

# FLARE REMOVAL WITH VISUAL PROMPT

**Anonymous authors**

Paper under double-blind review

## ABSTRACT

Flare removal methods remove the streak, shimmer, and reflective flare in flare-corrupted images while preserving the light source. Recent deep learning methods focus on flare extraction and achieve promising results. They accomplish the task by either viewing the flare equals to the residual information between the flare-corrupted image and the flare-free image and generating the flare-free image through subtracting the extracted flare image or generating the flare-free image and the flare image simultaneously. However, due to the gap between the flare image and the residual information and handling flare extraction and clear image generation process simultaneously will give the network too much pressure and cannot fully utilize the extracted flare, these methods tend to generate images with severe artifacts. To alleviate such a phenomenon, we propose a model-agnostic pipeline named Prompt Inpainting Pipeline (PIP). Specifically, instead of viewing the gap between the flare-free and flare corrupted image as the flare or generating the flare-free image and flare image simultaneously, our prompt inpainting pipeline provides a novel perspective. We borrow the idea from inpainting methods and remove the flare by masking the polluted area and rewriting image details within. Unlike inpainting methods, we first extract multi-scale features of flare-corrupted images as a visual prompt and rewrite missing textures with the visual prompt since we find out that directly writing the missing details based on the remaining area hardly generates promising image details with sufficient semantic and high-frequency information. To verify the function of our pipeline, we conduct comprehensive experiments and demonstrate its superiority.

## 1 INTRODUCTION

Flare removal methods aim to remove the unwanted lens flare in input images while preserving the light source. In theory, an ideal camera should be able to converge all rays from a single point source to a single focal point. However, in reality, due to the sophisticated imaging system and dirt on the lens, lenses reflect and scatter light along unintended paths, leading to the appearance of flare that produces brightness in radial areas of images. The flare severely degrades the qualities of images and hinders downstream visual tasks (*e.g.*, semantic segmentation, depth estimation). The industry commonly adopts Anti-Reflection (AR) coatings to suppress the reflective flare phenomenon, which is costly and only optimized for specific wavelengths and angles of light and hardly handles the scattering flare that frequently happens on smartphone cameras. Therefore, how to remove flare in images has attracted plenty of attention from industry and academia Asha et al. (2019); Vitoria & Ballester (2019); Feng et al. (2023); Sun et al. (2020); Wu et al. (2021); Dai et al. (2022).

Researchers have made great efforts in this field. By viewing the residual information between the flare-corrupted image and the flare-free image equals to the flare image, Wu et al. Wu et al. (2021), Dai et al. Dai et al. (2022), and Dai et al. Dai et al. (2023b) employ a pipeline which adopt existing image restoration methods Zamir et al. (2021); Chen et al. (2021); Wang et al. (2022); Zamir et al. (2022) to extract the flare image and generate flare-free image by subtracting the estimated flare image with flare-corrupted image. However, there exists an obvious gap between the residual image and the flare image, and artifacts often happen. Moreover, Dai et al. Dai et al. (2023a) propose another pipeline named Flare7k++ and generate the flare image and flare-free image simultaneously. Such a method avoids the loss caused by the gap between the residual information and the flare image, whereas gives the network too much pressure and cannot fully utilize the extracted flare which leads to annoying artifacts on generated images. Thus, how to efficiently remove flare from images without introducing artifacts becomes challenging.

054 In this paper, we propose a model-agnostic pipeline named **Prompt Inpainting Pipeline (PIP)** which  
055 can be equipped with any image restoration methods. To eliminate generated artifacts, especially  
056 when handling images with strong streaks, we provide a novel perspective and separate the flare  
057 removal process into a coarse flare removal stage and an image refinement stage. The coarse flare  
058 removal stage adopts a U-shape network to coarsely generate the flare-removed image and the flare.  
059 As for the image refinement stage, we borrow the idea of inpainting methods, which mask the area  
060 polluted by flare and rewrite the missing details within. Nevertheless, instead of compelling the  
061 network to infer the missing pixels based on the remaining area, we propose prompt calibration  
062 block which adopts features extracted during the coarse flare removal stage as a visual prompt to  
063 guide the rewriting process and generate promising flare-free images. In this way, our PIP pipeline  
064 manages to erase the flare and decrease artifacts.

065 We conduct comprehensive experiments on the PIP pipeline and experimental results on real-world  
066 benchmarks prove the effectiveness of our method. Our contributions can be summarized as follows:

- 067 • We propose a model-agnostic pipeline named PIP which suppresses artifacts by rewriting  
068 details corrupted by flare.
- 069 • The proposed prompt calibration block adopts features extracted during the coarse flare  
070 removal stage as a visual prompt to guide the rewriting process.
- 071 • Extensive experiments demonstrate the superiority of our method.

## 074 2 RELATED WORK

075  
076 Flare is an optical phenomenon, in which light is scattered and/or reflected in a sophisticated optical  
077 system. The flare may seriously degrade the image’s quality and hinder downstream visual tasks.  
078 Thus, numerous efforts have been made to mitigate this issue.

### 080 2.1 HARDWARE SOLUTIONS

081 To capture high-quality images, modern digital optical systems tend to employ multiple glass ele-  
082 ments. The increase of the glass elements raises the probability that light reflects from its surface  
083 and generates flare. To eliminate the reflective flare, most hardware solutions focus on improving  
084 the optical system (*e.g.*, sophisticated lens barrel design, lens hoods, reflective coating). Applying  
085 AR coating Nussberger et al. (2016); Raskar et al. (2008) to lens elements reduces internal reflection  
086 by destroying interference, which is widely adopted in practice. However, AR coatings are costly to  
087 add to all optical surfaces and the thickness of the coating is designed for specific wavelength and  
088 angle of incidence. Moreover, scattering flare always occurs when dust and fingerprints appear in  
089 front of the lens. Therefore, these efforts hardly eliminate the flare generated in the pre-process and  
090 cannot handle images with flare during the post-process.

### 092 2.2 SOFTWARE SOLUTIONS

093 Traditional flare removal methods Faulkner et al. (1989); Reinhard et al. (2010) have achieved good  
094 performance based on statistics prior. Since the development of deep learning, numerous methods  
095 have been proposed.

#### 097 2.2.1 DATASETS

098  
099 Aiming for a low-level task, the performance of flare removal methods depends on the qualities of  
100 the datasets. A dataset that contains a large amount of data pairs can greatly improve the performance  
101 of neural networks when handling real-world scenarios.

102 To this end, Wu et al. Wu et al. (2021) propose a semi-synthetic dataset that contains 2001 captured  
103 flare images and 3000 simulated flare images. However, they focus on daylight flare and the pro-  
104 posed dataset tends to be less generalization to real-world flare which can be captured by diverse  
105 lenses and light sources, especially at nighttime. Therefore, Dai et al. Dai et al. (2022) propose a  
106 nighttime flare removal dataset Flare7K, which contains 5000 scattering and 2000 reflective flare  
107 images. Furthermore, Dai et al. [Dai et al. (2023a)] propose Flare7K++, which is an extended ver-  
sion of Flare7K. On the basis of Flare7K, Flare7K++ gives each image an additional light source

108 annotation and proposes a new real-captured subset Flare-R which contains 962 flare images. These  
109 datasets exhibit high sensitivity to scattering flares while giving insufficient attention to reflective  
110 flares. Hence, Dai et al. Dai et al. (2023b) propose the first reflective flare removal dataset named  
111 BracketFlare dataset based on the prior that the reflective flare and light source are always symmet-  
112 rical around the lens’s optical center. They employ continuous bracketing to capture the reflective  
113 flare pattern in unexposed images and aggregate with exposed images to synthesize paired data.  
114 They conduct experiments and prove that neural networks trained on this dataset gain the capability  
115 of removing the ghosting effect in images.

### 116 2.2.2 NETWORK STRUCTURE

117  
118 As capturing large amounts of data pairs for flare removal is challenging and tedious, earlier deep  
119 learning methods He et al. He et al. (2010) tend to utilize unsupervised methods. Qiao et al. Qiao  
120 et al. (2021) propose a generative adversarial network-based learning framework to learn from un-  
121 paired data. They adopt the idea of cycleGAN Zhu et al. (2017) and separately detect the light source  
122 region and the flare region. The output is generated by blending the flare-removed image and the de-  
123 tected light source mask. Their method achieves promising results when handling tiny light sources  
124 and flares, whereas fail on images with strong light sources and large flares.

125 As multiple synthetic and real flare removal datasets have been proposed, Wu et al. Wu et al. (2021)  
126 and Dai et al. Dai et al. (2022) adopt many end-to-end image restoration methods Ronneberger et al.  
127 (2015); Chen et al. (2021); Zamir et al. (2022); Wang et al. (2022) to extract the flare image from  
128 the input flare-corrupted image. They surpass the unsupervised methods, whereas still generate  
129 artifacts during the flare removal process as they view the gap between the flare-corrupted image  
130 and the flare-free image equal to the flare image. Meanwhile, Dai et al. Dai et al. (2023a) propose  
131 a different pipeline named flare7k++, which separately extracts the flare and restores the image in  
132 a simultaneous manner and adopts image restoration networks to do both jobs instead of simply  
133 generating the flare-free image. However, such a method compels the network to analyze the flare  
134 pattern and generate flare-free images simultaneously, which some artifacts still happen, especially  
135 when a strong streak appears. To this end, we propose a model-agnostic PIP pipeline to reduce the  
136 flare and generate images with fewer artifacts. By borrowing the idea of inpainting methods and  
137 separating the flare removal process into a coarse flare removal stage and an image refinement stage,  
138 our pipeline surpasses state-of-the-art methods.

## 139 3 METHODOLOGY

140 To better illustrate our PIP pipeline and distinguish it from the previous pipeline, we introduce the  
141 Flare7k++ pipeline first. Sequentially, we illustrate the details of our PIP pipeline and demonstrate  
142 its superiority.

### 143 3.1 FLARE7K++ PIPELINE

144  
145  
146 As shown in Figure 1 (a), the Flare7k++ pipeline adopts a network with a U-net backbone to predict  
147 the flare-free image and the flare image which excludes the light source information. In this way,  
148 the network can preserve the light source image and better locate the flare image by individually  
149 estimating the flare image. However, due to the pixel limit of digital systems, some information is  
150 blocked by overexposed streaks. Such degradation is hard to compensate for as the blocked area  
151 contains little useful information and the Flare7k++ pipeline requires the network to extract the  
152 flare and generate flare-free images simultaneously, thereby giving too much burden for the neural  
153 network and leading to severe artifacts. To this end, we propose our two-stage pipeline named PIP  
154 to reduce the artifacts.

### 155 3.2 PROMPT INPAINTING PIPELINE (OURS)

156  
157  
158 Our PIP pipeline provides a novel perspective for the flare removal task. Existing flare removal  
159 pipelines accomplish this task under the idea of image-to-image translation. However, such a design  
160 may be suboptimal for the flare removal task as the streak and shimmer in the flare-corrupted image  
161 occasionally appear to be strong and fully block the content embedded and these methods may

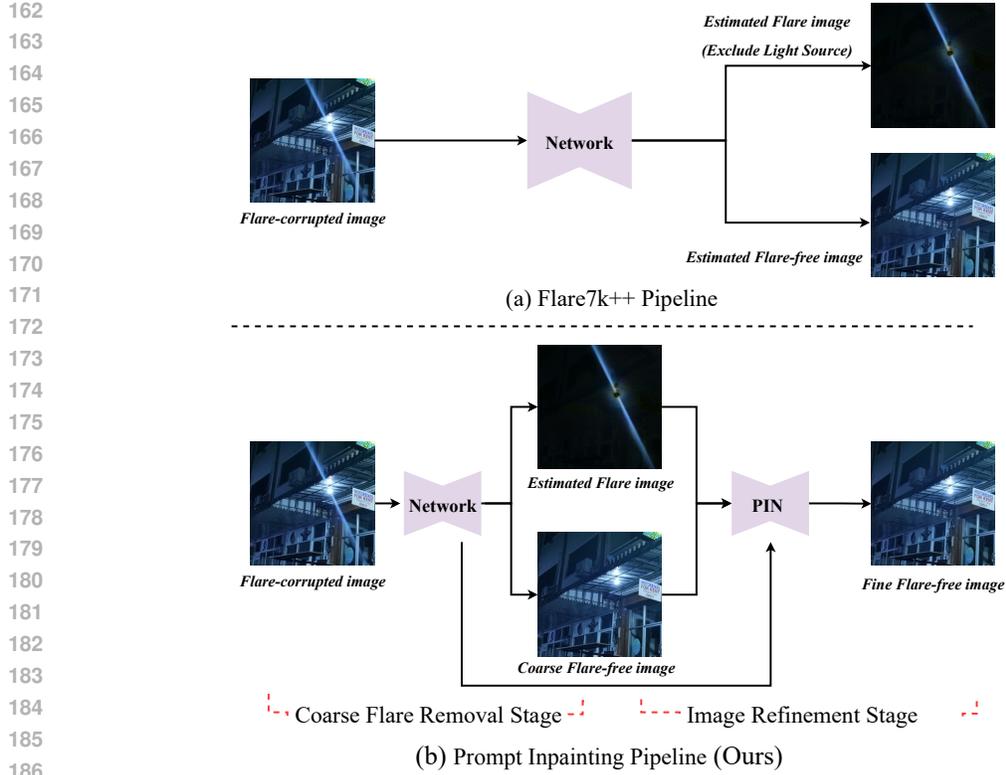


Figure 1: Comparison between Flare7k++ pipeline Dai et al. (2023a) and our PIP pipeline.

188 generate severe artifacts when handling such flare. Therefore, we propose the PIP pipeline, which  
189 accomplishes this mission by rewriting the image details occupied by the flare.

190 Figure 2 shows the details of the PIP pipeline. As a coarse-to-fine two-stage pipeline, the PIP  
191 pipeline consists of a coarse flare removal stage and an image refinement stage. Concretely, our PIP  
192 pipeline is a model-agnostic pipeline. During the coarse flare removal stage, any U-net flare removal  
193 method can be employed as multi-scale features are useful in such image translation tasks, and most  
194 methods in this task adopt one U-net structure as the backbone to extract multi-scale features. As for  
195 the image refinement stage, we propose the Prompt Inpainting Network (PIN) and employ a U-net  
196 backbone under the guidance of the multi-scale features extracted during the first stage.

### 198 3.2.1 COARSE FLARE REMOVAL STAGE

199 In this stage, we coarsely remove the flare from the flare-corrupted image by estimating the coarse  
200 flare-corrupted image and the entire flare image. The loss function is formulated as:

$$202 \mathcal{L} = \begin{cases} \frac{1}{N} \sum_{i=1}^N |I_{FF}^{i, Coarse} - I_{gt}^i| \\ \alpha \frac{1}{N} \sum_{i=1}^N |F_{DF}^i - F_{gt}^i| \end{cases} \quad (1)$$

208 where  $I_{FF}^{i, Coarse}$ ,  $F_{DF}^i$ , and  $F_{gt}^i$  represent the  $i^{th}$  pixel in the output coarse flare-free image, the  
209 output flare image, and the ground truth image.

### 211 3.2.2 IMAGE REFINEMENT STAGE

212 After obtaining the flare image which is added on top of the flare-free image and the coarse flare-free  
213 image, we introduce the image refinement stage to further remove the flare and artifacts from the  
214 coarse flare-free image. We do not adopt an image-to-image translation network as we argue that  
215 such a design hardly handles severe degradation. In the image restoration task, simply adopting such

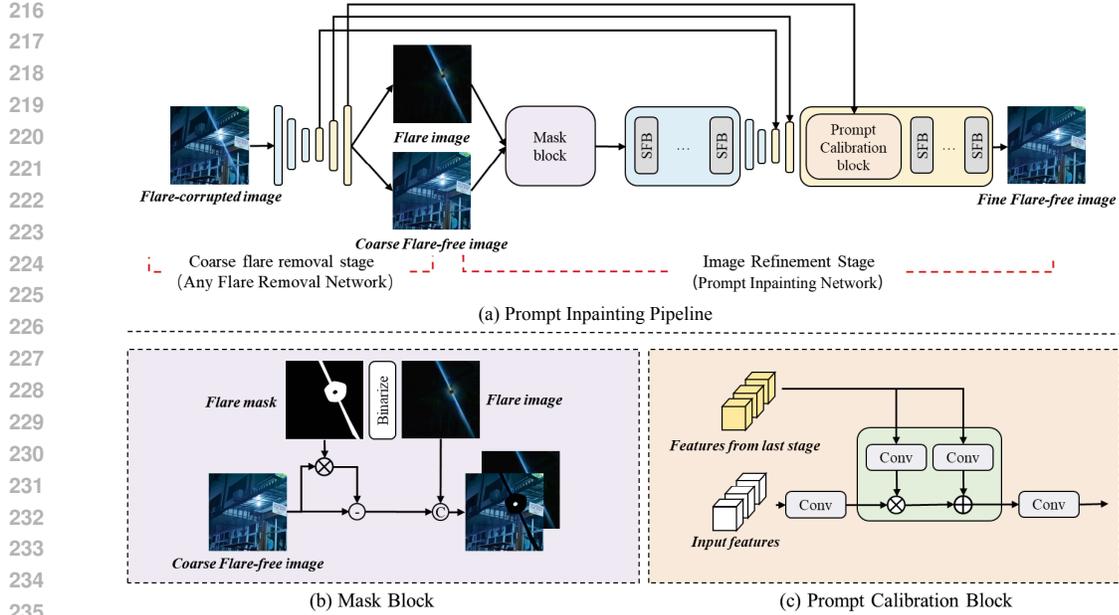


Figure 2: The overview of our PIP pipeline. Our PIP pipeline consists of a coarse flare removal stage and an image refinement stage. The coarse flare removal coarsely estimates the flare and removes the flare from the flare-corrupted image. The image refinement stage further removes the artifacts introduced during the coarse flare removal stage. We rewrite the missing details polluted by the flare image and employ a U-net structure as our backbone to eliminate the flare under the guidance of the multi-scale feature extracted from the last stage.

a network on image restoration suffers from unpromising results when handling severe degradation (*e.g.*, complicated motion blur, high-level noise). To this end, we accomplish this task from a novel view by extracting the semantic information from the unpolluted area, rewriting the details in the polluted area based on the extracted semantic information, and using multi-scale features extracted from the last stage as the visual prompt. Specifically, we propose the PIN network in this stage which adopts the mask block to generate the input and a U-net neural network to rewrite the missing details.

**The mask block:** The procedure of mask block is depicted in Figure 2 (b). Given the estimated flare image, we first binarize it to obtain the mask with a predefined threshold and multiply it with the coarse flare-free image  $I_{FF}^{Coarse}$  to obtain the masked image  $I_{FF}^{mask}$ . We select the coarse-generated flare-free images instead of the flare-corrupted image as the estimated mask omits the flare with slight luminance and using the flare-corrupted image may leave these flare on the final output. Secondly, we subtract  $I_{FF}^{mask}$  with  $I_{FF}^{Coarse}$ . Finally, we generate the input  $I_{FF}^{Input}$  by concatenating the image obtained in the last phase with the estimated flare image  $F_{DF}$ . The process is formulated as:

$$\begin{aligned} Mask &= \text{Binarize}(Mask > threshold), & I_{FF}^{mask} &= I_{FF}^{Coarse} \times Mask, \\ I_{FF}^{mask} &= I_{FF}^{Coarse} - I_{FF}^{mask}, & I_{FF}^{Input} &= \text{Concat}(I_{FF}^{mask}, F_{DF}) \end{aligned} \quad (2)$$

**U-shape Structure:** Given  $I_{FF}^{Input}$  and features obtained from the last stage  $X_{FF}^{Coarse}$ , we adopt a U-net structure for the inpainting task. Firstly, the encoder in the backbone employs 4 SFB blocks Zhang et al. (2023) at each level and extracts the multi-scale semantic feature of  $I_{FF}^{Input}$ . Thereafter, the decoder rewrites the missing details by using  $X_{FF}^{Coarse}$  as the visual prompt. Concretely, we adopt one prompt calibration block and 4 SFB blocks at each level in the decoder and the prompt calibration block is sequentially employed to utilize  $X_{FF}^{Coarse}$ .

**The prompt calibration block:** The prompt calibration block refines the image features by employing the  $X_{FF}^{Coarse}$  as the visual prompt and the structure is depicted in Figure 2 (c). The formulation of the prompt calibration block is presented as follows:

$$X^i = \text{Conv}(X_{FF}^{i,Coarse}) \times \text{Conv}(X^i) + \text{Conv}(X_{FF}^{i,Coarse}) \quad (3)$$

where  $i$  represents the  $i^{th}$  level in decoder.

The loss function of this stage is the L1 distance and perceptual loss between the output and the ground truth flare-free image. The formulation is shown as follows:

$$\mathcal{L} = \begin{cases} \frac{1}{N} \sum_{i=1}^N |I_{FF}^{i, fine} - I_{gt}^i| \\ \mathcal{L}_{per}(I_{FF}^{i, fine}, I_{gt}^i) \end{cases} \quad (4)$$

where  $I_{FF}^{i, fine}$  represents the output in refinement stage and  $\mathcal{L}_{per}$  means the perceptual loss.

## 4 EXPERIMENTS

In this section, we conduct ablation experiments on modules proposed within the PIP pipeline on real-world data. Moreover, we conduct comparison experiments on the Flare7K++ real dataset Dai et al. (2023a) and BracketFlare dataset Dai et al. (2023b).

### 4.1 DATASETS

For the training set, we adopt Flare7K++ Dai et al. (2023a) and BracketFlare Dai et al. (2023b) datasets. Notably, Flare7K and Flare7K++ provide a quantity of flare images, while giving no flare-free images in the training set. We follow the experimental setting in Flare7K which adopts the Flickr24k dataset Zhang et al. (2018) as flare-free images. To form the data pair for supervised training, we add compound flare images on top of flare-free images to generate nighttime flare-corrupted images. As for flare-free images, we add light source annotations on flare-free images. The same operation is implemented on the BracketFlare dataset.

To improve the robustness of trained networks, we further introduce a data augmentation strategy. Specifically, to allow the network to perform better in nighttime images, we alter the gamma of flare-corrupted images by recovering the linear luminance of flare images and flare-free images with an inverse gamma correction strategy ( $\gamma \sim U(1.8, 2.2)$ ). We also randomly multiply the RGB values with  $U(0.5, 1.2)$  and add a Gaussian noise with variance sampled from a scaled chi-square distribution  $\sigma^2 \sim 0.01\chi^2$ . Furthermore, we use horizontal and vertical flips to enlarge the training set. Nevertheless, to enhance the network’s abilities to deal with diverse flare, we carry out a series of affine transformations on flare images. The random blur on flare images with the kernel size in  $U(0.1, 3)$  and offsets in  $U(-0.02, 0.02)$  are also performed to control the brightness of flare images.

For the testing set, we adopt the real testing set provided by Flare7K++ and BracketFlare as the Flare7K++ real set mainly contains scattering flare and BracketFlare focuses on reflective flare. We conduct PIP pipeline on them for ablation experiments and comparison experiments.

### 4.2 IMPLEMENTATION DETAILS

Following Dai et al. Dai et al. (2022), we adopt PSNR, SSIM, and LPIPS as our metrics. Specifically, PSNR and SSIM evaluate the average pixel distance and the structure similarity between flare-corrupted images and flare-free images. Furthermore, LPIPS evaluates the feature-level distance between them. We set 4 levels in the PIN network. The channel of the immediate layer is 16. The optimizer for the PIP pipeline adopts the same setting as the FF-Former Zhang et al. (2023).

### 4.3 ABLATION RESULTS

In this section, we explore the function of the prompt calibration block and whether our pipeline is model-agnostic. Particularly, we do not conduct ablation experiments on our mask block. Without the mask block, our PIP pipeline will become a stacked U-net network. The pipeline will achieve better results by a deeper network with no doubt, whereas fails to suppress artifacts.

Table 1: Comparison of methods with or without prompt calibration block (PCB) Table 2: The number of parameters and Flops (512×512) with or without Prompt Inpainting Pipeline (PIP)

Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	Param(M)/Flops(G)	Unet	Uformer	FF-Former
w/o PCB	28.14	0.902	0.041	w/o PIP	9.0/62.36	20.4/162.1	46.5/407.8
w PCB	<b>28.44</b>	<b>0.906</b>	<b>0.039</b>	w PIP	11.8/84.9	23.6/194.3	49.6/440.0

324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377

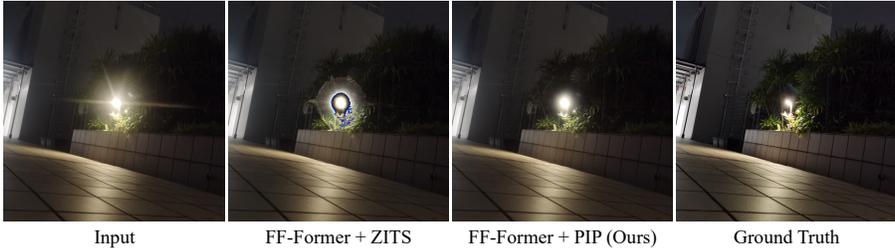


Figure 3: Visual comparison results on PIP pipeline VS inpainting network ZITS Dong et al. (2022). Severe artifacts happen in the image generated by FF-Former + ZITS.



Figure 4: Visual comparison results of ablation experiments on PIP pipeline. We choose different images to better illustrate the function of our PIP as different network has weaknesses in different scenarios. Our PIP manages to further suppress artifacts, especially with a weaker network (Unet).

#### 4.3.1 PROMPT CALIBRATION BLOCK

To verify the function of the proposed prompt calibration block and whether the visual prompt is necessary, we test how the pipeline will perform when no prompt calibration block are attended or other inpainting methods participate. To be fair, we adopt FF-Former as the base network in the coarse flare removal stage and test on the flare7k++ real set.

We show the comparison results on networks with or without prompt calibration block on Table 1. From the observation of Table 1, the network with the prompt calibration block surpasses the network without the prompt calibration block 0.3dB on PSNR, which proves that the visual prompt is non-substitutable and the prompt calibration block can provide significant PSNR score gains.

To further test whether other inpainting method can be useful for the flare removal task, we also compare it with ZITS Dong et al. (2022) method and adopt the official code for inference. The visual performance of the comparison is shown in Figure 3. Severe artifacts happen in the image generated by ZITS as the mask area is too complicated and uncertain to recover when much information has been discarded by the simple mask operation and no visual prompt can be used.

#### 4.3.2 MODEL AGNOSTIC

To demonstrate that our PIP is model-agnostic and can be applied with any other flare removal network, we conduct quantitative and qualitative ablation experiments on Unet Ronneberger et al. (2015), Uformer Wang et al. (2022), and FF-Former Zhang et al. (2023). The network in the coarse flare removal stage is pretrained for predicting the coarse flare-free image and the flare image on Flare7K++ Dai et al. (2023a) dataset.

For quantitative comparison, based on Table 3, we can observe that with PIP pipeline, Unet Ronneberger et al. (2015) gains 0.92dB, 0.007, 0.004 on PSNR, SSIM, and LPIPS scores which surpass the state-of-the-art method FF-Former with 0.21dB, 0.001 on PSNR and SSIM score and achieve the same LPIPS score with it. Moreover, our PIP pipeline helps Uformer with 0.61dB, 0.006, and 0.003 gains on PSNR, SSIM, and LPIPS scores. As for the state-of-the-art method FF-Former Zhang et al. (2023), equipped with our pipeline, it manages to obtain new state-of-the-art results, which achieve 28.44, 0.906, and 0.039 on PSNR, SSIM, and LPIPS scores and our pipeline provides 0.54, 0.006, and 0.002 promotions with a very small increase in the number of parameters, as shown in Table 2.

As for qualitative comparison, we show the visual comparisons on Figure 4. Compared with the original Unet, Unet + PIP pipeline removes more shimmer within the flare-corrupted images. As

Table 3: Comparison of flare removal methods on Flare7k++ real set and BracKetFlare dataset.

Method	Flare7K			BracKetflare		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
Wu et al. Wu et al. (2021)	24.61	0.871	0.060	26.13	0.895	0.055
Unet Ronneberger et al. (2015)	27.19	0.894	0.045	47.00	0.897	0.006
HINet Chen et al. (2021)	27.55	0.892	0.046	48.03	0.994	0.003
MPRNet* Zamir et al. (2021)	27.04	0.893	0.048	48.41	0.994	0.004
Restormer* Zamir et al. (2022)	27.60	0.897	0.045	48.11	0.994	0.004
Uformer Wang et al. (2022)	27.63	0.894	0.043	47.47	0.991	0.003
FF-Former Zhang et al. (2023)	27.90	0.900	0.041	49.03	0.992	0.003
Unet + PIP (Ours)	28.11	0.901	0.041	48.50	0.992	0.004
Uformer + PIP (Ours)	28.24	0.903	0.040	48.69	0.994	0.002
FF-former + PIP (Ours)	<b>28.44</b>	<b>0.906</b>	<b>0.039</b>	<b>49.18</b>	<b>0.994</b>	<b>0.002</b>

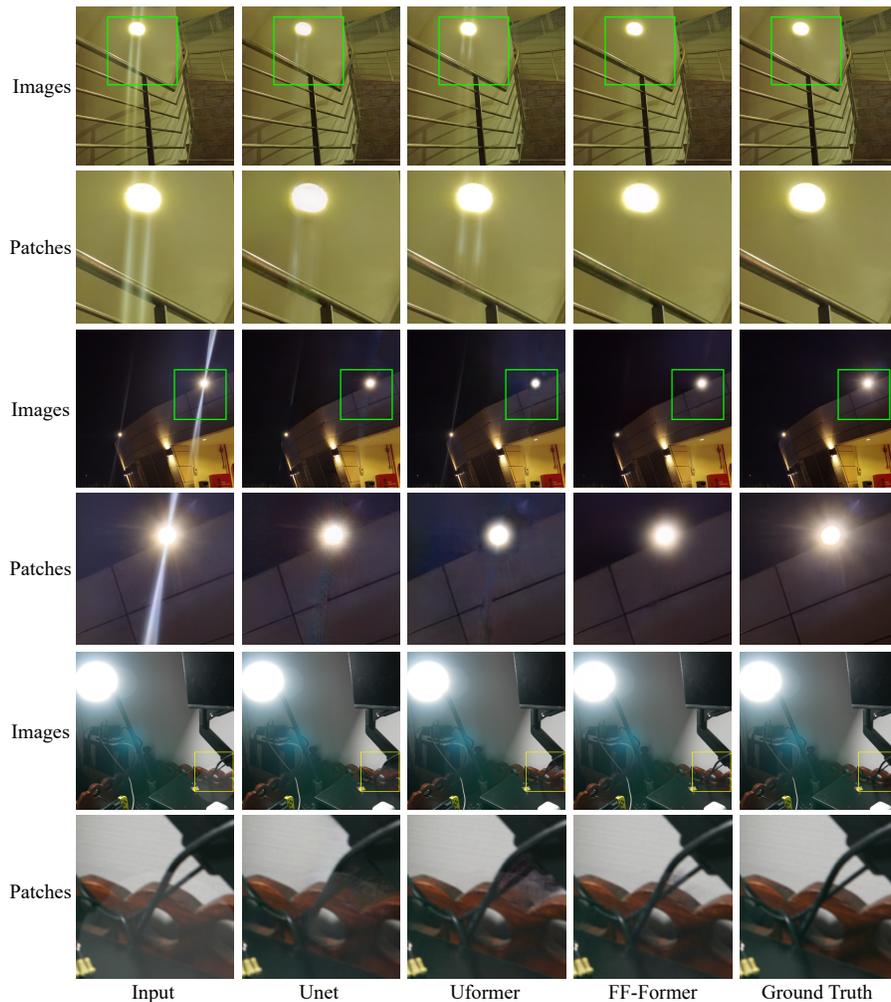


Figure 5: Visual comparison of the scattering flare and the reflective flare. Images in the first four rows are from the flare7k++ real set and mainly contain scattering flare. Images in the last two rows are from the BraceketFlare dataset and mainly contain reflective flare. Our method eliminates the most flare and generates images with the fewest artifacts.

for FF-Former, more streaks have been eliminated. Previous pipelines hardly identify and eliminate all the flares added to the image and tend to generate severe artifacts when handling strong flares with large overexposed streaks and shimmers. Our PIP pipeline suppresses this phenomenon by rewriting the image details polluted by flare with the extracted semantic information within the unpolluted area and the multi-scale features obtained in the coarse flare removal stage. The experiments demonstrate that our pipeline is model-agnostic.

#### 4.4 COMPARISON RESULTS

In this section, we conduct comparison results on the real testing set of the Flare7K++ dataset and BracketFlare dataset to demonstrate the functions of our PIP pipeline. We employ flare removal methods Wu et al. (2021) and FF-Former Zhang et al. (2023) and image restoration methods Unet Ronneberger et al. (2015), HINet Chen et al. (2021), MPRNet Zamir et al. (2021), Restormer Zamir et al. (2022), and Uformer Wang et al. (2022) for baselines.

##### 4.4.1 RESULTS ON FLARE7K++ DATASETS

To validate the performance of our method on complex scenarios which include both the scattering flare and reflective flare, we show the quantitative results (*e.g.*, PSNR, SSIM, LPIPS) on Table 3 for experiments on Flare7k++ real testing set, correspondingly. Furthermore, we also represent the visual comparison results on Figure 5 for experiments on the Flare7k++ real testing set (we employ FF-Former Zhang et al. (2023) in coarse flare removal network).

From the observation of Table 3, our PIP pipeline equipped with FF-Former Zhang et al. (2023) has achieved state-of-the-art PSNR and SSIM. Concretely, from the perspective of pixel level, we significantly outperform the state-of-the-art method FF-Former with 0.54dB on PSNR score, 0.006 on SSIM, and 0.002 on LPIPS. As for visual performance comparison, based on the first four rows in Figure 5, we figure that methods trained in a supervised manner achieve better results and our PIP pipeline equipped with FF-Former has essentially eliminated the flare and introduced the least artifacts in most situation. Despite the shape of the flare, round flare with a large radius, flare with a long streak, or the type of the flare, reflective flare, or scattering flare, our method can achieve more realistic and natural results than other baselines in most scenarios.

##### 4.4.2 RESULTS ON BRACKETFLARE DATASETS

To further estimate our method on reflective flare, we test our method on the BracKetFlare dataset Dai et al. (2023b). From the observation of Table 3, our PIP pipeline equipped with FF-Former surpasses MPRNet with 0.77dB on PSNR and 0.002 on LPIPS and obtains the same SSIM. Furthermore, based on the last two rows in Figure 5, we significantly outperform MPRNet by thoroughly eliminating the reflective flare with the fewest artifacts, which proves that some artifacts may occur during the reflective flare removal process and our PIP manages to depress such phenomenon.

## 5 DISCUSSION

To suppress the artifacts created during the flare removal process, we propose the two-stage pipeline named PIP. The PIP pipeline borrows the idea from the inpainting task and adopts a mask block to construct the input of the inpainting neural network. However, as shown in Figure 3, simply adopting the inpainting method will lead to essential content loss as plenty of the image content is hidden by the flare, which leads the current inpainting method hardly recovers them and the extracted flare mask is error-prone, which makes the inpainting method mistakenly rewrites the flare instead of image content. Therefore, we adjust the inpainting method for the flare removal task. By adopting the features extracted from the coarse flare removal stage as inference, the prompt calibration block in the image refinement stage manages to rewrite the blocked image content.

## 6 CONCLUSION

In this paper, we propose a model-agnostic two-stage pipeline named PIP pipeline. By giving the flare removal task a second thought, the PIP pipeline provides a novel view for the flare removal task by rewriting the pixels corrupted by the flare. The PIP pipeline consists of a coarse flare removal stage and an image refinement stage. The coarse flare removal stage generates coarse flare-free images and the estimated flare image and the image refinement stage alleviates the artifacts by adopting the idea of inpainting methods. However, instead of compelling the network to infer the masked image, we accomplish the mission by extracting the semantic information of the unmasked area and employing multi-scale features as visual prompts. Extensive results on real-world and synthetic benchmarks demonstrate the superiority of the PIP pipeline.

## REFERENCES

- 486  
487  
488 CS Asha, Sooraj Kumar Bhat, Deepa Nayak, and Chaithra Bhat. Auto removal of bright spot from  
489 images captured against flashing light source. In *2019 IEEE International Conference on Dis-*  
490 *tributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*, pp. 1–6. IEEE, 2019.
- 491 Liangyu Chen, Xin Lu, Jie Zhang, Xiaojie Chu, and Chengpeng Chen. Hinet: Half instance normal-  
492 ization network for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer*  
493 *Vision and Pattern Recognition*, pp. 182–192, 2021.
- 494 Yuekun Dai, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Flare7k: A  
495 phenomenological nighttime flare removal dataset. *Advances in Neural Information Processing*  
496 *Systems*, 35:3926–3937, 2022.
- 497 Yuekun Dai, Chongyi Li, Shangchen Zhou, Ruicheng Feng, Yihang Luo, and Chen Change Loy.  
498 Flare7k++: Mixing synthetic and real datasets for nighttime flare removal and beyond. *arXiv*  
499 *preprint arXiv:2306.04236*, 2023a.
- 500 Yuekun Dai, Yihang Luo, Shangchen Zhou, Chongyi Li, and Chen Change Loy. Nighttime  
501 smartphone reflective flare removal using optical center symmetry prior. In *Proceedings of the*  
502 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20783–20791, 2023b.
- 503 Qiaole Dong, Chenjie Cao, and Yanwei Fu. Incremental transformer structure enhanced image  
504 inpainting with masking positional encoding. In *Proceedings of the IEEE/CVF Conference on*  
505 *Computer Vision and Pattern Recognition*, pp. 11358–11368, 2022.
- 506 K Faulkner, CJ Kotre, and M Louka. Veiling glare deconvolution of images produced by x-ray  
507 image intensifiers. In *Third International Conference on Image Processing and its Applications,*  
508 *1989.*, pp. 669–673. IET, 1989.
- 509 Ruicheng Feng, Chongyi Li, Huaijin Chen, Shuai Li, Jinwei Gu, and Chen Change Loy. Generating  
510 aligned pseudo-supervision from non-aligned data for image restoration in under-display camera.  
511 In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.  
512 5013–5022, 2023.
- 513 Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE*  
514 *transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010.
- 515 Andreas Nussberger, Helmut Grabner, and Luc Van Gool. Robust aerial object tracking from an  
516 airborne platform. *IEEE Aerospace and Electronic Systems Magazine*, 31(7):38–46, 2016.
- 517 Xiaotian Qiao, Gerhard P Hancke, and Rynson WH Lau. Light source guided single-image flare  
518 removal from unpaired data. In *Proceedings of the IEEE/CVF International Conference on Com-*  
519 *puter Vision*, pp. 4177–4185, 2021.
- 520 Ramesh Raskar, Amit Agrawal, Cyrus A Wilson, and Ashok Veeraraghavan. Glare aware photog-  
521 raphy: 4d ray sampling for reducing glare effects of camera lenses. In *ACM SIGGRAPH 2008*  
522 *papers*, pp. 1–10. 2008.
- 523 Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol  
524 Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting*. Mor-  
525 gan Kaufmann, 2010.
- 526 Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomed-  
527 ical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention-*  
528 *MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceed-*  
529 *ings, Part III 18*, pp. 234–241. Springer, 2015.
- 530 Qilin Sun, Ethan Tseng, Qiang Fu, Wolfgang Heidrich, and Felix Heide. Learning rank-1 diffractive  
531 optics for single-shot high dynamic range imaging. In *Proceedings of the IEEE/CVF conference*  
532 *on computer vision and pattern recognition*, pp. 1386–1396, 2020.
- 533 Patricia Vitoria and Coloma Ballester. Automatic flare spot artifact detection and removal in pho-  
534 tographs. *Journal of Mathematical Imaging and Vision*, 61(4):515–533, 2019.

540 Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li.  
541 Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF*  
542 *conference on computer vision and pattern recognition*, pp. 17683–17693, 2022.

543  
544 Yicheng Wu, Qiurui He, Tianfan Xue, Rahul Garg, Jiawen Chen, Ashok Veeraraghavan, and  
545 Jonathan T Barron. How to train neural networks for flare removal. In *Proceedings of the*  
546 *IEEE/CVF International Conference on Computer Vision*, pp. 2239–2247, 2021.

547 Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-  
548 Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the*  
549 *IEEE/CVF conference on computer vision and pattern recognition*, pp. 14821–14831, 2021.

550 Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-  
551 Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceed-*  
552 *ings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5728–5739,  
553 2022.

554  
555 Dafeng Zhang, Jia Ouyang, Guanqun Liu, Xiaobing Wang, Xiangyu Kong, and Zhezhu Jin. Ff-  
556 former: Swin fourier transformer for nighttime flare removal. In *Proceedings of the IEEE/CVF*  
557 *Conference on Computer Vision and Pattern Recognition*, pp. 2823–2831, 2023.

558 Xuaner Zhang, Ren Ng, and Qifeng Chen. Single image reflection separation with perceptual losses.  
559 In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4786–  
560 4794, 2018.

561  
562 Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation  
563 using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference*  
564 *on computer vision*, pp. 2223–2232, 2017.

565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593