An improved method for Image Shadow Removal by Combining Deterministic and Stochastic Models

Hongjun Sheng*, Lanqing Guo†, Xinggan Peng‡§, Zhiping Lin* and Bihan Wen*
* School of EEE, Nanyang Technological University, Singapore
E-mail: {HSHENG002, EZPLin, bihan.wen}@e.ntu.edu.sg
† The University of Texas at Austin, US
E-mail: lanqing.guo@austin.utexas.edu
‡ CMCU Engineering Co., Ltd, China
E-mail: pxg@cmcu.cn

***Abstract*—This paper tackles shadow removal, a key low-level vision task aiming to recover shadow-free images from shadow-contaminated inputs. While Transformer-based methods achieve strong results, their high computational cost limits practical use. Diffusion models have recently shown superior performance in modeling complex image distributions for restoration tasks. We build upon the ShadowDiffusion framework—a diffusion-based model with a degradation prior and deep unrolling strategy—and propose two key improvements: (1) integrating the Nonlinear Activation Free Network (NAFNet) as the degradation estimator to combine efficient deterministic modeling with the generative power of diffusion, and (2) optimizing the penalty parameter in the unrolling process to enhance quality without added training cost. Our enhanced model, ShadowDiffusion+, is evaluated on two public datasets. On SRD, it improves PSNR from 34.73 dB to 36.01 dB and SSIM from 0.970 to 0.979, demonstrating both higher accuracy and efficiency.**

I. Introduction

Image restoration is a core task in computer vision, focused on recovering high-quality images from degraded inputs caused by various factors such as blur, rain, and shadows. Among these, shadow removal addresses illumination inconsistencies caused when opaque objects obstruct direct light, leading to regions of reduced brightness. Shadows not only degrade visual quality but also hinder the performance of downstream tasks such as object detection, tracking, semantic segmentation, and face recognition [1], [2].

Traditional shadow removal methods rely on hand-crafted priors such as illumination [3], gradient [4], and color cues [5] to restore shadow regions. One class of methods [21] focuses on removing image gradients along shadow edges, followed by gradient reintegration to reconstruct a shadow-free image. However, these approaches are based on idealized assumptions and often lead to visible artifacts near shadow boundaries in real-world conditions. Moreover, they require precise and fine-grained shadow edge detection as a strong prior, which is difficult to obtain in practice. Another line of work [3] formulates shadow removal as a relighting problem, aiming to estimate a scaling factor to brighten shadow pixels. The main challenge lies in assigning appropriate correction factors for different types of shadows (umbra and penumbra) and addressing non-uniform illumination degradation near shadow boundaries.

In recent years, deep learning-based shadow removal methods have made remarkable progress, driven by the availability of large-scale annotated datasets. Qu et al. [6] pioneered this direction by introducing a multi-branch fusion framework in 2017, marking the first application of CNNs to shadow removal. Subsequently, the success of GANs in low-level vision tasks led to the widespread adoption of GAN-based architectures, which not only reduce visual artifacts but also support unsupervised learning from unpaired shadow and shadow-free images. To further enhance restoration performance, several works have incorporated explicit priors such as illumination models and shadow masks into network designs. Optimization-based strategies like deep unrolling [7] and recurrent learning [8] have also been employed to improve feature refinement.

In this paper, we focus on shadow removal, as we believe that effectively eliminating shadows enhances the quality and usability of images for downstream tasks. By restoring images from shadow-induced degradation, our method can facilitate subsequent applications such as image segmentation and object tracking, ultimately improving their accuracy and performance. Building upon the state-of-the-art ShadowDiffusion framework [9], our work introduces key improvements aimed at increasing both performance and efficiency: (1) We adopt NAFNet [10] as a lightweight and accurate degradation prior, replacing Transformer-based architectures in ShadowDiffusion. By predicting precise degradation maps, NAFNet provides strong structural guidance for the diffusion model, which then refines the output via deep unrolling. This integration leverages the complementary strengths of deterministic and stochastic modeling, enabling fast, high-quality, and detail-preserving shadow removal. (2) We optimize the penalty parameter in the deep unrolling framework of ShadowDiffusion, which leads to improved reconstruction quality without incurring addi-

§Corresponding author: Xinggan Peng.

This work was supported by the Brain-Gain Plan of New Chongqing (No. CSTB2024YCJH-KYXM0118), and the fellowship of Chongqing Postdoctoral Program for Innovative Talents (No.CQBX202425).

979-8-3315-7206-8/25/$31.00 ©2025 IEEE

tional training cost. Our improved model, ShadowDiffusion+, achieves superior performance on two public datasets. Notably, on the SRD dataset [6], the PSNR is improved from 34.73 dB to 36.01 dB, and the SSIM is increased from 0.970 to 0.979, demonstrating the effectiveness of our proposed enhancements

II. Method

A. Problem Definition

Following the common formulation for deep learning–based shadow removal [11], we model a shadow image $I_s \in \mathbb{R}^{H\times W\times 3}$ (with height H and width W) as undergoing *localized degradation* that affects only shadowed regions. To direct restoration toward those regions while preserving the rest of the scene, we introduce an auxiliary binary shadow mask M, obtained either by manual annotation or a pre-trained detector. The mask provides spatial priors that guide the network's attention and has become integral to modern mask-guided frameworks. The restoration process is therefore formulated as:

$$\hat{I}_{\text{sf}} = G(I_s, M; \theta), \tag{1}$$

where $\hat{I}_{\text{sf}}$ denotes the restored shadow-free image, and $G(\cdot;\theta)$ represents the shadow removal network with learnable parameters θ.

B. Deep Unrolling

Several studies have employed deep unrolling to improve both interpretability and performance by embedding physical priors into deep networks. This technique has shown strong results across various image restoration tasks [12]. Zhu et al. [7], for instance, proposed a half-quadratic splitting–based unrolling scheme that enables a single network to handle multiple scale factors in super-resolution. In the context of shadow removal, the shadow formation model can be expressed as:

$$I_s = A \circ L_{sf}, \tag{2}$$

where A denotes the shadow degradation map and L_{sf} the shadow-free scene illumination. Based on this, shadow removal can be reformulated as a degradation-prior-guided optimization problem, where regularization $R(\cdot)$ is learned via deep CNNs [7] or generative diffusion models [9]:

$$\hat{I}_{sf} = \arg\min_{I_{sf}} \|I_s - A \circ I_{sf}\|_F^2 + R(I_{sf}). \tag{3}$$

C. Baseline Model - ShadowDiffusion

Guo et al. [9] first introduced a novel unrolling diffusion model, termed ShadowDiffusion. This approach significantly enhances shadow removal performance by progressively refining the target output through the integration of both a degradation prior and a diffusive generative prior, establishing a strong new baseline for image restoration. Moreover, ShadowDiffusion iteratively refines the predicted shadow mask as an auxiliary task within the diffusion process, resulting in more accurate and robust shadow-free image generation.

D. Our Improved Model - ShadowDiffusion+

We first employ NAFNet as the degradation estimation network to provide a more accurate and efficient degradation prior. Furthermore, we optimize the penalty parameter ρ to improve the performance of the unrolling process.

We get a local optimization formulation [9] for the image variable by minimizing the energy function with a joint image-mask regularizer, applying half-quadratic splitting to a variant of Eq. (3):

$$\begin{aligned}\min_{x,m,z,v} \quad & \frac{1}{2}\|h\cdot z - y\|_F^2 + \psi R([x|m]) + \frac{\phi}{2}\|v-\tilde{m}\|_F^2 \\ & + \frac{\rho}{2}\|x-z\|_F^2 + \frac{\rho}{2}\|m-v\|_F^2,\end{aligned} \tag{4}$$

where the degradation matrix h that is estimated by pre-trained degradation estimation network, the shadow image y, the diffsuion model $R(\cdot)$, the mask m, the auxiliary variables z and v, the trade-off parameters ψ and ϕ and the penalty coefficient ρ.

Then tractable sub-problems for image estimation:

$$z_{t-1} = \arg\min_{z} \frac{1}{2}\|h\cdot z - y\|_F^2 + \frac{\rho}{2}\|x_{t-1} - z\|_F^2, \tag{5}$$

then we get the update rule for z:

$$z_{t-1} = \frac{h\cdot y + \rho x_{t-1}}{h\cdot h + \rho}, \tag{6}$$

this formulation allows the image estimate x_{t-1} at diffusion timestep $t-1$ to be progressively refined while balancing the degradation consistency and the prior regularization through the diffsuion inference.

By combining the precision of deterministic modeling from NAFNet with the expressive power of stochastic generation from the diffusion framework, our method effectively leverages their complementary strengths to enhance both computational efficiency and reconstruction quality.

1) NAFNet: NAFNet (Nonlinear Activation Free Network) [10] is a lightweight yet powerful image restoration model. It challenges conventional designs by eliminating nonlinear activation functions like ReLU and GELU, replacing them with simpler gating mechanisms such as element-wise multiplication. NAFNet, developed based on a simplified U-Net architecture, delivers state-of-the-art performance in image denoising and deblurring tasks. For instance, NAFNet outperforms previous methods in various low-level tasks while requiring only a fraction of their computational cost, making it both efficient and effective.

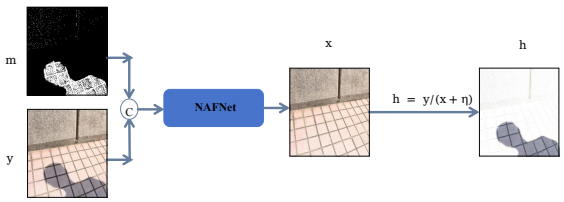

Fig. 1. Degradation estimation with NAFNet.

2) Degradation Estimation with NAFNet: As we can see in Fig. 1, we use NAFNet as our degradation estimation network . If a high-quality shadow-free image is produced, it can facilitate the generation of a more accurate degradation map, which in turn enhances the effectiveness of the unrolling functions described in Eq. (6). By integrating NAFNet's deterministic regression capability with the generative refinement process of the diffusion model, our framework leverages the complementary strengths of deterministic and stochastic modelling, leading to more stable optimization and finer detail reconstruction during shadow removal.

3) Loss Fuctions for Degradation Estimation with NAFNet: To guide the training of our NAFNet-based shadow removal model, we employ a composite loss function that balances multiple objectives, ensuring both pixel-wise accuracy and perceptual fidelity. The total loss is defined as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\ell_1} + 0.1 \cdot \mathcal{L}_{\text{SSIM}} + 0.05 \cdot \mathcal{L}_{\text{VGG}} + 0.2 \cdot \mathcal{L}_{\text{mask}-\ell_1} \quad (7)$$

Each component of this loss serves a distinct purpose:

- **L1 Loss ($\mathcal{L}_{\ell_1}$):**
  Also known as Mean Absolute Error (MAE), which penalizes the absolute difference between predicted and ground-truth pixel values, treating all deviations equally. Its robustness to outliers and ability to preserve high-frequency details, such as edges and textures, make it effective for visual fidelity.
- **SSIM Loss ($\mathcal{L}_{\text{SSIM}}$):**
  Structural Similarity Index (SSIM) loss measures perceptual similarity by comparing luminance, contrast, and structure between restored and reference images, encouraging visually pleasing and structurally consistent reconstructions.
- **VGG Perceptual Loss ($\mathcal{L}_{\text{VGG}}$):**
  This term compares high-level features from the predicted and ground-truth images, extracted via a pre-trained VGG network φ [13], encouraging semantic consistency and perceptual similarity beyond pixel-level losses.

$$\mathcal{L}_{\text{VGG}} = \left\| \varphi(I_{sf}) - \varphi(\hat{I}_{sf}) \right\|_F^2. \quad (8)$$

- **Masked L1 Loss ($\mathcal{L}_{\text{mask}-\ell_1}$):**
  This term focuses learning on shadow regions via L1 error weighted by a binary mask, enabling the model to better target and remove shadows.

$$\mathcal{L}_{\text{mask}-\ell_1} = \left\| (I_{sf} - \hat{I}_{sf}) \odot M \right\|. \quad (9)$$

Together, this multi-objective loss function ensures that the model produces results that are accurate in pixel values, structurally coherent, and visually compelling, while also placing targeted emphasis on regions affected by shadows.

III. Experiment

A. Datasets

The ISTD dataset [14] contains 1,330 training and 540 testing triplets, each comprising a shadow image, its manually annotated mask, and the corresponding shadow-free image. It primarily features outdoor scenes with hard shadows cast by various objects. To mitigate illumination inconsistencies between paired images in ISTD, the ISTD+ dataset [15] was proposed, offering improved lighting consistency. In this work, we adopt ISTD+ for evaluation.

The SRD dataset [6] includes 2,680 training and 408 testing pairs of shadow and shadow-free images, without manual mask annotations. Shadow-free images are obtained by removing occluding objects, as in ISTD. Compared to ISTD, SRD exhibits more diverse lighting conditions—e.g., different weather and times of day—and covers both hard and soft shadows. We use predicted masks from DHAN [16] during training and testing.

B. Implementation Details

All input images are resized to 256×256 as a standard preprocessing step, with data augmentation (horizontal flipping and random cropping) applied to enhance generalization. The proposed model adopts the NAFNet architecture with 4 input channels and a base width of 64, using $[2, 2, 4, 8]$ encoder blocks, 12 middle blocks, and $[2, 2, 2, 2]$ decoder blocks. We train the model using the AdamW optimizer (β_1=0.9, β_2=0.999), with a learning rate of 2×10^{-4} for ISTD+ and 1×10^{-3} for SRD to account for dataset scale. All experiments are conducted for 100 epochs on a single NVIDIA RTX 3090 GPU.

C. Evaluation Metrics

Peak Signal-to-Noise Ratio (PSNR) and Root Mean Square Error (RMSE) are standard metrics for image reconstruction. PSNR measures the ratio of maximum pixel intensity to distortion power (higher is better), while RMSE computes the square root of mean squared errors (lower is better).

Structural Similarity Index Measure (SSIM) assesses perceptual quality via luminance, contrast, and structure comparisons. Unlike pixel-wise metrics, it models degradation as structural distortion, ranging from -1 to 1, with 1 indicating perfect similarity.

D. Experimental Results and Analysis

Table I reports the quantitative results on the ISTD+ and SRD datasets, evaluated using PSNR (dB), SSIM, and RMSE. The best results are highlighted in bold red, and the second-best in bold blue. Our proposed ShadowDiffusion+ consistently achieves state-of-the-art performance across all metrics for the entire image.

On the SRD dataset, ShadowDiffusion+ demonstrates significant improvements. It achieves a PSNR of 36.01 dB on the full image, surpassing the second-best HomoFormer (35.37 dB) by 0.64 dB. In the shadow region, our method achieves 41.02 dB, outperforming HomoFormer's 38.81 dB. RMSE is also reduced from 3.33 to 2.95, indicating improved pixel-level accuracy. These results confirm the robustness of our method in handling complex shadow patterns and illumination. On ISTD+, ShadowDiffusion+ matches the highest PSNR (35.67 dB) achieved by the original ShadowDiffusion, while

TABLE I
QUANTITATIVE RESULTS ON ISTD+ AND SRD DATASETS. BEST RESULTS ARE IN **RED**, SECOND-BEST IN **BLUE**.

Dataset	Method	Shadow Region			Non-Shadow Region			All Image		
		PSNR↑	SSIM↑	RMSE↓	PSNR↑	SSIM↑	RMSE↓	PSNR↑	SSIM↑	RMSE↓
ISTD+	input	20.83	0.930	39.01	37.46	0.985	2.40	20.46	0.894	8.40
	BMNet [17]	37.87	0.991	5.23	37.51	0.985	2.45	33.98	0.972	2.97
	SG-ShadowNet [18]	36.80	0.990	5.13	35.98	0.978	2.92	32.46	0.962	3.41
	Inpaint4Shadow [19]	38.10	0.990	5.06	37.66	0.981	2.82	34.16	0.967	3.35
	ShadowFormer [20]	39.48	**0.992**	5.23	36.93	0.983	2.30	**35.46**	0.971	2.78
	ShadowDiffusion [9]	**39.69**	**0.992**	**4.97**	**38.89**	**0.987**	2.28	**35.67**	**0.975**	2.72
	HomoFormer [21]	**39.49**	**0.993**	**4.73**	38.75	**0.984**	**2.23**	35.35	**0.975**	**2.64**
	Ours	39.23	**0.993**	5.25	**39.67**	**0.987**	**2.16**	**35.67**	**0.976**	**2.63**
SRD	input	18.96	0.871	36.69	31.47	0.975	4.83	18.19	0.83	14.05
	DHAN [16]	33.67	0.978	8.94	34.79	0.979	4.80	30.51	0.949	5.67
	BMNet [17]	35.05	0.981	6.61	36.02	0.982	3.61	31.69	0.956	4.46
	ShadowFormer [20]	35.55	0.982	6.14	36.82	0.983	3.54	32.46	0.957	4.28
	ShadowDiffusion [9]	38.72	**0.987**	4.98	37.78	0.985	3.44	34.73	0.970	3.63
	Inpaint4Shadow [19]	36.73	0.985	5.70	36.70	0.985	3.27	33.27	0.967	3.81
	DeS3 [22]	37.91	0.986	5.27	37.45	0.984	3.03	34.11	0.968	3.56
	HomoFormer [21]	**38.81**	**0.987**	**4.25**	**39.45**	**0.988**	**2.85**	**35.37**	**0.972**	**3.33**
	Ours	**41.02**	**0.993**	**4.37**	**38.76**	**0.991**	**2.60**	**36.01**	**0.979**	**2.95**

further improving SSIM from 0.975 to 0.976 and reducing RMSE from 2.72 to 2.63. In non-shadow regions, it achieves the best results across all metrics—PSNR (39.67 dB), SSIM (0.987), and RMSE (2.16)—highlighting its ability to preserve fine details outside shadow areas.

In summary, ShadowDiffusion+ exhibits strong generalization across datasets with diverse shadow characteristics and achieves superior performance in both shadowed and non-shadowed regions, validating the effectiveness of the proposed degradation prior and iterative optimization framework.

E. Ablation Study

We conduct ablation studies to verify that the enhancements introduced in ShadowDiffusion+, compared to the original ShadowDiffusion, effectively contribute to performance improvements.

TABLE II
QUANTITATIVE COMPARISON ON SRD.

Method	**PSNR ↑**	**SSIM ↑**
ShadowDiffusion w/o unrolling	34.36	0.970
ShadowDiffusion (Uformer)	34.73	0.970
ShadowDiffusion+ (NAFNet)	**36.01**	**0.979**

1) Degradation Estimation Networks: Table II reports the performance of different degradation estimation networks within the unrolling framework, with best results in bold. Overall, the unrolling mechanism yields clear improvements over non-unrolling baselines, demonstrating its effectiveness in progressively refining predictions. We observe that stronger degradation estimators further enhance restoration quality. In particular, integrating NAFNet enables the model to capture finer structural details and generate informative degradation priors, which help iteratively suppress residual artifacts. This combination of deterministic regression and stochastic generation leads to faster, high-quality, and detail-preserving shadow removal.

As shown in Table III, the lightweight NAFNet consistently outperforms ShadowFormer in PSNR despite its lower computational cost. Since ShadowFormer is an improved variant of Uformer—the original degradation estimator in ShadowDiffusion—this comparison confirms NAFNet's superior capability in estimating degradation priors. On the ISTD+ dataset, NAFNet even surpasses HomoFormer, the previous state-of-the-art, demonstrating strong generalization and effectiveness. Given the pivotal role of degradation maps in the unrolling process, a more accurate prior with higher PSNR provides stronger deterministic guidance, leading to more effective shadow removal and overall performance gains.

TABLE III
QUANTITATIVE COMPARISON ON SRD AND ISTD+.

Method	**SRD (PSNR ↑)**	**ISTD+ (PSNR ↑)**
ShadowFormer	32.46	35.46
HomoFormer	**35.37**	35.35
NAFNet	33.78	**35.63**

2) Penalty parameter ρ in the unrolling function: To examine the effect of the penalty parameter in Eq. (6), we performed an ablation study aimed at evaluating its impact on the model's performance.

TABLE IV
QUANTITATIVE COMPARISON OF DIFFERENT PENALTY PARAMETERS FOR SHADOWDIFFUSION+ ON SRD.

Penalty Parameter ρ	**PSNR ↑**
0.2	35.91
0.5	**36.01**
0.8	**36.01**

Table IV shows the effect of the penalty parameter ρ, with best results in bold. Increasing ρ above the default 0.2 improves performance, but gains saturate near 0.8, indicating limited benefit from further unrolling. In Eq. (6), ρ weights the previous estimate x_{t-1}. As x_{t-1} refines, larger ρ accelerates convergence, but beyond $\rho = 0.8$, returns diminish as the prior estimate reaches optimal utility.

F. Visualization Analysis

To highlight the improvements of our proposed ShadowDiffusion+ over the original ShadowDiffusion, we present visual comparisons that qualitatively demonstrate enhanced shadow removal and overall image

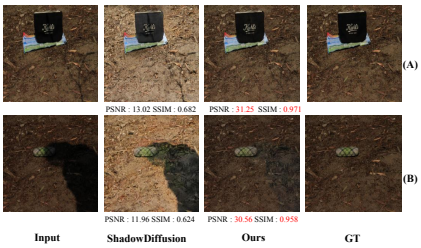

Fig. 2. SRD Visualization 1.

1) SRD Dataset: As illustrated in Fig. 2, a key limitation of the original ShadowDiffusion is its tendency to misclassify naturally dark background regions as shadows, resulting in over-correction, loss of scene realism, and reduced PSNR and SSIM. In contrast, our proposed ShadowDiffusion+ better distinguishes true shadows from dark content, preserving background integrity while more accurately removing shadows. Although shadow removal is not perfect, it avoids excessive brightening and achieves notably higher PSNR and SSIM, particularly because non-shadow regions—which dominate the image—contribute more heavily to these metrics.

Further qualitative comparisons are shown in Fig. 3, with red boxes highlighting representative cases. In (A), ShadowDiffusion leaves residual shadows, while ShadowDiffusion+ achieves more complete removal. In (B), ShadowDiffusion provides only partial restoration, whereas ShadowDiffusion+ yields results with greater visual fidelity. In (C), ShadowDiffusion fails to remove the shadow entirely, producing an output nearly identical to the input, while ShadowDiffusion+ successfully restores the shadow-free region. These examples highlight the superior restoration capability and perceptual quality of our method.

2) ISTD+ Dataset: Although the overall performance gains of ShadowDiffusion+ over the original ShadowDiffusion may appear modest in terms of quantitative metrics, the model introduces significant refinements at shadow boundaries. As shown in Fig. 4, ShadowDiffusion often produces unnatural "zebra stripe" artifacts, leading to abrupt transitions between shadowed and non-shadowed regions. In contrast, ShadowDiffusion+ yields smoother and more natural boundary transitions. These visual comparisons highlight improvements in perceptual quality that may not be fully captured by PSNR or SSIM, yet contribute substantially to the realism and visual fidelity of the restored images.

As shown in Fig. 5, we present additional visual comparisons between ShadowDiffusion and ShadowDiffusion+, with red boxes indicating zoomed-in regions for detailed analysis. In example (A), ShadowDiffusion shows partial shadow removal but introduces prominent "zebra stripe" artifacts along boundaries, resulting in unnatural transitions. ShadowDiffusion+ effectively suppresses these artifacts, producing smoother and more realistic edges. Although minor flaws remain, the output is visually closer to the ground truth (GT), contributing to improved PSNR and SSIM. In examples (B) and (C), ShadowDiffusion again leaves residual shadows and boundary artifacts, whereas ShadowDiffusion+ removes them more effectively, yielding cleaner, GT-aligned results. These comparisons highlight that, despite modest overall metric gains, the boundary refinements introduced by ShadowDiffusion+ significantly enhance perceptual quality in challenging regions.

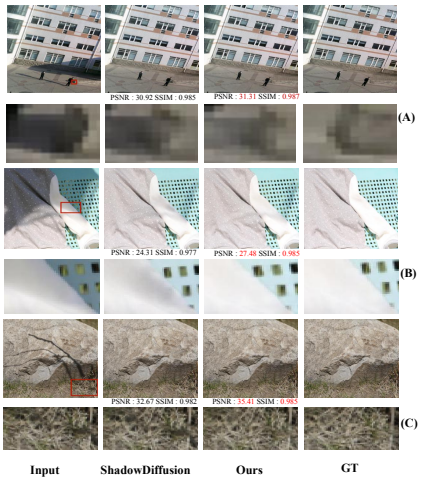

Fig. 3. SRD Visualization 2.

Fig. 4. ShadowDiffusion's "zebra stripe" problem.

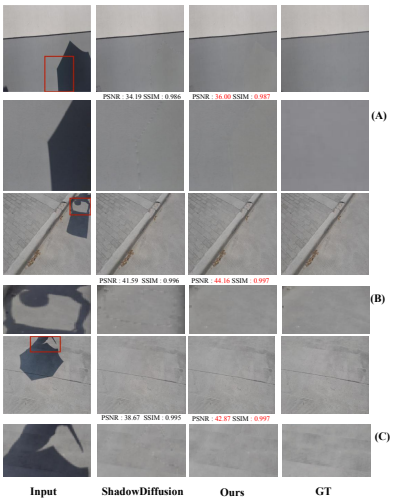

Fig. 5. ISTD+ Visualization.

IV. Conclusions

In this work, we enhance the ShadowDiffusion framework for shadow removal by integrating a lightweight degradation prior (NAFNet) and optimizing the unrolling strategy. This deterministic–stochastic synergy significantly improves restoration quality while reducing computational cost. Our improved model, ShadowDiffusion+, achieves state-of-the-art performance on public benchmarks (e.g., +1.28 dB PSNR on SRD), with clearer shadow boundaries and better perceptual consistency. These results validate the effectiveness of combining diffusion models, efficient priors, and deep unrolling for high-quality, efficient shadow removal.

References

[1] S. Nadimi and B. Bhanu, "Physical models for moving shadow and object detection in video," *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 8, pp. 1079–1087, 2004.

[2] W. Zhang, X. Zhao, J.-M. Morvan, and L. Chen, "Improving shadow suppression for illumination robust face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 3, pp. 611–624, 2018.

[3] L. Zhang, Q. Zhang, and C. Xiao, "Shadow remover: Image shadow removal based on illumination recovering optimization," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4623–4636, 2015.

[4] M. Gryka, M. Terry, and G. J. Brostow, "Learning to remove soft shadows," *ACM Transactions on Graphics (TOG)*, vol. 34, no. 5, pp. 1–15, 2015.

[5] G. D. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew, "On the removal of shadows from images," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 1, pp. 59–68, 2005.

[6] L. Qu, J. Tian, S. He, Y. Tang, and R. W. Lau, "Deshadownet: A multi-context embedding deep network for shadow removal," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4067–4075.

[7] Y. Zhu, Z. Xiao, Y. Fang, X. Fu, Z. Xiong, and Z.-J. Zha, "Efficient model-driven network for shadow removal," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, 2022, pp. 3635–3643.

[8] B. Ding, C. Long, L. Zhang, and C. Xiao, "Argan: Attentive recurrent generative adversarial network for shadow detection and removal," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 10 213–10 222.

[9] L. Guo, C. Wang, W. Yang, *et al.*, "Shadowdiffusion: When degradation prior meets diffusion model for shadow removal," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 14 049–14 058.

[10] L. Chen, X. Chu, X. Zhang, and J. Sun, "Simple baselines for image restoration," in *European conference on computer vision*, Springer, 2022, pp. 17–33.

[11] L. Guo, C. Wang, Y. Wang, *et al.*, "Single-image shadow removal using deep learning: A comprehensive survey," *arXiv preprint arXiv:2407.08865*, 2024.

[12] H. Zheng, H. Yong, and L. Zhang, "Unfolded deep kernel estimation for blind image super-resolution," in *European conference on computer vision*, Springer, 2022, pp. 502–518.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[14] H. Le and D. Samaras, "Shadow removal via shadow image decomposition," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 8578–8587.

[15] J. Wang, X. Li, and J. Yang, "Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1788–1797.

[16] X. Cun, C.-M. Pun, and C. Shi, "Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, 2020, pp. 10 680–10 687.

[17] Y. Zhu, J. Huang, X. Fu, F. Zhao, Q. Sun, and Z.-J. Zha, "Bijective mapping network for shadow removal," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5627–5636.

[18] J. Wan, H. Yin, Z. Wu, X. Wu, Y. Liu, and S. Wang, "Style-guided shadow removal," in *European Conference on Computer Vision*, Springer, 2022, pp. 361–378.

[19] X. Li, Q. Guo, R. Abdelfattah, *et al.*, "Leveraging inpainting for single-image shadow removal," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 13 055–13 064.

[20] L. Guo, S. Huang, D. Liu, H. Cheng, and B. Wen, "Shadowformer: Global context helps shadow removal," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 37, 2023, pp. 710–718.

[21] J. Xiao, X. Fu, Y. Zhu, *et al.*, "Homoformer: Homogenized transformer for image shadow removal," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 25 617–25 626.

[22] Y. Jin, W. Ye, W. Yang, Y. Yuan, and R. T. Tan, "Des3: Adaptive attention-driven self and soft shadow removal using vit similarity," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, 2024, pp. 2634–2642.