大湾区大学(筹) GREAT BAY UNIVERSITY
(东莞市大湾区高等研究院)

Difflare: Removing Image Lens Flare with Latent Diffusion Models

Tianwen Zhou<tianwenzhou0521@ieee.org>, Qihao Duan, Zitong Yu*

Abstract

High-quality image recovery from lens flare (Figure 1) is a key challenge in low-level vision. Current deep learning models are trained from scratch, often ignoring generative priors in pre-trained models, and overlook physical priors crucial for effective flare removal. We present **Difflare**, a novel approach that utilizes Pre-Trained Diffusion Models (PTDM) through a Structural Guidance Injection Module (SGIM) for guided restoration and operates in latent space for efficiency. An Adaptive Feature Fusion Module (AFFM) further integrates Luminance Gradient Prior (LGP) to manage feature extraction dynamically. Extensive experiments show Difflare achieves state-of-the-art results in real-world lens flare removal with high fidelity and perceptual quality.

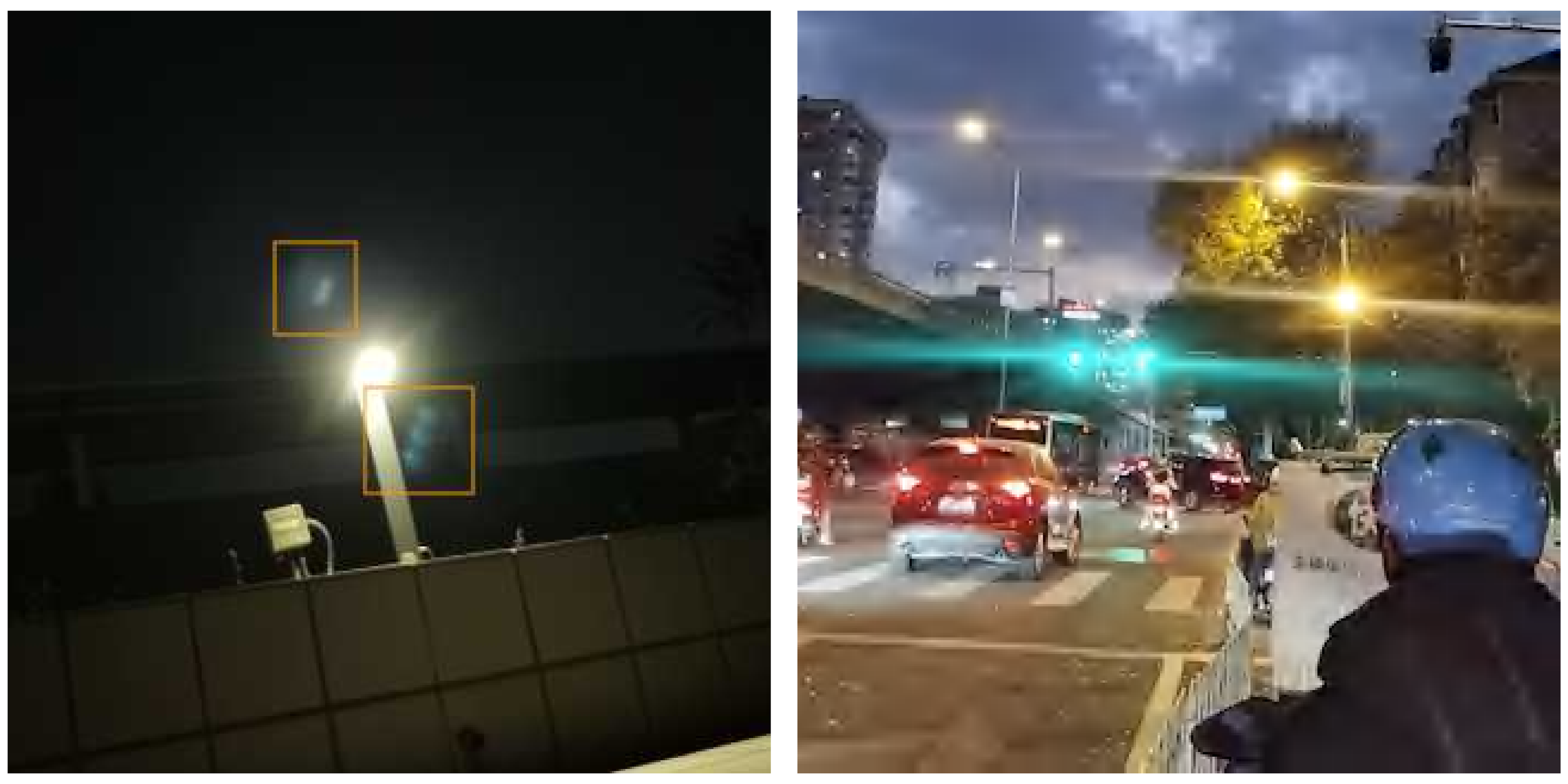

Figure1 Reflective Flare (left) and Scattering Flare (right) in the wild

Methods

Difflare leverages the generative prior in pre-trained latent diffusion models. To minimize training costs, we use a frozen VQ-GAN encoder to compress input images, allowing flare removal in latent space. Our method includes **PTDM fine-tuning** and **fidelity preservation** through lens flare physical priors. Figure 2 provides an overview.

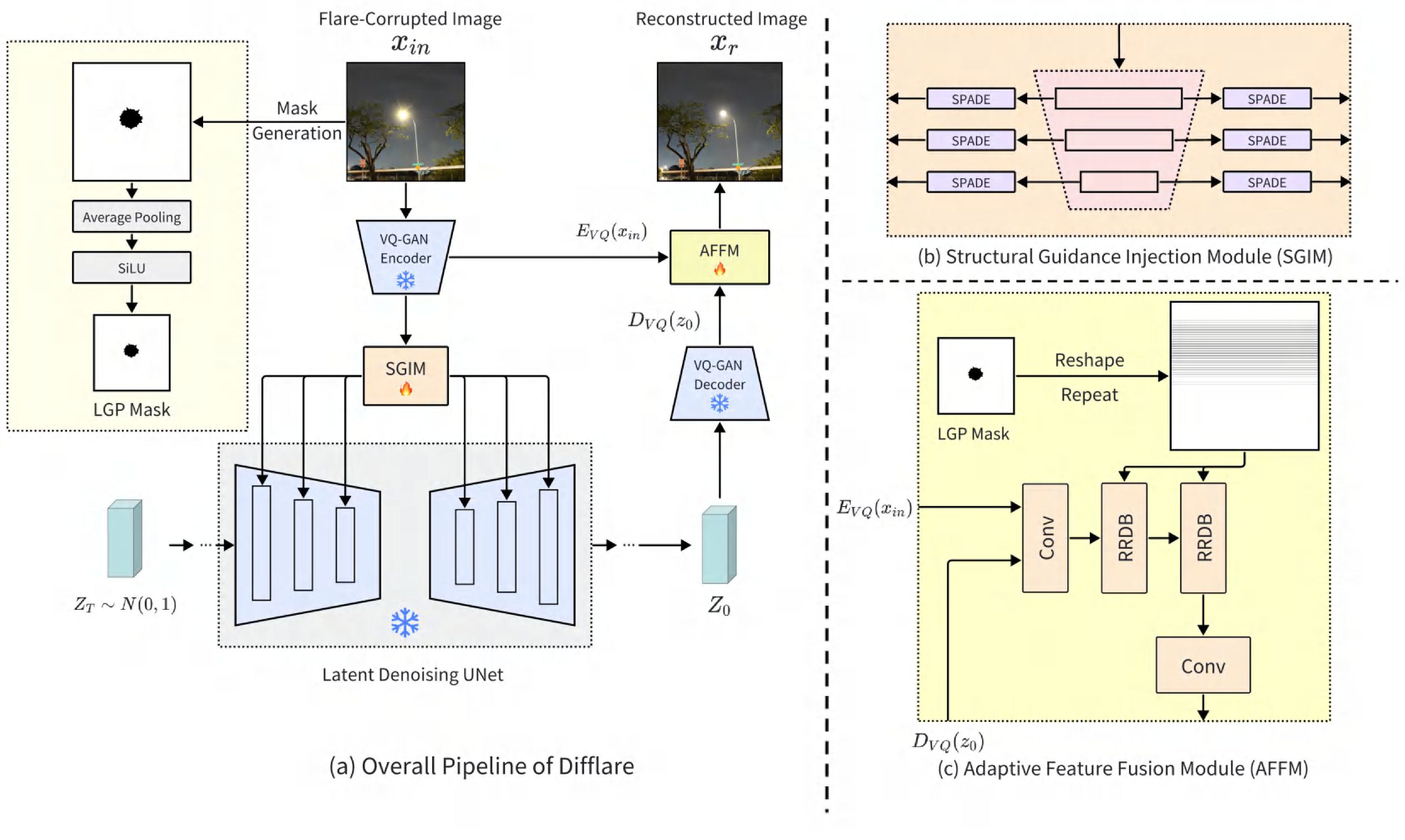

Figure2 Overview of Difflare

Injecting Structural Guidance To adapt pre-trained diffusion models (PTDM) for flare removal, Difflare fine-tunes PTDM to use flare-corrupted images as conditional inputs. Existing methods like ControlNet and LoRA only use sketches for guidance. Instead, we adopt a Structural Guidance Injection Module (SGIM) that extracts multi-scale structural information from the input, embedding semantic maps into PTDM's residual blocks. This maintains PTDM's generative prior while reducing fine-tuning costs.

$$Fea_i = E_{SGIM}(z_{in})_i, \quad Fea_i' = \beta_i \oplus (\gamma_i + 1) \otimes Fea_i, \quad i = 1, \dots, L$$

where γ_i, β_i stands for two trainable convolution layers, $\oplus$ stands for concatenation, and $\otimes$ stands for multiplication.

Adaptive Feature Fusion (AFFM) The Adaptive Feature Fusion Module (AFFM) preserves fidelity in flare-free areas. Leveraging a Luminance Gradient Prior (LGP), we generate a Luminance Mask that focuses on high-luminance regions for precise feature extraction. By integrating these guided features and adjusting self-attention modules, we enhance image restoration fidelity.

Results

Experimental Settings We use Stable-Diffusion v2.1 as the base model for Difflare, trained on the Flare7K dataset. Paired datasets are generated by randomly selecting background images B, reflective flares, compound scattering flares, and light sources. SGIM is trained for 85 epochs with a batch size of 192, while AFFM is trained for 11 epochs with a batch size of 48, evaluated on a test set of 100 real-world images, see Figure 3 for visual results.

(a) Input (b) Zhou *et al.* (c) Dai *et al.* (d) Ours (e) Groundtruth

Qualitative Comparison Difflare is compared quantitatively and qualitatively with methods by Zhang et al., Wu et al., Zhou et al., and Dai et al. We test publicly available pre-trained models or retrain others for fairness.

Quantitative Comparison We use PSNR, SSIM, MUSIQ, and CLIPIQA as metrics (Table 1). Zhang et al.'s method effectively removes lens flare but distorts colors. Wu et al. and Zhou et al. preserve light sources but yield oversharpened images. Dai et al. achieves high PSNR but may remove light sources, affecting quality. Our method excels in structural similarity and perceptual quality due to the AFFM and PTDM.

Metrics	Input	Wu	Zhang	Zhou	Dai	Difflare (Ours)
PSNR ↑	22.561	24.613	21.022	25.184	**26.978**	26.063
SSIM ↑	0.857	0.871	0.784	0.872	0.890	**0.898**
MUSIQ ↑	59.34	57.29	55.46	59.09	59.03	**59.48**
CLIPIQA ↑	0.332	0.312	0.279	0.281	0.337	**0.341**

Table1 Quantitative Comparison

Reference

- Yicheng Wu, Qiurui He, Tianfan Xue, Rahul Garg, Jiawen Chen, Ashok Veeraraghavan, and Jonathan T. Barron. How to train neural networks for flare removal. 2020.
- Yuekun Dai, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Flare7k: A phenomenological nighttime flare removal dataset. In Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track, 2022.

Homepage: https://zhoutianwen.com