGB Image Processing.

1 INTRODUCTION

PHOTOGRAPHING scenes with strong light sources (*e.g.*, sun, streetlights, headlights) often results in reflective and scattering flares, which are artifacts caused by reflections and scattering of lights, respectively, within the optical lens system. Such artifacts can become severe if the light sources are close to the camera field of view. While scattering flares are typically radial and often appear as diffuse glow around light sources, reflective flares manifest as bright ghosting artifacts with higher intensities, which are more likely to create visible artifacts or occlude important scene content. It is therefore desirable to remove them.

Previous non-deep learning-based methods typically use special hardware devices (*e.g.*, occlusion masks [39], [45]) and handcrafted low-level image statistics [8], [36], [47] to model certain types of lens flares, which limit their generalization capabilities to complex wild scenes. In recent years, a few deep-learning based methods are proposed to take advantage of the large-scale unpaired [38] and paired [11], [13], [18], [22], [25], [44], [50], [54], [55] datasets for learning flare representations. Qiao *et al.* [38] propose to detect light sources as a guide to help remove flares. However, their unpaired styleGAN-like learning scheme may not preserve image content fidelity. Other methods synthesize flares via the sampling Point Spread Functions [18], [44], [50], capturing under controlled settings [13], [50], using Adobe software for synthesis [11], and training advanced neural

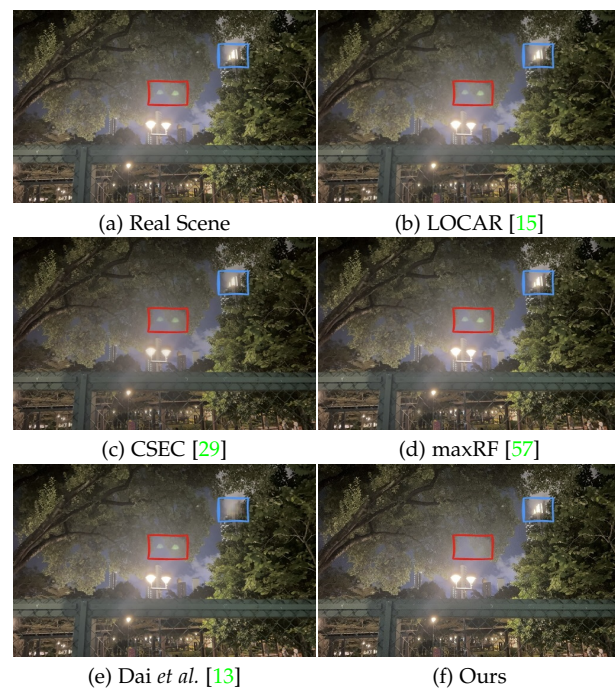


Fig. 1. Existing methods model reflective flares in the image spatial domain. They often fail to remove such flares (red boxes in (b-e)), and some methods may even mistakenly remove the light sources (blue box in (e)). We observe that in an image burst, reflective flares exhibit larger spatial-temporal disturbances than the light sources, as the camera lens system amplifies the disturbances of the reflective flares during camera shake. Our approach based on using image bursts can remove reflective flares well (f), while keeping all light sources intact.

- Ke Xu is with the School of AI&DS, University of Science and Technology of China, China.
- Zheng Dong is with the Zhejiang Sci-Tech University, China.
- Weiwei Xu is with the State Key Laboratory of CAD&CG, Zhejiang University, China.
- Gerhard Petrus Hancke is with the Department of Computer Science, City University of Hong Kong, Hong Kong SAR.
- Rynson W.H. Lau is with the Department of Computer Science, City University of Hong Kong (Dongguan Campus), China.

Manuscript received April 19, 2005; revised August 26, 2015.

networks [41], [42], [51], [53] in fully-supervised or adversarial manners for flare removal. While these methods may remove radial-shaped scattering flares well, they tend to fail in removing reflective flares, which usually have similar

intensities and shapes as those of the light sources. As shown in Figure 1(b-e), existing methods fail to remove the reflective flares (red boxes). Instead, they may sometimes mistakenly remove some light sources (blue boxes).

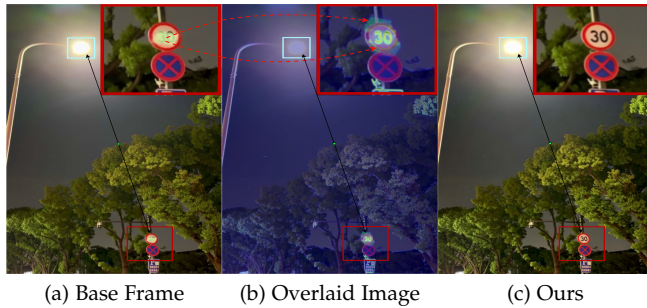


Fig. 2. (a) We show the base frame of a real-captured image burst, where the reflective flare (red box) is approximately symmetrical to the light source (blue box) with respect to the camera optical center as the symmetry center (green dot). (b) We calculate and accumulate the differences between the base frame and each subsequent frames in this burst, and overlay the accumulated difference map on the base frame (a) for visualization. The spatial displacement of reflective flare (red arrows) is larger than that of the light source (blue box) (c) Our result.

This ambiguity between reflective flares and light sources has not been explicitly addressed by existing methods, which mainly model flares in the image spatial domain in different ways. In this work, we observe that during camera shakes, reflective flares typically show larger displacements than the light sources in the spatial-temporal domain. This is because reflective flares are essentially ghost images formed by light that has traveled a different effective optical path (internal reflections between lens surfaces) than the primary image [43]. While the primary image is conjugate to the object plane, ghost images are conjugate to a pupil plane. For a camera rotation¹ (inevitable due to handshakes or camera vibrations when taking image bursts with a handheld camera [49] or even a tripod [27]), the displacement of the ghost image conjugate to a pupil on the sensor is larger than that of the primary image that conjugate to the object plane [40]. Therefore, under the same camera rotation, the displacement of the reflective flare will be larger than that of the primary light source. Visualization on a real-world image burst example is shown in Figure 2. This inspires us to model such spatial-temporal properties of reflective flares in image bursts² for reflective flare removal.

In this paper, we propose a novel neural approach for reflective flare removal based on image bursts. We first formulate an image formation model to synthesize the relative larger displacements of reflective flares (than the scene/light sources) under the same camera shake in a burst, which produces realistic data for training deep models. We then propose a novel neural network to capture such differential motion cue to remove reflective flares while keep the light sources intact. Specifically, our network includes a novel reflective flare localization (REFO) module to locate

1. Pure planar translation of camera results in the same displacement of the ghost image conjugate to a pupil on the sensor and the primary image that conjugate to the object plane [40].

2. We observe a similar phenomenon in videos. However, as videos are more likely to suffer from severe background object motions and/or camera motions, we focus on the image burst in this work.

the reflective flares by capturing larger displacements, a novel guided burst-feature alignment (GBFA) module to handle the global translative scene motion by registering misaligned burst features of non-flare regions, and an adaptive fusion (ADFU) module to remove reflective flares by fusing the multi-frame complementary information from the aligned non-flare regions. As shown in Figure 1(f) and Figure 2(c), our approach can handle reflective flares well.

The main contributions of this paper include:

- We propose the first neural approach to remove reflective flares from image bursts. We show that reflective flares typically have larger spatial displacements than light sources as a result of camera shakes, and by exploiting such cues, our approach can separate reflective flares from light sources naturally.
- We propose an effective image formation model to model the spatial disturbances of flares as well as the spatial and color misalignments in image bursts. Our formation model produces realistic images for learning the reflective flare removal process.
- We propose a novel network, with a novel REFO module for locating reflective flares, a novel GBFA module for handling spatial/color misalignments caused by camera shakes, and an ADFU module for removing reflective flares by exploiting the complementary information of non-flare regions across multi-frames.
- Extensive experiments show that the proposed approach outperforms state-of-the-art methods.

2 RELATED WORK

Lens Flares Removal Methods. In earlier years, two classic methods [39], [45] use occlusion masks (placed between the camera and the scene) to reduce lights that contribute to flares, and remove flares via direct-indirect separation [45] or lightfield-based reconstruction [39]. However, the dependence on special hardware devices limits their applications in wild scenes. Some other methods [8], [47] apply low-level image features (*e.g.*, intensity, shapes, colors, and/or sun position [36]) for detecting lens flares and removing them via exemplar-based inpainting [10]. However, they often fail in complex real-world scenes as handcrafted features may not represent the lens flares well.

Recently, a few deep methods [11], [12], [13], [38], [44], [50], [55] are proposed for lens flare removal. Wu *et al.* [50] propose a physics-based model to synthesize lens flares and train UNet- [41] and VGG-style [42], [53] networks for flare removal. Zhou *et al.* [55] propose an improved synthesis pipeline by extending the linear combination of flare and scene layers in [50] into a convex one and train a Uformer [48]-based network for flare removal. Qiao *et al.* [38] propose to detect light sources in the image to guide the flare removal process. They train the removal network using unpaired data and cycle-consistency constraints [56]. Some methods [11], [13] focus on flare removal in night-time scenes. Dai *et al.* [13] propose a static prior that reflective flares and light sources are nearly symmetrical with respect to the optical center. Based on this observation, they synthesize a night-time reflective flare dataset to train Transformer

networks for flare removal. Later, Lan and Chen [25] propose to analyze the effects of ISP operations (*e.g.*, denoising, compression, and sharpening) on flare removal. They curate a raw image dataset captured via mobile cameras and train Uformer [48] to achieve promising results on their dataset. He *et al.* [22] propose a self-supervised CNN-based method for joint glow and reflective flare removal based on the Point Spread Functions (PSF) rendering prior and the optical center symmetry prior [13]. They also create a synthetic dataset for training and evaluation. Very recently, Zhou *et al.* [54] propose to extend [55] by introducing the adversarial training paradigm based on Uformer [48] to improve the generalization ability.

All these existing methods model flares in the spatial domain. We observe that the motions of the reflective flares in the temporal domain are different from those of the light sources during camera shakes. We therefore propose to remove reflective flares based on image bursts, by modeling the “moving” characteristic of reflective flares.

Burst Image Processing. An image burst captures multiple shots in quick succession, which can be fused to produce a better image [20] or left for users to select a preferred one. The temporal information introduced by image bursts has been exploited for various low-level vision tasks, *e.g.*, motion blur removal [2], [6], image super-resolution [3], [4], [16], [17], [27], [49], image layer separation [9], image restoration and enhancement [16], [17], [19], [20], [28], [30], [32], [33], [34], and joint HDR synthesis and super-resolution [26]. These methods differ mainly in their ways of aligning burst images (*e.g.*, homography flow [2], [32], block-matching [20], and a combination of them [28], optical flow [3], [4], [49], kernels prediction [34], deformable alignment [16], [17], [30], [33], and warping estimations [26], [27]) and learning image priors for their respective tasks.

In this work, we propose to model the spatial-temporal disturbances of reflective flares using the readily available camera-output sRGB image bursts, with a new image formation model and a carefully-designed deep network, for reflective flare removal. This approach does not involve modeling sensor-specific camera operations, allowing the network to be camera-independent.

3 OUR APPROACH

The key insight of this work is that although both light sources and reflective flares undergo small translations due to camera shake, the displacement magnitudes differ: Reflective flares exhibit systematically larger displacements than the light sources because of the optical path difference (ghost images conjugate to a pupil plane *vs.* primary images conjugate to the object plane). To this end, we first propose a novel image formation model for modeling the reflective flare/scene differential displacements. We then propose a novel neural network for removing the reflective flares, which leverages the GBFA module to handle the global translative motion of the scene and the REFO module to specifically model the larger displacements of reflective flares relative to the light sources.

3.1 Our Image Formation Model

Since it is not practical to collect a pair of real image bursts (*i.e.*, with and without reflective flares) at the same time, we propose an image formation model to generate a burst of images with reflective flares in the intermediate high dynamic range (HDR) color space, from a low dynamic range (LDR) flare-free sRGB image.

Irradiance-to-Image Transformation. Digital cameras typically transform the scene irradiance $S \in \mathbb{R}^+$ into an LDR image $\mathcal{X}$, which values mainly depend on the camera response function $\mathcal{T}(\cdot)$, ISO gain, aperture size, and exposure duration Δt . However, adjusting the ISO gain may introduce noise, and adjusting the aperture size may introduce out-of-focus blur. Hence, we follow [14] to model the irradiance-to-image transformation as:

$$\mathcal{X} = \mathcal{T}(\Delta t \times S), \quad (1)$$

where Δt is often set to a small value to avoid object motion blur in photography.

Increasing the Dynamic Range. Scene irradiance is typically of high dynamic range ($\mathbb{R}^+$), but $\mathcal{T}(\cdot)$ reduces the dynamic range of the output image significantly for display purposes. As pixels representing light sources and reflective flares tend to be saturated, increasing the dynamic range helps produce more details and avoid color artifacts. On the other hand, when facing the light sources, the scene illuminance in images is typically high (> 10 Lux), resulting in a negligible noise level. Hence, we do not model sensor noise, nor do we directly use the camera raw data in this work. We increase the dynamic range of images by unprocessing the LDR images into the intermediate camera-independent linear color space [1]³, to avoid the modeling of sensor-specific camera operations (*e.g.*, denoising, demosaicing, and white balancing).

Spatial Disturbance. Wronski *et al.* [49] show that during burst imaging, hand movements are typically small and uniformly random (*i.e.*, no preferred directions). This suggests a “global-but-small” translation of images (including the light sources) in an image burst, which we approximate as a random 2D rigid motion (within ± 2 pixels and rotation angles of $\pm 2^\circ$) [34]. Meanwhile, we observe that reflective flares tend to have larger displacements (than light sources) across the image burst, while Dai *et al.* [13] show a static phenomenon that reflective flares are approximately symmetrical to the light sources with respect to camera focus in each image of the burst. This suggests that the movements of reflective flares (although larger than light sources) are still restricted in a local region. Hence, we model the global-but-small translation of the scene, and the local-but-large translation of reflective flares.

Image Formation Model. Based on the above analysis, we formulate our image formation model to generate a burst of N images, $\mathcal{Y} = \{y_1, \dots, y_N\}$, with reflective flares in the intermediate HDR color space, from an LDR non-flare sRGB image $\mathcal{X}$ and an LDR reflective flare map $\mathcal{R}$, as:

$$y_n = \delta_n^R \mathcal{T}^{-1}(\mathcal{R}) + \delta_n^S \mathcal{T}^{-1}(\mathcal{X}), \quad (2)$$

³In our preliminary implementation, we tried the camera raw reconstruction technique [5]. Unfortunately, it does not work on Dai *et al.*'s dataset [13], producing all-zero results.

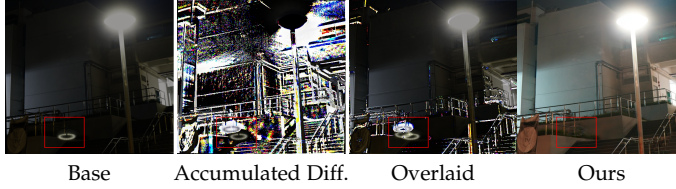


Fig. 3. We show a base frame (in linear color space) of the generated image burst, the accumulated difference (between subsequent frames and the base frame) map, the base frame overlaid with the difference map, and our result, respectively. Our image formation model synthesizes the “moving characteristics” of reflective flares (red boxes).

where δ_n^R and δ_n^S are the local-but-large and global-but-small translations, respectively, drawn from two 2D uniform integer distributions, for modeling the movements of flares and the scene. $\mathcal{T}^{-1}(\cdot)$ is the unprocessing function that converts the sRGB non-flare image $\mathcal{X}$ and the reflective flare $\mathcal{R}$ into the intermediate HDR color space. Here, we omit the exposure duration Δt , as it is typically set to a small constant value in burst imaging to avoid object motions. Using our image formation model, we have produced a total of 3,520 burst HDR images for training and 320 burst HDR images for evaluation.

Figure 3 shows an example of synthesized image burst. Specifically, we first show the base frame (in linear color space) of a generated image burst. We then show the accumulated difference (between subsequent frames and the base frame) map. We also superimpose the accumulated difference map on the base frame, where we can see that our image formation model synthesizes the “moving characteristics” of reflective flares (red boxes).

3.2 Our Neural Network

Given a burst of N HDR images generated by our image formation model, $\mathcal{Y} = \{y_1, \dots, y_N\}$, we now aim to seek a deep function $\mathcal{Z}_\theta$ with a learnable parameter θ to help remove reflective flares as well as handle spatial and color misalignments, to produce an LDR sRGB image $\hat{\mathcal{X}}$:

$$\hat{\mathcal{X}} = \mathcal{Z}_\theta(\mathcal{Y}). \quad (3)$$

Figure 4 shows the overview of our network, which has an encoder-decoder architecture with three modules for reflective flare localization, non-flare region alignment, and reflective flare removal. Specifically, given a burst of N HDR images $\mathcal{Y}$ as input, our network first extracts their corresponding features $\mathcal{F} = \{f_1, \dots, f_{N-1}, f_N\}$ in four scales. At each scale, we propose a novel REFO module to locate reflective flares by modeling their spatial-temporal displacements. It also guides another novel GBFA module in handling spatial and color misalignments in non-flare regions. An ADFU module then models the complementary multi-frame information in the aligned burst features. Our network restores the “non-flare” image $\hat{\mathcal{X}}$ progressively. In our implementation, we use the first frame in a burst as the base frame. We denote its corresponding features as f_b and the features of other misaligned frames as f_m (i.e., $f_m = \{f_2, \dots, f_{N-1}, f_N\}$).

3.2.1 The REFO Module

The goal of the REFO module is to model local but large spatial-temporal displacements of reflective flares. We apply REFO modules to process multi-scale burst features in a bottom-up manner to progressively locate the reflective flares. Figure 5 shows the structure of the REFO module at the i -th feature scale.

Given the extracted burst features $\mathcal{F}^i$, we first split them into features of the base frame f_b^i and features of the misaligned burst frames f_m^i . The features of f_m^i are then processed via a group of convolutions to align the feature dimensions with those of f_b^i . We then model the spatial displacements by subtracting f_m^i from f_b^i to produce $f_{d(r,m)}^i$, which accumulates the spatial differences of the unaligned burst features with respect to the base features.

To identify large-scale displacements to localize reflective flares, we use two groups of dilated convolutions (with kernel size/dilation rate of 3/1 and 5/2), denoted as $g_{d1}(\cdot)$ and $g_{d2}(\cdot)$, to extract features with different receptive fields, and then compute a contrast-aware attention map C_a between these two sets of features as:

$$C_a = \sigma(g_{d1}(f_{d(r,m)}^i) - g_{d2}(f_{d(r,m)}^i)), \quad (4)$$

where $\sigma(\cdot)$ represents the sigmoid function. C_a highlights the local relative contrast information, where pixels of high contrasts are regarded as having large displacements. We then use C_a to highlight the accumulated difference features to produce f_R^i via $f_R^i = C_a \otimes f_{d(r,m)}^i$. We further use f_R^i as well as the displacement features $f_{\delta R}^{i-1}$ from the previous scale guidance to compute the displacement features of reflective flares $f_{\delta R}^i$. Specifically, we first use convolution to align the feature dimension of $\mathcal{F}^i$ with f_R^i and $f_{\delta R}^{i-1}$, and then concatenate them together. We then apply the SE block [23] to adjust their channel-wise discrepancies and the offset generator [46] to produce $f_{\delta R}^i$. $f_{\delta R}^i$ are passed to the REFO module in the $i+1$ -th feature scale to progressively locate reflective flares in a bottom-up manner, while $\bar{C}_a = 1 - C_a$ are passed to the GBFA module in the i -th scale to guide the burst feature alignment.

Figure 6 shows the input and results of our method on a real scene (captured via an iPhone 13 camera) without and with REFO, where our method tends to remove the light source (blue box) instead of the reflective flare (red box) without REFO, which demonstrates the effectiveness of localizing the reflective flares.

3.2.2 The GBFA Module

The goal of the GBFA module is to handle spatial and color misalignments of the image burst. Since their translations are global but small, we apply GBFA modules to process multi-scale burst features in a top-down manner, where misalignments are progressively corrected from low spatial-resolution features to those of high spatial-resolution. Unlike existing methods [3], [4], [16], [17] that align all pixels of the displacements jointly, we leverage $\bar{C}_a$ from our REFO module to guide the GBFA module to focus on the alignment of non-flare pixels, which provide contextual information for reflective flare removal.

As shown in Figure 7, we first apply a group of convolutions to reduce the dimensions of the unaligned burst

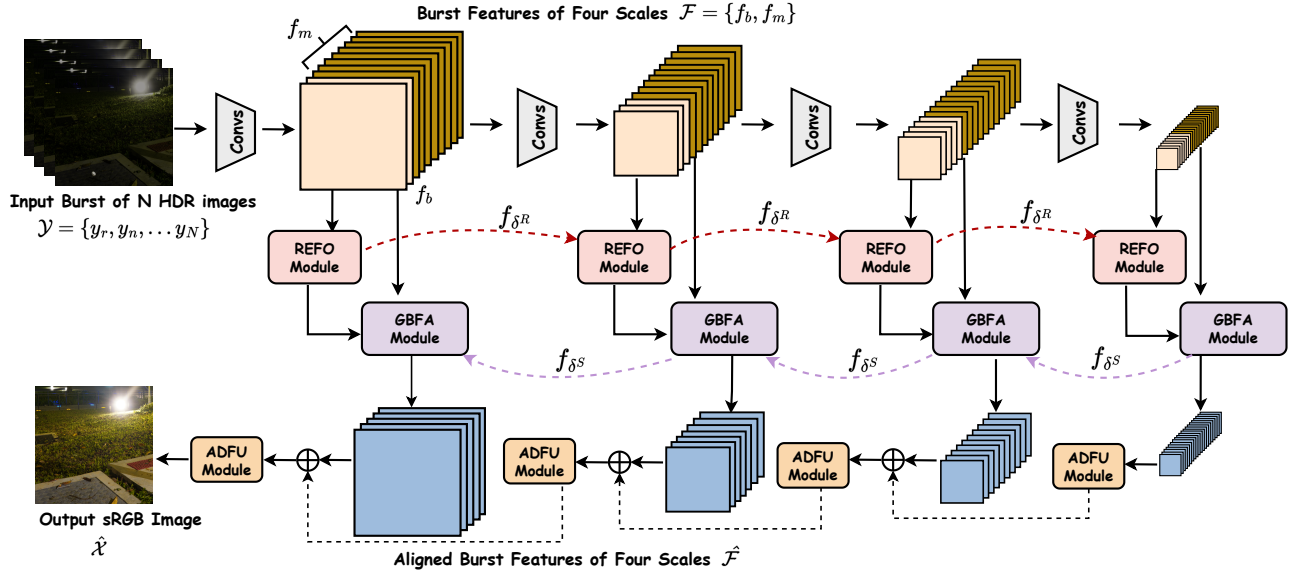


Fig. 4. Our network has an encoder-decoder architecture that takes a burst of N HDR images $\mathcal{Y}$ (produced by our image formation model) as input to extract their features $\mathcal{F} = \{f_b, f_m\}$ in four scales. In each scale, we propose a novel REFO module to model the spatial-temporal displacements of reflective flares for their localization, a novel GBFA module to handle the spatial and color misalignments of non-flare regions within the image burst, guided by the flare localization results, and an ADFU module to adaptively fuse the complementary multi-frame information in the aligned burst features. The REFO modules locate reflective flares in a bottom-up manner, while the GBFA modules align burst features in a top-down manner. The multi-scale fused burst features are used to produce the non-flare sRGB image $\hat{\mathcal{X}}$. During testing, we first feed the input sRGB image burst to [1] to obtain corresponding HDR images as input to our network.

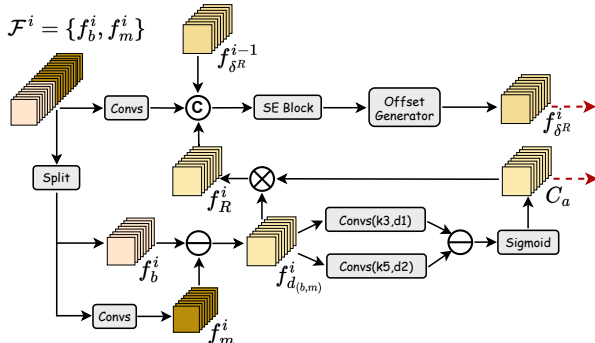


Fig. 5. The proposed REFO module aims to localize reflective flares by modeling their local but large spatial-temporal displacements.



Fig. 6. Visualization of our results on a real-world scene (captured by an iPhone 13 camera) with and without the REFO module.

features $\mathcal{F}^i$, and concatenate them with the displacement features from the higher feature level $f_{\delta^S}^{i+1}$. We then feed them into an offset generator to predict the displacement features $f_{\delta^S}^i$, and split the unaligned burst features $\mathcal{F}^i$ into the reference features f_b^i and the unaligned features f_m^i . f_m^i is first spatially weighted by $\hat{C}_a$ to focus on the non-flare

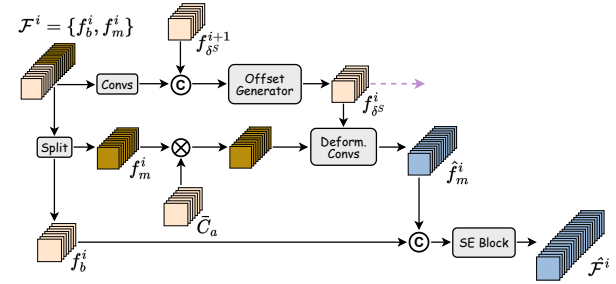


Fig. 7. The proposed GBFA module aims to align burst features by modeling their global but small spatial-temporal displacements.

regions and then fed into the deformable convolutions with the displacement features $f_{\delta^S}^i$ for alignment to produce $\hat{f}_m^i$. Finally, we concatenate f_b^i and $\hat{f}_m^i$, and apply a SE block to align them in the channel-wise dimension to produce the aligned burst image features $\hat{\mathcal{F}}^i$. $f_{\delta^S}^i$ are passed to the GBFA module in the $(i-1)$ -th feature scale for a progressive feature alignment in a bottom-up manner.

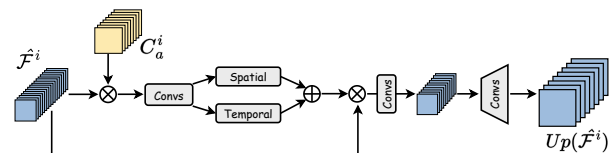


Fig. 8. The adaptive fusion (ADFU) module adaptively fuses the complementary aligned multi-frame information to produce non-flare images.

3.2.3 The ADFU Module

Given the aligned burst features, we exploit their complementary temporal information to remove reflective flares via an adaptive and progressive fusion, as shown in Figure 8. We start by processing the aligned burst features of the lowest spatial resolution. Given the aligned burst features $\hat{\mathcal{F}}^i$ at i -th scale, we first use $\mathcal{C}_a^i$ to highlight the reflective flare features, and then use paralleled spatial-wise and channel-wise attention mechanisms to aggregate spatial/temporal complementary information. We further use transpose convolutions to upsample the spatial resolution of the features and fuse them with fused features at $(i - 1)$ -th scale.

3.2.4 Implementation Details

We have implemented the proposed model under Pytorch [37], and trained/tested it on a PC with an i7 4GHz CPU and a RTX4090 GPU. We adopt the architecture [31] for our implementation of the multi-scale feature extractor. The network parameters are initialized using the truncated normal initializer (with the randomness seed set to 24). For loss minimization, we adopt the Adam optimizer [24]. The base learning rate, batch size, and the number of training epochs are $1e^{-4}$, 1, and 250, respectively, while the learning rate is reduced by 10 at the 150^{th} epoch. Input images are resized to 512×512 . In practice, for Eq. 2, we constrain the scenes to have maximum spatial displacements (δ_n^S) of ± 2 pixels [34] and rotation angles of $\pm 2^\circ$, while those of the reflective flares (δ_n^R) are set to the range of $(\pm 2, \pm d/4]$, where d is the distance between the flare and the light source. We use the unprocessing method [1] as our $\mathcal{T}^{-1}(\cdot)$, and the synthesized (symmetric to light sources with respect to focuses) reflective flares in [13] as our $\mathcal{R}$. During training, we adopt L_1 and $SSIM$ losses to measure the background accuracy between the estimated image $\hat{\mathcal{X}}$ and the ground truth image $\mathcal{X}$. We also adopt cosine similarity L_{cos} for background color alignment, which measures the similarity of two vectors by computing the angle between them. L_{cos} forces the predicted pixels to have the same directions as the ground truth pixels in the RGB space, without considering the magnitude. For inference, real sRGB bursts are first converted to the linear color space via [1], which are then fed to the network. It takes around 0.03s for our method to produce an image of 512×512 resolution.

4 EXPERIMENTS

Evaluation Methods. We compare our method to 13 state-of-the-art methods: BurstKPN [34], Godard *et al.* [19], Lecouat *et al.* [26], Burstormer [17], Wu *et al.* [50], LOCAR [15], maxRF [57], Dai *et al.* [13], RetinexFormer [7], CSEC [29], Uformer [48], CycleGAN [56] and Zhou *et al.* [54]. Among them, BurstKPN [34], Godard *et al.* [19], Lecouat *et al.* [26], and Burstormer [17] are four state-of-the-art burst image based enhancement methods, which are retrained on our burst flare data for a fair comparison to verify the effectiveness of our network designs. Other methods are single-image based. We re-train them with base frame of our burst data, with the symmetry prior [13] for reflective flare removal. Specifically, Wu *et al.* [50] use a UNet [41] for flare removal, Dai *et al.* [13] uses MPRNet [52], while Qiao [38] adopts CycleGAN [56]. Uformer [48] (originally proposed

for image restoration) is widely adopted for flare removal by previous flare removal methods [11], [12], [25], [55], while Zhou *et al.* [54] combines adversarial learning with Uformer. LOCAR [15] and maxRF [57] are two state-of-the-art single-image reflection removal methods, while RetinexFormer [7] and CSEC [29] are two state-of-the-art image enhancement methods. We train RetinexFormer [7] to remove flares in the illumination layer via Retinex-based decomposition, while leveraging the ability of CSEC [29] for exposure correction to remove flares (which are often saturated).

Evaluation Datasets and Metrics. We evaluate our method on two existing flare removal datasets, Dai *et al.* [13] and Wu *et al.* [50]. Among them, the dataset from Dai *et al.* [13] focuses on night-time reflective flares. We unprocess their sRGB images into burst linear images using our proposed image formation model, to train the burst image-based methods. As the dataset from Wu *et al.* [50] focuses on daytime scattering flares, we evaluate the generalization capability of our method on this dataset. Note that our image formation model is not designed for scattering flares (which are highly correlated with the light sources, *i.e.*, they tend to “move” in the same way as the light sources). We only unprocess the original images into the linear color space to train all burst-based models. For a fair comparison, all competing methods are re-trained on the flare removal datasets [13], [50]. Finally, we use the PSNR and SSIM metrics to measure the errors between flare removal results and ground truth images. We also report the NIQE [35] score (a popular subjective image quality metric) for reference.

4.1 Comparison with State-of-the-Art Methods

Removing Reflective Flares. We first evaluate the effectiveness of the proposed method in removing reflective flares on the dataset from Dai *et al.* [13]. Table 1(left) reports the quantitative comparison. We can see that Dai *et al.* [13] produce the best PSNR and SSIM results among existing methods. On the other hand, we can also see that directly training existing burst image-based enhancement methods may not work well. For example, despite a higher NIQE score, BurstKPN [34] still achieves lower PSNR and SSIM scores compared to Dai *et al.* [13]. In contrast, our method outperforms existing methods by progressively locating reflective flares, aligning burst features, and fusing multi-frame spatial-temporal information.

Figure 9 shows the visual comparison, where we can see that existing methods may fail to remove strong reflective flares (first example) and may also fail to remove small reflective flares (second example). Our method successfully removes all of them.

Generalizing to In-the-wild Scenes. Figure 10 shows reflective flare removal results from ours and existing methods on sRGB image bursts captured by a Sony A7m4 camera. Despite having high-performance multi-layer anti-reflective coatings (Nano AR Coating II) for lenses, these images still suffer from reflective flares⁴. We can see that existing methods may fail to detect the reflective flares and may mistakenly remove the light sources (blue boxes in the

4. These reflective flares tend to be greenish as the green wavelengths are reflected more efficiently than others in this lens system.

Methods/Datasets	N	Dai <i>et al.</i> [13]			Wu <i>et al.</i> [50]		
		PSNR↑	SSIM↑	NIQE↓	PSNR↑	SSIM↑	NIQE↓
Input	1	36.75	0.9833	5.291	19.80	0.8146	5.069
Wu <i>et al.</i> [50]	1	37.65	0.9881	5.330	23.56	0.8546	4.956
LOCAR [15]	1	38.85	0.9884	5.299	22.94	0.8509	5.043
maxRF [57]	1	32.01	0.9431	5.115	22.28	0.8276	4.820
Dai <i>et al.</i> [13]	1	<u>40.11</u>	<u>0.9887</u>	5.299	23.26	0.8569	5.113
RetinexFormer [7]	1	39.65	0.9886	5.113	23.65	0.8656	4.856
CSEC [29]	1	38.12	0.9880	5.310	22.47	0.8355	4.895
Uformer [48]	1	39.60	0.9884	5.389	23.83	0.8760	4.810
CycleGAN [56]	1	38.09	0.9840	5.411	21.91	0.8263	5.374
Zhou <i>et al.</i> [54]	1	40.10	<u>0.9887</u>	4.930	23.82	0.8675	4.772
BurstKPN [34]	4	38.95	0.9883	4.925	23.87	<u>0.8769</u>	4.809
Godard <i>et al.</i> [19]	4	34.70	0.9496	5.565	22.05	0.8245	5.415
Burstormer [17]	4	38.72	0.9880	4.964	23.81	0.8764	5.133
Lecouat <i>et al.</i> [26]	4	36.75	0.9853	<u>4.914</u>	22.76	0.8670	5.062
Ours	4	42.20	0.9917	4.913	24.79	0.9017	4.595

TABLE 1 Quantitative comparison on the reflective flare removal dataset [13] (left) and the scattering flare removal dataset [50] (right). N indicates the number of input images. The best and second best results are in **bold** and underlined, respectively.

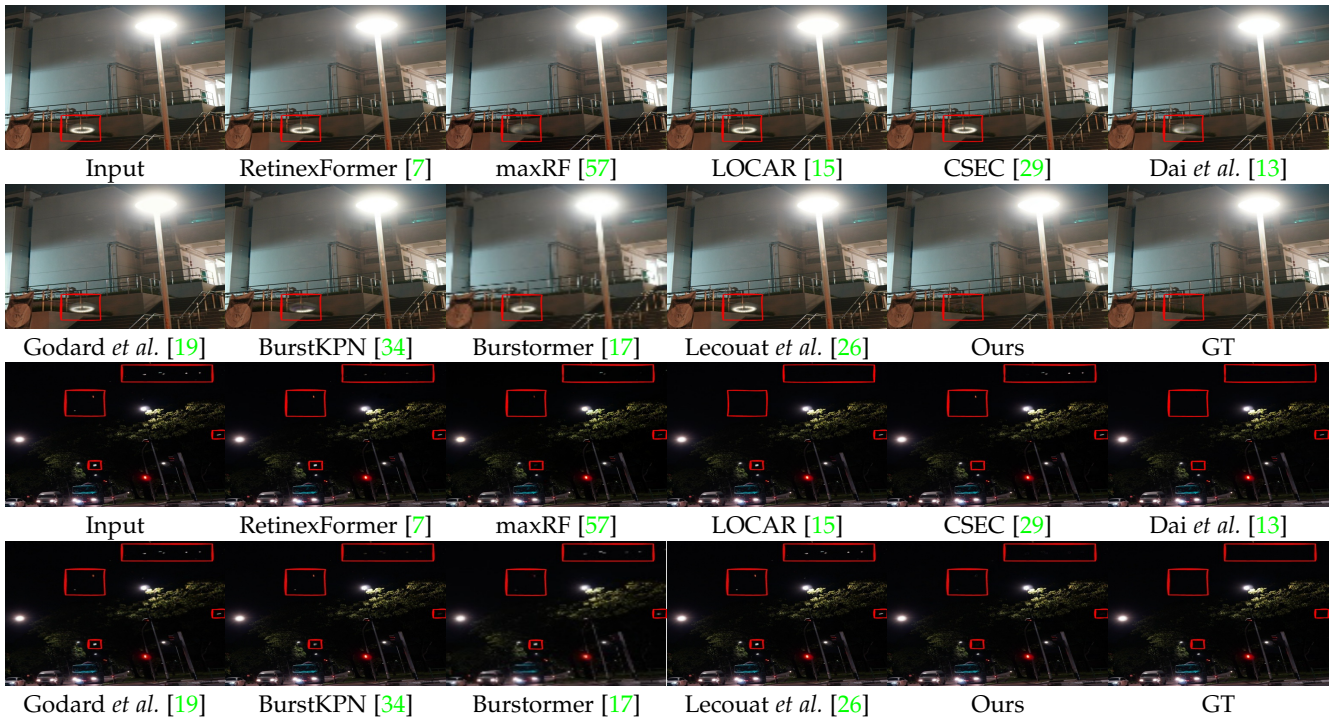


Fig. 9. Visual comparison of our method with existing methods on removing reflective flares.

first example). Our method can successfully detect and remove the reflective flares by modeling their spatial-temporal “moving” characteristics.

Generalizing to Scattering Flares. We evaluate the generalization capability of our method on removing scattering flares. Table 1(right) reports the comparison results. Although scattering flares do not exhibit the same spatial-temporal characteristics as the reflective flares, our method still benefits from processing in linear HDR space, where the intensities of scattering flares (tending to be weaker) are more separable. Our REFO module can localize them based on the contrast-aware attention mechanism, and the GBFA module aligns multi-frame non-flare information for their removal. Figure 11 shows two visual examples, demonstrating that our method can remove scattering flares while

recovering image contrast.

4.2 Internal Analysis

Ablation Study. Table 2 reports the ablation study of our network on the reflective flare removal dataset [13]. The 1st row shows that directly applying ConvNeXt (originally proposed for classification) does not work well in our task as it works on low-resolution feature maps for modeling image semantics. We replace the “patchify” stem with standard convolutions (while reducing the network width) to form our encoder (2nd row) for multi-scale feature extraction. We then remove the proposed REFO and GBFA modules, and report the corresponding results in 3rd and 4th rows, respectively. We also replace our GBFA module with the deformable convolutional networks (DCN) in the 5th row, and

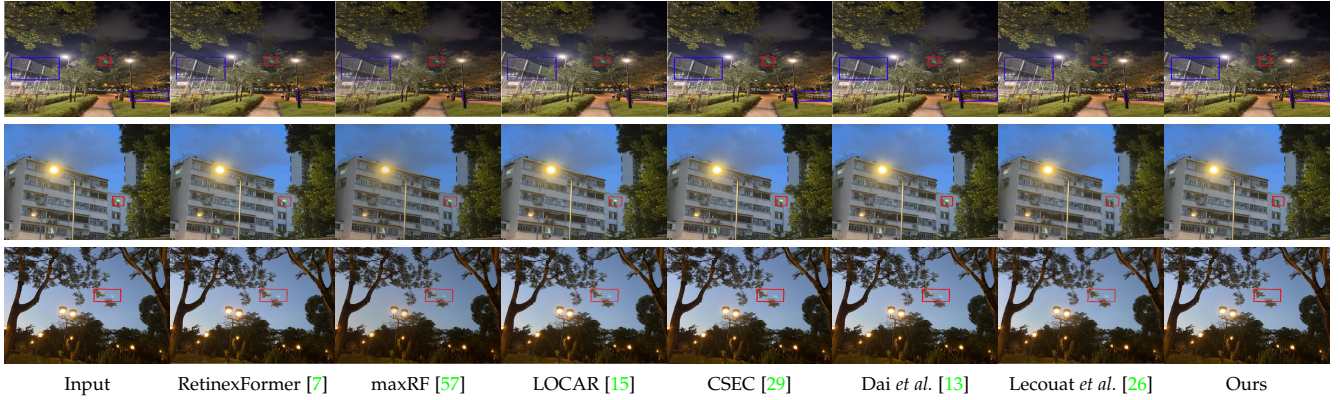


Fig. 10. Visual results on real image bursts captured by a Sony A7m4 camera (only the first frames are shown).

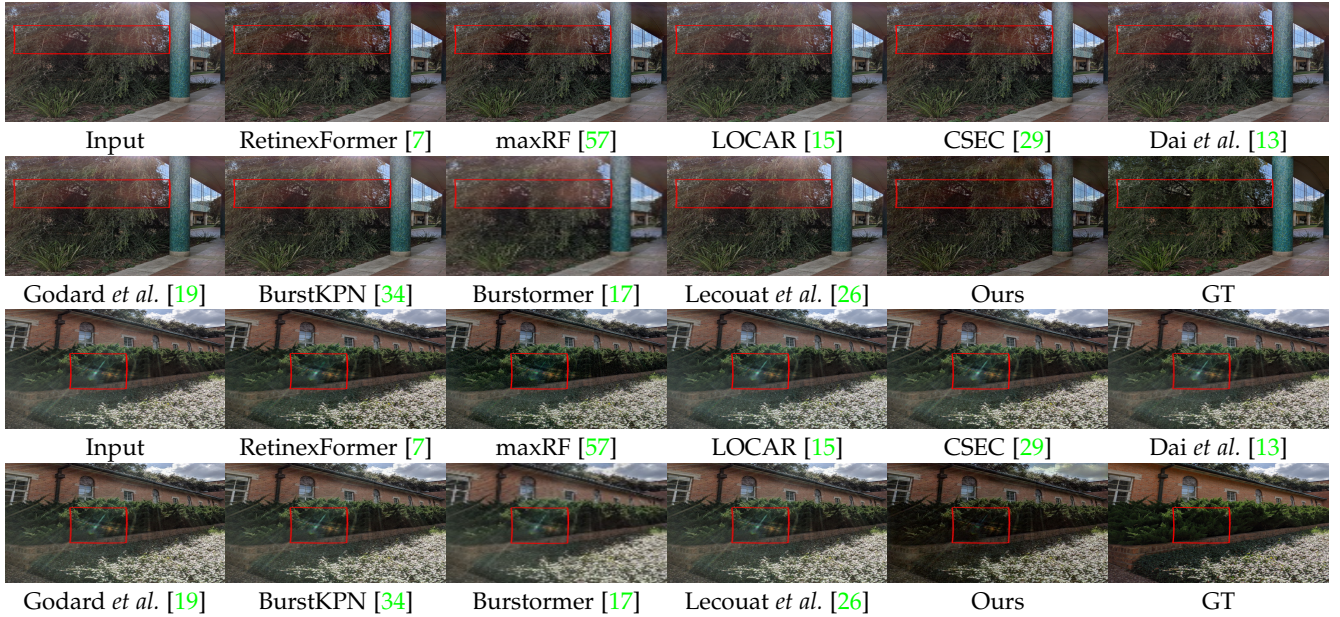


Fig. 11. Visual results of our method and existing approaches on removing the scattering flares.

replace the ADFU module with standard resBlocks [21] in the 6th row. Additionally, 7–8th rows show the effect of using different numbers of burst images. The results in 1–6th rows generally verify the effectiveness of our network design. For the selection of N , as using $N = 6$ may slightly degrade the PSNR and SSIM scores and introduce a higher computational cost, we set $N = 4$. We also includes comparisons with UNet [41] and Uformer [48] trained on our burst data⁵ in 9–10th rows. We can see that training on our data boosts the performance of both UNet [41] (i.e., PSNR/SSIM from 37.65/0.9881 to 38.42/0.9883) and Uformer [48] (i.e., PSNR/SSIM: from 39.60/0.9884 to 40.70/0.9888), which verifies the effectiveness of our image bursts synthesis pipeline. Meanwhile, our network still outperforms them by notable margins (e.g., +1.5 dB PSNR over Uformer [48]), demonstrating the effectiveness of our model.

We also report the ablation results of our network on the scattering flare removal dataset [50] in Table 3. The performance drop (by removing the unprocessing step) in the 1st

5. Restormer [51] cannot be trained on our single RTX 4090 GPU, as it causes out-of-memory error).

Methods	PSNR $\uparrow$	SSIM $\uparrow$
ConvNeXt [31]	31.80	0.9191
ConvNeXt [31]→Our Encoder	33.29	0.9493
<i>w/o</i> REFO	38.62	0.9851
<i>w/o</i> GBFA	39.26	0.9899
GBFA → DCN [17]	40.06	0.9896
<i>w/o</i> ADFU	40.92	0.9902
$N = 2$	41.24	0.9908
$N = 6$	42.12	0.9916
Ours→UNet [41] ($N = 4$)	38.42	0.9883
Ours→Uformer [48] ($N = 4$)	40.70	0.9888
Ours ($N = 4$)	42.20	0.9917

TABLE 2 Ablation study on the reflective flare removal dataset [13], which verifies the effectiveness of our method designs.

row indicates that handling the scattering flares in the linear color space helps. The middle two rows show that removing the REFO and GBFA modules degrades the removal performance. Although there are no “moving” characteristics to capture, the REFO modules locate the scattering flares based on the contrast information in a bottom-up manner, and the GBFA modules further refine the localization features of the

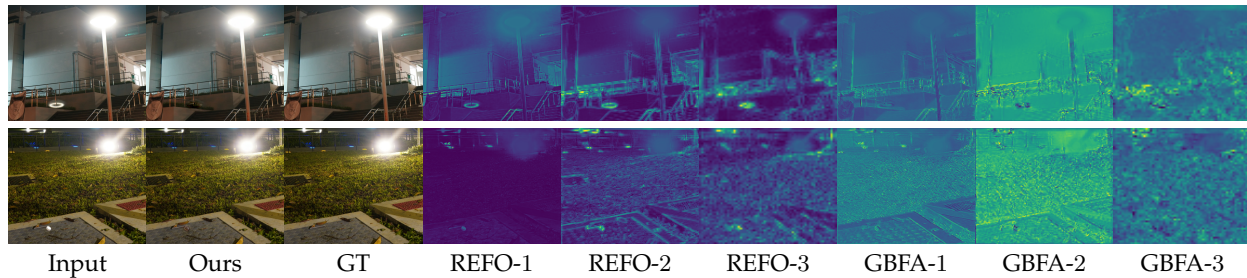


Fig. 12. Visualization of the features learned by the proposed REFO and GBFA modules at 1 – 3-th scale.

Methods	PSNR $\uparrow$	SSIM $\uparrow$
<i>w/o</i> Unprocessing	23.11	0.8641
<i>w/o</i> REFO	22.96	0.8508
<i>w/o</i> GBFA	23.28	0.8570
<i>w/o</i> ADFU	24.10	0.8829
Ours	24.79	0.9017

TABLE 3 Ablation on the scattering flare removal dataset [50], which demonstrates the effectiveness of our modules on scattering flares.

Methods	Numbers of Parameters	GFLOPs
BurstKPN [34]	30M	578.74
Burstormer [17]	3.54M	3.58M
Lecouat <i>et al.</i> [26]	17.02M	285.79
Godard <i>et al.</i> [19]	966.5K	1.98
Ours	1.8M	69.58

TABLE 5 Complexity comparison between burst imaging methods and ours, which shows that our method can achieve better results with fewer parameters and FLOPs.

REFO modules in a top-down manner. Finally, removing the ADFU modules (*i.e.*, the progressive flare restoration) produces relatively worse results.

Joint Training. We discuss the potential for enhancing performance by training our model with both synthesized scattering and reflective flare data in Table 4. Specifically, we explore two strategies: (1) first pre-training on scattering flares and then fine-tuning on reflective flares (denoted as “Ours $S \rightarrow R$ ”), and (2) first pre-training on reflective flares and then fine-tuning on scattering flares (denoted as “Ours $R \rightarrow S$ ”). Interestingly, compared with individually training on the reflective or scattering data (*i.e.*, “Ours (Individual)”), we can see that pre-training on scattering flares (*i.e.*, “Ours $S \rightarrow R$ ”) improves both performances, while pre-training on reflective flares (*i.e.*, “Ours $R \rightarrow S$ ”) does not. This is because pre-training on the scattering flares helps our model learn light-source-aware features, which further helps distinguish from reflective flares. In contrast, while our model first learns to differentiate reflective flares with light sources based on the spatial-temporal cues of reflective flares, it does not help differentiate light sources from scattering flares due to their similar spatial-temporal shifts.

Methods/Datasets	Dai <i>et al.</i> [13]		Wu <i>et al.</i> [50]	
	PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$
Ours ($S \rightarrow R$)	43.17	0.9930	25.57	0.9081
Ours ($R \rightarrow S$)	42.09	0.9908	24.80	0.9017
Ours (Individual)	42.20	0.9917	24.79	0.9017

TABLE 4 Joint training on reflective and scattering flare data. First pre-training on scattering flares helps our model learn light-source-aware features, which further facilitates the removal of reflective flares.

Feature Visualization. Figure 12 visualizes the learned multi-scaled features of the proposed REFO and GBFA modules. We can see that while the REFO modules focus on locating the reflective flares, the GBFA modules focus on the background alignment.

Method Complexity. Finally, we report the number of parameters and FLOPs for the proposed method and burst imaging methods in Table 5. Our method has relatively fewer parameters and FLOPs while achieving better results.

4.3 More Results

We provide more visual comparisons with prior works [11], [12], [13], [50], [55] in Figure 13, where our method achieves better reflective flare removal performance. We also evaluate our model’s generalization performance across different devices from the dataset of [55]. Since this dataset [55] provides single sRGB images, which does not provide spatial-temporal cues that our method aims to exploit, we inference on this dataset by unprocessing the target image into the linear color space and duplicating it as the input into our method. As shown in Fig. 14, our method can remove the reflective flares by leveraging the symmetry prior [12].

5 CONCLUSION

In this paper, we have proposed a novel approach for reflective flare removal based on image bursts. Our approach incorporates an effective image formation model to model the “moving” reflective flares in image bursts, based on which a novel network is proposed for reflective flare removal. Our network includes a novel reflective flare localization (REFO) module for locating the reflective flares, a novel guided burst feature alignment (GBFA) module for handling spatial and color misalignments caused by camera shakes, and an adaptive fusion (ADFU) module for removing reflective flares. We have conducted extensive experiments on standard datasets. Results show that our approach outperforms state-of-the-art methods, and generalizes well on real image bursts captured in the wild.

Our approach does have limitations. First, since our method does not model background object motions, although our method can remove the reflective flares (the



Fig. 13. Visual comparison of our method with existing methods on removing reflective flares.

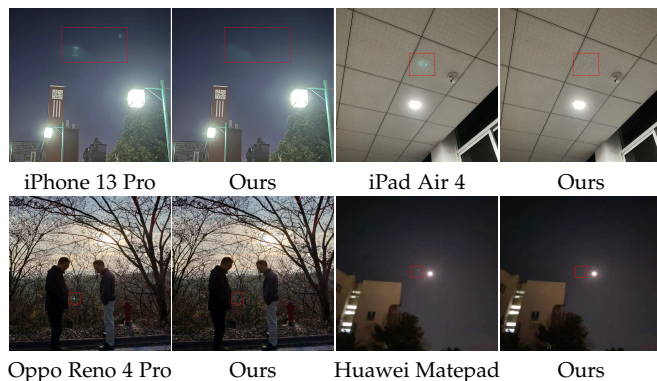


Fig. 14. Generalization evaluation on different devices from the dataset of [55].

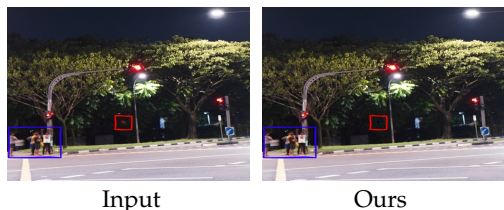


Fig. 15. Although our method can effectively remove the reflective flares (i.e., the red box), there may still be blurriness caused by background object motions (the blue box) in our result.

red box in Figure 15), our result shows blurriness caused by local motions (the blue box). Second, our model assumes moderate handshake during burst imaging, as large motions (e.g., > 8 pixels) may break the global alignment assumption. For such cases, a pre-alignment step (e.g., SIFT-based registration) may be integrated. Third, our method is designed for typical lens systems for smartphones and prime lenses (usually consisting of 5-10 lens elements). However, specialized lenses (e.g., ultra-wide or anamorphic lenses) may produce distinct (distorted) flare patterns. Zoom lenses (typically consisting of 10-25 elements) may produce other challenging flares (e.g., multi-layered flares). Meanwhile, our method may also fail for complex scenes containing multiple overlapping reflective flares from different light sources, genuine scene elements such as specular highlights or glass reflections that move differently from the rigid background under camera motion due to parallax, or restoring

semantically meaningful content covered by strong flares.

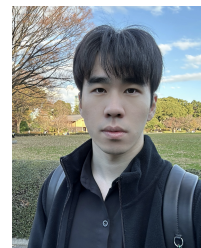
REFERENCES

- [1] Mahmoud Afifi, Abdelrahman Abdelhamed, Abdullah Abuolaim, Abhijith Punnappurath, and Michael S Brown. Cie xyz net: Unprocessing images for low-level computer vision tasks. *IEEE TPAMI*, 2021.
- [2] Miika Aittala and Frédo Durand. Burst image deblurring using permutation invariant convolutional neural networks. In *ECCV*, 2018.
- [3] Goutam Bhat, Martin Danelljan, Luc Van Gool, and Radu Timofte. Deep burst super-resolution. In *CVPR*, 2021.
- [4] Goutam Bhat, Martin Danelljan, Fisher Yu, Luc Van Gool, and Radu Timofte. Deep reparametrization of multi-frame super-resolution and denoising. In *CVPR*, 2021.
- [5] Tim Brooks, Ben Mildenhall, Tianfan Xue, Jiawen Chen, Dillon Sharlet, and Jonathan T Barron. Unprocessing images for learned raw denoising. In *CVPR*, 2019.
- [6] Jian-Feng Cai, Hui Ji, Chaoqiang Liu, and Zuowei Shen. Blind motion deblurring using multiple images. *Journal of Computational Physics*, 2009.
- [7] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *ICCV*, 2023.
- [8] Floris Chabert. Automated lens flare removal. Stanford Department of Electrical Engineering Technical Report, 2015.
- [9] Ilya Chugunov, David Shustin, Ruyu Yan, Chenyang Lei, and Felix Heide. Neural spline fields for burst image fusion and layer separation. In *CVPR*, 2024.
- [10] Antonio Criminisi, Patrick Pérez, and Kentaro Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE TIP*, 2004.
- [11] Yuekun Dai, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Flare7k: A phenomenological nighttime flare removal dataset. In *NeurIPS Datasets and Benchmarks Track*, 2022.
- [12] Yuekun Dai, Chongyi Li, Shangchen Zhou, Ruicheng Feng, Yihang Luo, and Chen Change Loy. Flare7k++: Mixing synthetic and real datasets for nighttime flare removal and beyond. *IEEE TPAMI*, 2024.
- [13] Yuekun Dai, Yihang Luo, Shangchen Zhou, Chongyi Li, and Chen Change Loy. Nighttime smartphone reflective flare removal using optical center symmetry prior. In *CVPR*, 2023.
- [14] Paul E Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH*, 2008.
- [15] Zheng Dong, Ke Xu, Yin Yang, Hujun Bao, Weiwei Xu, and Rynson WH Lau. Location-aware single image reflection removal. In *ICCV*, 2021.
- [16] Akshay Dudhane, Syed Zamir, Salman Khan, Fahad Khan, and Ming-Hsuan Yang. Burst image restoration and enhancement. In *CVPR*, 2022.
- [17] Akshay Dudhane, Syed Zamir, Salman Khan, Fahad Khan, and Ming-Hsuan Yang. Burstformer: Burst image restoration and enhancement transformer. In *CVPR*, 2023.
- [18] Ruicheng Feng, Chongyi Li, Huaijin Chen, Shuai Li, Chen Change Loy, and Jinwei Gu. Removing diffraction image artifacts in under-display camera via dynamic skip connection networks. In *CVPR*, 2021.

- [19] Clément Godard, Kevin Matzen, and Matt Uyttendaele. Deep burst denoising. In *ECCV*, 2018.
- [20] Samuel W. Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T. Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM TOG*, 2016.
- [21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.
- [22] Yuwen He, Wei Wang, Wanyu Wu, and Kui Jiang. Disentangle nighttime lens flares: self-supervised generation-based lens flare removal. In *AAAI*, 2025.
- [23] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *CVPR*, 2018.
- [24] P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015.
- [25] Fengbo Lan and Chang Wen Chen. Understanding and tackling scattering and reflective flare for mobile camera systems. In *ACM Multimedia*, 2024.
- [26] Bruno Lecouat, Thomas Eboli, Jean Ponce, and Julien Mairal. High dynamic range and super-resolution from raw image bursts. *ACM TOG*, 2022.
- [27] Bruno Lecouat, Jean Ponce, and Julien Mairal. Lucas-kanade reloaded: End-to-end super-resolution from raw image bursts. In *ICCV*, 2021.
- [28] Dasong Li, Yi Zhang, Ka Lung Law, Xiaogang Wang, Hongwei Qin, and Hongsheng Li. Efficient burst raw denoising with variance stabilization and multi-frequency denoising network. *IJCV*, 2022.
- [29] Yiyu Li, Ke Xu, Gerhard Petrus Hancke, and Rynson W.H. Lau. Color shift estimation-and-correction for image enhancement. In *CVPR*, 2024.
- [30] Zhetong Liang, Shi Guo, Hong Gu, Huaqi Zhang, and Lei Zhang. A decoupled learning scheme for real-world burst denoising from raw images. In *ECCV*, 2020.
- [31] Zhuang Liu, Hanzhi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *CVPR*, 2022.
- [32] Ziwei Liu, Lu Yuan, Xiaoou Tang, Matt Uyttendaele, and Jian Sun. Fast burst images denoising. *ACM TOG*, 2014.
- [33] Nancy Mehta, Akshay Dudhane, Subrahmanyam Murala, Syed Waqas Zamir, Salman Khan, and Fahad Shahbaz Khan. Gated multi-resolution transfer network for burst restoration and enhancement. In *CVPR*, 2023.
- [34] Ben Mildenhall, Jonathan Barron, Jiawen Chen, Dillon Sharlet, Ren Ng, and Robert Carroll. Burst denoising with kernel prediction networks. In *CVPR*, 2018.
- [35] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal processing letters*, 2012.
- [36] Andreas Nussberger, Helmut Grabner, and Luc Van Gool. Robust aerial object tracking in images with lens flare. In *ICRA*, 2015.
- [37] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In *NeurIPS Workshops*, 2017.
- [38] Xiaotian Qiao, Gerhard P Hancke, and Rynson WH Lau. Light source guided single-image flare removal from unpaired data. In *ICCV*, 2021.
- [39] Ramesh Raskar, Amit Agrawal, Cyrus A. Wilson, and Ashok Veeraraghavan. Glare aware photography: 4d ray sampling for reducing glare effects of camera lenses. *ACM TOG*, 2008.
- [40] S.F. Ray. *Applied Photographic Optics: Lenses and Optical Systems for Photography, Film, Video, Electronic and Digital Imaging*. Focal, 2002.
- [41] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015.
- [42] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR*, 2015.
- [43] W.J. Smith. *Modern Optical Engineering 4E (PB)*. McGraw Hill professional. McGraw Hill LLC, 2007.
- [44] Qilin Sun, Ethan Tseng, Qiang Fu, Wolfgang Heidrich, and Felix Heide. Learning rank-1 diffractive optics for single-shot high dynamic range imaging. In *CVPR*, 2020.
- [45] Eino-Ville Talvala, Andrew Adams, Mark Horowitz, and Marc Levoy. Veiling glare in high dynamic range imaging. *ACM TOG*, 2007.
- [46] Yapeng Tian, Yulun Zhang, Yun Fu, and Chenliang Xu. Tdan: Temporally-deformable alignment network for video super-resolution. In *CVPR*, 2020.
- [47] Patricia Vitoria and Coloma Ballester. Automatic flare spot artifact detection and removal in photographs. *Journal of Mathematical Imaging and Vision*, 2019.
- [48] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianshuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *CVPR*, 2022.
- [49] Bartłomiej Wronski, Ignacio Garcia-Dorado, Manfred Ernst, Damien Kelly, Michael Krainin, Chia-Kai Liang, Marc Levoy, and Peyman Milanfar. Handheld multi-frame super-resolution. *ACM TOG*, 2019.
- [50] 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 *ICCV*, 2021.
- [51] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, 2022.
- [52] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, 2021.
- [53] Xuaner Zhang, Ren Ng, and Qifeng Chen. Single image reflection separation with perceptual losses. In *CVPR*, 2018.
- [54] Yuyan Zhou, Dong Liang, Songcan Chen, and Sheng-Jun Huang. Image lens flare removal using adversarial curve learning. *IEEE TPAMI*, 2025.
- [55] Yuyan Zhou, Dong Liang, Songcan Chen, Sheng-Jun Huang, Shuo Yang, and Chongyi Li. Improving lens flare removal with general purpose pipeline and multiple light sources recovery. In *ICCV*, 2023.
- [56] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017.
- [57] Yurui Zhu, Xueyang Fu, Peng-Tao Jiang, Hao Zhang, Qibin Sun, Jinwei Chen, Zheng-Jun Zha, and Bo Li. Revisiting single image reflection removal in the wild. In *CVPR*, 2024.



Ke Xu is an Associate Professor in the School of AI&DS, University of Science and Technology of China. He was a Research Fellow at the City University of Hong Kong from 2021-2026. He received the dual Ph.D. degree from Dalian University of Technology and the City University of Hong Kong in 2021. His research interests include computer vision and interactive graphics. He serves as a (senior) program committee member/reviewer for several CV/AI conferences and journals, including CVPR, ICCV, ECCV, NeurIPS, ICLR, ICML, AAAI, IJCV, IEEE TPAMI, and ACM SIGGRAPH.



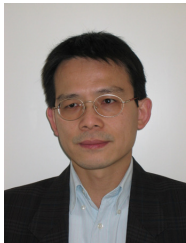
Zheng Dong is a Research Fellow at Zhejiang Sci-Tech University. He obtained the Ph.D. degree from the State Key Laboratory of CAD&CG, Zhejiang University in 2025. His research focuses on human-centric 3D vision and 2D image processing, including image-based 3D human reconstruction, human volumetric capture, and 2D image reflection removal. He serves as a reviewer for IJCV, NeurIPS, and CVPR.



Weiwei Xu is a Professor at the State Key Lab of CAD&CG in Zhejiang University. He was a Qianjiang Professor at Hangzhou Normal University and a researcher in Internet Graphics Group at Microsoft Research Asia from 2005 to 2012. He received his Ph.D. Degree in Computer Graphics from Zhejiang University. His research interests include 3D Reconstruction, Physical Simulation and Computer Animation, Computational Geometry, 3D Printing, and Virtual Reality. He has served as Program co-chair for ACM VRST 2013 and 2014, Committee member for Pacific Graphics, ACM Symposium on Geometry Processing, CASA, IEEE Virtual Reality, GMP, and SPM. He serves as a reviewer for ACM SIGGRAPH, ACM SIGGRAPH Asia, IEEE Visualization, and IEEE TVCG.



Gerhard Petrus Hancke is a Professor in the Department of Computer Science at City University of Hong Kong. He obtained his PhD in Computer Science with the Security Group at the University of Cambridge's Computer Laboratory in 2009. He is an IEEE Fellow since 2022, and is recognized as one of Stanford's top 2% most highly cited scientists since 2020.



Rynson W.H. Lau is a Professor in the Department of Computer Science at City University of Hong Kong (Dongguan Campus). He was on the faculty of Durham University and City University of Hong Kong. He received his Ph.D. degree from University of Cambridge.

Rynson serves on the Editorial Board of International Journal of Computer Vision (IJCV) and IET Computer Vision. He has served as the Guest Editor of a number of journal special issues, including ACM Trans. on Internet Technology, IEEE Trans. on Multimedia, IEEE Trans. on Visualization and Computer Graphics, and IEEE Computer Graphics & Applications. He has also served in the committee of a number of conferences, including Program Co-chair of ACM VRST 2004, ACM MTDL 2009, IEEE U-Media 2010, and Conference Co-chair of CASA 2005, ACM VRST 2005, ACM MDI 2009, ACM VRST 2014. His research interests include computer graphics and computer vision.