

Fourier Priors-Guided Diffusion for Zero-Shot Joint Low-Light Enhancement and Deblurring

Xiaoqian Lv¹, Shengping Zhang^{*,1,2}, Chenyang Wang¹, Yichen Zheng³,
 Bineng Zhong⁴, Chongyi Li^{5,6}, Liqiang Nie¹

¹Harbin Institute of Technology ²Peng Cheng Laboratory

³Huazhong University of Science and Technology ⁴Guangxi Normal University

⁵NKIARI, Shenzhen Futian ⁶VCIP, CS, Nankai University

{xiaoqian.lv, c.wang}@stu.hit.edu.cn, s.zhang@hit.edu.cn,

{yichenzheng02, nieliqiang}@gmail.com, bnzhong@gxnu.edu.cn, lichongyi@nankai.edu.cn

Abstract

Existing joint low-light enhancement and deblurring methods learn pixel-wise mappings from paired synthetic data, which results in limited generalization in real-world scenes. While some studies explore the rich generative prior of pre-trained diffusion models, they typically rely on the assumed degradation process and cannot handle unknown real-world degradations well. To address these problems, we propose a novel zero-shot framework, *FourierDiff*, which embeds Fourier priors into a pre-trained diffusion model to harmoniously handle the joint degradation of luminance and structures. *FourierDiff* is appealing in its relaxed requirements on paired training data and degradation assumptions. The key zero-shot insight is motivated by image characteristics in the Fourier domain: most luminance information concentrates on amplitudes while structure and content information are closely related to phases. Based on this observation, we decompose the sampled results of the reverse diffusion process in the Fourier domain and take advantage of the amplitude of the generative prior to align the enhanced brightness with the distribution of natural images. To yield a sharp and content-consistent enhanced result, we further design a spatial-frequency alternating optimization strategy to progressively refine the phase of the input. Extensive experiments demonstrate the superior effectiveness of the proposed method, especially in real-world scenes. The code is available at <https://github.com/aipixel/FourierDiff>.

1. Introduction

In night photography, long exposure is commonly used to capture more light, which inevitably causes motion blurs

* Corresponding author.

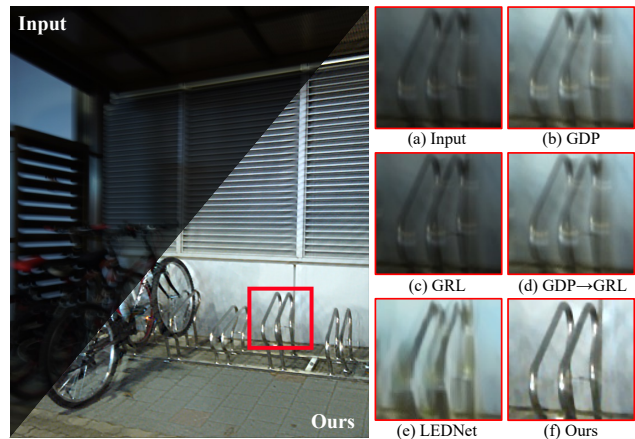


Figure 1. Visual comparisons of the state-of-the-art diffusion-based low-light enhancement method GDP [17], deblurring method GRL [33], and joint low-light enhancement and deblurring method LEDNet [77] on a low-light blurry image. Existing methods (b)-(e) fail to cope with the real-world night blurry image. In contrast, the proposed *FourierDiff* (f) yields a visually pleasing result with more natural brightness and sharper textures. Moreover, our method does not require paired training data.

due to camera shake and object motion. Even under long exposure settings, images taken in low-light conditions still suffer from limited visibility, low contrast, and distorted color. Therefore, both low light and motion blur naturally co-exist in images captured in low-light environments, which not only affects the visual quality but also limits the performance of high-level tasks such as object detection [14, 34] and action recognition [9, 10].

With the recent advances in deep learning, numerous low-light enhancement [3, 19, 30, 39] and deblurring [11, 26, 33, 71] methods have been proposed. Although these methods cope well with their specific task individually, they are still far from satisfactory in handling the joint degra-

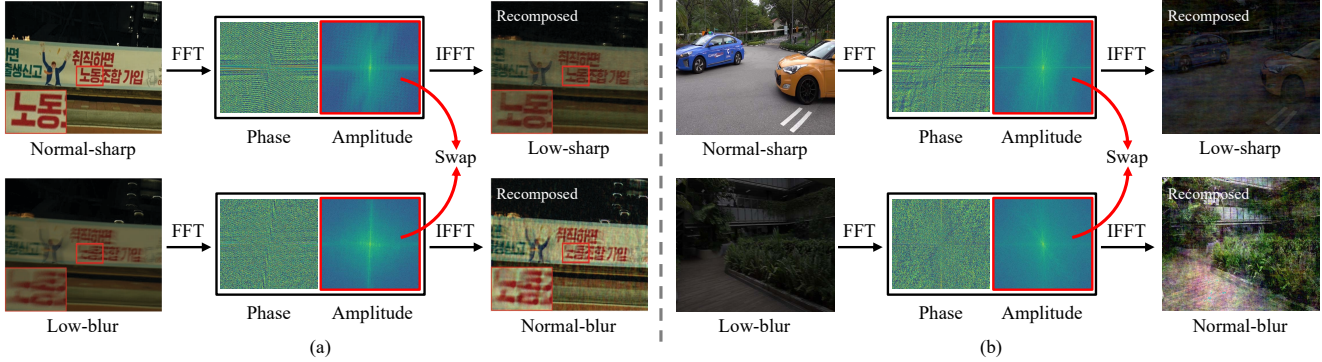


Figure 2. Motivations. (a) Swapping the amplitude of a normal-light sharp (normal-sharp) image with that of its corresponding low-light blurry (low-blur) image produces a low-light sharp (low-sharp) image and a normal-light blurry (normal-blur) image. This implies that most luminance information concentrates on amplitudes while structure information is closely related to phases. (b) Swapping the amplitude of two images with different content changes the appearance (e.g., brightness) of the images but preserves their main content. This means image content can be preserved in phases. The amplitude and phase are produced by Fast Fourier Transform (FFT) and the recomposed images are obtained by Inverse FFT (IFFT).

dation of luminance and structures. Specifically, low-light enhancement methods usually focus on improving brightness and denoising but ignore the spatial degradation caused by motion and therefore remain blurry in the enhanced results as shown in Fig. 1(b). Besides, deblurring methods are vulnerable in low-light environments as shown in Fig. 1(c), since they typically assume that blurred images are captured in well-lit conditions. An intuitive idea for tackling the joint task is to cascade these two kinds of methods. However, the process of light enhancement may lose the informative clues for blur removal, resulting in the failure of deblurring (see Fig. 1(d)). Recently, Zhou *et al.* [77] propose the first joint low-light enhancement and deblurring network, called LEDNet. Unfortunately, due to the difficulty of collecting low-light blurry and normal-light sharp image pairs in real-world scenes, LEDNet relies on the pixel-wise mapping learned from synthetic data, limiting its generalization in real-world scenes, as shown in Fig. 1(e).

Additionally, diffusion models have shown impressive performance in generating realistic and detailed images. Some pioneering studies [17, 23, 59] attempt to explore the rich generative prior of pre-trained diffusion models for zero-shot image restoration. However, these methods typically rely on the assumed degradation process, either in the form of a fixed linear matrix [23, 59] or a learnable degradation model [17]. These limitations impede the application of these diffusion-based methods in real-world low-light scenes, where the degradations are unknown and cannot be precisely modeled.

Unlike existing methods that rely on paired data or degradation assumptions, this paper addresses the joint task from a new perspective. As shown in Fig. 2, the core insight is motivated by our observation in the Fourier domain: most luminance information concentrates on amplitudes while structure and content information are closely

related to phases. We extend these characteristics to the pre-trained diffusion model as three-fold: (1) The amplitude of the sampled results in the diffusion process contains luminance priors from large-scale real-world data, which motivates us to utilize the amplitude to obtain reasonable brightness. (2) The phase of the input image preserves the original content, which can guide the pre-trained diffusion model to generate data-consistent results. (3) The blurry structure can be processed in the phase of the input image during the diffusion process.

With these characteristics at hand, we propose a novel zero-shot framework, FourierDiff, which embeds Fourier priors into a pre-trained diffusion model to simultaneously enhance image brightness and sharpness. FourierDiff performs amplitude-phase decomposition in the reverse diffusion process to handle the joint degradation. Specifically, we recombine the amplitude of the sampled results and the phase of the input image in each step to guide the diffusion sampling process, which progressively aligns the enhanced brightness with the distribution of natural images while preserving content consistency. Meanwhile, we present a spatial-frequency alternating optimization strategy to iteratively refine the phase of the input image to further facilitate the sampling process to generate pleasing results with sharp structures. The combination of the enhanced amplitude and the refined phase ensures FourierDiff considers the synergy between the inner-connected degradations in each step, therefore yielding harmonious results.

Our contributions are summarized as follows:

- We propose the first zero-shot method, FourierDiff, for joint low-light enhancement and deblurring through embedding Fourier priors into a pre-trained diffusion model, which allows harmoniously handling the inner-connected degradations without any paired training data.
- We leverage Fourier characteristics to distill the lumi-

nance priors conforming to human visual perception from pre-trained diffusion models, which enables FourierDiff to process the complex luminance degradations for yielding pleasing results with natural brightness.

- We present a spatial-frequency alternating optimization strategy to progressively refine the phase of degraded images, providing the diffusion process with structure-clear and content-consistent guidance. Extensive experimental results demonstrate that FourierDiff outperforms state-of-the-art methods.

2. Related Work

2.1. Low-Light Image Enhancement

Deep learning-based methods have become the mainstream in low-light image enhancement (LLIE) [28]. Inspired by the Retinex theory [27], several deep Retinex-based methods are proposed [3, 56, 63, 64, 69, 72]. They usually design multiple sub-networks for image decomposition, reflection restoration, and illumination adjustment. With the availability of paired datasets [4, 30, 34, 63], various methods [30, 38, 57, 58, 66, 67, 73] design effective networks to predict normal-light images directly from low-light ones through supervised learning. Despite their success, supervised methods suffer from limited generalization capability.

In recent years, unsupervised methods have attracted increasing attention [19, 29, 35, 39, 68, 75]. Zero-DCE [19] formulates light enhancement as a task of image-specific curve estimation. Zhao *et al.* [75] present a unified zero-reference framework based on Deep image prior (DIP) [55]. Ma *et al.* [39] propose a self-calibrated illumination learning framework to achieve fast, flexible, and robust LLIE. However, existing LLIE methods usually focus on brightness enhancement and noise reduction, while ignoring the spatial degradation of motion blurs. Moreover, LLIE methods may lose the informative clues for blur removal due to over-smoothing during the denoising process.

2.2. Image Deblurring

Image deblurring has been studied for a long time, with numerous deep learning-based methods proposed [5, 11, 26, 31, 33, 54, 62, 70, 71]. In the early stage, researchers often use networks to predict the blur kernels followed by non-blind methods [18, 50, 53]. In the past few years, end-to-end kernel-free networks have dominated image deblurring. Several novel components and techniques are proposed to improve the accuracy of deblurring, such as multi-scale strategies [11, 45], GAN-based structures [25, 26], and attention modules [33, 71]. Besides, some studies [2, 47] use image generation prior to achieve unsupervised deblurring.

Due to the poor visibility and noticeable noise, existing methods exhibit degraded performance when processing images captured in low-light conditions. Hence, some

methods [6, 7, 7, 8, 21, 74] have been specifically designed for deblurring low-light images, but they cannot deal with the joint degradation of luminance and structures. Most recently, Zhou *et al.* [77] propose the first joint low-light enhancement and deblurring network, named LEDNet, which considers the synergy between the two interconnected tasks. However, LEDNet relies on the pixel-wise mapping learned from paired synthetic data, resulting in limited generalization in diverse real-world scenes.

2.3. Diffusion-based Image Restoration

Recently, diffusion models have shown impressive performance in image restoration [32]. Existing methods can be roughly divided into two categories: supervised methods and zero-shot methods. The former [22, 37, 49, 60, 76] often uses the degraded image as the condition and entails training the diffusion model from scratch. Zero-shot methods [12, 13, 17, 23, 36, 59] exploit the generative prior of pre-trained diffusion models for image restoration. How to satisfy both data consistency and realness is the essential challenge of zero-shot methods. DDRM [23] uses singular value decomposition (SVD) to decompose the degradation operators of linear reverse problems. DDNM [59] introduces the range-null space decomposition to ensure content consistency. GDP [17] adopts a blind degradation estimation strategy, where the degradation parameters are optimized during the diffusion process. However, since these methods typically rely on the assumed degradation process, they tend to fall short when dealing with complex degradations, especially unknown and mixed degradations.

In contrast, we utilize the characteristics in the Fourier domain to guide the reverse diffusion process without any extra training or degradation estimation. Despite DiffLL [22] also decomposing images into the frequency domain, it focuses on accelerating the diffusion process and relies on end-to-end training. The role of frequency priors in achieving diffusion-based zero-shot image restoration has not been explored in previous works.

3. Preliminary

Diffusion models [20, 51, 52] are generative models with a Markov chain structure, which consists of a forward process and a reverse process. The forward process gradually adds Gaussian noise to an input image \mathbf{x}_0 through T steps. The present state \mathbf{x}_t is only dependent on the previous state \mathbf{x}_{t-1} , which can be formulated as the following Gaussian distribution

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I}) \quad (1)$$

where β_t is the predefined variance at step t . After reparameterization, it becomes

$$q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t} \mathbf{x}_0, (1 - \bar{\alpha}_t) \mathbf{I}) \quad (2)$$

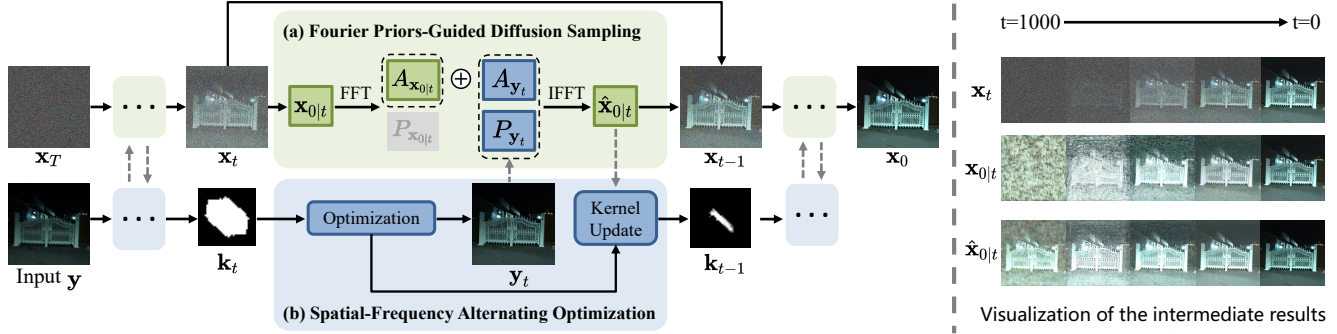


Figure 3. The architecture of FourierDiff consists of two components: **(a) Fourier priors-guided diffusion sampling**: we decompose the sampled results of the reverse diffusion process in the Fourier domain and leverage the amplitude of the generative prior to make the enhanced brightness progressively satisfy the distribution of natural images. **(b) Spatial-frequency alternating optimization**: The phase of the input image is iteratively refined to further provide sharper guidance for the sampling process. The synergy between the sampling process and the optimization process enables FourierDiff to yield pleasing and realistic results with natural brightness and sharp structures.

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=0}^t \alpha_i$.

The reverse process constructs the clean image from the Gaussian noise step by step, which can be formulated as

$$p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)) \quad (3)$$

where $\boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)$ is a time-dependent constant, and $\boldsymbol{\mu}_\theta(\mathbf{x}_t, t)$ can be written as

$$\boldsymbol{\mu}_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right) \quad (4)$$

where $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$ is the estimated noise derived from a deep neural network. According to Ho *et al.* [20], the model randomly picks a clean image \mathbf{x}_0 from the dataset and samples a noise $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$, then optimizes the network parameters θ with the following objective function

$$L_{\text{diff}}(\theta) = \|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}, t)\|_2^2 \quad (5)$$

By iteratively sampling \mathbf{x}_{t-1} from $p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t)$, clean images $\mathbf{x}_0 \sim q(\mathbf{x})$ can be generated from random noises $\mathbf{x}_T \sim \mathcal{N}(0, \mathbf{I})$, where $q(\mathbf{x})$ represents the original distribution in the dataset.

4. Methodology

The proposed FourierDiff is a zero-shot framework that only applies the reverse diffusion process to restore natural brightness and sharp structures from low-light blurry images. Thanks to the real-world data priors from the pre-trained diffusion model, FourierDiff does not need any paired synthetic data. As shown in Fig. 3, given a low-light blurry image \mathbf{y} , FourierDiff progressively produces its corresponding normal-light sharp version \mathbf{x}_0 from the pure noise \mathbf{x}_T under the guidance of the input \mathbf{y} . In each step of the reverse diffusion process, we adopt a Fourier priors-guided diffusion sampling to leverage the amplitude of lu-

minance priors for harmonious and natural light enhancement. Meanwhile, a spatial-frequency alternating optimization is proposed to refine the phase of the input \mathbf{y} as the sampling process. The refined phase offers sharp structures to guide the generated content during the sampling process. We detail the key components as follows.

4.1. Fourier Priors-Guided Diffusion Sampling

As described in Sec. 3, the pre-trained diffusion models are devoted to preserving the data distribution rather than pixel-wise content consistency. Therefore, how to extract the corresponding generative priors while maintaining the content from the input is the essential challenge faced by diffusion-based zero-shot image restoration methods. Motivated by the observations that luminance and structure can be decoupled in the Fourier domain as highlighted in Sec. 1, we utilize the Fourier characteristics to guide the sampling process to yield realistic and data-consistent results.

As shown in Fig. 3, we perform amplitude-phase decomposition on the sampled result in each step of the reverse diffusion process. Specifically, to produce clean sampled results for amplitude-phase decomposition, we estimate \mathbf{x}_0 in each step and denote it as $\mathbf{x}_{0|t}$ following [59], which can be formulated as

$$\mathbf{x}_{0|t} = \frac{1}{\sqrt{\alpha_t}} (\mathbf{x}_t - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \sqrt{1 - \bar{\alpha}_t}) \quad (6)$$

Then, we perform Fast Fourier Transform (FFT) on both the input image \mathbf{y} and the sampled result $\mathbf{x}_{0|t}$ to obtain the amplitude and phase as follows

$$(A_{\mathbf{y}}, P_{\mathbf{y}}) = \text{FFT}(\mathbf{y}) \quad (7)$$

$$(A_{\mathbf{x}_{0|t}}, P_{\mathbf{x}_{0|t}}) = \text{FFT}(\mathbf{x}_{0|t}) \quad (8)$$

where $A_{\mathbf{y}}$, $A_{\mathbf{x}_{0|t}}$, $P_{\mathbf{y}}$ and $P_{\mathbf{x}_{0|t}}$ represent the amplitude and phase of \mathbf{y} and $\mathbf{x}_{0|t}$, respectively. As mentioned before,

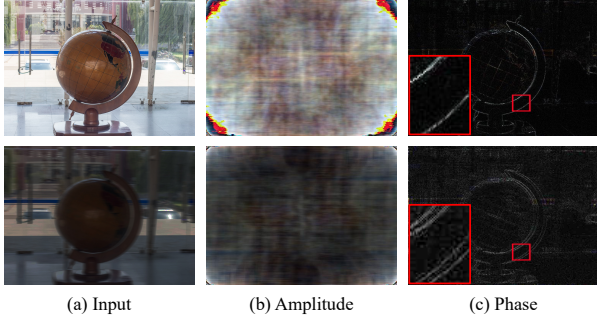


Figure 4. Visualization of the amplitude and phase of a low-light blurry image and its corresponding normal-light sharp image. The amplitudes differ significantly between different brightness. The motion information is encoded as repeated edges in phases.

$A_{x_{0|t}}$ contains luminance priors conforming to the distribution of natural images, but $P_{x_{0|t}}$ extracts random content from the sampled result and disturbs the specific content generation. To leverage the generative prior while preserving the content from the input image \mathbf{y} , we combine $A_{x_{0|t}}$ and $A_{\mathbf{y}}$ to update the amplitude and replace $P_{x_{0|t}}$ with $P_{\mathbf{y}}$ to guide the content of the diffusion process. Next, we can obtain the updated sampled result $\hat{\mathbf{x}}_{0|t}$ by Inverse Fast Fourier Transform (IFFT), which is defined as

$$\hat{\mathbf{x}}_{0|t} = \text{IFFT}(A_{x_{0|t}} + A_{\mathbf{y}}, P_{\mathbf{y}}) \quad (9)$$

As suggested in [59], the next state \mathbf{x}_{t-1} can be sampled from a joint distribution, which is formulated as

$$p_{\theta}(\mathbf{x}_{t-1} | \mathbf{x}_t, \hat{\mathbf{x}}_{0|t}) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_t(\mathbf{x}_t, \hat{\mathbf{x}}_{0|t}), \sigma_t^2 \mathbf{I}) \quad (10)$$

where $\boldsymbol{\mu}_t(\mathbf{x}_t, \hat{\mathbf{x}}_{0|t}) = \frac{\sqrt{\bar{\alpha}_t - 1} \beta_t}{1 - \bar{\alpha}_t} \hat{\mathbf{x}}_{0|t} + \frac{\sqrt{\bar{\alpha}_t(1 - \bar{\alpha}_t - 1)}}{1 - \bar{\alpha}_t} \mathbf{x}_t$ and $\sigma_t^2 = \frac{1 - \bar{\alpha}_t - 1}{1 - \bar{\alpha}_t} \beta_t$. By using Fourier priors to guide the sampling process in each step, we ultimately obtain the result \mathbf{x}_0 with natural brightness and consistent content.

Although the luminance distribution in the generated result \mathbf{x}_0 aligns with that of natural images, human perception of brightness is highly subjective. The optimal brightness level varies among individuals. Therefore, to adapt to user-specific requirements, we introduce a learnable adaptive factor γ to control the brightness level of the sampling process. Based on the factor, Eq. (9) can be reformulated as

$$\hat{\mathbf{x}}_{0|t} = \text{IFFT}(\gamma A_{x_{0|t}} + A_{\mathbf{y}}, P_{\mathbf{y}}) \quad (11)$$

To optimize γ , we introduce a non-reference brightness control constraint L_{bri} in each step, which is defined as

$$L_{\text{bri}} = \frac{1}{R} \sum_{n=1}^R |I_t^n - E| \quad (12)$$

where R is the number of non-overlapping local regions of size 16×16 . I_t^n represents the average intensity value of

Algorithm 1 Fourier Priors-Guided Diffusion Sampling.

Input: The degraded image \mathbf{y} , the total diffusion step T and the alternating optimization interval step N .

- 1: $(A_{\mathbf{y}}, P_{\mathbf{y}}) = \text{FFT}(\mathbf{y})$
- 2: $\mathbf{k} = \mathcal{A}(|P_{\mathbf{y}}|)$
- 3: $\hat{\mathbf{y}} = \mathbf{y}$
- 4: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
- 5: **for** $t = T, \dots, 1$ **do**
- 6: $\mathbf{x}_{0|t} = \frac{1}{\sqrt{\bar{\alpha}_t}} (\mathbf{x}_t - \epsilon_{\theta}(\mathbf{x}_t, t)) \sqrt{1 - \bar{\alpha}_t}$
- 7: $\mathbf{y}_t = \hat{\mathbf{y}}$
- 8: $(A_{x_{0|t}}, P_{x_{0|t}}) = \text{FFT}(\mathbf{x}_{0|t})$
- 9: $(A_{\mathbf{y}_t}, P_{\mathbf{y}_t}) = \text{FFT}(\mathbf{y}_t)$
- 10: $\hat{\mathbf{x}}_{0|t} = \text{IFFT}(\gamma A_{x_{0|t}} + A_{\mathbf{y}_t}, P_{\mathbf{y}_t})$
- 11: $\mathbf{x}_{t-1} \sim p_{\theta}(\mathbf{x}_{t-1} | \mathbf{x}_t, \hat{\mathbf{x}}_{0|t})$
- 12: **if** $t \bmod N == 0$ **then**
- 13: $\mathbf{k}_t = \mathcal{A}(|P_{\hat{\mathbf{x}}_{0|t}}|)$
- 14: $\mathbf{k} = (1 - \frac{1}{t})\mathbf{k} + \frac{1}{t}\mathbf{k}_t$
- 15: $\min_{\hat{\mathbf{y}}, \mathbf{k}} \|\mathbf{k} \otimes \hat{\mathbf{y}} - \mathbf{y}\|_2^2 + \lambda_1 \|\mathbf{k}\|_2^2 + \lambda_2 h(\nabla \hat{\mathbf{y}})$
- 16: **end if**
- 17: **end for**

Output: \mathbf{x}_0

the local region n in the rectified sampled result $\hat{\mathbf{x}}_{0|t}$. E represents the brightness level and is set to the gray level in the RGB color space following [42]. As shown in Fig. 8 of the ablation study, the brightness of the generated images can be adjusted by setting different E values.

4.2. Spatial-Frequency Alternating Optimization

Although Fourier priors-guided diffusion sampling enables the generation of content-consistent images, the degraded input image brings indistinct structural guidance that makes the enhanced results still blurry. It is well known that the phase of blurry images can provide faithful information about the blur pattern [41, 46]. As illustrated in Fig. 4, the appearance of the phase is similar to the structure of the image and the motion information is encoded as repeated image edges in phases, which to some extent reflects the shape and size of the blur kernel. Therefore, to provide sharp structural guidance for the sampling process, we design a spatial-frequency alternating optimization strategy to refine the phase of the blurry input image \mathbf{y} .

Specifically, we first calculate the autocorrelation $\mathcal{A}(|P_{\mathbf{y}}|)$ of the image $|P_{\mathbf{y}}|$ reconstructed from the absolute phase of the input image \mathbf{y} , which is formulated as

$$\mathcal{A}(|P_{\mathbf{y}}|) = \text{IFFT} \left(\text{FFT}(|P_{\mathbf{y}}|) \odot \overline{\text{FFT}(|P_{\mathbf{y}}|)} \right) \quad (13)$$

Then, the blur kernel \mathbf{k} of the input image can be calculated from the autocorrelation following [46]. Given the initial kernel \mathbf{k} and the blurry image \mathbf{y} , we can acquire sharp guidance by solving the optimization-based deblurring problem,

Table 1. Quantitative comparisons with state-of-the-art methods on the LOL-Blur dataset. The symbol * indicates the network is trained on the LOL-Blur dataset. The proposed FourierDiff is a zero-shot method without requiring any paired training data and degradation assumptions. The best and the second-best scores are shown in bold and underlined, respectively.

Method	Enhancement → Deblurring			Deblurring → Enhancement						Joint	
	Zero-DCE++ [29] → GRL [33]	RetinexDIP [39] → GRL [33]	GDP [17] → GRL [33]	Chen [6] →Zero-DCE++ [29]	Chen [6] →GDP [17]	W-DIP [2] →Zero-DCE++ [29]	W-DIP [2] →GDP [17]	GRL [33] →Zero-DCE++ [29]	GRL [33] →GDP [17]	LEDNet* [77]	FourierDiff (Ours)
NIQE ↓	4.27	4.59	4.31	4.76	4.87	4.82	5.03	4.28	4.32	<u>3.99</u>	3.80
PI ↓	5.05	5.38	4.81	4.97	4.70	4.32	<u>4.10</u>	5.13	4.90	5.07	3.88
BRISQUE ↓	42.21	47.45	41.03	47.10	49.83	37.58	<u>35.60</u>	44.36	43.25	42.59	33.13
MUSIQ ↑	53.36	49.61	56.42	51.64	55.19	47.04	50.10	55.96	58.92	<u>59.64</u>	62.46
PSNR ↑	18.45	13.65	17.72	17.43	16.52	16.52	15.69	18.90	18.16	25.74	<u>20.53</u>
SSIM ↑	0.59	0.55	0.66	0.51	0.56	0.42	0.46	0.64	0.70	0.85	<u>0.71</u>

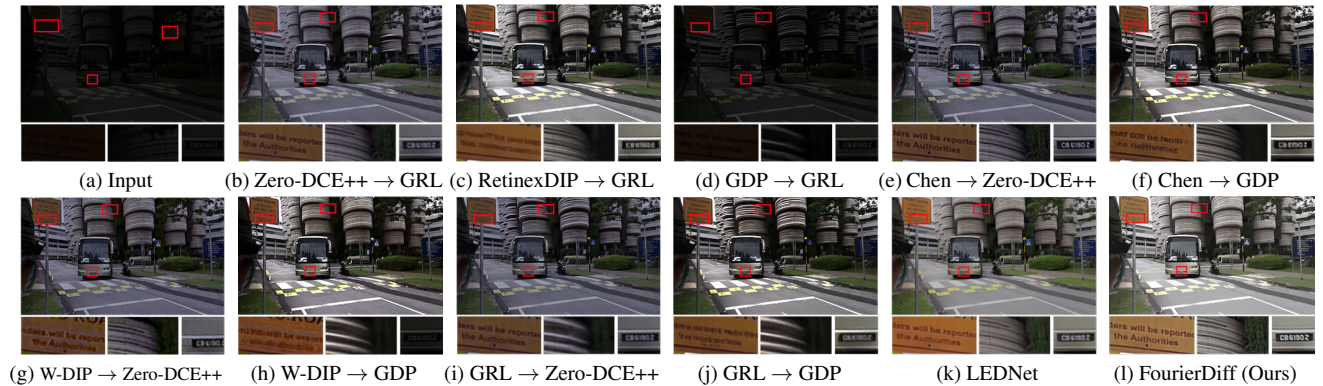


Figure 5. Qualitative comparisons with state-of-the-art methods on the LOL-Blur dataset. (Zoom in for best view)

which is generally formulated as

$$\min_{\hat{\mathbf{y}}, \mathbf{k}} \|\mathbf{k} \otimes \hat{\mathbf{y}} - \mathbf{y}\|_2^2 + \lambda_1 \|\mathbf{k}\|_2^2 + \lambda_2 h(\nabla \hat{\mathbf{y}}) \quad (14)$$

where $\hat{\mathbf{y}}$ is the latent sharp image. λ_1 and λ_2 are weight parameters. $h(\cdot)$ is a truncated-quadratic gradient [65] regularization term used to prevent over-sharpening.

However, inaccurate estimation of the blur kernel \mathbf{k} in low-light environments leads to the sub-optimal deblurring result $\hat{\mathbf{y}}$, making it difficult for the phase of $\hat{\mathbf{y}}$ to provide sharp structural guidance for the sampling process. With Fourier priors-guided diffusion sampling, $\hat{\mathbf{x}}_{0|t}$ progressively becomes clear in content and natural in brightness, which can provide more visible details for blur kernel estimation. The right of Fig. 3 shows the visualization of $\hat{\mathbf{x}}_{0|t}$ during the sampling process. Therefore, to refine the content guidance, we present to leverage $\hat{\mathbf{x}}_{0|t}$ to update the optimized blur kernel. We decompose $\hat{\mathbf{x}}_{0|t}$ in the Fourier domain and use its autocorrelation $\mathcal{A}(|P_{\hat{\mathbf{x}}_{0|t}}|)$ from $P_{\hat{\mathbf{x}}_{0|t}}$ to estimate the blur kernel \mathbf{k}_t . Then, the blur kernel \mathbf{k} is updated according to the following strategy

$$\mathbf{k} = (1 - \frac{1}{t})\mathbf{k} + \frac{1}{t}\mathbf{k}_t \quad (15)$$

with the iterative updating of the blur kernel, the phase of the input image gradually becomes sharper.

Utilizing sampling results to update the blur kernel improves the robustness of deblurring algorithms when processing low-light images. Meanwhile, the progressively refined input image provides sharper structural guidance

for the sampling process. The optimization process and the sampling process work together and complement each other. To improve the sampling efficiency, we conduct the spatial-frequency alternating optimization at intervals of N steps. Algorithm 1 shows the detailed process. The synergy between the optimization and sampling processes enables FourierDiff to generate pleasing and realistic results with natural brightness and sharp structures.

5. Experiments

5.1. Datasets and Evaluation Metrics

We evaluate the proposed method on the LOL-Blur [77] and RealBlur [48] datasets. The LOL-Blur dataset is the first large-scale dataset for joint low-light enhancement and deblurring, which consists of 12,000 synthetic low-blur/normal-sharp pairs with diverse darkness and motion blurs. We use the same training/test separation as LOL-Blur. The RealBlur dataset is the first real-world image deblurring dataset, which contains 4,738 pairs of images in 232 different scenes. Following [77], we use 482 real-world night blurry images selected from RealBlur as the test set to verify the generalization of the proposed method. Since there is no normal-sharp ground truth corresponding to low-blur images in RealBlur, we use four commonly-used no-reference image quality metrics to perform quantitative comparisons, including Natural Image Quality Evaluator (NIQE) [44], Perceptual Index (PI) [1], Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [43], and Multi-Scale Image Quality Transformer (MUSIQ) [24].

Table 2. Quantitative comparisons with state-of-the-art methods on the RealBlur dataset.

Method	Enhancement \rightarrow Deblurring			Deblurring \rightarrow Enhancement						Joint	
	Zero-DCE++ [29] \rightarrow GRL [33]	RetinexDIP [39] \rightarrow GRL [33]	GDP [17] \rightarrow GRL [33]	Chen [6] \rightarrow Zero-DCE++ [29]	Chen [6] \rightarrow GDP [17]	W-DIP [2] \rightarrow Zero-DCE++ [29]	W-DIP [2] \rightarrow GDP [17]	GRL [33] \rightarrow Zero-DCE++ [29]	GRL [33] \rightarrow GDP [17]	LEDNet [77]	FourierDiff (Ours)
NIQE \downarrow	3.33	3.35	3.26	4.88	4.67	4.23	4.06	3.70	3.58	3.72	3.25
PI \downarrow	4.55	4.29	4.54	4.88	4.95	4.31	4.20	4.71	4.61	5.03	3.36
BRISQUE \downarrow	30.46	30.90	28.96	45.89	45.60	35.60	33.00	34.80	33.10	42.31	26.39
MUSIQ \uparrow	42.88	44.79	39.35	49.96	47.19	41.43	38.68	45.50	43.22	49.45	52.24



Figure 6. Qualitative comparisons with state-of-the-art methods on the RealBlur dataset. (Zoom in for best view)

Larger MUSIQ indicates more naturalistic and perceptually favored quality. Contrary to MUSIQ, smaller NIQE, PI and BRISQUE mean better perceptual quality. On the LOL-Blur dataset, we also use well-known full-reference metrics Peak Signal-to-Noise Ratio (PSNR) and Structural SIMilarity (SSIM) [61] to measure the difference between the enhanced results and ground truth.

5.2. Implementation Details

We implement our framework with Pytorch on a single NVIDIA GeForce RTX 3090 GPU. We use the released unconditional 256x256 diffusion model [16] pre-trained on ImageNet [15]. The total diffusion step T and the alternating optimization interval step N are set to 1000 and 200, respectively. The brightness level E is flexible for users, and we use 0.5 by default. For spatial-frequency alternating optimization, we set $\lambda_1 = 2$, $\lambda_2 = 0.005$. For extremely dark images, we use PEC [40] with a small exposure parameter to warm-start the input image to prevent the complete disappearance of content guidance.

5.3. Comparison with State-of-the-art Methods

To comprehensively evaluate our method, we construct the following three types of baselines for comparisons.

1. Joint Enhancement and Deblurring. We choose LEDNet [77] trained on the LOL-Blur dataset as the baseline since it is the only work dedicated to the joint task.

2. Enhancement \rightarrow Deblurring. We choose the recent

representative unsupervised low-light enhancement methods Zero-DCE++ [29], RetinexDIP [75] and the diffusion-based method GDP [17] followed by a state-of-the-art deblurring method GRL [33] trained on the RealBlur dataset.

3. Deblurring \rightarrow Enhancement. For deblurring, we choose an optimization-based method [6] specifically designed for low-light deblurring, a recent zero-shot deblurring method W-DIP [2] and a state-of-the-art deblurring method GRL [33]. Since RetinexDIP [75] tends to produce halo artifacts that may obscure previous deblurring results, we use Zero-DCE++ [29] and GDP [17] for low-light enhancement in this type of baseline.

Quantitative Comparison. Tables 1 and 2 report the quantitative results on the LOL-Blur and RealBlur datasets, respectively. As we can see, FourierDiff outperforms state-of-the-art methods in terms of all no-reference metrics, indicating that our results are perceptually best. Such results demonstrate the stability of the proposed method when handling real-world images with various lighting conditions and blur patterns. In addition, we also evaluate full-reference metrics PSNR and SSIM on the LOL-Blur dataset. As shown in Table 1, FourierDiff achieves comparable results to LEDNet trained with ground truth and exceeds other baseline methods by a large margin, which further suggests the effectiveness of the proposed method.

Qualitative Comparison. In Figs. 5 and 6, we show the visual comparison results on the LOL-Blur and RealBlur datasets, respectively. It can be seen that the results en-

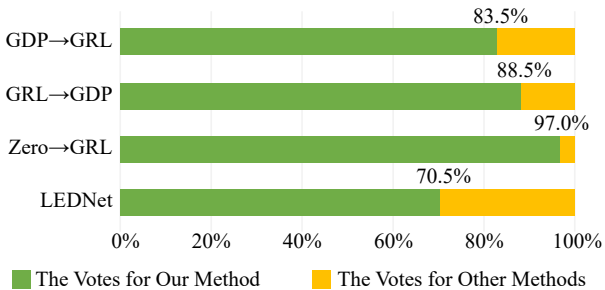


Figure 7. The results of the user study.

hanced by cascade methods suffer from noticeable color deviation and severe blur artifacts. Although the method of Chen *et al.* [7] is specifically designed for low-light deblurring, their cascading baselines still cannot eliminate blur as shown in Figs. 5(e), 5(f), 6(e), 6(f). LEDNet can handle the joint degradation of luminance and structures, but it exhibits degraded performance when processing real-world images as shown in Fig. 6(k). In contrast, FourierDiff produces visually pleasing results with more natural brightness and sharper textures in various scenes.

5.4. User Study

Furthermore, we conduct a user study to evaluate the subjective perception of different methods. Specifically, we randomly select 20 testing images from the LOL-Blur and RealBlur datasets and choose 4 baselines based on the rank of the average NIQE scores on LOL-Blur and RealBlur. For each image, we provide the input degraded image, the corresponding images enhanced by our method and a baseline. A total of 40 participants are invited to select their preferred image. As shown in Fig. 7, our method is more favored by human subjects.

5.5. Ablation Study

Effect of Brightness Adjustment. To validate the effectiveness of the brightness adjustment strategy, we adopt different E in Eq. (12) to control the brightness level of outputs. As shown in Fig. 8, the brightness of the generated images can be adjusted by setting different E values. Note that even without using brightness adjustment (denoted as w/o BA), our method still yields reasonable brightness because the diffusion model trained on ImageNet contains the luminance priors of natural images. The effectiveness of the strategy allows FourierDiff to generate images with various brightness according to the user-specific requirements.

Effect of the Spatial-Frequency Alternating Optimization. To verify the effectiveness of the spatial-frequency alternating optimization strategy (denoted as SFA), we conduct five different settings of the alternating optimization interval step N . “w/o SFA” indicates that we refine the phase of the input image before diffusion sampling and use the refined phase as guidance during the whole sampling pro-

Table 3. Ablation study of the spatial-frequency alternating optimization on the RealBlur dataset.

	w/o SFA	$N = 500$	$N = 200$	$N = 100$	$N = 1$
NIQE ↓	3.62	3.37	<u>3.25</u>	3.26	3.19
PI ↓	3.87	3.61	<u>3.36</u>	3.39	3.34
BRISQUE ↓	31.65	29.72	26.39	<u>26.12</u>	25.14
MUSIQ ↑	48.81	49.87	<u>52.24</u>	52.41	51.11



Figure 8. Visual results of the effect of brightness adjustment.

cess. As shown in Table 3, the performance shows a consistent improvement as $N=500 \rightarrow 1$. When N substantially decreases (e.g., $N=500 \rightarrow 200$), the performance improves significantly. When N comes to relatively small changes (e.g., $N=200 \rightarrow 100$), the model has negligible performance gains. To achieve a trade-off between efficiency and performance, we choose $N=200$ as the interval step.

6. Conclusion

This paper proposes the first zero-shot joint low-light enhancement and deblurring method, FourierDiff, which is able to harmoniously handle the inner-connected degradations without any paired training data and degradation assumptions. The success of our method is inspired by the characteristics of real low-light blurry images in the Fourier domain. This is the first time to extend Fourier characteristics to the diffusion model, which offers a new perspective on utilizing the generative prior of pre-trained diffusion models to restore degraded images while maintaining content consistency. Thanks to the unique design of our framework that processes luminance and blur in amplitudes and phases respectively, FourierDiff outperforms state-of-the-art methods on the joint task.

Limitations. Although our method achieves impressive results in enhancing real low-light blurry images, it does not work as well in extremely dark environments because of the severe loss of content guidance. Furthermore, FourierDiff inherits the limited inference speed of diffusion models and is not yet efficient enough for real-time image processing.

Acknowledgement. This work is supported by the National Natural Science Foundation of China (Nos. 62272134, 62236003 and 62072141), Shenzhen College Stability Support Plan (Grant No. GXWD20220817144428005) and The Major Key Project of PCL (PCL2023A10-2).

References

- [1] Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *CVPR*, pages 6228–6237, 2018. 6
- [2] Gustav Bredell, Ertunc Erdil, Bruno Weber, and Ender Konukoglu. Wiener guided DIP for unsupervised blind image deconvolution. In *WACV*, pages 3046–3055, 2023. 3, 6, 7
- [3] 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*, pages 12504–12513, 2023. 1, 3
- [4] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *CVPR*, pages 3291–3300, 2018. 3
- [5] Liangyu Chen, Xin Lu, Jie Zhang, Xiaojie Chu, and Chengpeng Chen. HINet: Half instance normalization network for image restoration. In *CVPRW*, pages 182–192, 2021. 3
- [6] Liang Chen, Jiawei Zhang, Songnan Lin, Faming Fang, and Jimmy S. Ren. Blind deblurring for saturated images. In *CVPR*, pages 6308–6316, 2021. 3, 6, 7
- [7] Liang Chen, Jiawei Zhang, Jinshan Pan, Songnan Lin, Faming Fang, and Jimmy S. Ren. Learning a non-blind deblurring network for night blurry images. In *CVPR*, pages 10542–10550, 2021. 3, 8
- [8] Liang Chen, Jiawei Zhang, Zhenhua Li, Yunxuan Wei, Faming Fang, Jimmy Ren, and Jinshan Pan. Deep richardson-lucy deconvolution for low-light image deblurring. *IJCV*, pages 1–18, 2023. 3
- [9] Rui Chen, Jiajun Chen, Zixi Liang, Huaian Gao, and Shan Lin. Darklight networks for action recognition in the dark. In *CVPRW*, pages 846–852, 2021. 1
- [10] Zhi Chen, Zijun Fan, Yongjie Li, Huaian Gao, and Shan Lin. Z-domain entropy adaptable flex for semi-supervised action recognition in the dark. In *CVPRW*, pages 4258–4265, 2022. 1
- [11] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *ICCV*, pages 4621–4630, 2021. 1, 3
- [12] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. ILVR: conditioning method for denoising diffusion probabilistic models. In *ICCV*, pages 14347–14356, 2021. 3
- [13] Hyungjin Chung, Jeongsol Kim, Michael Thompson McCann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *ICLR*, 2023. 3
- [14] Ziteng Cui, Guo-Jun Qi, Lin Gu, Shaodi You, Zenghui Zhang, and Tatsuya Harada. Multitask AET with orthogonal tangent regularity for dark object detection. In *ICCV*, pages 2533–2542, 2021. 1
- [15] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *CVPR*, pages 248–255, 2009. 7
- [16] Prafulla Dhariwal and Alexander Quinn Nichol. Diffusion models beat gans on image synthesis. In *NeurIPS*, pages 8780–8794, 2021. 7
- [17] Ben Fei, Zhaoyang Lyu, Liang Pan, Junzhe Zhang, Weidong Yang, Tianyue Luo, Bo Zhang, and Bo Dai. Generative diffusion prior for unified image restoration and enhancement. In *CVPR*, pages 9935–9946, 2023. 1, 2, 3, 6, 7
- [18] Dong Gong, Jie Yang, Lingqiao Liu, Yanning Zhang, Ian D. Reid, Chunhua Shen, Anton van den Hengel, and Qinfeng Shi. From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur. In *CVPR*, pages 3806–3815, 2017. 3
- [19] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *CVPR*, pages 1777–1786, 2020. 1, 3
- [20] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, pages 6840–6851, 2020. 3, 4
- [21] Zhe Hu, Sunghyun Cho, Jue Wang, and Ming-Hsuan Yang. Deblurring low-light images with light streaks. In *CVPR*, pages 3382–3389, 2014. 3
- [22] Hai Jiang, Ao Luo, Haoqiang Fan, Songchen Han, and Shuaicheng Liu. Low-light image enhancement with wavelet-based diffusion models. *ACM TOG*, 42(6):1–14, 2023. 3
- [23] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In *NeurIPS*, 2022. 2, 3
- [24] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. MUSIQ: multi-scale image quality transformer. In *ICCV*, pages 5128–5137, 2021. 6
- [25] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiri Matas. DeblurGAN: Blind motion deblurring using conditional adversarial networks. In *CVPR*, pages 8183–8192, 2018. 3
- [26] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better. In *ICCV*, pages 8877–8886, 2019. 1, 3
- [27] Edwin H Land. The retinex theory of color vision. *Sci. Amer.*, 237(6):108–129, 1977. 3
- [28] Chongyi Li, Chunle Guo, Linghao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. Low-light image and video enhancement using deep learning: A survey. *IEEE TPAMI*, 44(12):9396–9416, 2022. 3
- [29] Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE TPAMI*, 44(8):4225–4238, 2022. 3, 6, 7
- [30] Chongyi Li, Chun-Le Guo, Man Zhou, Zhixin Liang, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Embedding fourier for ultra-high-definition low-light image enhancement. In *ICLR*, 2023. 1, 3
- [31] Haoying Li, Ziran Zhang, Tingting Jiang, Peng Luo, Huajun Feng, and Zhihai Xu. Real-world deep local motion deblurring. In *AAAI*, pages 1314–1322, 2023. 3
- [32] Xin Li, Yulin Ren, Xin Jin, Cuiling Lan, Xingrui Wang, Wenjun Zeng, Xinchao Wang, and Zhibo Chen. Diffusion models for image restoration and enhancement—a comprehensive survey. *arXiv preprint arXiv:2308.09388*, 2023. 3

- [33] Yawei Li, Yuchen Fan, Xiaoyu Xiang, Denis Demandolx, Rakesh Ranjan, Radu Timofte, and Luc Van Gool. Efficient and explicit modelling of image hierarchies for image restoration. In *CVPR*, pages 18278–18289, 2023. 1, 3, 6, 7
- [34] Jiaying Liu, Dejia Xu, Wenhan Yang, Minhao Fan, and Haofeng Huang. Benchmarking low-light image enhancement and beyond. *IJCV*, 129(4):1153–1184, 2021. 1, 3
- [35] Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *CVPR*, pages 10561–10570, 2021. 3
- [36] Andreas Lugmayr, Martin Danelljan, Andrés Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *CVPR*, pages 11451–11461, 2022. 3
- [37] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Refusion: Enabling large-size realistic image restoration with latent-space diffusion models. In *CVPR*, pages 1680–1691, 2023. 3
- [38] Xiaoqian Lv, Shengping Zhang, Qinglin Liu, Haozhe Xie, Bineng Zhong, and Huiyu Zhou. BacklitNet: A dataset and network for backlit image enhancement. *CVIU*, 218:103403, 2022. 3
- [39] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *CVPR*, pages 5627–5636, 2022. 1, 3, 6, 7
- [40] Long Ma, Tianjiao Ma, Xinwei Xue, Xin Fan, Zhongxuan Luo, and Risheng Liu. Practical exposure correction: Great truths are always simple. *arXiv preprint arXiv:2212.14245*, 2022. 7
- [41] Xintian Mao, Yiming Liu, Fengze Liu, Qingli Li, Wei Shen, and Yan Wang. Intriguing findings of frequency selection for image deblurring. In *AAAI*, pages 1905–1913, 2023. 5
- [42] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion. In *Proc. Pacific Conf. Comput. Graph. Appl.*, pages 382–390, 2007. 5
- [43] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE TIP*, 21(12):4695–4708, 2012. 6
- [44] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a “completely blind” image quality analyzer. *IEEE Sign. Process. Letters*, 20(3):209–212, 2013. 6
- [45] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, pages 257–265, 2017. 3
- [46] Liyuan Pan, Richard I. Hartley, Miaomiao Liu, and Yuchao Dai. Phase-only image based kernel estimation for single image blind deblurring. In *CVPR*, pages 6034–6043, 2019. 5
- [47] Dongwei Ren, Kai Zhang, Qilong Wang, Qinghua Hu, and Wangmeng Zuo. Neural blind deconvolution using deep priors. In *CVPR*, pages 3338–3347, 2020. 3
- [48] Jaesung Rim, Haeyun Lee, Jucheol Won, and Sunghyun Cho. Real-world blur dataset for learning and benchmarking deblurring algorithms. In *ECCV*, pages 184–201, 2020. 6
- [49] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J. Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE TPAMI*, 45(4):4713–4726, 2023. 3
- [50] Christian J. Schuler, Michael Hirsch, Stefan Harmeling, and Bernhard Schölkopf. Learning to deblur. *IEEE TPAMI*, 38(7):1439–1451, 2016. 3
- [51] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *ICML*, pages 2256–2265, 2015. 3
- [52] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *ICLR*, 2021. 3
- [53] Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *CVPR*, pages 769–777, 2015. 3
- [54] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *CVPR*, pages 8174–8182, 2018. 3
- [55] Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. Deep image prior. In *CVPR*, pages 9446–9454, 2018. 3
- [56] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *CVPR*, pages 6849–6857, 2019. 3
- [57] Tao Wang, Kaihao Zhang, Tianrun Shen, Wenhan Luo, Björn Stenger, and Tong Lu. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In *AAAI*, pages 2654–2662, 2023. 3
- [58] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex C. Kot. Low-light image enhancement with normalizing flow. In *AAAI*, pages 2604–2612, 2022. 3
- [59] Yinhuai Wang, Jiwen Yu, and Jian Zhang. Zero-shot image restoration using denoising diffusion null-space model. In *ICLR*, 2023. 2, 3, 4, 5
- [60] Yufei Wang, Yi Yu, Wenhan Yang, Lanqing Guo, Lap-Pui Chau, Alex C Kot, and Bihan Wen. Exposediffusion: Learning to expose for low-light image enhancement. In *ICCV*, pages 12438–12448, 2023. 3
- [61] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004. 7
- [62] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *CVPR*, pages 17662–17672, 2022. 3
- [63] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *BMVC*, page 155, 2018. 3
- [64] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *CVPR*, pages 5891–5900, 2022. 3
- [65] Li Xu, Shicheng Zheng, and Jiaya Jia. Unnatural L0 sparse representation for natural image deblurring. In *CVPR*, pages 1107–1114, 2013. 6
- [66] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *CVPR*, pages 17693–17703, 2022. 3

- [67] Xiaogang Xu, Ruixing Wang, and Jiangbo Lu. Low-light image enhancement via structure modeling and guidance. In *CVPR*, pages 9893–9903, 2023. 3
- [68] Shuzhou Yang, Moxuan Ding, Yanmin Wu, Zihan Li, and Jian Zhang. Implicit neural representation for cooperative low-light image enhancement. In *ICCV*, pages 12918–12927, 2023. 3
- [69] Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *ICCV*, pages 12302–12311, 2023. 3
- [70] Syed Waqas Zamir, Aditya Arora, Salman H. Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, pages 14821–14831, 2021. 3
- [71] 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*, pages 5718–5729, 2022. 1, 3
- [72] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *IJCV*, 129(4): 1013–1037, 2021. 3
- [73] Zhao Zhang, Huan Zheng, Richang Hong, Mingliang Xu, Shuicheng Yan, and Meng Wang. Deep color consistent network for low-light image enhancement. In *CVPR*, pages 1889–1898, 2022. 3
- [74] Zhihong Zhang, Yuxiao Cheng, Jinli Suo, Liheng Bian, and Qionghai Dai. INFWIDE: image and feature space wiener deconvolution network for non-blind image deblurring in low-light conditions. *IEEE TIP*, 32:1390–1402, 2023. 3
- [75] Zunjin Zhao, Bangshu Xiong, Lei Wang, Qiaofeng Ou, Lei Yu, and Fa Kuang. RetinexDIP: A unified deep framework for low-light image enhancement. *IEEE TCSVT*, 32(3):1076–1088, 2022. 3, 7
- [76] Dewei Zhou, Zongxin Yang, and Yi Yang. Pyramid diffusion models for low-light image enhancement. In *IJCAI*, pages 1795–1803, 2023. 3
- [77] Shangchen Zhou, Chongyi Li, and Chen Change Loy. LED-Net: Joint low-light enhancement and deblurring in the dark. In *ECCV*, pages 573–589, 2022. 1, 2, 3, 6, 7