Published: 27 October 2025

Citation in BibTeX format

MM '25: The 33rd ACM International Conference on Multimedia
October 27 - 31, 2025
Dublin, Ireland

Conference Sponsors:
SIGMM

RESEARCH-ARTICLE

Severe Light, Textureless Sight: A Benchmark for Extreme Exposure Correction

BO WANG, Beijing University of Posts and Telecommunications, Beijing, Beijing, China

JIN LIU, Beijing University of Posts and Telecommunications, Beijing, Beijing, China

HUIYUAN FU, Beijing University of Posts and Telecommunications, Beijing, Beijing, China

XIN WANG, Stony Brook University, Stony Brook, NY, United States

HENG ZHANG, Xiaomi, Inc., Beijing, China

HUADONG MA, Beijing University of Posts and Telecommunications, Beijing, Beijing, China

Open Access Support provided by:

Stony Brook University

Xiaomi, Inc.

Beijing University of Posts and Telecommunications

MM '25: Proceedings of the 33rd ACM International Conference on Multimedia (October 2025)
https://doi.org/10.1145/3746027.3755353
ISBN: 9798400720352

Severe Light, Textureless Sight: A Benchmark for Extreme Exposure Correction

Bo Wang
Beijing University of Posts and Telecommunications
Beijing, China
wbkumiko@bupt.edu.cn

Jin Liu
Beijing University of Posts and Telecommunications
Beijing, China
ljin@bupt.edu.cn

Huiyuan Fu*
Beijing University of Posts and Telecommunications
Beijing, China
fhy@bupt.edu.cn

Xin Wang
State University of New York at Stony Brook
Stony Brook, New York, United States
x.wang@stonybrook.edu

Heng Zhang
Xiaomi Corporation
Chengdu, China
kazenokizi@live.com

Huadong Ma
Beijing University of Posts and Telecommunications
Beijing, China
mhd@bupt.edu.cn

Abstract

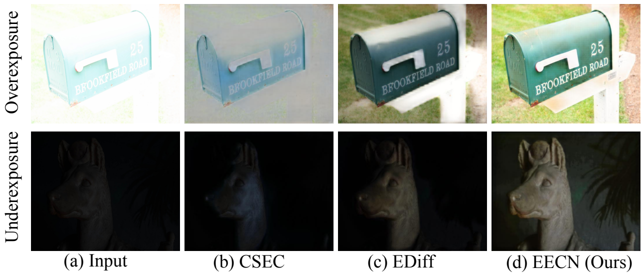
Exposure correction aims to restore underexposed and overexposed images to normal exposed images in a single network. However, conventional methods primarily focus on correcting non-extreme exposure cases and struggle to accurately restore lightness and structure information in extreme exposure scenarios. Through a thorough investigation, we observe that the extreme exposure correction task is limited by the lack of high-quality benchmark datasets. To address the above challenges, in this paper, we construct the first Extreme Exposure Dataset named EED by manually collecting a large number of diverse scenes. By introducing probabilistic blur kernel, EED not only ensures the rich diversity and brightness distribution of scenes but also approaches the degradation of the real world. To achieve exposure correction in extreme conditions, we propose a novel Extreme Exposure Correction Network by leveraging the mask-aware Fourier transform prior, which decouples lightness and structure components precisely. To restore severe abnormal lightness and lost structure information in extreme exposure scenes, we introduce a well-exposed referenced image to guide the coarse restoration and employ a Timestep-guided Frequency Diffusion Module for further refinement. Extensive experiments demonstrate the superiority of our dataset and method. The dataset will be available at https://github.com/juvenoia/EED.

CCS Concepts

• **Computing methodologies → Machine learning**.

Keywords

Exposure Correction, Fourier Transform, Dataset



(a) Input (b) CSEC (c) EDiff (d) EECN (Ours)

Figure 1: Extreme exposure scenarios often encounter severe abnormal lightness and lost structure information. Conventional non-extreme exposure correction methods CSEC [16] and EDiff [32] struggle to recover lightness and structure information, while our proposed EECN performs well.

ACM Reference Format:
Bo Wang, Jin Liu, Huiyuan Fu, Xin Wang, Heng Zhang, and Huadong Ma. 2025. Severe Light, Textureless Sight: A Benchmark for Extreme Exposure Correction. In *Proceedings of the 33rd ACM International Conference on Multimedia (MM '25), October 27–31, 2025, Dublin, Ireland.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3746027.3755353

1 Introduction

With rapid development of electronic devices, we frequently take photos in our daily lives. However, abnormally exposed images are often captured under unsatisfactory lighting conditions, which causes unreasonable lightness and structure information loss. In non-extreme exposure scenarios, the degraded images can be enhanced through professional adjustment or conventional exposure correction algorithms. Traditional methods utilize histogram equalization [2, 28, 37] and gamma correction [6, 11] to make the luminance distribution uniform. Some learning-based methods utilize the exposure-related priors or assumptions such as Retinex theory [20, 33, 34, 36], curve mapping [7, 13, 14], Fourier transform [8, 15, 21, 38], and customized approaches [16–18, 22, 30, 35]. Other methods [1, 9, 10, 18, 29] concentrate on the variations of representations between under/over-exposure and introduce improved optimization strategies to promote optimization consistency.

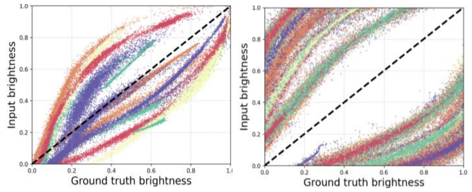
Designed to work under simple non-extreme conditions with moderate brightness and well-preserved structural information, the

---

[*] Corresponding author.

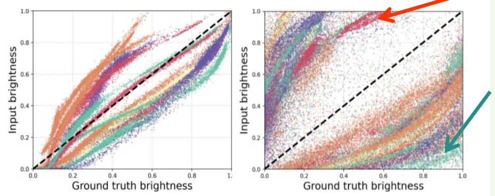
MM '25, Dublin, Ireland

ACM ISBN 979-8-4007-2035-2/2025/10
https://doi.org/10.1145/3746027.3755353



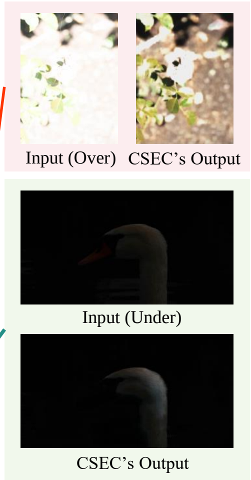
(a) Luminance mapping of MSEC Dataset

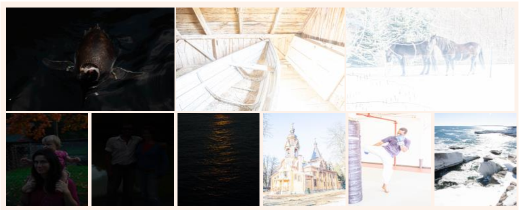
(c) Luminance mapping of EED Dataset (ours)



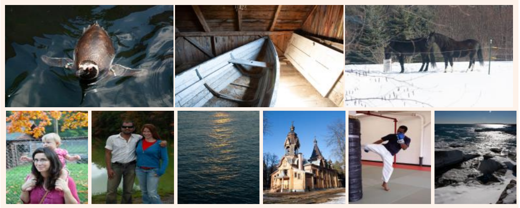
(b) Luminance mapping of DIME Dataset

(d) Input-Output luminance mapping of CSEC model





(e) Input samples from EED



(f) Ground Truth samples from EED

Figure 2: Luminance mapping curves for non-extreme exposure datasets MSEC (a) and DIME (b) are concentrated in the center, while ours (c) exhibits a more dispersed curve, indicating our input luminance is extreme. The luminance mapping learned by the non-extreme method CSEC is shown in (d), demonstrating limited recovery capability. (e) and (f) show some samples from our dataset.

performance of these schemes often suffers in extreme exposure scenarios. For example, as shown in Fig. 1, with the extremely over-/under-exposed images taken under intense light, drastic lightness changes or almost no light conditions, CSEC [16] and EDiff [32] fail to recover proper lightness and detailed structural information.

To overcome above challenges, in this paper, we focus on the task of correcting images under extreme exposure for the first time. To our best knowledge, there is currently an absence of well-constructed methods or benchmark datasets. We make two contributions in this paper. To create a benchmark to facilitate the algorithm design, we propose the first Extreme Exposure Dataset (EED) where images are taken under more severe abnormal lightness with realistic structure information and in abundant scenes. We manually collect RAW images from over 20,000 scenes from various online archives. Subsequently, we utilize image editing software to adjust the brightness of RAW images based on exposure value (EV). We also propose a probabilistic blur kernel based on statistics from multi-exposure real-world scenes to make the lightness adjustment of images more realistic. Additionally, we construct a test set EED-Real to evaluate the model's ability of restoring images from the degradations in real-world extreme exposure scenarios.

As a second contribution, we propose an Extreme Exposure Correction Network (EECN) to recover distorted images from severely abnormal lightness and large structural information degradation. Due to the incomplete disentanglement of lightness and structural information based on the vanilla Fourier prior, we leverage a mask-aware Fourier prior to precisely decouple lightness and structure components. To restore images from severe lightness abnormality and structural information loss in extreme exposure scenes, we introduce the Dual Frequency Cross-attention Module to fuse and match additional information decoupled from the well-exposed reference image in the frequency domain. For further lightness refinement and structural information reproduction, we propose a Timestep-guided frequency diffusion module that emphasizes frequency-specific information according to the time-step variation. In brief, our contributions can be summarized as follows:

- We manually collect over 20,000 scenes and construct the first Extreme Exposure Dataset with varying and severe lightness and realistic structural information degradation.
- To achieve extreme exposure correction, we introduce a novel method that accurately restores lightness and structural information via mask-aware Fourier prior and timestep-guided frequency Diffusion module.
- Extensive experiments on general and extreme exposure correction tasks demonstrate the superiority of our dataset and method.

2 Dataset

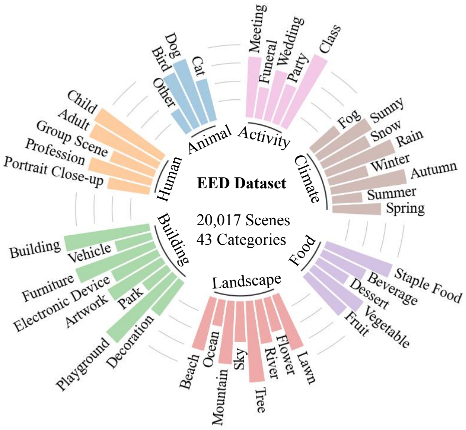
Through thorough investigation, the development of extreme exposure correction is limited by the absence of a benchmark dataset.

We first revisit two conventional exposure datasets, the Multi-Scale Exposure Correction (MSEC) dataset [1] and the Dynamic scenes In Multiple Exposure dataset (DIME) [19]. Proposed to solve multi-exposure problems, their lightness intensity is not extreme. From the luminance mapping (luminance calculated according to ITU-R 709[1]) in Fig. 2 (a) and (b), the curves concentrate at the center, indicating that the luminance mappings are closed to an identical function. We also show the luminance mapping of a recent non-extreme exposure correction SOTA method CSEC [16] in Fig. 2 (d), where we can see that CSEC fails to retouch the image in extreme exposure scenarios (see the red and green box in the middle part of Fig. 2). This demonstrates a new dataset on extreme exposure scenarios is desired.

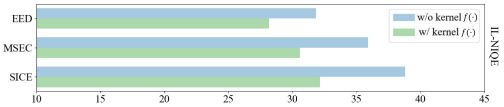
We collect RAW images of over 20,000 scenes from various websites and quantify the categories to ensure the scene quantity and diversity, as shown in Fig. 3 (a). For each RAW-RGB image, we employ Adobe Photoshop Lightroom to emulate different EVs as

---

[1] ITU-R Recommendation 709, https://www.itu.int/rec/r-rec-bt.709



(a) Category statistics in our dataset.



(b) $f(\cdot)$ reduces the IL-NIQE metric on each dataset.

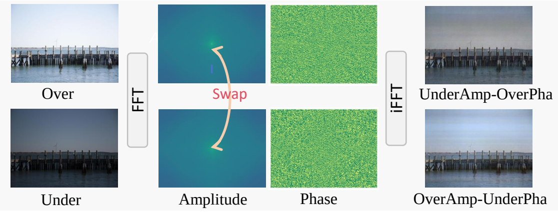
Figure 3: (a) Dataset category statistics of EED. (b) We apply the probabilistic blur kernel $f(\cdot)$ to the MSEC, SICE, and EED datasets. The IL-NIQE (↓) metric comparison on the origin datasets and their $f(\cdot)$ applied counterpart indicates that $f(\cdot)$ make the images appear more natural.

would be applied by a camera. Lightroom leverages the metadata embedded in various RAW image formats (such as .dng) to simulate the nonlinear camera rendering process. Specifically, we apply exposure adjustments of -2.5 EV, -2 EV, 2 EV, and 2.5 EV to generate extremely overexposed and underexposed image samples.

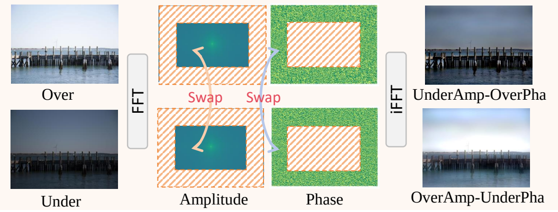
Regarding the ground truth (GT) images, we prioritize selecting those captured at EV = 0. However, if an EV = 0 image contains overexposed or underexposed regions, we utilize Lightroom's highlight and shadow adjustment functions to refine these areas, ensuring that local textures remain visible. Furthermore, for the generated extreme exposure samples, we also apply highlight and shadow adjustments in Lightroom to mitigate large-scale saturation or clipping artifacts. This step helps prevent severe model hallucination caused by excessive overexposure or underexposure.

However, there is a significant gap between synthetic and real-world data. Real-world data collection is limited by the variation in the natural ambient light or shooting angles, thus lacking rich scene diversity. To make our data more realistic, we revisit the real-world multi-exposure dataset DIME [19]. We utilize edge detection tool [4] to identify edges for the above image pairs and record the pixels where degradations occur. Then we apply a convolution with a kernel of 5 × 5 to compute the weighted average lightness of each pixel. Finally, we fit a probability curve f that represents the relation between the luminance and the structure degradation, thereby obtaining the probabilistic blur kernel $f(\cdot)$ for pixel $I(i,j)$ on sRGB images:

$$P(f(I(i,j)) = K \cdot I(i,j)) = 1 - \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(\phi(I(i,j))-\mu)^2}{2\sigma^2}}, \quad (1)$$



(a) Vanilla Fourier Prior



(b) Mask-aware Fourier Prior

Figure 4: Vanilla Fourier prior (a) swaps all parts of amplitude while keeping phase identical, resulting in a significant difference in exposure level compared to the input images. Mask-aware Fourier prior (b) is presented as a more general case of (a), bringing the exposure results closer to the input image. The complementary parts of the amplitude and phase are swapped, while the masked area (indicated by diagonal lines) is unchanged.

where $\phi(\cdot)$ transforms $I(i,j)$ from sRGB to the YCbCr format and obtains the average convolution value of the Y channel in the 5 × 5 window, f indicates the pixel-wise mapping function, μ and σ denote mean and variance. We use a Gaussian blur kernel K to approximate edge degradation. The probability of $f(I(i,j))$ follows the Bernoulli distribution, and the complementary case is $f(I(i,j)) = I(i,j)$, indicating that no degeneration occurs.

Then we apply the probabilistic blur kernel to images generated by Lightroom. In total, we collect 56,244 paired multi-exposure images, of which 39,370 are used for training, 11,249 for validation, and 5,625 for testing. Fig. 2 (c) shows that the luminance mapping of our EED dataset is more dispersed, indicating the extreme lightness intensity. Some dataset samples are shown in Fig. 2 (e) and (f). In summary, our dataset is not only diverse and rich in scenes with extreme ill-exposure, but also approaches the degradation of the real world, as shown in Fig. 3 (b).

Furthermore, we utilize different devices to capture an unpaired real-world extreme exposure test set of 200 photos, namely EED-Real. The photos are taken with four devices, including Canon EOS R10, Huawei P40, iPhone XS, and OPPO Reno 12. The examples can be found in the supplementary materials.

3 Method

3.1 Preliminary

Fourier transform is widely used to analyze the frequency content of images due to unique global receptive field property [26]. Given a single channel in an image $x \in \mathbb{R}^{H \times W}$, the 2D discrete Fourier transform is described as:

$$\mathcal{F}(x)(u,v) = \frac{1}{HW} \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} x(h,w) e^{-j2\pi(\frac{h}{H}u + \frac{w}{W}v)}, \quad (2)$$

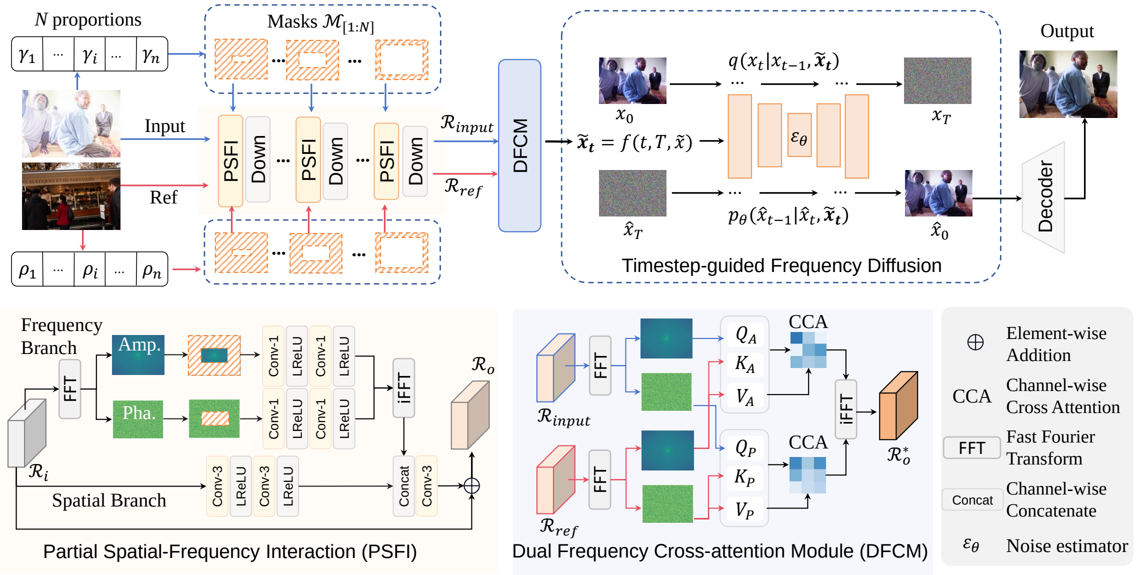


Figure 5: Overall architecture of EECN. We introduce mask-aware Fourier prior to generating mask sequence, guiding PSFI to precisely decouple amplitude and phase components. Then DFCM leverages well-exposed reference images to supplement lightness and structure information for coarse restoration. TFD further refines the image based on the timestep-dependent characteristic of the denoising process.

where h, w are the coordinates in the spatial space and u, v are the coordinates in the Fourier space. We denote $\mathcal{F}(x)(u,v)$ as $X(u,v)$. X consists of complex values, where the real part and imaginary part are $R(X)$ and $I(X)$. The amplitude $\mathcal{A}$ and phase $\mathcal{P}$ components are obtained from the real and imaginary parts of complex X:

$$\mathcal{A}(X) = \sqrt{R^2(X) + I^2(X)}, \mathcal{P}(X) = arctan\frac{I(X)}{R(X)}. \tag{3}$$

According to the previous Fourier-based method [8], the amplitude and phase components in exposure correction are associated with lightness and structural representation respectively as shown in Fig. 4 (a). However, vanilla Fourier prior simply swaps the amplitude component, which leads to ill-exposed problems in the target image. This is due to the incomplete disentanglement of the lightness and structure characteristics. We observe that in the frequency spectrum obtained after Fourier transform and shift operation, there is more low-frequency information near the center, while high-frequency information is more abundant farther from the center. Additionally, the amplitude primarily captures low-frequency illumination information, while the phase mainly captures high-frequency structural information. Therefore, we propose a mask-aware Fourier prior, which concentrates low-frequency information from the amplitude and high-frequency information from the phase through the controllable mask, achieving more precise decoupling of lightness and structural information:

$$x' = iFFT(\mathcal{A}(x) \cdot \mathcal{M} + \mathcal{P}(x) \cdot (1 - \mathcal{M})), \tag{4}$$

where $\mathcal{M}$ denotes the mask with a value of 0 or 1, filtering information of diverse frequencies at different positions. As shown in Fig. 4 (b), the generated over-/under-exposed images exhibit exposure levels closer to the original ill-exposed images before the transformation. More discussion is shown in Sec. 4.3.

3.2 Partial Frequency-Spatial Interaction Module

Based on the above analysis of mask-aware Fourier Prior, as shown in Fig. 5, we propose a Partial Frequency-Spatial Interaction Module to precisely decouple lightness and structure components in extremely ill-exposed scenarios. Furthermore, considering that convolution acts as a high-pass filter [24], this suggests that more emphasis should be placed on encoding low-frequency representations in the amplitude at the shallow layers of the network. To this end, as the network depth increases, we design the proportion of the mask area to change monotonically.

We leverage a learnable predictor consisting of a series of Conv-LeakyReLU blocks, MLPs, and ending with a Softplus function (see supplementary material for details) to predict the proportion sequence of the mask:

$$\begin{aligned}
&[d_1, d_2, \ldots, d_n] = \mathrm{Softplus}(\mathrm{MLP}(C_3(C_3(I)))), \\
&\Delta_i = \frac{d_i}{\sum_{j=1}^{N} d_j}, \gamma_i = \sum_{k=1}^{i} \Delta_k, i = 1, 2, ..., n,
\end{aligned} \tag{5}$$

where C_3 denotes the Conv3 × 3 and LeakyReLU function, The N values of γ are between [0, 1], arranged in ascending order, which indicates that as the network depth increases, it enables a more precise focus on both high- and low-frequency information.

The above proportion sequence is used to generate the masks, which are then sent to the PSFI module. Each PSFI at level i takes feature $\mathcal{R}_i$ and Fourier mask $\mathcal{M}_i$ as the input. The spatial branch is processed by a series of convolutions. For the frequency branch, we filter amplitude and phase with complementary masks:

$$\begin{aligned}
&\mathcal{A}(\mathcal{R}_i) = \mathcal{A}(\mathcal{R}_i) \cdot \mathcal{M}_i, \mathcal{P}(\mathcal{R}_i) = \mathcal{P}(\mathcal{R}_i) \cdot (1 - \mathcal{M}_i), \\
&\mathcal{R}_o = C_3(Concat(\mathcal{R}_f, \mathcal{R}_s)) + \mathcal{R}_i,
\end{aligned} \tag{6}$$

where $\mathcal{A}(\mathcal{R}_i), \mathcal{P}(\mathcal{R}_i)$ are the outputs of $\mathcal{R}_i$ after Fourier transform operation. $\cdot$ represents element-wise multiplication. $\mathcal{R}_f, \mathcal{R}_s$ denote the output of the frequency branch and the spatial branch, respectively.

3.3 Dual Frequency Cross-attention Module

In extreme exposure scenarios, directly enlightening the image cannot restore structure information well, and lightness may appear unnatural. Hence, we guide the restoration using lightness and structure information from an unpaired well-exposed reference image.

We first encode input and reference images into features $\mathcal{R}_{input}$ and $\mathcal{R}_{ref}$ utilizing PSFIs. To achieve better alignment and complementarity of lightness and structural information, we decomposed these features into amplitude $\mathcal{A}_{input}, \mathcal{A}_{ref}$ and phase component $\mathcal{P}_{input}, \mathcal{P}_{ref}$. To learn each $\mathcal{A}$ and $\mathcal{P}$, several convolutions are applied:

$$\hat{\mathcal{A}} = C_1(\xi(C_1(\mathcal{A}))), \hat{\mathcal{P}} = C_1(\xi(C_1(\mathcal{P}))), \tag{7}$$

where ξ is ReLU function. Then we apply cross-attention to $\mathcal{A}$ and $\mathcal{P}$ to provide additional lightness and structure information independently, resulting in modified $\mathcal{A}^*$ and $\mathcal{P}^*$:

$$\mathcal{A}^* = \mathrm{CCA}(\hat{\mathcal{A}}_{input}, \hat{\mathcal{A}}_{ref}), \mathcal{P}^* = \mathrm{CCA}(\hat{\mathcal{P}}_{input}, \hat{\mathcal{P}}_{ref}), \tag{8}$$

where *CCA* denotes channel-wise cross attention operation. The final feature $\mathcal{R}_o^*$ is obtained by the inverse Fourier transform with the combination of $\mathcal{A}^*$ and $\mathcal{P}^*$.

3.4 Timestep-guided Frequency Diffusion Module

Feature $\mathcal{R}_o^*$ exhibits coarsely restored lightness and structure information. However, there might still be slight lightness mismatch and structure information loss. We propose a Timestep-guided Frequency Diffusion Module (TFDM), leveraging the powerful generative capabilities of the Diffusion Model to refine the image. The forward process transforms the input $\mathbf{x}_0$ into a noise sample $\mathbf{e} \in \mathcal{N}(\mathbf{0}, \mathbf{I})$ in step T:

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_{t-1}}\mathbf{x}_{t-1}, \beta_t \mathbf{I}, \tilde{\mathbf{x}}_t) \tag{9}$$

Here, $\beta_1, \beta_2, \ldots, \beta_T$ is a sequence of pre-defined variables related to the time step. By defining $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=0}^{t} \alpha_i$, we can derive the relationship between $\mathbf{x}_t$ and $\mathbf{x}_0$ as $\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t$, where ϵ_t is from $\mathcal{N}(\mathbf{0}, \mathbf{I})$ In the forward process of the TFDM, we utilize the modified features $\mathcal{R}_o^*$ as condition $\tilde{\mathbf{x}}$ to guide the training of the model's forward process.

For further low-frequency lightness refinement and high-frequency structural information reproduction, we observe that during the denoising process in Diffusion, the restoration of high- and low-frequency information exhibits a timestep-dependent characteristic: low-frequency information is prioritized in the earlier timesteps, while high-frequency information is prioritized in the later timesteps. (More discussion is shown in Supplementary Material.) Thus we utilize the timestep variation to guide the learning of frequency-specific information. We modify the fixed condition $\tilde{\mathbf{x}}$ to time-aware $\tilde{\mathbf{x}}_t$ by defining $\tilde{\mathbf{x}}_t = f(t, T, \tilde{\mathbf{x}})$ as follows: For each time step t, we generate a mask $\mathcal{M}$ with ratio $\gamma = \frac{t}{T}$, where T denotes the total timestep, then apply to $\tilde{\mathbf{x}}$ to generate $\tilde{\mathbf{x}}_t$. The reverse process recovers $\mathbf{x}_0$ from random sampled noise $\hat{x}_T$ with condition $\tilde{\mathbf{x}}_t$:

$$p_\theta(\hat{\mathbf{x}}_{t-1} \mid \hat{\mathbf{x}}_t, \tilde{\mathbf{x}}_t) = \mathcal{N}(\hat{\mathbf{x}}_{t-1}; \mu_\theta(\hat{\mathbf{x}}_t, \tilde{\mathbf{x}}_t, t), \sigma_t^2 \mathbf{I}), \tag{10}$$

where $\sigma_t^2 = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$ denotes variance and $\mu_\theta(\hat{\mathbf{x}}_t, \tilde{\mathbf{x}}_t, t) = \frac{1}{\sqrt{\alpha_t}}(\hat{\mathbf{x}}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\hat{\mathbf{x}}_t, \tilde{\mathbf{x}}_t, t))$ denotes mean value.

3.5 Network Training

Our method adopts a two-stage strategy for network training. The first training stage consists of PFSI, DFCM, and decoder $\mathcal{D}$. The parameters of the diffusion module (TFDM) are frozen. We select ill-exposed image I_{in} and well-exposed image I_{gt} pair on the MSEC dataset for training. We utilize the reconstruction loss as supervision:

$$\mathcal{L}_{rec} = ||I_{gt} - \mathcal{D}(\mathcal{E}(I_{in})||_2 + ||I_{gt} - \mathcal{D}(\mathcal{E}(I_{gt})||_2, \tag{11}$$

where $\mathcal{E}$ represents a series of PSFI. The predicted amplitude and phase of DFCM($\mathcal{E}(I_{in}), \mathcal{E}(I_{gt})$) aim to align with the amplitude of $\mathcal{E}(I_{gt})$ and the phase of $\mathcal{E}(I_{in})$. Let DFCM($\mathcal{E}(I_{in}), \mathcal{E}(I_{gt})$) denote $\mathcal{Z}$, then the frequency loss is expressed as:

$$\mathcal{L}_f = ||\mathcal{A}(\mathcal{Z}) - \mathcal{A}(\mathcal{E}(I_{gt}))||_2 + ||\mathcal{P}(\mathcal{Z}) - \mathcal{P}(\mathcal{E}(I_{in}))||_2, \tag{12}$$

The total loss for first stage is $\mathcal{L}_{s1} = \lambda_1 \mathcal{L}_{rec} + \lambda_2 \mathcal{L}_f$. In the second stage, the parameters of the above modules are frozen. We barely optimize the diffusion model. The total loss for TFDM is $\mathcal{L}_{s2} = \mathcal{L}_{df} + \lambda_3 \mathcal{L}_{con}$:

$$\mathcal{L}_{df} = ||\epsilon_t - \epsilon_\theta(x_t, \tilde{x}_t, t)||, \mathcal{L}_{con} = ||\mathcal{F}_{gt} - \mathcal{F}_{pred}||, \tag{13}$$

where θ indicates Gaussian noise predicting network, and following [12], $\mathcal{F}_{pred}$ means the output of reverse process with the condition $\tilde{x}_t$.

4 Experiments

4.1 Experimental Settings

Implementation Details. We implement the proposed method with PyTorch on two NVIDIA RTX 3090Ti GPUs, where the batch size and patch size are set to 4 and 512×512. We employ the Adam optimizer for optimization with the initial learning rate set to 1e-4 in the encoder training and decay by a factor of 0.8, and fixed 1e-4 for the TFDM training. The encoder-decoder modules in the first stage converge after 54k iterations, while the diffusion with 48k iterations. The PFSI layers is set to 3, and $\lambda_1, \lambda_2, \lambda_3$ are set to $1, 0.1, 0.1$, respectively. We utilize U-Net architecture for TFDM to estimate Gaussian noise.

Dataset and Metrics We evaluate our method and baseline methods on SICE [3], MSEC [1], EED and unpaired real-world dataset EED-Real. We use supervised metrics PSNR and SSIM to assess the similarity between the generated images and the ground truth. For the unpaired dataset, we choose unsupervised metrics NIQE [23] and PI [31] to evaluate the quality of the model performance.

4.2 Comparison with Current Methods

Comparison Methods We select existing methods, including physical methods HE [25] and CLAHE [27], supervised methods URetinexNet [34], LCDP [29], FECNet [8], SMG [35], RECNet [18],

Table 1: Quantitative PSNR/SSIM (↑) and NIQE/PI (↓) results of methods on EED and EED-Real dataset. The best results are shown in bold while the suboptimal results are underlined.

Method	Publication	EED						EED-Real					
		Under		Over		Average		Under		Over		Average	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	NIQE	PI	NIQE	PI	NIQE	PI
HE [25]	CVGIP’87	14.05	0.7044	12.85	0.6617	13.45	0.6831	5.121	4.779	5.318	5.246	5.220	5.013
CLAHE [27]	VLSI’04	11.38	0.5845	13.79	0.6033	12.59	0.5939	5.203	4.855	5.676	5.353	5.439	5.104
URetinexNet [34]	CVPR’22	10.22	0.6494	13.63	0.7527	11.93	0.7011	4.810	4.937	5.210	4.452	5.019	4.695
LCDP [29]	ECCV’22	21.48	0.8424	20.21	0.8171	20.85	0.8298	4.483	4.650	4.153	3.948	4.318	4.299
FECNet [8]	ECCV’22	21.72	0.8089	19.47	0.7972	20.60	0.8031	4.563	4.511	4.312	3.952	4.438	4.232
SMG [35]	CVPR’23	20.49	0.8367	21.44	0.8419	20.97	0.8393	3.957	3.603	4.121	4.142	4.039	3.873
LightenDiffusion [12]	ECCV’24	17.01	0.7184	17.76	0.7261	17.38	0.7223	4.041	3.825	3.787	4.200	3.920	4.013
ZeroDCE++ [7]	CVPR’20	14.94	0.5432	13.00	0.4802	13.97	0.5117	4.808	4.286	4.478	4.372	4.643	4.329
SCI [22]	CVPR’22	10.92	0.5448	9.34	0.6468	10.13	0.5958	4.875	5.120	4.907	5.112	4.891	5.116
GDP [5]	CVPR’23	15.57	0.5982	13.97	0.6776	14.77	0.6379	4.705	4.545	4.437	4.779	4.571	4.662
FourierDiff [21]	CVPR’24	17.98	0.7012	15.79	0.5854	16.89	0.6433	4.470	3.551	4.131	3.641	4.301	3.596
CoTF [17]	CVPR’24	21.01	0.7993	21.50	0.8285	21.26	0.8139	3.909	3.957	4.371	3.784	4.140	3.871
RECNet [18]	AAAI’24	21.28	0.8197	21.67	0.8439	21.48	0.8318	4.030	3.890	4.066	3.793	4.048	3.842
CSEC [16]	CVPR’24	21.50	0.8250	21.61	0.8487	21.56	0.8369	3.719	3.448	3.872	3.553	3.796	3.501
EDiff [32]	ICCV’23	21.80	0.8479	21.22	0.8397	21.51	0.8438	3.980	3.383	3.781	3.568	3.881	3.476
EECN (Ours)	/	**22.01**	**0.8508**	**21.94**	**0.8574**	**21.98**	**0.8541**	**3.626**	**3.081**	**3.755**	**3.488**	**3.691**	**3.265**

Table 2: Quantitative PSNR and SSIM (↑) results on SICE and MSEC dataset. The best results are shown in bold while the suboptimal results are underlined.

Method	MSEC						SICE					
	Under		Over		Average		Under		Over		Average	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
HE [25]	16.52	0.6918	16.53	0.6991	16.53	0.6959	14.69	0.5651	12.87	0.4991	13.78	0.5376
CLAHE [27]	16.77	0.6211	14.45	0.5842	15.38	0.5990	12.69	0.5037	10.21	0.4847	11.45	0.4942
URetinexNet [34]	13.85	0.7371	9.81	0.6733	11.42	0.6988	17.39	0.6448	7.40	0.4543	12.40	0.5496
LCDP [29]	22.35	0.8650	22.17	0.8476	22.30	0.8552	17.45	0.5622	17.04	0.6463	17.25	0.6043
FECNet [8]	22.96	0.8598	23.22	0.8748	23.12	0.8688	22.01	0.6737	19.91	0.6961	20.96	0.6849
SMG [35]	23.28	0.8653	22.17	0.8622	22.73	0.8638	22.73	0.8168	22.10	0.7566	22.42	0.7867
LightenDiffusion [12]	18.25	0.6382	18.01	0.6011	18.13	0.6197	19.08	0.6443	18.86	0.6722	18.97	0.6583
ZeroDCE++ [7]	13.82	0.5887	9.74	0.5142	11.37	0.5583	11.93	0.4755	6.88	0.4088	9.41	0.4422
SCI [22]	9.97	0.6681	5.84	0.5190	7.49	0.5786	17.86	0.6401	4.45	0.3629	12.49	0.5051
GDP [5]	14.18	0.7213	13.68	0.5429	13.93	0.6321	15.78	0.5899	10.30	0.5167	13.04	0.5533
FourierDiff [21]	18.90	0.7630	18.54	0.5641	18.72	0.6636	19.57	0.7022	15.28	0.5271	17.43	0.6147
CoTF [17]	23.36	0.8630	23.49	0.8793	23.44	0.8728	22.90	0.7029	20.13	0.7274	21.51	0.7151
RECNet [18]	23.57	0.8655	23.81	0.8765	23.69	0.8699	22.50	0.7254	20.81	0.7363	21.66	0.7309
CSEC [16]	23.28	0.8550	23.67	0.8564	23.62	0.8557	23.04	0.8730	22.40	0.8546	22.72	0.8638
EDiff [32]	23.51	0.8691	23.63	0.8801	23.57	0.8746	23.38	0.8741	21.95	0.8416	22.67	0.8579
EECN (Ours)	**24.09**	**0.8693**	**23.96**	**0.8885**	**24.03**	**0.8789**	**23.41**	**0.8830**	**22.61**	**0.8677**	**23.12**	**0.8754**

CSEC [16], ExposureDiff [21], and unsupervised methods LightenDiffusion [12], SCI [22], GDP [5], FourierDiff [21] and ExposureDiffusion [32].

Quantitative Comparison We first evaluate PSNR, SSIM on EED, and NIQE, PI metrics on the EED-Real. As shown in Tab. 1, our method outperforms existing methods by 0.5 PSNR and 0.02 SSIM on EED, and all NIQE and PI metrics. Then, we compare the methods on the conventional non-extreme ill-exposed SICE and MSEC datasets with PSNR and SSIM. As shown in Tab. 2, our model also performs better across these evaluations.

Qualititive Comparison We select several competitive methods for qualitative comparison. Fig. 6 shows the visual results on the SICE, MSEC, and EED datasets. Our results exhibit clearer textures and more vivid colors on previous datasets, and our method achieves better texture restoration and more natural brightness on the EED dataset. Furthermore, for the EED-Real dataset shown in Fig. 7, existing methods struggle to restore out-of-distribution samples, while our method performs well. In addition, we also conduct a user study to evaluate the subjective perceptions of different methods in supplementary material.

4.3 Mask-aware Fourier Prior

Amplitude and Phase Filters We analyze 50 paired under-/over-exposed images. We swap amplitude with a mask $\mathcal{M}$ at different scales generated by γ sampled from 0 to 1 with step length 0.01, and keep phase identical, calculating the PSNR/SSIM between over-exposed image and amplitude-swapped underexposed image. The result is shown in Fig. 9 (a), as γ reaches the peak at low frequency, and declines fast as γ grows. This indicates that lightness information is concentrated in the low-frequency portion of the center of the spectrogram. As for phase, we conduct the same experimental setting except we apply the complementary mask $1 - \mathcal{M}$ to phase,

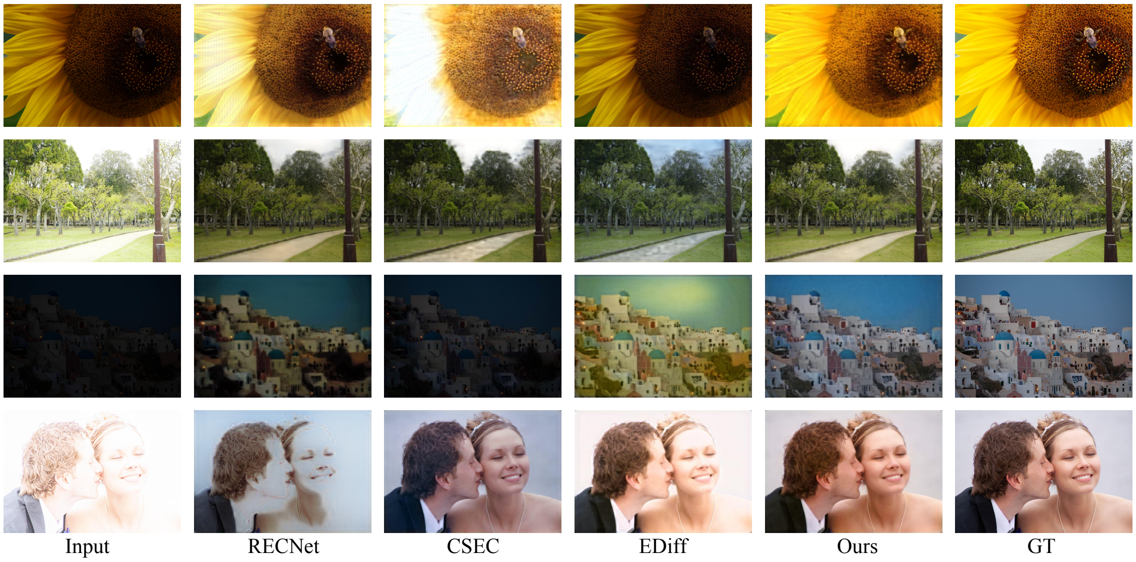


Figure 6: Quantitative test with several competitive methods. The first and second rows present under-/over-exposed cases on MSEC/SICE datasets, while the third and fourth rows are under-/over-exposed cases on EED. More example comparisons can be found in the supplementary material.

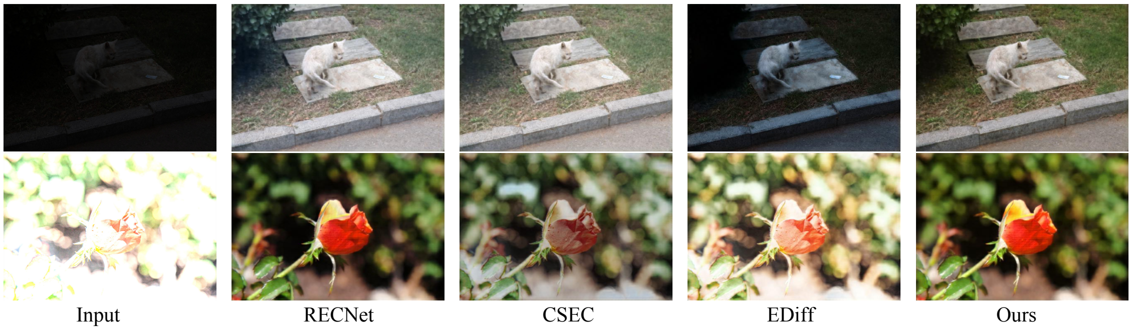


Figure 7: Quantitative test with several competitive methods on the unpaired real dataset EED-real. The first and second rows present under-/over-exposed cases, respectively. More example comparisons can be found in the supplementary material.

keep amplitude identical, and compare the overexposed image to the phase-swapped overexposed image. As shown in Fig. 9 (b), the γ works better at high frequency, which indicates the structural information is concentrated in the high-frequency part far away from the center of the spectrogram.

Investigate on Mask Value We perform complementary mask exchange between amplitude and phase based on the prior in Eq. 4, with a step size of 0.005 and values ranging between [0, 1]. As shown in Fig. 9 (c), the best result appears when $\gamma = 0.04$. Qualitative comparisons in Fig. 9 (f) and (g) illustrate that vanilla Fourier prior produces artifact and noise, while mask-aware Fourier prior performs better.

In this section, We conduct ablation studies to explore the effectiveness of the proposed PFSI, DFCM, TFDM, and losses on EED and EED-Real datasets. The quantitative and qualitative results are shown in Tab. 4 and Fig. 8. We evaluate model running time with one 512×512 image.

4.4 Ablation Studies

Partial Frequency-Spatial Interaction Module We first validate the number of PSFI modules. As shown in Tab. 4, $n = 3$ ((d) ours) achieves the best performance, with affordable model running time. To validate mask-aware Fourier decomposition, we set: (a) PSFI without mask, (b) PSFI with mask fixed $\gamma = 0.04$. As shown in Fig. 8 (a), without a mask-aware Fourier prior, the model struggles to learn effective lightness and structure information encoding representations. Applying a fixed mask enhances model performance (Fig. 8 (b)); however, the quantitative results still lag behind the learnable prior.

Dual Frequency Cross-attention Module We investigate the impact of different reference image selection strategies. Specifically, we evaluate their effects on the EED dataset through a 5-fold cross-validation, where 20% of the images are selected as reference images in each fold. As shown in Tab. 3, our EECN surpasses existing SOTA methods even with arbitrary reference images, and achieves further improvement when selecting reference images based on

Input (a) w/o mask (b) fixed mask (c) w/o ref (d) linear (e) $\mathcal{L}_{rec}$ (f) Ours (n = 3)

Input (h) n = 1 (i) n = 2 (j) n = 4 (k) vanilla DDPM (l) fixed modification (m) Ours (n = 3)

Figure 8: Ablation studies on the outputs of our model under different configurations on the displayed input.

Table 3: We investigate reference selection and fusion strategies using PSNR (↑). Our EECN outperforms SOTA even with arbitrary references and gains further with VGG-similarity selection. References also benefit other models—simple concatenation boosts performance, while DFCM enables more effective utilization.

	No Ref	Simple Concat		DFCM	
		Arbitary	VGG	Arbitary	VGG
RECNet [18]	21.48	21.50	21.55	21.59	21.63
CSEC [16]	21.51	21.52	21.57	21.63	21.73
EDiff [32]	21.56	21.59	21.64	21.67	21.76
EECN (Ours)	21.59	21.62	21.65	21.74	21.98

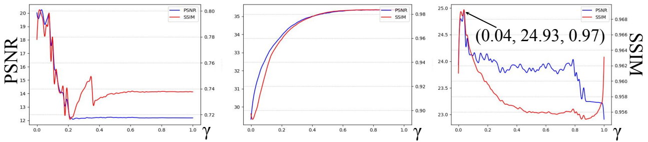
Table 4: Ablation studies. We investigate mask-aware Fourier decomposition on (a) and (b), Dual Frequency Cross-attention Module on (c) and (d), loss functions on (e). We compare the number of PSFI layers n from (g) to (j).

	Settings	EED		EED-Real		Time (s)
		PSNR	SSIM	NIQE	PI	
(a)	w/o mask	21.12	0.8013	4.231	4.129	0.298
(b)	fixed mask	21.37	0.8275	3.942	3.643	0.304
(c)	$\mathcal{L}_{rec}$	19.79	0.7791	4.623	4.348	0.311
(d)	Ours	21.98	0.8541	3.691	3.265	0.310
(e)	$n = 0$	16.19	0.6031	5.834	5.439	3.981
(f)	$n = 1$	17.23	0.7132	4.941	5.129	0.693
(g)	$n = 2$	20.12	0.8023	4.012	3.812	0.472
(h)	$n = 4$	21.65	0.8421	3.821	3.399	0.246

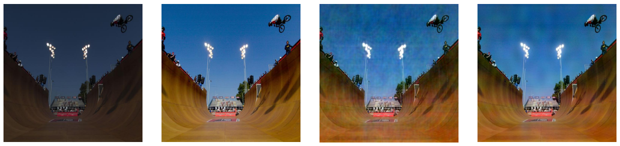
Table 5: Ablation study on TFDM with different conditions.

Method	PSNR	SSIM	Converge
(i) vanilla DDPM	21.39	0.8481	62000
(j) fixed modification	20.97	0.8309	86000
(k) timestep guidance	21.98	0.8541	48000

the highest VGG similarity. Additionally, incorporating reference images also boosts the performance of other baseline methods. Simply concatenating the reference image with the input leads to noticeable gains, while integrating the DFCM module results in even greater improvements[2]. The performance of EECN without reference images and with simple concatenation is also visualized in Fig. 8 (c) and (d).



(a) Swap amplitude (b) Swap phase (c) Swap both



(d) Under (e) Over (f) γ = 0 (g) γ = 0.04

Figure 9: (a) and (b) show the amplitude and phase mainly focus on low and high frequency, respectively. (c) to (g) demonstrate that the mask-aware Fourier prior outperforms the vanilla Fourier prior both quantitatively and qualitatively.

Timestep-guided Frequency Diffusion Module To validate the effectiveness of TFDM, we conduct experiments with the following settings: (i) $\tilde{x}$ without modification; (j) modification of $\tilde{x}$ with $\alpha = 0.04$; and (k) timestep-guided setting. The results in Tab. 5 indicate that (k) converged to better performance in fewer iterations.

Losses To verify the effectiveness of $\mathcal{L}_f$, we conduct experiments with: (c) only using $\mathcal{L}_{rec}$. As shown in Fig. 8 (e), without $\mathcal{L}_f$, it is difficult for the model to further decompose amplitude and phase for the subsequent attention mechanism. This prevents the model from leveraging reference images to guide lightness and structure information.

5 Conclusion

We build the first extreme exposure dataset that ensures scene variation and realistic degradation. A benchmark dataset is provided for both training and evaluation of extreme exposure correction methods. To accurately restore lightness and structure information in extreme exposure scenarios, we propose a novel Extreme Exposure Correction Network by leveraging the mask-aware Fourier transform prior, which decouples lightness and structure components precisely. Then we introduce a well-exposed reference image to coarsely enhance structural information and correct lightness, then apply a Timestep-guided Frequency Diffusion Module for further refinement. Experiments demonstrate that the proposed method outperforms current non-extreme exposure correction methods.

Acknowledgements. This work is supported in part by the National Key R&D Program of China under No.2023YFF0904800,

---

[2] In the no-reference setting, the input serves also as the reference image. For simple concat, we replace DFCM with the concat operation. We fuse features of input and reference images via concatenation or attention for other models.

the NSFC under No.62272059 and No.U24B20176, the Beijing Natural Science Foundation under No.JQ24020, and the Beijing Nova Program under No.20230484406.

References

[1] Mahmoud Afifi, Konstantinos G Derpanis, Bjorn Ommer, and Michael S Brown. 2021. Learning Multi-Scale Photo Exposure Correction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 9157–9167.

[2] Tarik Arici, Salih Dikbas, and Yucel Altunbasak. 2009. A Histogram Modification Framework and Its Application for Image Contrast Enhancement. *IEEE Transactions on Image Processing* 18, 9 (2009), 1921–1935.

[3] Jianrui Cai, Shuhang Gu, and Lei Zhang. 2018. Learning a Deep Single Image Contrast Enhancer From Multi-Exposure Images. *IEEE Transactions on Image Processing* 27, 4 (2018), 2049–2062.

[4] John Canny. 1986. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8, 6 (1986), 679–698.

[5] Ben Fei, Zhaoyang Lyu, Liang Pan, Junzhe Zhang, Weidong Yang, Tianyue Luo, Bo Zhang, and Bo Dai. 2023. Generative Diffusion Prior for Unified Image Restoration and Enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 9935–9946.

[6] Xu Guan, Su Jian, Pan Hongda, Zhang Zhiguo, and Gong Haibin. 2009. An Image Enhancement Method Based on Gamma Correction. In *International Symposium on Computational Intelligence and Design*, Vol. 1. IEEE, 60–63.

[7] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. 2020. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 1780–1789.

[8] Jie Huang, Yajing Liu, Feng Zhao, Keyu Yan, Jinghao Zhang, Yukun Huang, Man Zhou, and Zhiwei Xiong. 2022. Deep Fourier-Based Exposure Correction Network with Spatial-Frequency Interaction. In *European Conference on Computer Vision.* Springer, 163–180.

[9] Jie Huang, Feng Zhao, Man Zhou, Jie Xiao, Naishan Zheng, Kaiwen Zheng, and Zhiwei Xiong. 2023. Learning Sample Relationship for Exposure Correction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 9904–9913.

[10] Jie Huang, Man Zhou, Yajing Liu, Mingde Yao, Feng Zhao, and Zhiwei Xiong. 2022. Exposure-Consistency Representation Learning for Exposure Correction. In *ACM International Conference on Multimedia.* 6309–6317.

[11] Shih-Chia Huang, Fan-Chieh Cheng, and Yi-Sheng Chiu. 2012. Efficient Contrast Enhancement Using Adaptive Gamma Correction with Weighting Distribution. *IEEE Transactions on Image Processing* 22, 3 (2012), 1032–1041.

[12] Hai Jiang, Ao Luo, Xiaohong Liu, Songchen Han, and Shuaicheng Liu. 2024. LightenDiffusion: Unsupervised Low-Light Image Enhancement with Latent-Retinex Diffusion Models. *arXiv preprint arXiv:2407.08939* (2024).

[13] Chongyi Li, Chunle Guo, Ruicheng Feng, Shangchen Zhou, and Chen Change Loy. 2022. CUDi: Curve Distillation for Efficient and Controllable Exposure Adjustment. *arXiv preprint arXiv:2207.14273* (2022).

[14] Chongyi Li, Chunle Guo, and Chen Change Loy. 2021. Learning to Enhance Low-Light Image via Zero-Reference Deep Curve Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 8 (2021), 4225–4238.

[15] Chongyi Li, Chun-Le Guo, Man Zhou, Zhexin Liang, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. 2023. Embedding Fourier for Ultra-High-Definition Low-Light Image Enhancement. *arXiv preprint arXiv:2302.11831* (2023).

[16] Yiyu Li, Ke Xu, Gerhard Petrus Hancke, and Rynson WH Lau. 2024. Color Shift Estimation-and-Correction for Image Enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 25389–25398.

[17] Ziwen Li, Feng Zhang, Meng Cao, Jinpu Zhang, Yuanjie Shao, Yuehuan Wang, and Nong Sang. 2024. Real-Time Exposure Correction via Collaborative Transformations and Adaptive Sampling. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2984–2994.

[18] Jin Liu, Huiyuan Fu, Chuanming Wang, and Huadong Ma. 2024. Region-Aware Exposure Consistency Network for Mixed Exposure Correction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 3648–3656.

[19] Jin Liu, Bo Wang, Chuanming Wang, Huiyuan Fu, and Huadong Ma. 2024. Learning Exposure Correction in Dynamic Scenes. In *ACM International Conference on Multimedia.* 3858–3866.

[20] Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. 2021. Retinex-Inspired Unrolling with Cooperative Prior Architecture Search for Low-Light Image Enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 10561–10570.

[21] Xiaoqian Lv, Shengping Zhang, Chenyang Wang, Yichen Zheng, Bineng Zhong, Chongyi Li, and Liqiang Nie. 2024. Fourier Priors-Guided Diffusion for Zero-Shot Joint Low-Light Enhancement and Deblurring. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 25378–25388.

[22] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. 2022. Toward Fast, Flexible, and Robust Low-Light Image Enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 5637–5646.

[23] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. 2012. Making a "Completely Blind" Image Quality Analyzer. *IEEE Signal Processing Letters* 20, 3 (2012), 209–212.

[24] Namuk Park and Songkuk Kim. 2022. How Do Vision Transformers Work? *arXiv preprint arXiv:2202.06709* (2022).

[25] Stephen M Pizer, E Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderveld. 1987. Adaptive Histogram Equalization and Its Variations. *Computer Vision, Graphics, and Image Processing* 39, 3 (1987), 355–368.

[26] Yongming Rao, Wenliang Zhao, Zheng Zhu, Jiwen Lu, and Jie Zhou. 2021. Global Filter Networks for Image Classification. *Advances in Neural Information Processing Systems* 34 (2021), 980–993.

[27] Ali M Reza. 2004. Realization of the Contrast Limited Adaptive Histogram Equalization (CLAHE) for Real-Time Image Enhancement. *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology* 38 (2004), 35–44.

[28] Adin Ramirez Rivera, Byungyong Ryu, and Oksam Chae. 2012. Content-Aware Dark Image Enhancement Through Channel Division. *IEEE Transactions on Image Processing* 21, 9 (2012), 3967–3980.

[29] Haoyuan Wang, Ke Xu, and Rynson WH Lau. 2022. Local Color Distributions Prior for Image Enhancement. In *European Conference on Computer Vision.* Springer, 343–359.

[30] Wenjing Wang, Huan Yang, Jianlong Fu, and Jiaying Liu. 2024. Zero-Reference Low-Light Enhancement via Physical Quadruple Priors. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 26057–26066.

[31] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. 2018. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In *European Conference on Computer Vision Workshops.* 0–0.

[32] Yufei Wang, Yi Yu, Wenhan Yang, Lanqing Guo, Lap-Pui Chau, Alex C Kot, and Bihan Wen. 2023. ExposureDiffusion: Learning to Expose for Low-Light Image Enhancement. In *IEEE/CVF International Conference on Computer Vision.* 12438–12448.

[33] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. 2018. Deep Retinex Decomposition for Low-Light Enhancement. *arXiv preprint arXiv:1808.04560* (2018).

[34] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. 2022. Uretinex-Net: Retinex-Based Deep Unfolding Network for Low-Light Image Enhancement. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 5901–5910.

[35] Xiaogang Xu, Ruixing Wang, and Jiangbo Lu. 2023. Low-Light Image Enhancement via Structure Modeling and Guidance. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 9893–9903.

[36] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. 2021. Sparse Gradient Regularized Deep Retinex Network for Robust Low-Light Image Enhancement. *IEEE Transactions on Image Processing* 30 (2021), 2072–2086.

[37] Lu Yuan and Jian Sun. 2012. Automatic Exposure Correction of Consumer Photographs. In *European Conference on Computer Vision.* Springer, 771–785.

[38] Tongshun Zhang, Pingping Liu, Ming Zhao, and Haotian Lv. 2024. DMFourLLIE: Dual-Stage and Multi-Branch Fourier Network for Low-Light Image Enhancement. In *ACM International Conference on Multimedia.*