A Diffusion Model with Physically Plausible Gradient for Remote Sensing Shadow Removal

Kaichen Chi, Junjie Li, Jun Chu, Qiang Li, *Member, IEEE*, and Qi Wang, *Senior Member, IEEE*

***Abstract*—Remote sensing imagery is important for geographical object exploration, but shadow contamination consistently challenges the image formation quality and subsequent applications. Although the diffusion model significantly advances the shadow removal field, current paradigms ignore the physical property of shadow images and thus lose the desired interpretability. To bridge this gap, we propose SR-Diffusion, a shadow removal diffusion model collaborating infrared thermal distribution, chromaticity, and illumination intensity regulations. The core insight of SR-Diffusion is to inject nearly all available physical priors into the noise during the reverse process, thus providing desirable generative paths in noisy environments. Specifically, we leverage a modal translation (visible $\mapsto$ Infrared) scheme to explore the cross-domain mapping, thus providing the thermal spectrum. Simultaneously, we introduce a novel horizontal/vertical-intensity (HVI) space to decouple the visible modality into chromaticity and illumination. Coupled with a gradient guidance, the above physical constraints are embedded into the noise, which contributes to generate stable shadow-free images. Comprehensive experiments demonstrate that SR-Diffusion outperforms state-of-the-art shadow removal methods.**

***Index Terms*—Shadow removal, remote sensing, denoising diffusion model, physical property.**

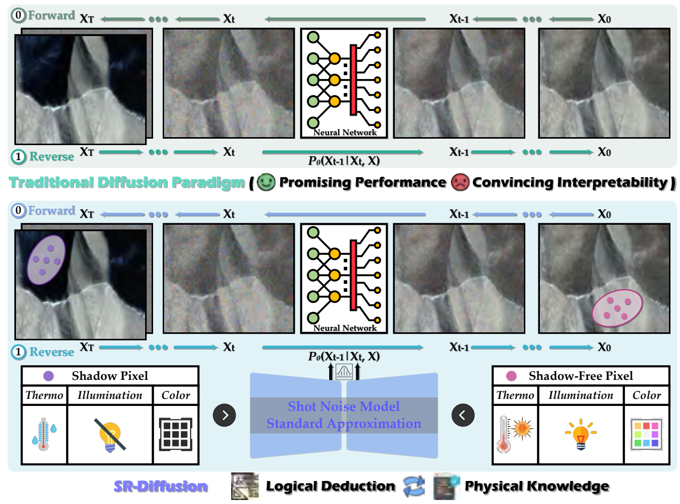

Fig. 1. Schematic of our basic idea. SR-Diffusion incorporates the domain knowledge of thermodynamics, optics, and chromatology into the traditional Diffusion paradigm to accomplish the win-win of promising shadow removal performance and convincing interpretability.

I. Introduction

THE shadow is prevalent in nature and typically forms when the light source is blocked. Unfortunately, such ubiquitous physical phenomena not only distort human perception, but also challenge the vision tasks, such as object detection [1]–[3], segmentation [4], [5], [82], [88], and counting [6]. Thus, it is essential to develop a shadow removal framework to recover the conveyed information of shadow regions.

Currently, shadow removal methods are roughly classified into three groups: illumination model-based, non-illumination model-based, and deep learning-based methods. The illumination model-based method divides shadows into the umbra region, the penumbra region, and the shadow boundary. Subsequently, the illumination of the umbra region is assumed as the ambient illumination, while the illumination of penumbra region and shadow boundary is assumed as a weighted sum of ambient illumination and direct illumination [7]. Based on that, a series of attempts aim to compute illumination intensity to removal shadow contamination, such as texture similarity [7], chromaticity invariance [8], and 3D intensity surface modeling [9]. Nevertheless, it is not reasonable to simplify shadows and illumination intensity to a linear relationship. Such a manner ignores the spatial-variant property of shadows [10]. In contrast, the non-illumination model-based method relies on image intrinsic priors to mitigate shadow effects, such as gradient [11] and user interaction [12]. However, hand-crafted assumptions typically fail in real-world shadow detection and removal, since the diversity of shadows in terms of intensity, shape, and size [13].

Recently, the deep learning-based method flourishes, crediting to the high generalization capability of neural networks. Pioneering ideas have been elaborated from different perspectives, such as homogenized distribution [14], shadow generation [15], light field [16], [81], Retinex theory [17], fourier transform [18], and multiple task decoupling [19], [87]. Unfortunately, these methods compromise on the brute-force mapping from the shadow domain to the shadow-free domain, ignoring convincing physical intuitiveness and interpretability. Therefore, existing deep learning-based methods face three significant challenges: ❶ Overlooking the thermal spectrum distribution. Shadow-free regions with strong thermal energy due to solar radiation, leading to visually attractive contrast and visibility of the infrared modality. In contrast, shadow regions receive only scattered and ambient radiation, displaying a uniform thermal spectrum distribution. This leads to a lack of

This work was supported in part by the National Natural Science Foundation of China under Grant 62301385, 62471394, and U21B2041, and in part by the Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University under Grant CX2024107. Corresponding author: Qi Wang.

The authors are with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, China (e-mail: chikaichen@mail.nwpu.edu.cn, junjieli@mail.nwpu.edu.cn, junchu@mail.nwpu.edu.cn, liqmges@gmail.com, crabwq@gmail.com).

high-frequency components in shadow regions [20]. ❷ Overlooking the chromaticity spectrum inconsistency. Compared with shadow-free regions, the color of shadow regions displays significant degradation. This is because the ratios of standard RGB (sRGB) spectrums have changed. Existing work typically mitigates color deviation by computing the attenuation ratio between shadow and shadow-free pixels in sRGB space [15], [21]. However, they produce color artifacts because of the inherent color sensitivity in sRGB space [22]. ❸ Overlooking the illumination entanglement. The illumination property is typically used to indicate the degree of quality degradation [17], [23], [24]. Nevertheless, the illumination renders a strong coupling with the chromaticity of the sRGB space, which leads to black plane noise [22]. Afterward, a natural question arises, "Why not embed diverse physical properties into the shadow removal model to accomplish the best of both worlds between competitiveness and interpretability?"

To this end, we propose SR-Diffusion that optimizes the diffusion process through gradient-based guidance. The key intuition is to inject the gradient of physical loss into the noise estimation at the training step, thus constructing a rigorous measurement consistency projection step. As shown in Fig. 1, the guidance of SR-Diffusion consists of three components. First, we leverage the infrared thermal spectral distribution to accomplishes high-frequency consistency between shadow and shadow-free pixels. Such a manner ensures detail and texture alignment between the enhanced and shadow-free versions. Later on, we introduce the HVI color space, consisting of a polarized hue-saturation plane and an intensity axis. On the one hand, we alleviate reddish color casts through the Euclidean distance constraint on HV maps with color information. On the other hand, we dynamically compress low-light pixels to remove black noise. In summary, the main contributions are as follows.

- **Perspective contribution.** We rethink the shadow removal task from the perspective of physical rules, injecting gradients from heat, illumination, and chromaticity priors into the noise to ensure the deductive logic of the visual reconstruction.
- **Technical contribution.** We propose an a SR-Diffusion that decouples multiple physical properties through modal and spatial transformations, while treating them as constraints to guide visual fidelity.
- **Practical contribution.** Numerical and visual experiments on representative benchmarks demonstrate the superiority of SR-Diffusion. More importantly, our method is able to generalize to lunar, natural and underwater scenarios with strong generalizability and adaptability.

II. Related Work

In this section, we outline the development of shadow removal, and then discuss the diffusion model towards low-level vision tasks.

Shadow removal. Early attempts typically explore shadow removal based on the simplified illumination formulation. Liu *et al.* [25] decoupled the shadow image into the shadow-free image and the illumination change surface. Subsequently, they employed a gradient field to remove shadow contamination in umbra and penumbra regions. Arbel *et al.* [26] modeled shadows as a product of albedo, illumination field without shadows, and shadow scale factors, while using the intensity surface to compute scale factors to cope with shadow effects. However, the assumptions in these illumination models are typically too restrictive for real-world shadow samples. Moreover, some classical methods take advantage of prior information, *e.g.*, Guo *et al.* [27] employed a graph-cut technique to distinguish shadow and shadow-free regions, and then estimated shadow coefficient values to relight shadow pixels. Yang *et al.* [28] derived a 3D intrinsic version based on chromaticity and bilateral filtering, and then combined the base layer of the shadow version with the detail layer of the intrinsic version to recover illumination. Unfortunately, the above schemes suffer from unsatisfactory performance and poor generalization ability when information transfer between shadow and shadow-free regions fails.

The deep learning-based method achieves a significant breakthrough in shadow removal and dominates this field. Guo *et al.* [30] proposed a shadow degradation model that treats the shadow image as a coupling of the shadow-free image and the pixel-wise illumination degradation map. Subsequently, they embedded the shadow degradation model into the diffusion model to progressively remove shadows through degradation prior and diffusion generation prior. Yu *et al.* [31] designed a three-stage network to accomplish the mitigation of illumination discrepancies between shadow and shadow-free regions, the prediction of illumination characteristics in shadow regions, and shadow removal, respectively. Chen *et al.* [32] developed a progressive aggregation module to incorporate physical property, spatial relation, and temporal coherence to remove shadows from dynamic scenarios. Based on the cycle consistency constraint, Wang *et al.* [33] decomposed shadow removal into more manageable illumination recovery and color transfer, leading to brightness and color consistency in both shadow and shadow-free regions. Nevertheless, these methods inevitably ignore the potential of physical intuitiveness.

Diffusion model. The diffusion-based generative model displays non-trivial attractiveness in low-level vision tasks, such as super-resolution [34], [35], [83]–[85], inpainting [36], [86], low-light enhancement [37], [38], and underwater image enhancement [39], [40]. Chen *et al.* [34] embedded spatial adaptation and temporal alignment into the latent space of the diffusion model to achieve high-resolution appearance and temporal consistency of the video. Li *et al.* [35] leveraged the diffusion model to provide the high-frequency prior, and then designed the large window transformer to accept the prior while expanding the receptive field. Such a manner ensured that super-resolution magnetic resonance imaging details were undistorted. Liu *et al.* [36] proposed a structure-guided diffusion model that learns semantic discrepancies between masked and unmasked regions with the guidance of edge maps. Li *et al.* [37] utilized reinforcement learning to guide the diffusion process for low-light enhancement. Specifically, they treated responses from depth maps, low-light images, and text captions as rewards, ensuring that the diffusion process was consistent with human perception. Given that luminance

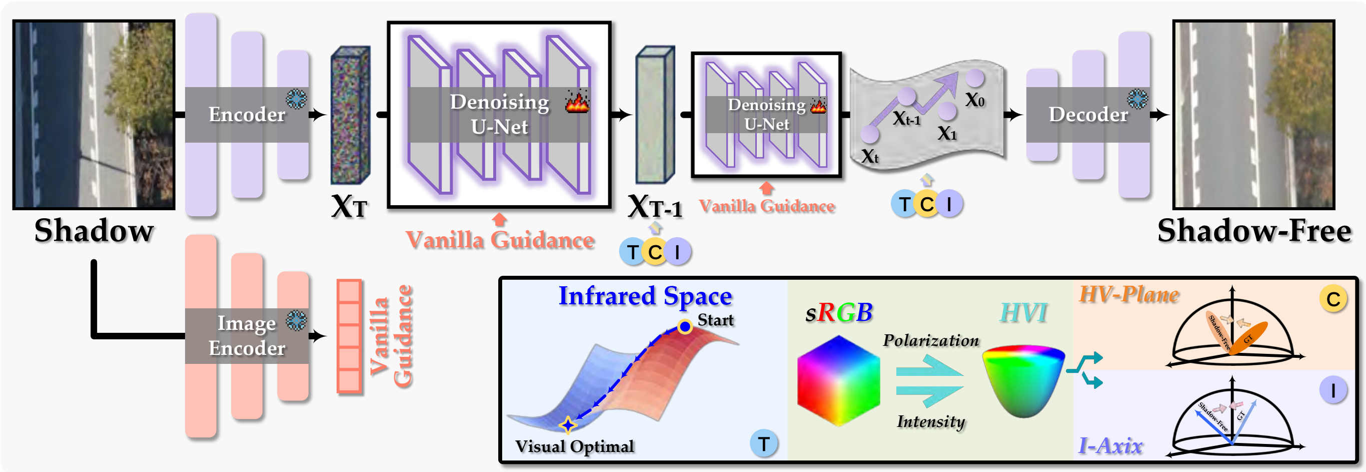

Fig. 2. Schematic diagram of SR-Diffusion. Our method introduces the physical guidance of the shadows themselves (*i.e.*, thermo $\mathcal{T}$, color $\mathcal{C}$, and illumination $\mathcal{I}$) to conquer shadow degradation and logical deduction.

information concentrates on the amplitude while structure information concentrates on the phase, Lv *et al.* [38] embedded the fourier transform into the diffusion model to harmoniously cope with low-light and blur. Similarly, Zhao *et al.* [39] embedded the wavelet transform and fourier transform into the diffusion model to progressively enhance the frequency information. Such a manner integrates frequency optimization and diffusion adjustment to improve the visual quality of underwater scenarios. Zhang *et al.* [40] combined the luminance channel in CIELab space and the RGB channel as a color guidance, forcing the diffusion trajectory to align with visually pleasing underwater images. Although the effectiveness of the diffusion model has been demonstrated in several low-level vision tasks, the implementation of the diffusion model for shadow removal is still a challenge.

III. Methodology

In this section, we first review the diffusion model formulation. Subsequently, we introduce the SR-Diffusion pipeline.

Preliminaries. The diffusion model is a generative paradigm with variational inference, aiming at generating high quality samples x_0 from latent variables x_T with a simple distribution. Specifically, the diffusion model consists of forward and reverse procedures. Based on a variance schedule $\{\beta_1, ..., \beta_T\}$, the forward procedure progressively adds noise perturbations following a normal distribution to the shadow sample:

$$\begin{aligned} \alpha_t &= \Pi_{t=1}^{T}(1-\beta_t), \\ q(x_t \mid x_0) &= \mathcal{N}(x_t; \sqrt{\alpha_t}x_0, (1-\alpha_t)\mathrm{I}), \end{aligned} \tag{1}$$

where I represents the identity matrix. Afterward, the denoising network ϵ_ϕ optimizes the evidence lower bound to predict the noise ϵ:

$$\mathcal{L}_{\text{original}}(\phi) = \mathbb{E}_{x_0,t,\epsilon}\left[\|\epsilon_\phi(x_t,t)-\epsilon\|^2\right], \tag{2}$$

where $\epsilon_\phi(x_t,t)$ represents the predicted noise. Given $\epsilon_\phi(x_t,t)$, the reverse procedure converts the isotropic noise into the distribution form of the shadow-free sample:

$$p_\theta(x_{t-1} \mid x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t,t), \Sigma_\theta(x_t,t)), \tag{3}$$

where $\mu_\theta(x_t,t)$ and $\Sigma_\theta(x_t,t)$ represent the mean and covariance at x_{t-1}, respectively.

In view of the sluggish sampling of the diffusion denoising probabilistic model [41], the denoising diffusion implicit model [42] employs a non-Markovian diffusion process to improve the sampling speed:

$$x_{t-1} = \sqrt{\alpha_{t-1}}\tilde{x}_0(x_t) + \sqrt{1-\alpha_{t-1}-\sigma_t^2}\epsilon_\phi(x_t,t) + \sigma_t z, \tag{4}$$

where $\tilde{x}_0$ represents the predicted x_0 from x_t and σ_t represents the variance. In addition, z follows a standard normal distribution. Assuming σ_t is 0, the sampling process of the denoising diffusion implicit model is considered as a deterministic process, leading to quick sampling from noise:

$$\tilde{x}_0(x_t) = \frac{1}{\sqrt{\alpha_t}}(x_t - \sqrt{1-\alpha_t}\epsilon_\phi(x_t,t)). \tag{5}$$

Pipeline. The reverse procedure of the denoising diffusion implicit model typically relies on diverse constraints to generate stable, visually pleasing samples, *e.g.*, supplementing the loss function with weighted constraints is the mainstream scheme. In contrast, we tackle the constraint issue from the posterior sampling perspective. Inspired by [20], we explore physically plausible priors and inject their gradients into Gaussian noise, as shown in Fig. 2. Such a manner forces the diffusion model to provide trustworthy fidelity according to thermodynamics, optics, and chromatology.

In general, the noise predicted by the denoising U-Net at time t is related to the score from the denoising U-Net at the current moment:

$$\epsilon_\phi(x_t,t) = -\sqrt{1-\alpha_t}\nabla_{x_t}\log p(x_t), \tag{6}$$

where $\log p(\cdot)$ represents the probability density function. We embed the gradient term $\mathcal{G}$ in the probability density function, thus the score of the denoising U-Net at time t is converted to $\nabla_{x_t}\log p(x_t \mid \mathcal{G})$. Unfortunately, $\nabla_{x_t}\log p(x_t \mid \mathcal{G})$ is unknown. How to derive the unknown $\nabla_{x_t}\log p(x_t \mid \mathcal{G})$ from the known $\nabla_{x_t}\log p(x_t)$ is a direction to explore.

According to the Bayes rule, *i.e.*, $p(A \mid B) = \frac{p(B \mid A)}{p(B)} \cdot p(A)$, $\nabla_{x_t} \log p(x_t \mid \mathcal{G})$ can be expressed as:

$$\nabla_{x_t} \log p(x_t \mid \mathcal{G}) = \nabla_{x_t} \log p(x_t) + \nabla_{x_t} \log p(\mathcal{G} \mid x_t), \tag{7}$$

where $\nabla_{x_t} \log p(x_t)$ is known. Therefore, the solution of the unknown $\nabla_{x_t} \log p(\mathcal{G} \mid x_t)$ becomes critical. According to [43], we factorize $p(\mathcal{G} \mid x_t)$ as:

$$\begin{aligned} p(\mathcal{G} \mid x_t) &= \int p(\mathcal{G} \mid x_0, x_t) p(x_0 \mid x_t) dx_0, \\ &= \int p(\mathcal{G} \mid x_0) p(x_0 \mid x_t) dx_0, \end{aligned} \tag{8}$$

where the second equation applies to $\mathcal{G}$ and x_t conditionally independent of x_0. In view of the interpretation $p(\mathcal{G} \mid x_t) = \mathbb{E}_{x_0 \sim p(x_0 \mid x_t)}[p(\mathcal{G} \mid x_0)]$ from Eq. 8, we encourage a reasonable approximation [43] for $p(\mathcal{G} \mid x_t)$:

$$\begin{aligned} \hat{x}_0 &= \frac{1}{\sqrt{\bar{\alpha}(t)}} (x_t + (1 - \bar{\alpha}(t)) \nabla_{x_t} \log p_t(x_t)), \\ p(\mathcal{G} \mid x_t) &\simeq p(\mathcal{G} \mid \hat{x}_0), \end{aligned} \tag{9}$$

Such a manner provides approximate posterior sampling by maximizing the likelihood. According to the Jensen gap, the upper bound of the approximation error can be expressed as:

$$\begin{aligned} \mathcal{J} &\le \frac{d}{\sqrt{2\pi\sigma^2}} e^{-1/2\sigma^2} \|\nabla_x \mathcal{M}(x)\| m, \\ m &= \int \|x_0 - \hat{x}_0\| p(x_0 | x_t) dx_0, \end{aligned} \tag{10}$$

where $\mathcal{M}(\cdot)$ represents a forward operator. Since m is finite for most distributions, the Jensen gap tends to 0 as $\sigma \mapsto \infty$ [44]. Therefore, we employ the approximate gradient of the log likelihood:

$$\nabla_{x_t} \log p(\mathcal{G} \mid x_t) \simeq \nabla_{x_t} \log p(\mathcal{G} \mid \hat{x}_0), \tag{11}$$

The likelihood function can be expressed as:

$$p(\mathcal{G} \mid x_0) = \frac{1}{\sqrt{(2\pi)^n \sigma^{2n}}} \exp\left[-\frac{\|\mathcal{G} - \mathcal{M}(x_0)\|_2^2}{2\sigma^2} \right], \tag{12}$$

where n represents the dimension of $\mathcal{G}$. Based on Eq. 11, We differentiate $p(\mathcal{G} \mid x_t)$ with respect to x_t:

$$\nabla_{x_t} \log p(\mathcal{G} \mid x_t) \simeq -\rho \nabla_{x_t} \|\mathcal{G} - \mathcal{M}(\hat{x}_0(x_t))\|_2^2, \tag{13}$$

where $\rho \triangleq 1/\sigma^2$ represents the step size. Besides, $\nabla_{x_t} \|\mathcal{G} - \mathcal{M}(\hat{x}_0(x_t))\|_2^2$ also can be expressed as $\nabla \mathcal{L}_{\text{gradient}}$ [20]. Based on Eq. 6, the predicted noise adjusted for condition $\mathcal{G}$ can be expressed as:

$$\epsilon'_\phi = \epsilon_\phi(x_t, t) + \rho \sqrt{1 - \alpha_t} \nabla \mathcal{L}_{\text{gradient}}. \tag{14}$$

In summary, we impose reasonable constraints on the reverse procedure by summing the physical gradient with the intrinsic noise prediction. As shown in Fig. 3, such a manner prompts the diffusion model to search for the optimal solution along the manifold, thus avoiding the error accumulation caused by sample falling off the data manifold.

Gradient Guidance. Our gradient guidance consists of three parts, *i.e.*, thermo guidance $\mathcal{L}_\mathcal{T}$, illumination guidance $\mathcal{L}_\mathcal{I}$, and color guidance $\mathcal{L}_\mathcal{C}$.

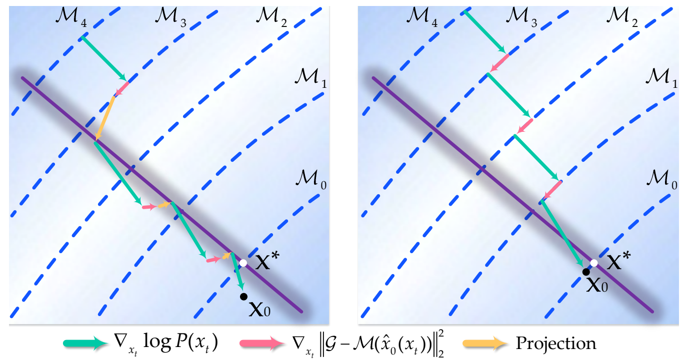

Fig. 3. Comparison of sampling trajectories between the projection strategy and SR-Diffusion.

$\mathcal{L}_\mathcal{T}$ aims to render consistent infrared characteristics between shadow and shadow-free regions. First, we employ a pre-trained modal translation network (U2Fusion [45], use $\mathcal{U}$ for short) to provide infrared appearances for SR-Diffusion and ground truth:

$$\mathbf{T} = \mathcal{U}(\mathbf{V}), \quad \mathbf{T}_{\text{GT}} = \mathcal{U}(\mathbf{V}_{\text{GT}}), \tag{15}$$

where $\mathbf{V}$ represents the output item of SR-Diffusion, *i.e.*, the shadow-free image. Besides, $\mathbf{V}_{\text{GT}}$ represents the ground truth image. Subsequently, we repair contrast and visibility through the mean squared error between $\mathbf{I}$ and $\mathbf{I}_{\text{GT}}$:

$$\mathcal{L}_\mathcal{T} = \|\mathbf{T}_{\text{GT}} - \mathbf{T}\|_2^2. \tag{16}$$

Existing color spaces are not ideal for shadow removal, this is because they fail to decouple luminance and color [22]. Therefore, we introduce the HVI color space [22], which consists of the HV-plane and the I-axis. Based on the Max-RGB theory, the I-axis can be expressed as:

$$\mathbf{I} = \max \mathbf{V}_c, \quad c \in \{R, G, B\} \tag{17}$$

where $\mathbf{I}$ corresponds to the illumination intensity. The HV-plane evolves from the Hue, Saturation and Value (HSV) space:

$$\begin{aligned} \mathbf{H} &= \mathcal{K} \odot \mathcal{S} \odot \mathcal{H}, \\ \mathbf{V} &= \mathcal{K} \odot \mathcal{S} \odot \mathcal{V}, \end{aligned} \tag{18}$$

where $\mathcal{S}$ represents the saturation map, $\mathcal{H}$ represents the hue map, and $\mathcal{V}$ represents the value map. Besides, $\mathcal{K}$ is the intensity collapse function that collapses low-light pixels to remove the black plane noise:

$$\mathcal{K} = \sqrt[k]{\sin(\frac{\pi \cdot \mathbf{I}_{\max}}{2})}, \tag{19}$$

where k represents the parameter to control the dark point density. The HV-plane inherits color information while removing sensitive noise, making it the best choice for shadow removal. Similar to $\mathcal{L}_\mathcal{T}$, we employ the mean square error to maintain illumination and color consistency:

$$\begin{aligned} \mathcal{L}_\mathcal{I} &= \|\mathbf{I}_{\text{GT}} - \mathbf{I}\|_2^2, \\ \mathcal{L}_\mathcal{C} &= \|\text{Cat}[\mathbf{H}_{\text{GT}}, \mathbf{V}_{\text{GT}}] - \text{Cat}[\mathbf{H}, \mathbf{V}]\|_2^2. \end{aligned} \tag{20}$$

Fig. 4. Qualitative comparison on UAV-SC [46].

where $\mathrm{Cat}[;]$ represents the dimension concatenation operator.

$$\mathcal{L} = \tau\mathcal{L}_{\mathcal{T}} + \upsilon\mathcal{L}_{\mathcal{I}} + \zeta\mathcal{L}_{\mathcal{C}}. \tag{21}$$

where τ, υ, and ζ are responsible for balancing the gradient guidance. Following [20], they are set to 0.2, 0.5, and 0.3, respectively. The incorporation of thermo, illumination, and color guidance infuses the diffusion procedure with diverse physical constraints for simultaneous optimization in visibility, luminance, and chromaticity domains.

IV. Experiment

A. Experimental Setting

Implementation. We implement SR-Diffusion through PyTorch. SR-Diffusion is trained with the Adam optimizer on an NVIDIA RTX A6000 GPU. The initial learning rate is set to 1×10^{-4} and subsequently stabilized to 1×10^{-7} through the cosine annealing scheme. In addition, a batch-mode learning scheme with a batch size of 32 is employed.

Benchmark. We perform qualitative and quantitative evaluation on the UAV-SC [46] dataset. UAV-SC consists of shadow, shadow-free, and shadow mask triplets, where 6924 triplets and 30 triplets are divided into training and testing sets, respectively. Notably, UAV-SC is captured from Wuhan through a drone platform. Therefore, the shadow samples are challenging because the location, intensity, shape, and size of cast shadow are variable.

Competitor. Three traditional shadow removal methods, including Guo [27], Gong [12], and Silva [47]. Nine deep learning-based shadow removal methods, including Mask-ShadowGAN [48], DC-ShadowNet [21], LG-ShadowNet [49], G2R-ShadowNet [50], ST-CGAN [51], DHAN [52], ShadowFormer [53], DMTN [19], and TBRNet [54]. Notably, these competitors retrained on UAV-SC and achieved the best quantitative scores.

Metric. For UAV-SC, we perform the full-reference evaluation employing the root mean square error (RMSE) in Lab color space and the peak signal-to-noise ratio (PSNR)

TABLE I

QUANTITATIVE COMPARISONS ON UAV-SC [46]. "↑" REPRESENTS THAT THE LARGER SCORE IS BETTER, WHILE "↓" REPRESENTS THAT THE LOWER SCORE IS BETTER. THE BEST SCORE IS IN RED, THE SECOND-BEST SCORE IS IN BLUE, AND THE THIRD-BEST SCORE IS IN GREEN.

Methods	RMSE(↓)			PSNR(↑)			SSIM(↑)		
	S.	N.S.	All	S.	N.S.	All	S.	N.S.	All
Guo [27]	29.8381	17.9412	20.3919	23.5991	18.3971	16.6687	0.9042	0.8093	0.6945
Gong [12]	26.0850	18.8162	20.3155	25.1448	18.1944	16.8290	0.9179	0.7926	0.6984
Silva [47]	32.0696	18.5089	21.3024	22.8902	18.1034	16.2621	0.8899	0.8035	0.6898
Mask-ShadowGAN [48]	19.7786	17.5036	17.9722	27.5729	20.4507	19.1357	0.9450	0.8142	0.7374
DC-ShadowNet [21]	16.1495	13.6091	14.1324	29.2519	22.4207	21.0712	0.9543	0.8479	0.7852
LG-ShadowNet [49]	23.0353	16.6409	17.9581	25.4526	19.6434	18.0556	0.9352	0.8352	0.7629
G2R-ShadowNet [50]	22.0473	18.6220	19.3276	24.6951	17.9899	16.8325	0.9015	0.8189	0.7183
ST-CGAN [51]	9.4592	9.2458	9.2898	32.4301	24.8846	23.8219	0.9563	0.8679	0.8317
DHAN [52]	9.1524	9.0602	9.1086	33.0654	25.1760	24.1399	0.9752	0.9010	0.8620
ShadowFormer [53]	9.6703	9.2937	9.5927	32.7140	24.3702	23.4432	0.9711	0.8804	0.8350
DMTN [19]	11.5195	9.9282	10.2560	31.6098	24.1577	23.0287	0.9664	0.8696	0.8195
TBRNet [54]	11.4115	10.8361	10.9546	31.2298	23.6601	22.5798	0.9619	0.8313	0.7722
SR-Diffusion	9.0891	8.9012	8.9560	33.3171	25.8113	24.5611	0.9791	0.9153	0.8815

TABLE II

QUANTITATIVE COMPARISONS ON LOL [55] AND UIEB [68]. "↑" REPRESENTS THAT THE LARGER SCORE IS BETTER, WHILE "↓" REPRESENTS THAT THE LOWER SCORE IS BETTER. THE BEST SCORE IS IN RED, THE SECOND-BEST SCORE IS IN BLUE, AND THE THIRD-BEST SCORE IS IN GREEN.

Methods	LOL		Methods	UIEB			
	PSNR(↑)	SSIM(↑)		PSNR(↑)	SSIM(↑)	UCIQE(↑)	UIQM(↑)
EnlightenGAN [56]	18.9005	0.7627	IBLA [69]	13.8129	0.3712	0.6277	1.5450
RUAS [57]	16.4792	0.6083	GDCP [70]	14.0055	0.3625	0.6253	1.5351
Zero-DCE [58]	15.2586	0.5367	Fusion [71]	15.1764	0.4382	0.5930	1.4982
Zero-Restore [59]	12.6676	0.3371	Haze-lines [72]	14.5129	0.4366	0.6566	1.6313
DUNP [60]	13.8080	0.4772	ERH [73]	14.3570	0.4021	0.5402	1.4652
SCI [61]	14.7617	0.5525	MMLE [74]	13.5866	0.3655	0.6153	1.8579
SGZ [62]	14.9161	0.4264	Water-Net [68]	15.8373	0.4302	0.5943	1.4814
NeRCo [63]	15.0944	0.5376	Ucolor [75]	16.4070	0.4637	0.5673	1.3558
LLFormer [64]	23.9799	0.8305	TOPAL [76]	14.5744	0.4244	0.5701	1.4660
PairLIE [65]	18.4691	0.7567	UICoE-Net [77]	15.3769	0.6023	0.5718	1.4790
RQ-LLIE [66]	25.5871	0.8521	TACL [78]	17.2567	0.5154	0.6279	1.5981
Zero-IG [67]	18.1603	0.6198	CLUIE-Net [79]	14.2944	0.4061	0.5869	1.5363
SR-Diffusion	26.2613	0.8609	SR-Diffusion	17.8531	0.6511	0.6264	1.6035

and the structural similarity (SSIM) in RGB color space, respectively. A lower RMSE score represents better shadow removal performance, while the opposite is true for PSNR and SSIM. Besides, the full-reference evaluation is accomplished on the shadow region (S.), the shadow-free region (N.S.), and the whole sample (All).

B. Qualitative Comparison

We display a visual comparison on UAV-SC in Fig. 4. Guo [27], Gong [12], Silva [47], Mask-ShadowGAN [48], DC-ShadowNet [21], LG-ShadowNet [49], and G2R-ShadowNet [50] are helpless against both large-area umbra shadows and small-scale cast shadows. Some of them even introduce over-exposure in the shadow samples #1 and #25 (*e.g.*, Gong [12]) and artifacts in the shadow samples #20 and #24 (*e.g.*, Mask-ShadowGAN [48], DC-ShadowNet [21], and LG-ShadowNet [49]). The main reason is that Guo [27], Gong [12], and Silva [47] locate shadow pixels through illumination conditions, user labels, and spectrum ratios, respectively. Unfortunately, these prior assumptions require accurate parameter settings, which fail to meet the requirements of the real world. Besides, unsupervised or self-supervised methods can remove shadows under cycle-consistency constraints, but they fail to learn the underlying relationship between shadow and shadow-free domains. Therefore, their performance is unsatisfactory. Although DMTN [19] and TBRNet [54] provide superior shadow removal performance, they produce streaky patches in the shadow sample #20. This obviously changes the scenario content and is unacceptable for geographical object observation. This is because the multi-branch structure of DMTN [19] and TBRNet [54] fails to reasonably restore the illumination intensity. For the shadow sample #20, ShadowFormer [53] misidentifies the roof as shadows, leading to a color deviation of the bottom-left roof. This is because ShadowFormer [53] locates shadow regions through Retinex theory, which lacks robustness. Additionally, ST-CGAN [51] and DHAN [52] retain tiny shadow remnants towards the shadow sample #20. The main reason is that they employ dilated convolutions

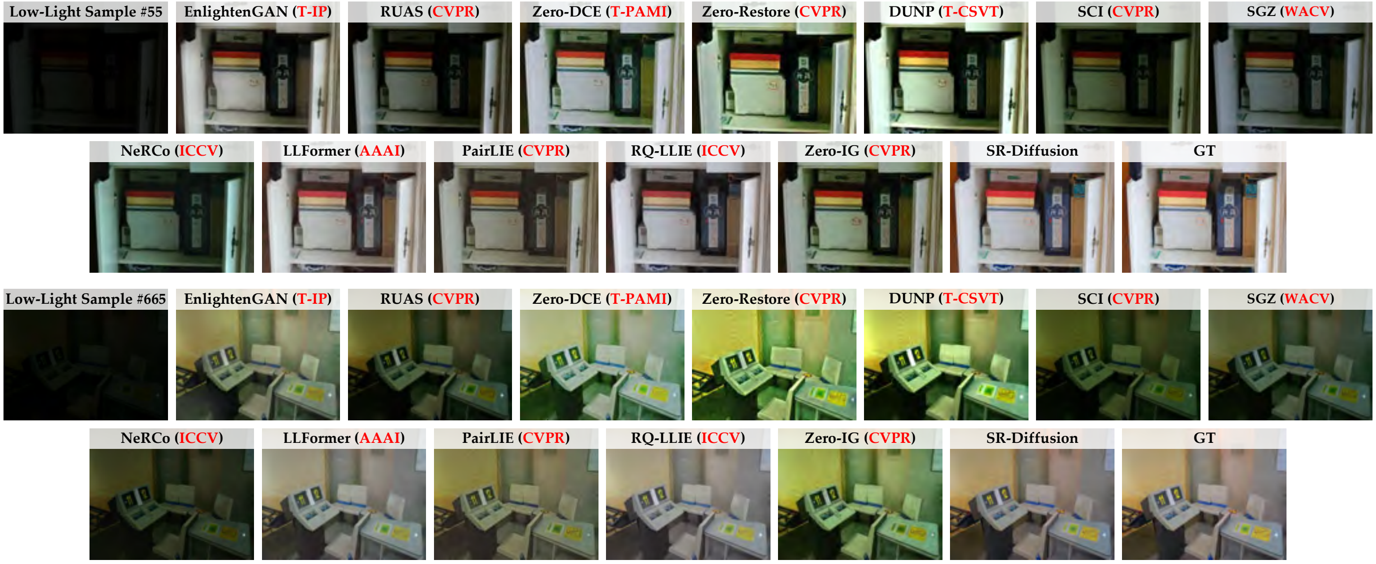

Fig. 5. Qualitative comparison on LOL [55].

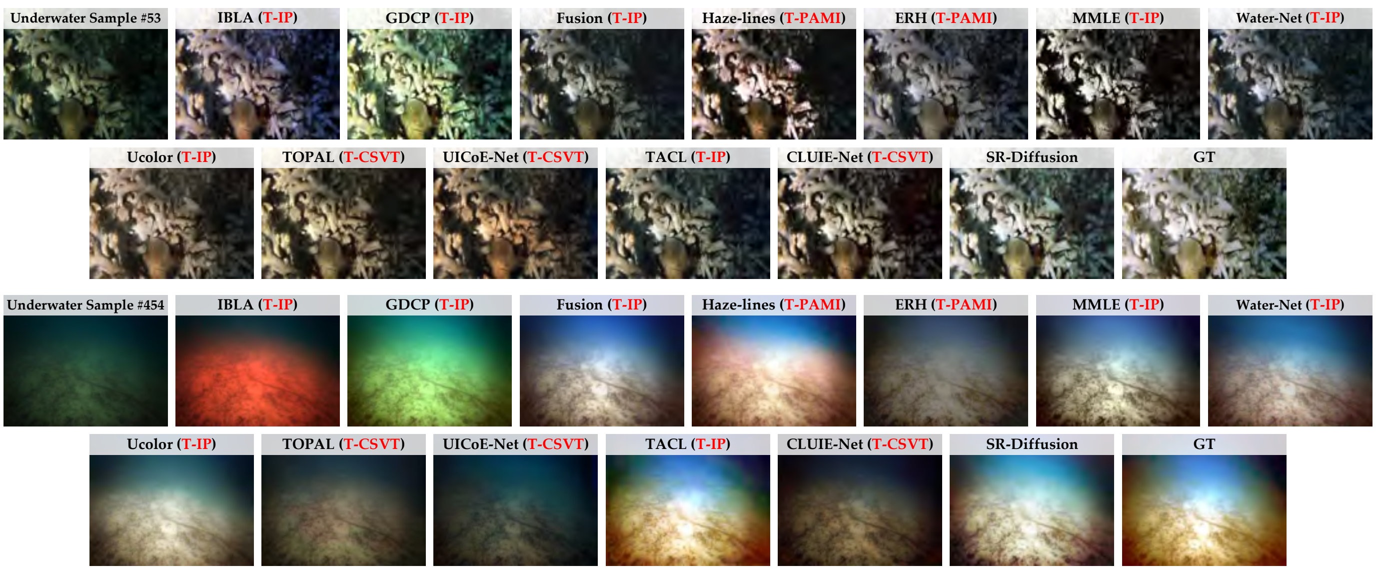

Fig. 6. Qualitative comparison on UIEB [68].

as the backbone, thereby neglecting details. In contrast, SR-Diffusion conquers shadow contamination with diverse intensities, shapes, and sizes, demonstrating convincing effectiveness.

C. Quantitative Comparison

We report the RMSE, PSNR, and SSIM scores of SR-Diffusion and competitors in Table I. Our method achieves the best full-reference scores on UAC-SC. Compared with the top-performing competitor DHAN [52], the RMSE, PSNR, and SSIM scores of SR-Diffusion achieve the percentage gain of 0.7%/ 0.8%/ 0.4% in shadow regions, the percentage gain of 1.8%/ 2.5%/ 1.8% in shadow-free regions, and the percentage gain of 1.7%/ 1.7%/ 2.3% in whole images. Therefore, the precise numerical evaluation likewise demonstrates the superiority of SR-Diffusion. This is credited to the introduction of triple physical constraints. We achieve excellent shadow removal performance by compressing the differences in heat, luminance, and color between shadow and shadow-free domains during the diffusion process. Besides, such a manner provides interpretable constraints for the diffusion process.

D. Scenario Adaptability

To demonstrate the scenario adaptability of SR-Diffusion, we generalize it to the natural scene and the underwater scene. Likewise, the competitors retrained on LOL and UIEB, respectively, and achieved the best quantitative scores.

Natural Scene. Qualitative and quantitative comparisons of the natural scene are performed on LOL [55]. The competitors are EnlightenGAN [56], RUAS [57], Zero-DCE [58], Zero-Restore [59], DUNP [60], SCI [61], SGZ [62], NeRCo [63], LLFormer [64], PairLIE [65], RQ-LLIE [66], and Zero-IG [67]. In Fig. 5, SCI [61], SGZ [62], and NeRCo [63] fail to

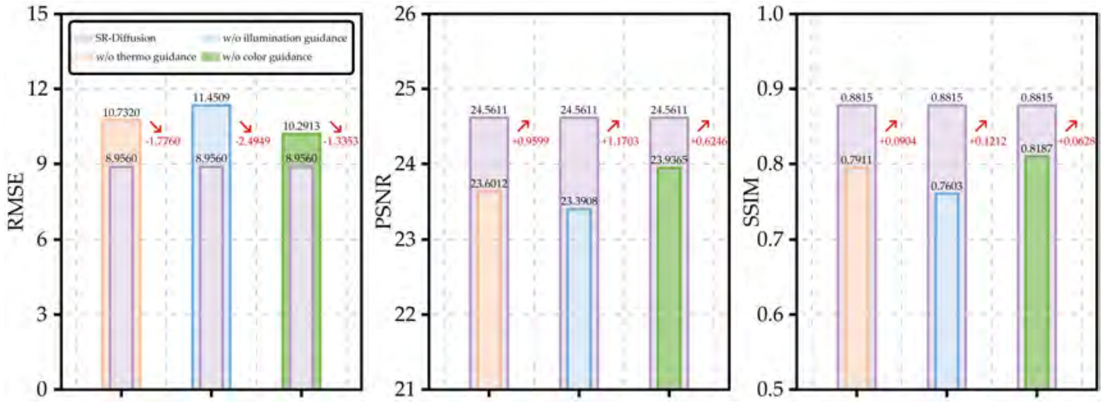

Fig. 7. Quantitative score of the ablation study on UAV-SC [46].

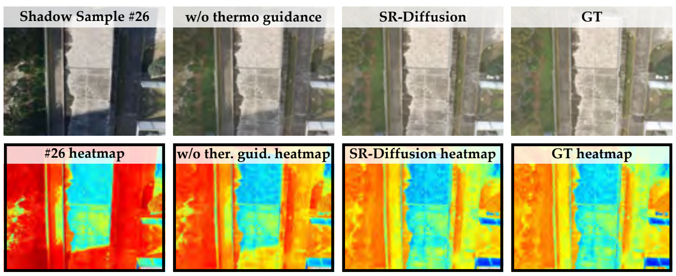

Fig. 8. Ablation study toward the thermo guidance.

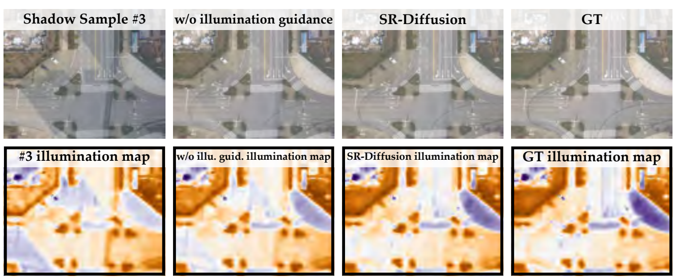

Fig. 9. Ablation study toward the illumination guidance.

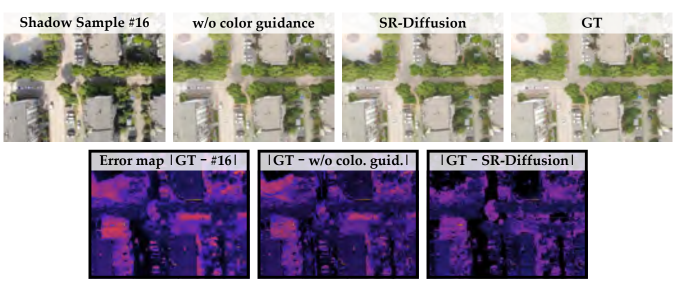

Fig. 10. Ablation study toward the color guidance.

cope with the low-light issue. For low-light samples #55 and #665, Zero-Restore [59], DUNP [60], and Zero-IG [67] introduce obvious greenish color deviations. EnlightenGAN [56], RUAS [57], Zero-DCE [58], and PairLIE [65] introduce undesired noise perturbations, which are particularly noticeable in the low-light sample #665. LLFormer [64], RQ-LLIE [66], and SR-Diffusion demonstrate convincing illumination recovery capability, but our results are closest to ground truthes in terms of color. Besides, the quantitative comparison is shown in Table II. Compared with the top-performing competitor RQ-LLIE [66], SR-Diffusion accomplishes the percentage gain of 2.6%/ 1.0% in terms of PSNR/ SSIM. Therefore, we are able to seamlessly generalize SR-Diffusion to the natural scene.

Underwater Scene. Qualitative and quantitative comparisons of the underwater scene are performed on UIEB [68]. Notably, we select 30 low-light samples from UIEB as a testing set, so the quantitative scores differ from other works. The competitors are IBLA [69], GDCP [70], Fusion [71], Haze-lines [72], ERH [73], MMLE [74], Water-Net [68], Ucolor [75], TOPAL [76], UICoE-Net [77], TACL [78], and CLUIE-Net [79]. As shown in Fig. 6, IBLA [69], Haze-lines [72], and TACL [78] introduce reddish color deviations in the underwater sample #454, this is because the red channel is overcompensated. Similarly, GDCP [70] introduces greenish color deviations that are likewise unacceptable. The main reason is that the green channel is overcompensated. Fusion [71], MMLE [74], Water-Net [68], and Ucolor [75] are able to improve the luminance of the underwater sample #454, but fail to cope with the other samples. The other competitors compromise on the low-light issue. In contrast, our method improves illumination while maintaining a realistic tone. This is because we improve color deviation in the hue space, avoiding the high color sensitivity of the sRGB space. Besides, the quantitative comparison is shown in Table II. SR-Dffusion achieves two best scores and one third-best score. Since both UCIQE [80] and UIQM [29] employ colorfulness in evaluating image quality, images with color casts typically have higher scores. Therefore, we can still conclude that SR-Diffusion has excellent scenario adaptability.

E. Ablation Study

In this subsection, we explore the contribution of the gradient guidance.

Influence of the thermo guidance. As shown in Fig. 8, the heatmap of SR-Diffusion is closer to the ground truth in shadow regions, compared with the ablated model w/o thermo guidance. This benefits from the thermal spectrum of the infrared modality guiding the contrast and visibility interactions between shadow and shadow-free regions. In addition, as shown in Fig. 7, SR-Diffusion achieves the percentage gain of 16.5%/ 4.1%/ 11.4% in terms of RMSE/ PSNR/ SSIM, compared with the ablated model w/o thermo guidance. This likewise suggests that injecting thermo guidance into the reverse procedure is critical.

Influence of the illumination guidance. In Fig. 9, we employ the illumination visualization to reflect the luminance intensity. With the guidance of the illumination map, SR-Diffusion is able to understand the degradation degree of shadowed regions, thus relighting shadow pixels. In contrast, the ablated model w/o illumination guidance suffers from shadow remnants, which are obvious in the illumination map. Besides, SR-Diffusion achieves the percentage gain of 21.8%/ 4.8%/ 13.7% in terms of RMSE/ PSNR/ SSIM, compared with the ablated model w/o illumination guidance. Therefore, the design of the illumination guidance is reasonable and effective.

Influence of the color guidance. In Fig. 10, visualize the color difference through the error map. By ignoring the hue and saturation of the HV-plane, the ablated model w/o color guidance produces non-consistent tones between shadow and shadow-free regions, and therefore has a large error value. In contrast, SR-Diffusion removes shadow traces while maintaining the realistic tone. Meanwhile, SR-Diffusion achieves the percentage gain of 13.0%/ 2.5%/ 7.1% in terms of RMSE/ PSNR/ SSIM, compared with the ablated model w/o color guidance. Therefore, the effectiveness of the color guidance has been demonstrated.

V. Conclusion

In this paper, we propose a physical property-driven shadow removal diffusion model. Specifically, we employ thermodynamic, optical, and chromatological information as guidance, thus ensuring the prior transfer of visibility, luminance, and chroma between shadow and shadow-free regions. Such a manner optimizes the step-wise iteration of the diffusion process, refining the model at each denoising step and accomplishing a nontrivial shadow removal performance. In addition, SR-Diffusion is able to be generalized to multiple scenarios, which further enhances its utility.

References

[1] C. Zhang, K.-M. Lam, T. Liu, Y.-L. Chan, and Q. Wang, "Structured adversarial self-supervised learning for robust object detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–20, Mar. 2024.

[2] C. Zhang, J. Su, Y. Ju, K.-M. Lam, and Q. Wang, "Efficient inductive vision transformer for oriented object detection in remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–20, Jul. 2023.

[3] C. Zhang, J. Xiao, C. Yang, J. Zhou, K.-M. Lam, and Q. Wang, "Integrally mixing pyramid representations for anchor-free object detection in aerial imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, May. 2024.

[4] W. Jing, K. Chi, Q. Li, and Q. Wang, "3-D neighborhood cross-differencing: A new paradigm serves remote sensing change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–11, Jul. 2024.

[5] Z. Li, Z. Yuan, L. Li, D. Liu, X. Tang, and F. Wu, "Object segmentation-assisted inter prediction for versatile video coding," *IEEE Trans. Broadcast.*, vol. 70, no. 4, pp. 1236–1253, Dec. 2024.

[6] H. Guo, J. Gao, and Y. Yuan, "Balanced density regression network for remote sensing object counting," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–13, May. 2024.

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

[8] W. Zhang, X. Zhao, J.-M. Morvan, and L. Chen, "Improving shadow suppression for illumination robust face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 3, pp. 611–624, Mar. 2019.

[9] K. He, R. Zhen, J. Yan, and Y. Ge, "Single-image shadow removal using 3D intensity surface modeling," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 6046–6060, Dec. 2017.

[10] C. Tang, Z. Li, Y. Bian, L. Li, and D. Liu, "Neural video compression with context modulation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2025, pp. 12553–12563.

[11] G. D. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew, "On the removal of shadows from images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 1, pp. 59–68, Jan. 2006.

[12] H. Gong and D. Cosker, "Interactive shadow removal and ground truth for variable scene categories," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2014, pp. 1–11.

[13] T. Hu, Q. Yan, Y. Qi, and Y. Zhang, "Generating content for HDR deghosting from frequency view," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 25732–25741.

[14] J. Xiao *et al.*, "HomoFormer: Homogenized transformer for image shadow removal," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 25617–25626.

[15] Y. Zhu, J. Huang, X. Fu, F. Zhao, Q. Sun, and Z.-J. Zha, "Bijective mapping network for shadow removal," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 5617–5626.

[16] Z. Xiao, Z. Li, and W. Jia, "Occlusion-embedded hybrid transformer for light field super-resolution," in *Proc. AAAI Conf. Artif. Intell.*, vol. 39, no. 8, 2025, pp. 8700–8708.

[17] K. Chi, S. Guo, J. Chu, Q. Li, and Q. Wang, "RSMamba: Biologically plausible retinex-based Mamba for remote sensing shadow removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, pp. 1–10, Jan. 2025.

[18] K. Chi, J. Li, W. Jing, Q. Li, and Q. Wang, "Neural implicit fourier transform for remote sensing shadow removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–10, Jun. 2024.

[19] J. Liu, Q. Wang, H. Fan, W. Li, L. Qu, and Y. Tang, "A decoupled multi-task network for shadow removal," *IEEE Trans. Multimedia*, vol. 25, pp. 9449–9463, Mar. 2023.

[20] X. Li *et al.*, "DifIISR: A diffusion model with gradient guidance for infrared image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2025, pp. 7534–7544.

[21] Y. Jin, A. Sharma, and R. T. Tan, "DC-ShadowNet: Single-image hard and soft shadow removal using unsupervised domain-classifier guided network," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 5007–5016.

[22] Q. Yan *et al.*, "HVI: A new color space for low-light image enhancement," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2025, pp. 5678–5687.

[23] Z. Li, J. Li, Y. Li, L. Li, D. Liu, and F. Wu, "In-loop filtering via trained look-up tables," in *Proc. IEEE Int. Conf. Vis. Commun. Image Process. (VCIP)*, Dec. 2024, pp. 1–5.

[24] Y. Xu, S. Huang, H. Zhang, and X. Li, "Why does dropping edges usually outperform adding edges in graph contrastive learning?," in *Proc. AAAI Conf. Artif. Intell.*, Feb. 2025, pp. 21824–21832.

[25] F. Liu and M. Gleicher, "Texture-consistent shadow removal," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2008, pp. 437–450.

[26] E. Arbel and H. Hel-Or, "Shadow removal using intensity surfaces and texture anchor points," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 6, pp. 1202–1216, Jun. 2011.

[27] R. Guo, Q. Dai, and D. Hoiem, "Paired regions for shadow detection and removal," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 2956–2967, Dec. 2013.

[28] Q. Yang, K.-H. Tan, and N. Ahuja, "Shadow removal using bilateral filtering," *IEEE Trans. Image Process.*, vol. 21, no. 10, pp. 4361–4368, Oct. 2012.

[29] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE J. Ocean. Eng.*, vol. 41, no. 3, pp. 541–551, Jul. 2016.

[30] L. Guo *et al.*, "ShadowDiffusion: When degradation prior meets diffusion model for shadow removal," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 14049–14058.

[31] J. Yu *et al.*, "Portrait shadow removal using context-aware illumination restoration network," *IEEE Trans. Image Process.*, vol. 34, pp. 1–15, Jan. 2015.

[32] Z. Chen, L. Wan, Y. Xiao, L. Zhu, and H. Fu, "Learning physical-spatio-temporal features for video shadow removal," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 7, pp. 5830–5842, Jul. 2024.

[33] Q. Wang, K. Chi, W. Jing, and Y. Yuan, "Recreating brightness from remote sensing shadow appearance," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–11, May. 2024.
[34] Z. Chen *et al.*, "Learning spatial adaptation and temporal coherence in diffusion models for video super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 9232–9241.
[35] G. Li, C. Rao, J. Mo, Z. Zhang, W. Xing, and L. Zhao, "Rethinking diffusion model for multi-contrast MRI super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 11365–11374.
[36] H. Liu, Y. Wang, B. Qian, M. Wang, and Y. Rui, "Structure matters: Tackling the semantic discrepancy in diffusion models for image inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 8038–8047.
[37] J. Li *et al.*, "Light the night: A multi-condition diffusion framework for unpaired low-light enhancement in autonomous driving," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 15205–15215.
[38] X. Lv *et al.*, "Fourier priors-guided diffusion for zero-shot joint low-light enhancement and deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 25378–25388.
[39] C. Zhao, W. Cai, C. Dong, and C. Hu, "Wavelet-based fourier information interaction with frequency diffusion adjustment for underwater image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 8281–8291.
[40] Y. Zhang, J. Yuan, and Z. Cai, "DCGF: Diffusion-color-guided framework for underwater image enhancement," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, pp. 1–12, Jan. 2025.
[41] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Dec. 2020, pp. 6840–6851.
[42] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, May. 2021, pp. 1–12.
[43] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye, "Diffusion posterior sampling for general noisy inverse problems," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, May. 2023, pp. 1–12.
[44] S. Simic, "On a global upper bound for jensen's inequality," *J. Math. Anal. Appl.*, vol. 343, no. 1, pp. 414–419, Jul. 2008.
[45] H. Xu, J. Ma, J. Jiang, X. Guo, and H. Ling, "U2Fusion: A unified unsupervised image fusion network," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 502–518, Jan. 2022.
[46] S. Luo, H. Li, Y. Li, C. Shao, H. Shen, and L. Zhang, "An evolutionary shadow correction network and a benchmark UAV dataset for remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–14, Jul. 2023.
[47] G. F. Silva, G. B. Carneiro, R. Doth, L. A. Amaral, and D. F. G. de Azevedo, "Near real-time shadow detection and removal in aerial motion imagery application," *ISPRS J. Photogramm. Remote Sens.*, vol. 140, pp. 104–121, Jun. 2018.
[48] X. Hu, Y. Jiang, C.-W., Fu, and P.-A. Heng, "Mask-ShadowGAN: Learning to remove shadows from unpaired data," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2019, pp. 2472–2481.
[49] Z. Liu, H. Yin, Y. Mi, M. Pu, and S. Wang, "Shadow removal by a lightness-guided network with training on unpaired data," *IEEE Trans. Image Process.*, vol. 31, pp. 1853–1865, Jan. 2021.
[50] Z. Liu, H. Yin, X. Wu, Z. Wu, Y. Mi, and S. Wang, "From shadow generation to shadow removal," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 4925–4934.
[51] J. Wang, X. Li, and J. Yang, "Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 1788–1797.
[52] X. Cun, C.-M. Pun, and C. Shi, "Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting GAN," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 7, 2020, pp. 10680–10687.
[53] L. Guo, S. Huang, D. Liu, C. Hao, and B. Wen, "ShadowFormer: Global context helps shadow removal," in *Proc. AAAI Conf. Artif. Intell.*, vol. 37, no. 1, 2023, pp. 710–718.
[54] J. Liu, Q. Wang, H. Fan, J. Tian, and Y. Tang, "A shadow imaging bilinear model and three-branch residual network for shadow removal," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 11, pp. 15857–15871, Nov. 2024.
[55] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2018, pp. 1–12.
[56] Y. Jiang *et al.*, "EnlightenGAN: Deep light enhancement without paired supervision," *IEEE Trans. Image Process.*, vol. 30, pp. 2340–2349, Jan. 2021.
[57] R. Liu, L. Ma, J. Zhang, X. Fan, and Z. Luo, "Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10556–10565.
[58] C. Li, C. Guo, and C. C. Loy, "Learning to enhance low-light image via zero-reference deep curve estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 8, pp. 4225–4238, Aug. 2022.
[59] A. Kar, S. K. Dhara, D. Sen, and P. K. Biswas, "Zero-shot single image restoration through controlled perturbation of koschmieder's model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 16200–16210.
[60] J. Liang, Y. Xu, Y. Quan, B. Shi, and H. Ji, "Self-supervised low-light image enhancement using discrepant untrained network priors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 11, pp. 7332–7345, Nov. 2022.
[61] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward fast, flexible, and robust low-light image enhancement," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5627–5636.
[62] S. Zheng and G. Gupta, "Semantic-guided zero-shot learning for low-light image/video enhancement," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops (WACVW)*, Jan. 2022, pp. 581–590.
[63] S. Yang, M. Ding, Y. Wu, Z. Li, and J. Zhang, "Implicit neural representation for cooperative low-light image enhancement," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 12872–12881.
[64] T. Wang, K. Zhang, T. Shen, W. Luo, B. Stenger, and T. Lu, "Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method," in *Proc. AAAI Conf. Artif. Intell.*, Feb. 2023, pp. 2654–2662.
[65] Z. Fu, Y. Yang, X. Tu, Y. Huang, X. Ding, and K.-K. Ma, "Learning a simple low-light image enhancer from paired low-light instances," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 22252–22261.
[66] Y. Liu, T. Huang, W. Dong, F. Wu, X. Li, and G. Shi, "Low-light image enhancement with multi-stage residue quantization and brightness-aware attention," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 12106–12115.
[67] Y. Shi, D. Liu, L. Zhang, Y. Tian, X. Xia, and X. Fu, "ZERO-IG: Zero-shot illumination-guided joint denoising and adaptive enhancement for low-light images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 3015–3024.
[68] C. Li *et al.*, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, Feb. 2020.
[69] Y.-T. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1579–1594, Apr. 2017.
[70] Y.-T. Peng, K. Cao, and P. C. Cosman, "Generalization of the dark channel prior for single image restoration," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2856–2868, Jun. 2018.
[71] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 379–393, Jan. 2018.
[72] D. Berman, D. Levy, S. Avidan, and T. Treibitz, "Underwater single image color restoration using haze-lines and a new quantitative dataset," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 8, pp. 2822–2837, Aug. 2021.
[73] H. Song, L. Chang, Z. Chen, and P. Ren, "Enhancement-registration-homogenization (ERH): A comprehensive underwater visual reconstruction paradigm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6953–6967, Oct. 2022.
[74] W. Zhang, P. Zhuang, H.-H. Sun, G. Li, S. Kwong, and C. Li, "Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement," *IEEE Trans. Image Process.*, vol. 31, pp. 3997–4010, Jun. 2022.
[75] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, "Underwater image enhancement via medium transmission-guided multi-color space embedding," *IEEE Trans. Image Process.*, vol. 30, pp. 4985–5000, May. 2021.
[76] Z. Jiang, Z. Li, S. Yang, X. Fan, and R. Liu, "Target oriented perceptual adversarial fusion network for underwater image enhancement," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 10, pp. 6584–6598, Oct. 2022.
[77] Q. Qi *et al.*, "Underwater image co-enhancement with correlation feature matching and joint learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1133–1147, Mar. 2022.
[78] R. Liu, Z. Jiang, S. Yang, and X. Fan, "Twin adversarial contrastive learning for underwater image enhancement and beyond," *IEEE Trans. Image Process.*, vol. 31, pp. 4922–4936, Jul. 2022.

[79] K. Li *et al.*, "Beyond single reference for training: Underwater image enhancement via comparative learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 6, pp. 2561–2576, Jun. 2023.
[80] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6062–6071, Dec. 2015.
[81] Z. Xiao and Z. Xiong, "Incorporating degradation estimation in light field spatial super-resolution," *Comput. Vis. Image Underst.*, vol. 252, pp. 104295, Feb. 2025.
[82] Y. Liu, Z. Xiong, Y. Yuan, and Q. Wang, "Transcending pixels: Boosting saliency detection via scene understanding from aerial imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–16, Jul. 2023.
[83] Z. Xiao, D. Kai, Y. Zhang, X. Sun, and Z. Xiong, "Asymmetric event-guided video super-resolution," in *Proc. ACM Int. Conf. Multimedia*, Oct. 2024, pp. 2409–2418.
[84] Z. Xiao and X. Wang, "Event-based video super-resolution via state space models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 12564–12574.
[85] Z. Xiao, Z. Lu, and X. Wang, "P-bic: Ultra-high-definition image moiré patterns removal via patch bilateral compensation," in *Proc. ACM Int. Conf. Multimedia*, Oct. 2024, pp. 8365–8373.
[86] H. Zhang, S. Huang, Y. Guo, and X. Li, "Variational positive-incentive noise: How noise benefits models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 47, no. 9, pp. 8313–8320, Sep. 2025.
[87] H. Zhang, Y. Zhu, and X. Li, "Decouple graph neural networks: Train multiple simple GNNs simultaneously instead of one," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 11, pp. 7451–7462, Nov. 2024.
[88] H. Zhang, Y. Zhu, and X. Li, "Toward projected clustering with aggregated mapping," *IEEE Trans. Image Process.*, vol. 32, pp. 4103–4113, Jul. 2023.

Kaichen Chi received the B.E. degree in electronic and information engineering and the M.E. degree in communication and information system from Liaoning Technical University, Huludao, China, in 2019 and 2022 respectively. He is currently working toward the Ph.D. degree in the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. His research interests include image processing and deep learning.

Junjie Li received the B.E. degree in software engineering from Zhengzhou University, Zhengzhou, China, in 2024. He is currently working toward the M.S. degree in the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision, pattern recognition and remote sensing.

Jun Chu received the B.E. degree in automation from Northwestern Polytechnical University, Xi'an, China, in 2024. He is currently pursuing the M.S. degree with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. His research interests include deep learning and computer vision.

Qiang Li (Member, IEEE) is currently with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University. His research interests include remote sensing image processing, particularly for image quality enhancement, object/change detection.

Qi Wang (Senior Member, IEEE) received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively. He is currently a Professor with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision, pattern recognition and remote sensing. For more information, visit the link (https://crabwq.github.io/).