ShadowDiffusion: When Degradation Prior Meets Diffusion Model for Shadow Removal

Lanqing Guo[1], Chong Wang[1], Wenhan Yang[2], Siyu Huang[3], Yufei Wang[1], Hanspeter Pfister[3], Bihan Wen[1]*
[1]Nanyang Technological University, Singapore
[2]Peng Cheng Laboratory, China [3]Harvard University, USA
{lanqing001, wang1711, yufei001, bihan.wen}@ntu.edu.sg,
yangwh@pcl.ac.cn, huang@seas.harvard.edu, pfister@g.harvard.edu

Abstract

Recent deep learning methods have achieved promising results in image shadow removal. However, their restored images still suffer from unsatisfactory boundary artifacts, due to the lack of degradation prior embedding and the deficiency in modeling capacity. Our work addresses these issues by proposing a unified diffusion framework that integrates both the image and degradation priors for highly effective shadow removal. In detail, we first propose a shadow degradation model, which inspires us to build a novel unrolling diffusion model, dubbed ShandowDiffusion. It remarkably improves the model's capacity in shadow removal via progressively refining the desired output with both degradation prior and diffusive generative prior, which by nature can serve as a new strong baseline for image restoration. Furthermore, ShadowDiffusion progressively refines the estimated shadow mask as an auxiliary task of the diffusion generator, which leads to more accurate and robust shadow-free image generation. We conduct extensive experiments on three popular public datasets, including ISTD, ISTD+, and SRD, to validate our method's effectiveness. Compared to the state-of-the-art methods, our model achieves a significant improvement in terms of PSNR, increasing from 31.69dB to 34.73dB over SRD dataset.[1]

1. Introduction

Shadow removal aims to enhance visibility of the image shadow regions, pursuing a consistent illumination distribution between shadow and non-shadow regions. Deep learning-based methods [4, 7, 10, 20, 54] achieved superior performance recently by fully utilizing the power of large collections of data. While most of the existing methods focused on learning the discriminative models for shadow removal, modeling the underlying distribution of nature images is overlooked in their restoration process. Consequently, the shadow removal results usually contain severe boundary artifacts and remaining shadow patterns, as shown in Figure 1(b).

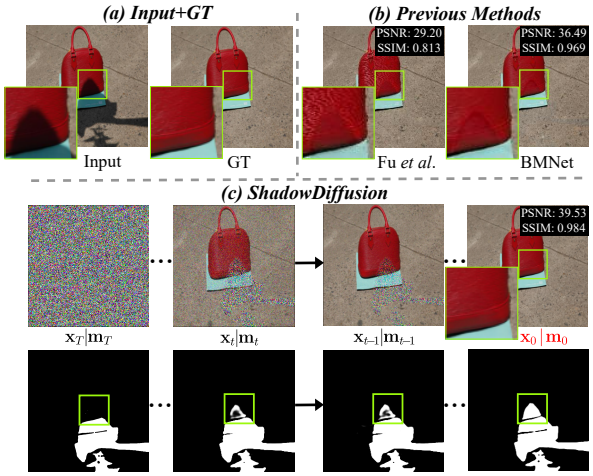

Figure 1. (a) Input shadow image and corresponding ground truth shadow-free image, (b) shadow removal results of two most recent competing methods Fu *et al.* [7] and BMNet [54], and (c) our proposed ShadowDiffusion iteratively ($T \rightarrow 0$) restores the shadow-free image and refines the shadow mask, in which the $\mathbf{x}_0$ and $\mathbf{m}_0$ are the final enhanced result and refined mask, respectively.

Though the adversarial loss can alleviate this issue, these approaches [19, 38] require careful adjustment during training, might overfit certain visual features or data distribution, and might hallucinate new content and artifacts. Very recently, various diffusion models, such as the popular diffusion denoising diffusion probability model (DDPM) [17], have gained wide interest in the field of low-level vision [33, 34]. Comparing to other deep generative mod-

*Corresponding author: Bihan Wen.
This work was carried out at ROSE Lab, supported in part by the MOE AcRF Tier 1 (RG61/22) and Start-Up Grant.

[1]https://github.com/GuoLanqing/ShadowDiffusion

els, diffusion models are more powerful for modeling image pixel distribution, which provides great potential for significantly improving visual quality and benefits high-quality image restoration. However, no work to-date has exploited diffusion models for shadow removal tasks.

Moreover, there are two major limitations in existing shadow removal methods: First, the shadow degradation prior that reflects its corresponding physical properties has not been well exploited in deep learning. Though recent work [25] attempted to incorporate simple shadow model as a linear and uniform degradation, such an assumption is too restrictive for restoring real shadow images subjective to complicated lighting conditions. Second, most of the deep shadow removal methods requires an estimated shadow mask as the inputs, which are either provided by the benchmark datasets [38] or generated by a pre-trained shadow detector [5]. However, these mask estimates are usually inaccurate, *e.g.*, wrong indicators near the boundary or small shadow objects. Even the carefully hand-crafted masks sometimes contain coarse boundaries. Since existing methods blindly rely on the estimated masks without exploiting their correlation to the actual shadow images for refinement, there are usually severe boundary artifacts in their shadow removal results [7, 54], as shown in Figure 1(b).

To alleviate the challenges in shadow removal, we first introduce a general shadow model of spatially-variant degradation, by decomposing the degradation matrix into the shadow mask and shadow intensities. Based on the new shadow model, we propose a novel unrolling diffusion-based shadow removal framework, called ShadowDiffusion, which integrates both the generative and degradation priors. Specifically, we formulate the shadow removal problem as to *jointly* pursue the shadow-free image and refined shadow mask. Mask refinement is designed as an auxiliary task of the diffusion generator to progressively refine the shadow mask along with shadow-free image restoration in an interactive manner as shown in Figure 1(c). After that, we further propose an unrolling-inspired diffusive sampling strategy to explicitly integrate the degradation prior into the diffusion framework. Experimental results show that ShadowDiffusion can achieve superior performance consistently over the three widely-used shadow removal datasets and significantly outperform the state-of-the-art methods. Besides, our model can be applied to other image enhancement tasks, *e.g.*, low-light image enhancement and exposure correction. Our main contributions are summarized as follows:

- We propose the first diffusion-based model for shadow removal. A novel dynamic mask-aware diffusion model (DMDM) is introduced to jointly pursue a shadow-free image and refined shadow mask, which leads to robust shadow removal even with an inaccurate mask estimate.
- We further propose an unrolling-inspired diffusive sampling strategy to explicitly integrate the shadow degradation prior into the intrinsic iterative process of DMDM.
- Extensive experimental results on the public ISTD, ISTD+, and SRD datasets show that the proposed ShadowDiffusion outperforms the state-of-the-art shadow removal methods by large margins. Besides, our method can be generalized to a series of image enhancement tasks.

2. Related Work

Shadow removal. Different from the global corrupted problems, *e.g.*, low-light enhancement [12, 13, 22, 40, 51], deraining [45, 48], denoising [11, 14, 56], and compression [28,41,47], the shadow is a partially corrupted problem with abundant non-corrupted information. Deep learning-based methods [6, 19, 25] achieved superior performance in recent years relying on large-scale training data. One group of works still reconstructed the shadow-free image under a physical illumination model with global degradation. For instance, Le *et al.* [25] ideally applied the physical linear transformation model to enhance the shadow region and reconstructed the shadow-free image by image decomposition. Fu *et al.* [7] proposed an over-exposure fusion way for shadow removal, where the proposed model can smartly blend a series of over-enhanced shadow images as well as the original shadow image by a learnable pixel-wise weighting map. However, such a global degradation model is too strict since real-world illumination degradation is always non-uniform. Besides, generative adversarial network techniques [19, 29, 38] are applied to enhance the reality of enhanced results. However, the results of these methods always suffered from color distortions and might hallucinate new content and artifacts.

Diffusion model for image restoration. Diffusion-based generative models [36] recently produced amazing results with improvements adopted in denoising diffusion probabilistic models [17], which becomes increasingly influential in the field of low-level vision tasks, such as super-resolution [23, 34], inpainting [30], and colorization [33]. Saharia *et al.* [34] introduced a denoising diffusion probabilistic model to image super-resolution and achieved better performance compared with the state-of-the-art Generative Adversarial Network (GAN) [8] based methods. Ozan *et al.* [31] presented a patch-based diffusion model for weather removal that enables the size-agnostic processing, which employed a guided denoising process across overlapping patches during inference.

However, most of these methods focus on synthetic degradation, such as image colorization, image inpainting, and super-resolution, in which it is very easy to simulate

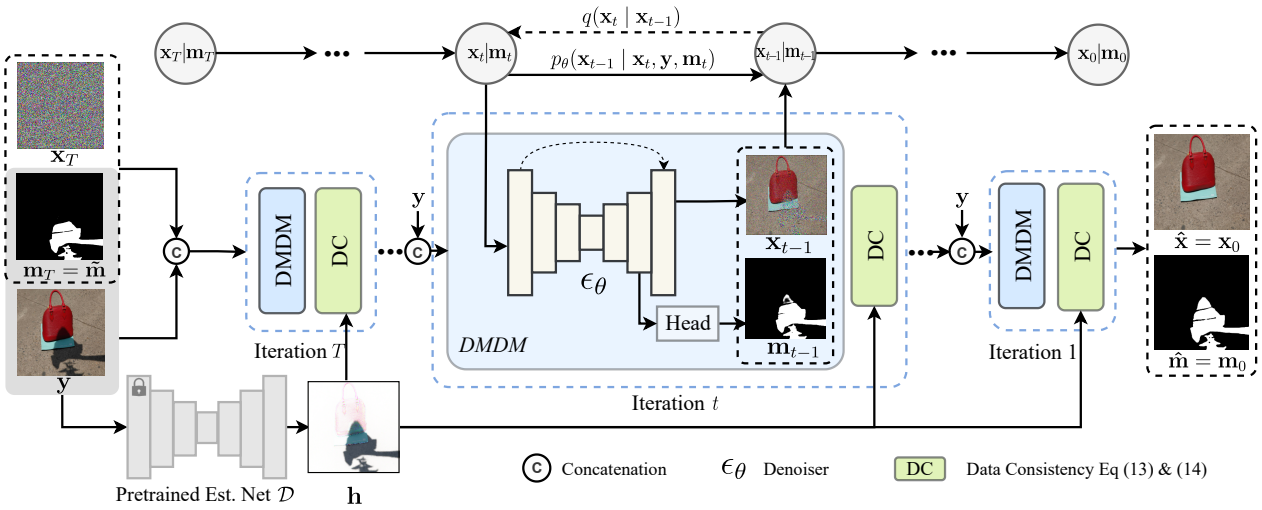

Figure 2. Illustration of the proposed ShadowDiffusion, in which the training (dashed line) and sampling (solid line) processes are detailed in Algorithm 1 and Algorithm 2, respectively. Each sampling iteration consists of the sampling of dynamic mask-aware diffusion model (DMDM) and data consistency (DC) steps (Here DC corresponds to Eq (13) & (14)).

large-scale training pairs on existing natural image datasets to train diffusion-based models. In this paper, we explore the real-world shadow removal problem with limited training pairs. A novel dynamic mask-aware diffusion model with the stricter and iteratively refined conditions is proposed to address the above problem.

Deep unrolling methods. Consistency of the predictions with respect to the degradation model is crucial for reliably solving ill-posed restoration tasks. Deep unrolling, by incorporating the known degradation model into the deep networks via an iterative optimization algorithm, has demonstrated remarkable performance on various inverse problems. For example, Karol *et al.* [9] proposed to unroll the iterative shrinkage thresholding algorithm (ISTA) for sparse coding, which demonstrated promising results on super-resolution [27]. Yang *et al.* [46] introduced an unrolling network describing the data flow graphs in the iterative procedures of Alternating Direction Method of Multipliers (ADMM) for magnetic resonance imaging (MRI) reconstruction. Based on half-quadratic splitting, Zhang *et al.* [50] proposed an unfolding scheme that enables a single network to address different scale factors in the super-resolution task. Compared with model-free learning based methods, the deep unrolling scheme integrates the degradation constraint into the learning model by iteratively regularizing the network output according to the model prior.

3. ShadowDiffusion

We present the proposed ShadowDiffusion, by first introducing our shadow degradation model. Then, the dynamic mask-aware diffusion model (DMDM) and its training process are presented, which can predict the shadow-free image jointly with progressive mask refinement. Finally, we introduce the unrolling-inspired diffusive sampling based on the DMDM which integrates the diffusive generative model and shadow degradation prior.

3.1. Shadow Degradation Model

A shadow region of an image $\mathbf{y}$ is caused by partial or complete occlusions. Inspired by Retinex theory [24], classic methods adopted a simple shadow degradation, in which shadow images $\mathbf{y}$ is formed by applying an illumination change surface $\mathbf{a}$ to the shadow-free image $\mathbf{x}$ as follows:

$$\mathbf{y} = \mathbf{a} \cdot \mathbf{x} \,, \tag{1}$$

where $\cdot$ denotes the element-wise multiplication. Here $\mathbf{a}$ is strictly assumed to be 1 in the lit area, and a constant $a \in (0, 1)$ in the umbra area. (1) is usually too restrictive as the natural lighting are mostly non-uniform in practice.

In general, illumination degradation should be spatially-variant and highly dependent on shadow mask information. Thus, we propose a new shadow degradation model as

$$\mathbf{y} = \mathbf{h} \cdot \mathbf{x} = \mathbf{w} \cdot \mathbf{m} \cdot \mathbf{x} + (\mathbf{1} - \mathbf{m}) \cdot \mathbf{x} \,. \tag{2}$$

Here, $\mathbf{h}$ denotes the pixel-wise illumination degradation map, which can be decomposed into the shadow mask $\mathbf{m}$ and illumination weight $\mathbf{w}$. The shadow mask $\mathbf{m}$ indicates the shadow locations that shadow regions are 1 and the rest are 0. Our Model (2) owns the following advantages:

- The shadow image $\mathbf{y}$ can be modeled as a non-uniform illumination transformation on shadow-free image $\mathbf{x}$ under the pixel-wise degradation map $\mathbf{h}$. This degradation prior provides richer information than the uniform degradation for shadow removal.

- The shadow mask $\mathbf{m}$ provides shadow location information, which have a direct and critical effect on the shadow degradation $\mathbf{h}$, which therefore significantly affect the estimation of the desired shadow-free image $\hat{\mathbf{x}}$. As a wrong shadow mask leads to an inaccurate degradation map, which is quite common in practice, our degradation model flexibly enables to embed the mask refinement as an auxiliary task, and make the refinement and shadow-free image restoration mutually beneficial.

3.2. Dynamic Mask-Aware Diffusion Model

The shadow mask is crucial for shadow removal. It indicates the exact location of shadow regions according to (2). In other words, inaccurate shadow mask inputs will directly affect the shadow removal outcomes. Thus, we remodel the shadow removal as a joint task to pursue a shadow-free image and refined mask, in which mask refinement would be an auxiliary task of the diffusion generator to progressively refine the shadow mask along with shadow-free image generation. Different from previous conditional diffusion-based image restoration works [33, 34] generating the underlying image with an invariable condition, we propose a dynamic mask-aware diffusion model (DMDM) to progressively generate the shadow-free image and refined conditions (masks).

We first revisit the previous conditional diffusion model [34], which learns a conditional reverse process $p_\theta(\mathbf{x}_{0:T}|\mathbf{y})$ without modifying the diffusion process $q(\mathbf{x}_{1:T}|\mathbf{x}_0)$ for $\mathbf{x}$, such that the sampled image has high fidelity to the data distribution conditioned on $\mathbf{y}$. During training, we sample $(\mathbf{x}_0, \mathbf{y}, \tilde{\mathbf{m}}) \sim q(\mathbf{x}, \mathbf{y}, \tilde{\mathbf{m}})$ from a triplet data distribution (*e.g.*, a shadow-free image $\mathbf{x}$, shadow image $\mathbf{y}$, and corresponding initial shadow mask $\tilde{\mathbf{m}}$). Our training approach is outlined in Algorithm 1, in which we learn the dynamic mask-aware reverse process:

$$p_\theta\left(\mathbf{x}_{0:T}\,|\,\mathbf{y}, \mathbf{m}_{0:T}\right) = p\left(\mathbf{x}_T\right) \prod_{t=1}^{T} p_\theta\left(\mathbf{x}_{t-1}\,|\,\mathbf{x}_t, \mathbf{y}, \mathbf{m}_t\right) \ . \tag{3}$$

We can marginalize the Gaussian diffusion process to sample intermediate $\mathbf{x}_t$ terms directly from shadow-free image $\mathbf{x}_0$ through $\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}$, where β_t is the noise schedule, $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$, and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ has the same dimensionality as $\mathbf{x}_0$. The denoiser $\boldsymbol{\epsilon}_\theta$ takes the shadow image $\mathbf{y}$, the intermediate variable $\mathbf{x}_t$, and the time step t as input to predict the noise map $\mathbf{e}_t$ and the refined mask $\mathbf{m}_t$ as follows:

$$\mathbf{e}_t, \mathbf{m}_t = \boldsymbol{\epsilon}_\theta\left(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, \mathbf{y}, \tilde{\mathbf{m}}, t\right) \ . \tag{4}$$

As the shadow mask information is highly dependent on the shadow-free image generation, we build a model to perform the shadow-free image prediction and mask refinement jointly. We add a mask prediction head after the last layer of $\boldsymbol{\epsilon}_\theta$, with one 1×1 convolution layer and one Sigmoid function to predict the refined mask. Following [17], the diffusive objective function is

$$\mathcal{L}_{diff} = \mathbb{E}_{\mathbf{x}_0, t, \boldsymbol{\epsilon}} \left\| \mathbf{e}_t - \boldsymbol{\epsilon} \right\|_F^2 . \tag{5}$$

Besides, according to the shadow and shadow-free image pairs in the training stage, we can adopt a ground truth shadow mask as a reference to constrain the rationality of the refined mask

$$\mathcal{L}_{mask} = \mathbb{E}_{t\sim[1,T]} \left\| \mathbf{m}_t - \mathbf{m}_{gt} \right\|_F^2 , \tag{6}$$

where the ground truth shadow mask $\mathbf{m}_{gt}$ can be obtained by binarizing the residual map between shadow and shadow-free images $\mathbf{m}_{gt} = \begin{cases} 1 & \mathbf{x} - \mathbf{y} > 0.1 \\ 0 & \text{otherwise} \end{cases}$.

The hybrid objective function $\mathcal{L}_{total}$ is obtained by combining the above losses, which guides the training of the denoiser $\boldsymbol{\epsilon}_\theta$ in our DMDM as follows,

$$\mathcal{L}_{total} = \mathcal{L}_{diff} + \lambda \mathcal{L}_{mask} \ , \tag{7}$$

where λ are the weighting coefficient to balance the influence of each term.

Algorithm 1 Dynamic mask-aware diffusion training.

Input: shadow image $\mathbf{y}$, shadow-free image $\mathbf{x}$, and initial mask $\tilde{\mathbf{m}}$.

while not converged **do**
  $t \sim \text{Uniform}\{1, \ldots, T\}$
  $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  $\mathbf{e}_t, \mathbf{m}_t = \boldsymbol{\epsilon}_\theta\left(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, \mathbf{y}, \tilde{\mathbf{m}}, t\right)$
  Perform Gradient descent steps on $\nabla_\theta\, \mathcal{L}_{total}(\theta)$
end while
return θ

Algorithm 2 Unrolling-inspired diffusive sampling.

Input: shadow image $\mathbf{y}$, initial mask $\tilde{\mathbf{m}}$, diffusion model $\boldsymbol{\epsilon}_\theta$, number of implicit sampling iterations T, $\mathbf{z}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\mathbf{v}_T = \tilde{\mathbf{m}}$, and initial parameters ψ, ϕ, and ρ.

for $t = T, \ldots, 1$ **do**
  $\mathbf{e}_{t-1}, \mathbf{m}_{t-1} = \boldsymbol{\epsilon}_\theta\left(\mathbf{z}_t, \mathbf{y}, \mathbf{v}_t, t\right)$
  $\mathbf{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\left(\frac{\mathbf{z}_t - \sqrt{1-\bar{\alpha}_t}\cdot\mathbf{e}_{t-1}}{\sqrt{\bar{\alpha}_t}}\right) + \sqrt{1-\bar{\alpha}_{t-1}}\cdot\mathbf{e}_{t-1}$
  update $\mathbf{z}_{t-1}$ with Eq (13).
  update $\mathbf{v}_{t-1}$ with Eq (14).
end for
return $\mathbf{x}_t$, $\mathbf{m}_t$

3.3. Unrolling-Inspired Diffusive Sampling

Based on the shadow degradation model (2), we formulate the shadow removal as a degradation prior guided

model, where regularization terms are inferred by a learnable conditional generative diffusion model instead of using hand-crafted priors under Maximum A Posteriori (MAP) framework. By considering the provided initial mask $\tilde{\mathbf{m}}$ may be coarse or inaccurate, the shadow mask $\mathbf{m}$ would be iteratively refined along with desired shadow-free image $\mathbf{x}$ optimization, which can be obtained by minimizing the following energy function with joint image-mask regularizer:

$$\min_{\mathbf{x},\mathbf{m},\mathbf{z},\mathbf{v}} \frac{1}{2}\|\mathbf{h}\cdot\mathbf{z}-\mathbf{y}\|_F^2+\psi\mathcal{R}([\mathbf{x}|\mathbf{m}])+\frac{\phi}{2}\|\mathbf{v}-\tilde{\mathbf{m}}\|_F^2 \quad \text{s.t. } \mathbf{x}=\mathbf{z},\ \mathbf{m}=\mathbf{v}\ , \tag{8}$$

where $\mathbf{z}$ and $\mathbf{v}$ are the auxiliary variables that convert (8) into a constrained problem and $\mathcal{R}(\cdot)$ is the regularizer capturing assumed joint image and mask priors. $[\cdot|\cdot]$ denotes the concatenation operation. ψ and ϕ are trade-off parameters. Here we assume that the degradation matrix $\mathbf{h}$ can be estimated by a pre-trained degradation estimation network as $\mathbf{h}=\mathcal{D}(\mathbf{y},\tilde{\mathbf{m}})$. To deal with the equality constraints, two quadratic penalty terms are introduced, and the problem is rewritten as follows:

$$\min_{\mathbf{x},\mathbf{m},\mathbf{z},\mathbf{v}} \frac{1}{2}\|\mathbf{h}\cdot\mathbf{z}-\mathbf{y}\|_F^2+\psi\mathcal{R}([\mathbf{x}|\mathbf{m}])+\frac{\phi}{2}\|\mathbf{v}-\tilde{\mathbf{m}}\|_F^2 + \frac{\rho_1}{2}\|\mathbf{x}-\mathbf{z}\|_F^2+\frac{\rho_2}{2}\|\mathbf{m}-\mathbf{v}\|_F^2\ , \tag{9}$$

where ρ_1 and ρ_2 are penalty parameters (we set the $\rho=\rho_1=\rho_2$ for simpler solutions). By employing variable splitting algorithms such as half-quadratic splitting (HQS) [2], the optimization problem (9) can be addressed by iteratively solving three sub-problems[2] as follows:

$$[\mathbf{x}|\mathbf{m}]_{t-1}=\arg\min_{\mathbf{x},\mathbf{m}} \psi\mathcal{R}([\mathbf{x}|\mathbf{m}])+\frac{\rho}{2}\|[\mathbf{x}|\mathbf{m}]-[\mathbf{z}_t|\mathbf{v}_t]\|_F^2\ , \tag{10}$$

$$\mathbf{z}_{t-1}=\arg\min_{\mathbf{z}} \frac{1}{2}\|\mathbf{h}\cdot\mathbf{z}-\mathbf{y}\|_F^2+\frac{\rho}{2}\|\mathbf{x}_{t-1}-\mathbf{z}\|_F^2\ , \tag{11}$$

$$\mathbf{v}_{t-1}=\arg\min_{\mathbf{v}} \frac{\phi}{2}\|\mathbf{v}-\tilde{\mathbf{m}}\|_F^2+\frac{\rho}{2}\|\mathbf{m}_{t-1}-\mathbf{v}\|_F^2\ . \tag{12}$$

Here (11) and (12) are least-squares problems with quadratic penalty terms, which have closed-form solutions

$$\mathbf{z}_{t-1}=(\mathbf{h}\cdot\mathbf{y}+\rho\mathbf{x}_{t-1})/(\mathbf{h}\cdot\mathbf{h}+\rho\mathbf{1})\ , \tag{13}$$

$$\mathbf{v}_{t-1}=(\phi\tilde{\mathbf{m}}+\rho\mathbf{m}_{t-1})/(\phi+\rho)\ . \tag{14}$$

Note that, (13) and (14) commonly refer to the data consistency (DC) steps [35] by sharing the information between the input and reconstructed variable. The update process of $\mathbf{x}$ and $\mathbf{m}$ can be solved by the sampling process of DMDM, denoted as $\mathcal{G}_\theta(\cdot)$ (details refer to Section 3.2), yielding the iterates

$$[\mathbf{x}|\mathbf{m}]_{t-1}=\mathcal{G}_\theta(\mathbf{z}_t,\mathbf{y},\mathbf{v}_t,t)\ . \tag{15}$$

Algorithm 2 summarizes the whole process of diffusion-based unrolling, where the $\mathcal{G}_\theta(\cdot)$ corresponds to Lines 2-3. The sampling of $\mathcal{G}_\theta$ follows the diffusive sampling strategy of DDIM [31, 37] to accelerate the inference stage.

It is noted that, in contrast to repeatly forwarding the single-stage model, diffusion model is a natural architecture to solve the unrolling optimization problem via the progressive generation process. ***Our framework can incorporate the degradation priors into the diffusion model with almost no additional inference time.*** The shadow-free image is slowly restored based on the diffusion model, while the extra degradation prior can largely accelerate the shadow-free image generation and make the iterations close to the truly shadow-free data manifold. The effectiveness of our design is manifested in Section 4.3.

4. Experiments

4.1. Experimental Setups

Implementation details. The proposed method is implemented using PyTorch, which is trained using one NVIDIA RTX A5000 GPU. The training epoch is set as 1000. We use Adam optimizer with the momentum as $(0.9, 0.999)$. The initial learning rate is 3×10^{-5}. Following [34], we use the Kaiming initialization technique [16] to initialize the weights of the proposed model and use 0.9999 Exponential Moving Average (EMA) for all our experiments. We followed the similar U-Net architecture as denoiser $\boldsymbol{\epsilon}_\theta$ of [34]. We used 1000 diffusion steps T and noise schedule β_t linearly increasing from 0.0001 to 0.02 for training, and 25 steps for inference. We select the most recent transformer-based image-to-image backbone [43] as the degradation estimation network $\mathcal{D}$, in which we pre-train $\mathcal{D}$ with the concatenation of shadow image and mask as input and regard the $\mathbf{h}_{gt}=\mathbf{y}/(\mathbf{x}+\eta)$, with $\eta=1\times e^{-4}$, as the ground truth degradation map. We set the $\lambda=0.5$ in our experiments. The detailed architectures and hyper-parameter settings can be found in the **supplementary**.

Benchmark datasets. We work with three benchmark datasets for the various shadow removal experiments: (1) ISTD [38] dataset includes 1330 training and 540 testing triplets (shadow images, masks and shadow-free images). (2) Adjusted ISTD (ISTD+) dataset [25] reduces the illumination inconsistency between the shadow and shadow-free image of ISTD. (3) SRD [32] dataset consists of 2680 training and 408 testing pairs of shadow and shadow-free images. We use the predicted masks that are provided by DHAN [5] for training and testing following most previous methods [5, 7, 54, 55].

[2]Note that we follow the $t\rightarrow t-1$ iteration update order to preserve the order consistency between unrolling and the diffusion sampling process.

	Method	Shadow Region (S)			Non-Shadow Region (NS)			All Image (ALL)		
		PSNR↑	SSIM↑	RMSE↓	PSNR↑	SSIM↑	RMSE↓	PSNR↑	SSIM↑	RMSE↓
ISTD	Input Image	22.40	0.936	32.10	27.32	0.976	7.09	20.56	0.893	10.88
	Guo *et al.* [15]	27.76	0.964	18.65	26.44	0.975	7.76	23.08	0.919	9.26
	MaskShadow-GAN [19]	-	-	12.67	-	-	6.68	-	-	7.41
	ST-CGAN [38]	33.74	0.981	9.99	29.51	0.958	6.05	27.44	0.929	6.65
	DSC [18]	34.64	0.984	8.72	31.26	0.969	5.04	29.00	0.944	5.59
	G2R [29]	31.63	0.975	10.72	26.19	0.967	7.55	24.72	0.932	7.85
	DHAN [5]	35.53	0.988	7.49	31.05	0.971	5.30	29.11	0.954	5.66
	Fu *et al.* [7]	34.71	0.975	7.91	28.61	0.880	5.51	27.19	0.945	5.88
	DC-ShadowNet [21]	31.69	0.976	11.43	28.99	0.958	5.81	26.38	0.922	6.57
	Zhu *et al.* [55]	36.95	0.987	8.29	31.54	0.978	4.55	29.85	0.960	5.09
	BMNet [54]	35.61	0.988	7.60	32.80	0.976	4.59	30.28	0.959	5.02
	Ours	**40.15**	**0.994**	**4.13**	**33.70**	**0.977**	**4.14**	**32.33**	**0.969**	**4.12**
SRD	Input Image	18.96	0.871	36.69	31.47	0.975	4.83	18.19	0.830	14.05
	Guo *et al.* [15]	-	-	29.89	-	-	6.47	-	-	12.60
	DeshadowNet [32]	-	-	11.78	-	-	4.84	-	-	6.64
	DSC [18]	30.65	0.960	8.62	31.94	0.965	4.41	27.76	0.903	5.71
	DHAN [5]	33.67	0.978	8.94	34.79	0.979	4.80	30.51	0.949	5.67
	Fu *et al.* [7]	32.26	0.966	9.55	31.87	0.945	5.74	28.40	0.893	6.50
	DC-ShadowNet [21]	34.00	0.975	7.70	35.53	0.981	3.65	31.53	0.955	4.65
	Zhu *et al.* [55]	34.94	0.980	7.44	35.85	0.982	3.74	31.72	0.952	4.79
	BMNet [54]	35.05	0.981	6.61	36.02	0.982	3.61	31.69	0.956	4.46
	Ours	**38.72**	**0.987**	**4.98**	**37.78**	**0.985**	**3.44**	**34.73**	**0.970**	**3.63**

Table 1. The quantitative results of shadow removal using our ShadowDiffusion and recent methods on ISTD [38] and SRD [32] datasets.

Method	Shadow		Non-Shadow		All	
	PSNR↑	RMSE↓	PSNR↑	RMSE↓	PSNR↑	RMSE↓
Input Image	20.83	40.2	37.46	2.6	20.46	8.5
DeshadowNet [32]	-	15.9	-	6.0	-	7.6
ST-CGAN [38]	-	13.4	-	7.7	-	8.7
Param-Net [26]	-	9.7	-	3.0	-	4.0
SP+M-Net [25]	37.59	5.9	36.02	3.0	32.94	3.5
DHAN [5]	32.92	11.2	27.15	7.1	25.66	7.8
Fu *et al.* [7]	36.04	6.6	31.16	3.8	29.45	4.2
BMNet [54]	-	5.6	-	2.5	-	3.0
Ours	**39.82**	**4.9**	**38.90**	**2.3**	**35.72**	**2.7**

Table 2. The quantitative results of shadow removal using our ShadowDiffusion and recent methods on ISTD+ [25] dataset.

Evaluation measures. Following the previous works [5, 7, 15, 25, 32, 38], we utilize the root mean square error (RMSE) in the LAB color space as the **quantitative** evaluation metric of the shadow removal results, comparing to the ground truth shadow-free images. Besides, we also adopt the Peak Signal-to-Noise Ratio (PSNR) and the structural similarity (SSIM) [42] to measure the performance of various methods in the RGB color space. For the PSNR and SSIM metrics, higher values represent better results.

4.2. Comparison with State-of-the-Art

We compare the proposed method with the popular or state-of-the-art shadow removal algorithms, including one traditional method, *i.e.*, Guo *et al.* [15], and several deep learning-based methods, *i.e.*, MaskShadow-GAN [19], DeshadowNet [32], ST-CGAN [38], DSC [18], G2R [29], DHAN [5], Param-Net [26], SP+M-Net [25], Fu *et al.* [7], DC-ShadowNet [21], Zhu *et al.* [55], and BMNet [54]. All of the shadow removal results by the competing methods are quoted from the original papers or reproduced using their official implementations. Following the previous methods [7, 54, 55] to evaluate the shadow removal performance, we evaluate the shadow removal results with a resolution of 256×256.

Quantitative evaluation. Tables 1&2 show the quantitative results on the testing sets over ISTD, SRD, and ISTD+, respectively. It is clear that our methods outperform all competing methods by large margins in the shadow area, non-shadow area and the whole image over all of the three datasets. It significantly improves the PSNR from 31.69dB to 34.73dB over SRD dataset, compared to the most recent method BMNet, especially for the shadow region from 35.05dB to 38.72dB. The shadow scenarios from SRD dataset are much more complicated than ISTD dataset, even including some object-correlated shadows, *e.g.*, the shadows are caused by the trees and building in the second row of Figure 3. Existing methods may ignore the effective degradation and generative priors, failing in such complicated textures. Conversely, with the merits of the joint image-mask modeling, our methods can better deal with object-correlated shadow cases. Besides, by incorporating the useful illumination degradation assumption and learning the desired shadow-free image distribution, our method can handle complicated cases and produce artifact-free results.

Qualitative evaluation. To further demonstrate the advantage of our method against other competing methods, Figures 3 & 4 present the visual examples of the shadow removal results on SRD and ISTD datasets, respectively. More visual examples can be found in the **supplementary**. Note that the images from the SRD dataset have more complicated textures and color distributions. In these samples of the SRD dataset, previous works fail to enhance the illumination of the background and suppress the boundary artifacts in a complicated and colorful region, *e.g.*, the blue poster of the third example in Figure 3. Almost all competing methods cannot preserve the illumination consistencies between shadow and non-shadow regions, which seriously destroys the image structures and patterns as shown in the wall of the first and third examples in Figure 3. Some small shadow regions, *e.g.*, the shadow caused by the street lamp of the second example of Figure 3, are easily ignored by ex-

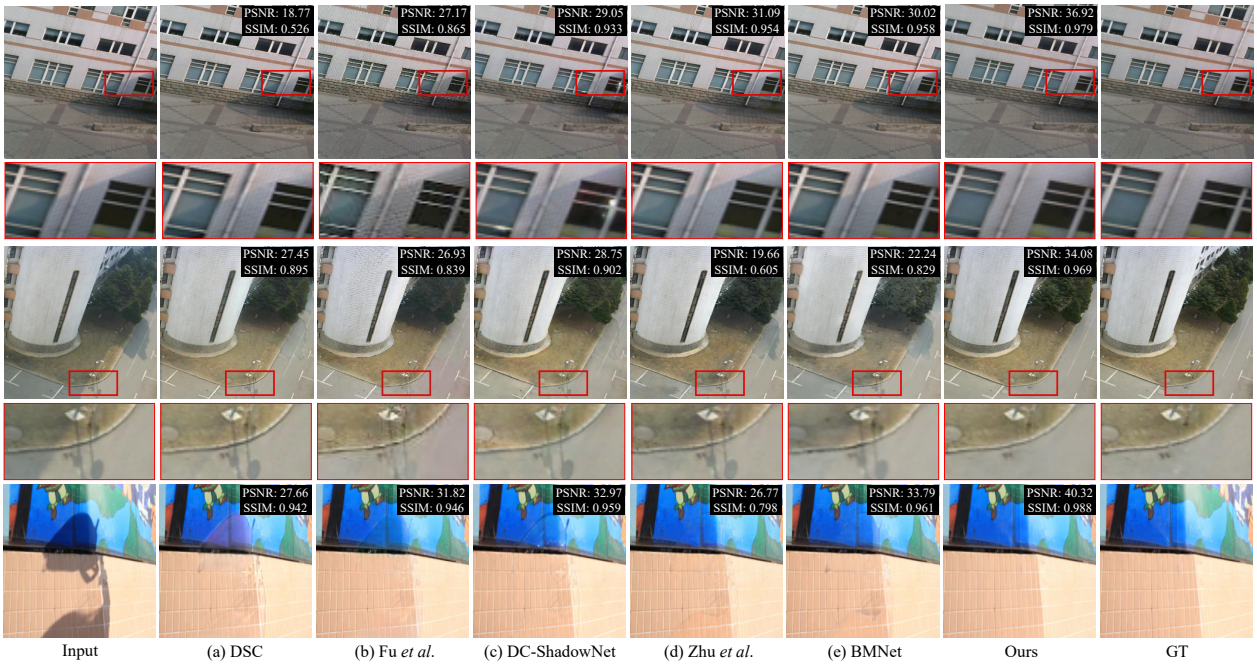

Figure 3. Examples of shadow removal results on the SRD [32] dataset. The input shadow image, the estimated results of (a) DSC [18], (b) Fu *et al.* [7], (c) DC-ShadowNet [21], (d) Zhu *et al.* [55], (e) BMNet [54], and Ours, as well as the ground truth, respectively. Please **zoom in** to see the details.

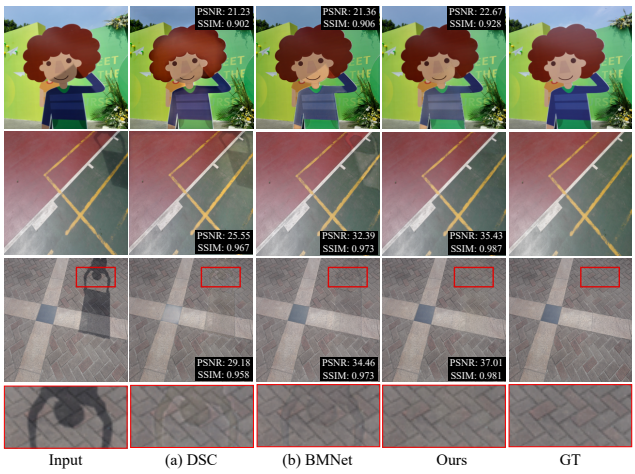

Figure 4. Examples of shadow removal results on the ISTD [38] dataset. The input shadow image, the estimated results of (a) DSC [18], (b) BMNet [54], and Ours, as well as the ground truth, respectively. Please **zoom in** to see the details.

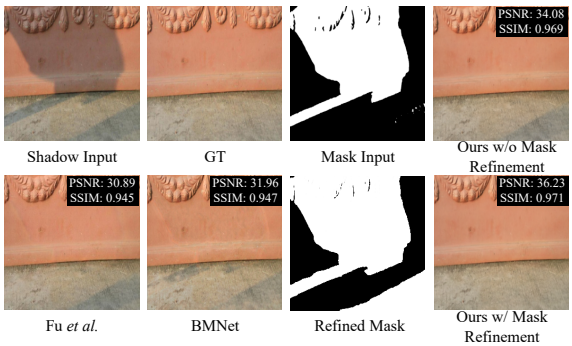

Figure 5. Visual examples of the results of our model w/ and w/o mask refinement, as well as two recent competing methods, *i.e.*, Fu *et al.* [7] and BMNet [54].

isting methods. Instead, thanks to the auxiliary of the mask refinement, the shadow mask of fine areas can be more accurate, achieving better results for different sizes of shadows. On the other hand, the image from ISTD dataset has high context similarity and the scene is relatively simple, while the shadow residual would be more visible. In these samples of the ISTD dataset, previous works usually produce illumination inconsistencies and wrongly-enhanced shadow boundaries. The DSC [18] would wrongly enlighten some regions with insufficient lightness, *e.g.*, the black floor tile in the third example in Figure 4, leading to many ghosts. Moreover, due to the limited dataset, methods without effective prior always have poor scene understanding and generalizability. The performance would largely degrade when generalizing to unusual cases, *e.g.*, the first example in Figure 4, where images with stronger contrast and face patterns are scarce in the training set. However, with the merits of employed shadow degradation and generative priors, it is clear that our methods can successfully produce natural shadow-free images without boundary artifacts and shadow patterns.

Method	Detector1 [5]		Detector2 [53]	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
w/o mask refine	34.55	0.968	33.61	0.943
w/ mask refine	**34.73**	**0.970**	**34.27**	**0.952**

Table 3. Quantitative comparison between results produced by w/ and w/o mask refinement on SRD [32] dataset using the detected shadow mask from different shadow detectors as initial mask $\tilde{m}$.

Method	ISTD		ISTD+		SRD	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Ours w/o unrolling	32.12	0.964	35.21	0.967	34.36	**0.970**
Ours (Complete model)	**32.33**	**0.969**	**35.72**	**0.969**	**34.73**	**0.970**

Table 4. Quantitative comparison between the results produced by w/o unrolling, and the complete model over ISTD [38], ISTD+ [25], and SRD [32] datasets.

Method	Shadow		Non-Shadow		All	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
SR3 [34]	35.44	0.980	34.35	0.970	31.29	0.946
WeatherDiffusion [31]	33.38	0.981	31.15	0.972	28.45	0.951
DMDM Only	38.39	**0.987**	37.21	0.982	34.36	**0.970**
Ours (Complete model)	**38.72**	**0.987**	**37.78**	**0.985**	**34.73**	**0.970**

Table 5. Quantitative comparisons with different diffusion-based models over SRD dataset [32].

Low-light enhancement				Exposure correction		
Method	PSNR↑	SSIM↑	LPIPS↓	Method	PSNR↑	SSIM↑
KinD++ [52]	21.30	0.82	0.16	Deep UPE [39]	14.25	0.64
URetinex-Net [44]	21.33	0.83	0.12	DPE (HDR) [3]	16.21	0.62
MIRNet [49]	24.14	0.84	0.13	Afifi *et al.* [1]	19.48	0.74
Ours	**27.36**	**0.93**	**0.10**	Ours	**22.33**	**0.84**

Table 6. A comparison of the recent state-of-the-art methods for (left) low-light enhancement and (right) exposure correction.

4.3. Ablation Study

The effect of iterative mask refinement. Previous methods [4, 5, 7, 54] employed the detected shadow masks as additional auxiliary information of the network to provide the shadow locations. However, the detected masks might contain inaccurate regions or coarse boundaries as shown in Figure 5. Such wrong guidance makes almost all mask-guided shadow removal algorithms fail, especially for the SRD dataset [32], since the complicated scenes and diverse shadow shapes. With the merits of the proposed dynamic mask-aware diffusion model (DMDM), the coarse or inaccurate mask can be iteratively corrected along with the shadow-free image generation. Thus, the effect of the wrong mask can be effectively alleviated and our model has better robustness in practical applications. Table 3 shows that the shadow removal results of our model with and without mask refinement under the guidance of masks generated by different shadow detectors [5,53] over SRD dataset. The shadow removal results will be better if the initial mask is more accurate as shown in Table 3, and the decline with a worse mask input will not be too noticeable with the proposed iterative mask refinement.

The effect of unrolling framework. In order to illustrate the effectiveness of our unrolling framework, we first investigate the performance of only our dynamic mask-aware diffusion model (DMDM) without the unrolling optimization as shown in the first row in Table 4. We observe that the performance of the complete model is better than the separate diffusion model over all testing sets. The unrolling optimization can provide reliable guidance in the sampling stage of the diffusion model, producing more accurate exposure results than the model without unrolling.

The effect of different diffusion models. We select two recent diffusion-based image restoration methods, *i.e.*, SR3 [34] and WeatherDiffusion [31] as the competing methods to verify the effectiveness of the proposed dynamic mask-aware diffusion model (DMDM) and the proposed diffusion-based unrolling framework as shown in Table 5. For a fair evaluation, we re-train these two methods and change the original three-channel condition (shadow image) into four-channel (the concatenation of shadow image and mask). Obviously, the existing diffusion model performs much worse for shadow removal than our proposed DMDM. Only the concatenation of shadow image and mask as a condition cannot provide a sufficient prior for shadow-free image generation, especially for real-world shadow removal with limited training pairs, comparing the first and fourth rows in Table 5.

4.4. Extension to Other Image Enhancement Tasks

Our ShadowDiffusion can be easily applied to other image enhancement tasks, *e.g.*, low-light enhancement and exposure correction, whose degradation can be regarded as the special case of proposed degradation Model (2) and the m will be the all 1 matrix. Note that we remove the mask refinement and mask conditions in the framework since other enhancement tasks are globally corrupted without the provided mask. Our revised ShadowDiffusion also achieves new state-of-the-art performances among these two tasks as shown in Table 6. More comparison results can be found in the **supplementary**.

5. Conclusion

In this paper, we propose a spatially-variant shadow degradation model, decomposing the shadow degradation map into the shadow mask and shadow intensity. Inspired by that, we propose an unrolling diffusion framework, dubbed as ShadowDiffusion, to explicitly integrate degradation prior and diffusive generative prior. Moreover, we further consider mask refinement as an auxiliary task of the diffusion generator to progressively refine the shadow mask. Finally, comprehensive experiments demonstrate the superiority of our ShadowDiffusion, which achieves significant improvement compared to the state-of-the-art methods over ISTD, ISTD+, and SRD datasets.

References

[1] Mahmoud Afifi, Konstantinos G Derpanis, Bjorn Ommer, and Michael S Brown. Learning multi-scale photo exposure correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9157–9167, 2021. 8

[2] M. V Afonso, J. M Bioucas-Dias, and M. AT Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE transactions on image processing*, 19(9):2345–2356, 2010. 5

[3] Yu-Sheng Chen, Yu-Ching Wang, Man-Hsin Kao, and Yung-Yu Chuang. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6306–6314, 2018. 8

[4] Zipei Chen, Chengjiang Long, Ling Zhang, and Chunxia Xiao. Canet: A context-aware network for shadow removal. In *ICCV*, pages 4743–4752, 2021. 1, 8

[5] Xiaodong Cun, Chi-Man Pun, and Cheng Shi. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan. In *AAAI*, pages 10680–10687, 2020. 2, 5, 6, 8

[6] Bin Ding, Chengjiang Long, Ling Zhang, and Chunxia Xiao. Argan: Attentive recurrent generative adversarial network for shadow detection and removal. In *ICCV*, pages 10213–10222, 2019. 2

[7] Lan Fu, Changqing Zhou, Qing Guo, Felix Juefei-Xu, Hongkai Yu, Wei Feng, Yang Liu, and Song Wang. Auto-exposure fusion for single-image shadow removal. In *CVPR*, pages 10571–10580, 2021. 1, 2, 5, 6, 7, 8

[8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020. 2

[9] K. Gregor and Y. LeCun. Learning fast approximations of sparse coding. In *Proceedings of the 27th international conference on international conference on machine learning*, pages 399–406, 2010. 3

[10] Lanqing Guo, Siyu Huang, Ding Liu, Hao Cheng, and Bihan Wen. Shadowformer: Global context helps image shadow removal. *arXiv preprint arXiv:2302.01650*, 2023. 1

[11] Lanqing Guo, Siyu Huang, Haosen Liu, and Bihan Wen. Fino: Flow-based joint image and noise model. *arXiv preprint arXiv:2111.06031*, 2021. 2

[12] Lanqing Guo, Renjie Wan, Guan-Ming Su, Alex C Kot, and Bihan Wen. Multi-scale feature guided low-light image enhancement. In *2021 IEEE International Conference on Image Processing (ICIP)*, pages 554–558. IEEE, 2021. 2

[13] Lanqing Guo, Renjie Wan, Wenhan Yang, Alex Kot, and Bihan Wen. Enhancing low-light images in real world via cross-image disentanglement. *arXiv preprint arXiv:2201.03145*, 2022. 2

[14] Lanqing Guo, Zhiyuan Zha, Saiprasad Ravishankar, and Bihan Wen. Self-convolution: A highly-efficient operator for non-local image restoration. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1860–1864, 2021. 2

[15] Ruiqi Guo, Qieyun Dai, and Derek Hoiem. Paired regions for shadow detection and removal. 35(12):2956–2967, 2012. 6

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015. 5

[17] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020. 1, 2, 4

[18] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection and removal. 42(11):2795–2808, 2020. 6, 7

[19] Xiaowei Hu, Yitong Jiang, Chi-Wing Fu, and Pheng-Ann Heng. Mask-shadowgan: Learning to remove shadows from unpaired data. In *ICCV*, pages 2472–2481, 2019. 1, 2, 6

[20] Yeying Jin, Ruoteng Li, Wenhan Yang, and Robby T Tan. Estimating reflectance layer from a single image: Integrating reflectance guidance and shadow/specular aware learning. *arXiv preprint arXiv:2211.14751*, 2022. 1

[21] Yeying Jin, Aashish Sharma, and Robby T Tan. Dc-shadownet: Single-image hard and soft shadow removal using unsupervised domain-classifier guided network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5027–5036, 2021. 6, 7

[22] Yeying Jin, Wenhan Yang, and Robby T Tan. Unsupervised night image enhancement: When layer decomposition meets light-effects suppression. In *European Conference on Computer Vision*, pages 404–421. Springer, 2022. 2

[23] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *arXiv preprint arXiv:2201.11793*, 2022. 2

[24] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977. 3

[25] Hieu Le and Dimitris Samaras. Shadow removal via shadow image decomposition. In *ICCV*, pages 8578–8587, 2019. 2, 5, 6, 8

[26] Hieu Le and Dimitris Samaras. From shadow segmentation to shadow removal. In *ECCV*, 2020. 6

[27] D. Liu, Z. Wang, B. Wen, J. Yang, W. Han, and T. S Huang. Robust single image super-resolution via deep networks with sparse prior. *IEEE Transactions on Image Processing*, 25(7):3194–3207, 2016. 3

[28] Jiaying Liu, Wenhan Yang, Shuai Yang, and Zongming Guo. Erase or fill? deep joint recurrent rain removal and reconstruction in videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3233–3242, 2018. 2

[29] Zhihao Liu, Hui Yin, Xinyi Wu, Zhenyao Wu, Yang Mi, and Song Wang. From shadow generation to shadow removal. In *CVPR*, 2021. 2, 6

[30] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11461–11471, 2022. 2

[31] Ozan Özdenizci and Robert Legenstein. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *arXiv preprint arXiv:2207.14626*, 2022. 2, 5, 8

[32] Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, and Rynson WH Lau. Deshadownet: A multi-context embedding deep network for shadow removal. In *CVPR*, pages 4067–4075, 2017. 5, 6, 7, 8

[33] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–10, 2022. 1, 2, 4

[34] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 1, 2, 4, 5, 8

[35] Jo Schlemper, Jose Caballero, Joseph V. Hajnal, Anthony N. Price, and Daniel Rueckert. A deep cascade of convolutional neural networks for dynamic mr image reconstruction. *IEEE Transactions on Medical Imaging*, 37(2):491–503, 2018. 5

[36] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265. PMLR, 2015. 2

[37] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020. 5

[38] Jifeng Wang, Xiang Li, and Jian Yang. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *CVPR*, pages 1788–1797, 2018. 1, 2, 5, 6, 7, 8

[39] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6849–6857, 2019. 8

[40] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex Kot. Low-light image enhancement with normalizing flow. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 2604–2612, 2022. 2

[41] Yufei Wang, Yi Yu, Wenhan Yang, Lanqing Guo, Lap-Pui Chau, Alex Kot, and Bihan Wen. Raw image reconstruction with learned compact metadata. *arXiv preprint arXiv:2302.12995*, 2023. 2

[42] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 6

[43] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17683–17693, 2022. 5

[44] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5901–5910, 2022. 8

[45] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1357–1366, 2017. 2

[46] Yan Yang, Jian Sun, Huibin Li, and Zongben Xu. Deep admm-net for compressive sensing mri. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. 3

[47] Yi Yu, Yufei Wang, Wenhan Yang, Shijian Lu, Yap-peng Tan, and Alex C Kot. Backdoor attacks against deep image compression via adaptive frequency trigger. *arXiv preprint arXiv:2302.14677*, 2023. 2

[48] Yi Yu, Wenhan Yang, Yap-Peng Tan, and Alex C Kot. Towards robust rain removal against adversarial attacks: A comprehensive benchmark analysis and beyond. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6013–6022, 2022. 2

[49] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *ECCV*, 2020. 8

[50] Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding network for image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3217–3226, 2020. 3

[51] Rongkai Zhang, Lanqing Guo, Siyu Huang, and Bihan Wen. Rellie: Deep reinforcement learning for customized low-light image enhancement. In *Proceedings of the 29th ACM international conference on multimedia*, pages 2429–2437, 2021. 2

[52] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *International Journal of Computer Vision*, 129(4):1013–1037, 2021. 8

[53] Lei Zhu, Ke Xu, Zhanghan Ke, and Rynson WH Lau. Mitigating intensity bias in shadow detection via feature decomposition and reweighting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4702–4711, 2021. 8

[54] Yurui Zhu, Jie Huang, Xueyang Fu, Feng Zhao, Qibin Sun, and Zheng-Jun Zha. Bijective mapping network for shadow removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5627–5636, 2022. 1, 2, 5, 6, 7, 8

[55] Yurui Zhu, Zeyu Xiao, Yanchi Fang, Xueyang Fu, Zhiwei Xiong, and Zheng-Jun Zha. Efficient model-driven network for shadow removal. 2022. 5, 6, 7

[56] Wangmeng Zuo, Kai Zhang, and Lei Zhang. *Convolutional Neural Networks for Image Denoising and Restoration*, pages 93–123. Springer International Publishing, Cham, 2018. 2