RSMamba: Biologically Plausible Retinex-Based Mamba for Remote Sensing Shadow Removal

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

***Abstract*—Shadow removal is an essential task for remote sensing imagery analysis, which is tricky due to spatial irregular and inhomogeneous degradation distribution. Unfortunately, current shadow removal pipelines face challenges with suboptimal performance and insufficient interpretability. To this end, we unleash the long-sequence modeling potential of State Space Models (SSMs) in the context of shadow removal. Coupled with the accurate perception of traditional Retinex decomposition towards illumination, the well-designed RSMamba enjoys the best of both worlds between superior competitiveness and theoretical intuitiveness. Specifically, RSMamba mimics the retina and cerebral cortex to explore illumination and reflectance. The former drives the selective scan mechanism to enhance the response towards contamination, while the latter serves as a tool to preserve illumination fidelity. In addition, contour and gradient regularizations of illumination and reflectance components reflect the spatial opponency of shadows, which are consistent with the center-surround opponent receptive field of the human visual system. Such a manner incorporates the domain knowledge of neurophysiological mechanisms into neural networks, providing new insights into shadow removal. Extensive experiments demonstrate that RSMamba outperforms state-of-the-art methods.**

***Index Terms*—Shadow removal, Mamba, Retinex, biologically inspired vision.**

I. Introduction

SHADOWS are ubiquitous physical phenomenon, typically formed due to illumination occlusion [1], [49]–[52], [63]. Unfortunately, such undesired visual degradation inevitably brings great challenges to downstream tasks, such as object detection [2], [54]–[57], target segmentation [3], object counting [4], and classification [62]. Thus, shadow removal is a crucial research topic in graphics.

Conventional cognition methods rely on hand-crafted priors (such as gradient [5], morphology [6], and user interaction [7]) to analyze illumination statistics, thus progressively detecting and removing shadows. Nevertheless, they typically invalid when prior assumptions cannot be met. With the development of deep learning, pioneering researches have been explored from different perspectives, such as multiple task decoupling [8], shadow generation [9], exposure fusion [10], imagery decomposition [11], and diffusion [60], [61].

This work was supported in part by the National Natural Science Foundation of China under Grant 62301385, Grant 62471394, and Grant U21B2041, and in part by the Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University, PR China 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, saiguo@mail.nwpu.edu.cn, junchu@mail.nwpu.edu.cn, liqmges@gmail.com, crabwq@gmail.com).

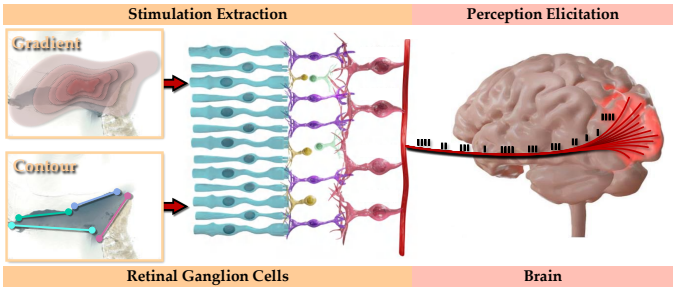

Fig. 1. Schematic illustration of our basic idea. Physical properties of shadows, *i.e.*, contours and gradients, trigger spike trains of retinal ganglion cells, thus activating the occipital lobe to perceive the location, intensity, shape, and size of shadows. Such a manner is consistent with neural networks learning the most discriminative representations from shadows.

However, these methods directly leverage convolutional neural networks (CNNs) or Transformer to learn a brute-force mapping function from shadow version to corresponding shadow-free version, thus ignoring interpretability and theoretically proven properties. More importantly, the limited receptive field or quadratic computational overhead leads to unsatisfactory global representation and unacceptable computational demands.

Recent popular Mamba [12], a novel State Space Model heralds a new era of win-win model accuracy and computational efficiency. On the one hand, Mamba accomplishes promising performance in dense data domains through selection scanning mechanism and hardware optimization. On the other hand, the computational complexity of Mamba is linearly related to the number of tokens, compared with quadratic computational overhead for Transformer [13]. Nevertheless, Mamba falls short in interpretability. Retinex theory could be a practical answer to this question. Physiological studies indicate that the image exits as contrast in the human visual system [14]. In particular, illumination is discontinuous on shadow lines, rendering a striking contrast. Coincidentally, Retinex theory simulates such center-surround difference through lightness and color perception [15]. In summary, a natural question arises: can Retinex theory be integrated into Mamba to simulate the human visual system capturing and understanding shadows?

In this work, we introduce a purely data-driven Retinex decoupling network to split shadow samples into illumination and reflectance, which escapes the highly ill-posed constraints from hand-crafted priors. Illumination, as a pre-processing condition, is incorporated into the state space sequence to relight dark pixels. Besides, since illumination gradient describes

the global variation of lightness while reflectance contour reflects the point-wise variation of the spectral surface [15], we suggest both as regularization terms to maintain illumination consistency and texture richness. Such a manner coincides with the visual cortex (retinal cells to curvature cells to shape-sensitive cells) capturing shadow physical properties in terms of gradients and contours, as depicted in Fig. 1. In summary, the main contributions are as follows:

- **Perspective contribution.** We rethink the shadow removal task from the perspective of retinal perception formation, simulating stimulus encoding in the visual system toward illumination and reflectance aberrations to remove shadow traces. Such a manner bridges the gap between incomprehensible black box and biologically plausible explanation.
- **Technical contribution.** We propose the Retinex-guided Mamba for remote sensing imagery shadow removal, liberating the visually pleasing quality from complex shadow distributions through contour and gradient regularizations. To our best knowledge, this is the first attempt to remove shadows via Mamba.
- **Practical contribution.** RSMamba achieves the superior performance in terms of visual quality, inference efficiency, and computational cost.

II. Related Work

In this section, we review Retinex theory and State Space Models, and then discuss representative shadow removal methods, both traditional and deep learning-based.

A. Retinex Theory

According to Retinex theory [16], a shadow sample $\mathcal{S} \in \mathbb{R}^{H\times W\times 3}$ is able to be decoupled into an illumination component $\mathcal{L} \in \mathbb{R}^{H\times W}$ and a reflectance component $\mathcal{R} \in \mathbb{R}^{H\times W\times 3}$:

$$\mathcal{S} = \mathcal{R} \circ \mathcal{L}, \tag{1}$$

where $\circ$ represents the element-wise multiplication, $\mathcal{L}$ reflects the intensity and distribution of illumination, and $\mathcal{R}$ reflects the inherent properties of the object, such as texture and color. Notably, the consistent scene towards shadow and shadow-free conditions has diverse illumination components, but sharing the same reflectance component. Inspired by this, Wu *et al.* [66] formulated the Retinex decomposition as implicit prior regularization and employed a parallel neural network for data-dependent initialization, unfolding optimization, and illumination enhancement. Jiang *et al.* [67] used histogram equalization counterparts as inter-consistency constraints to guide a Retinex decomposition and correction network to correct illumination and remove noise. Ma *et al.* [68] proposed a fidelity term with self-reinforced function to acquire an ideal illumination with smoothing property, while a reflectance optimization mechanism with projection was designed to suppress noise and artifacts in the enhancement process. In summary, we are able to remove shadow remnants when a reasonable illumination component is available.

B. State Space Models

The State Space Model, a family architecture encapsulating sequential modeling, displays promise in computer vision. Zhao *et al.* [13] embedded an omnidirectional selective scan module in the State Space Model to acquire spatial context from horizontal, vertical, diagonal, and anti-diagonal axes, thus capturing the spatial distribution of land covers on very-high-resolution remote sensing images. Such a manner accomplishes the trade-off between efficiency and accuracy for dense prediction tasks such as semantic segmentation and change detection. Li *et al.* [17] employed spatial Mamba and spectral Mamba to extract multi-grained features. Then, a close cooperation between spatial and spectral signals was dedicated to hyperspectral image classification. Chen *et al.* [18] incorporated spatio-temporal relationship modeling mechanisms into Mamba to explore the spatio-temporal interaction of multi-temporal surface features, thus mining change information. Notably, spatio-temporal sequential modeling, spatio-temporal cross modeling, and spatio-temporal parallel modeling respectively cope with a variety of change detection tasks, which contribute to scene adaptability. To rule out visual quality degradation in mobile devices, Ju *et al.* [19] embedded a multi-scale progressive fusion module into Mamba to bridge global semantic loss. Ma *et al.* [20] employed a Mamba-based U-shaped configuration for medical image segmentation. In addition, a self-supervised contrast learning technique was introduced to enhance the segmentation performance. Ge *et al.* [21] proposed a visual State Space Model with multi-stream patch embedding to aggregate global associations of traffic signs, thus enabling efficient traffic sign recognition. Liao *et al.* [58] regarded Mamba as a multi-modal interaction architecture, thus fusing hyperspectral and LiDAR data for superior classification. Despite these advances, the potential of Mamba towards shadow removal is left unexplored. Designing a Mamba-based architecture with interpretability is undoubtedly valuable for subsequent research in the community.

C. Shadow Removal

Early methods commonly focus on mining a range of physical shadow properties. He *et al.* [22] employed 3D intensity surface modeling to remove shadow traces while preserving texture edges, especially for non-uniform and curved surface contamination. Zhang *et al.* [23] constructed the correspondence between shadow and shadow-free regions according to structural similarity, then proposed an illumination recovery operator to remove sharp and soft shadows. Guo *et al.* [24] calculated relative brightness coefficients between shadow and shadow-free patches to recover attractive illumination. Arbel *et al.* [25] utilized cubic smoothing splines to compute per-pixel scale factors of shadows, thus handling non-uniform shadow effects. Unfortunately, traditional methods are impractical for real world shadow detection and shadow removal because of dependence on particular physically modeling.

Recently, deep learning-based approaches have become mainstream benefiting from the availability of large-scale benchmarks [26], [27]. To accommodate all types of shadows, Jin *et al.* [28] introduced adaptive attention into the diffusion

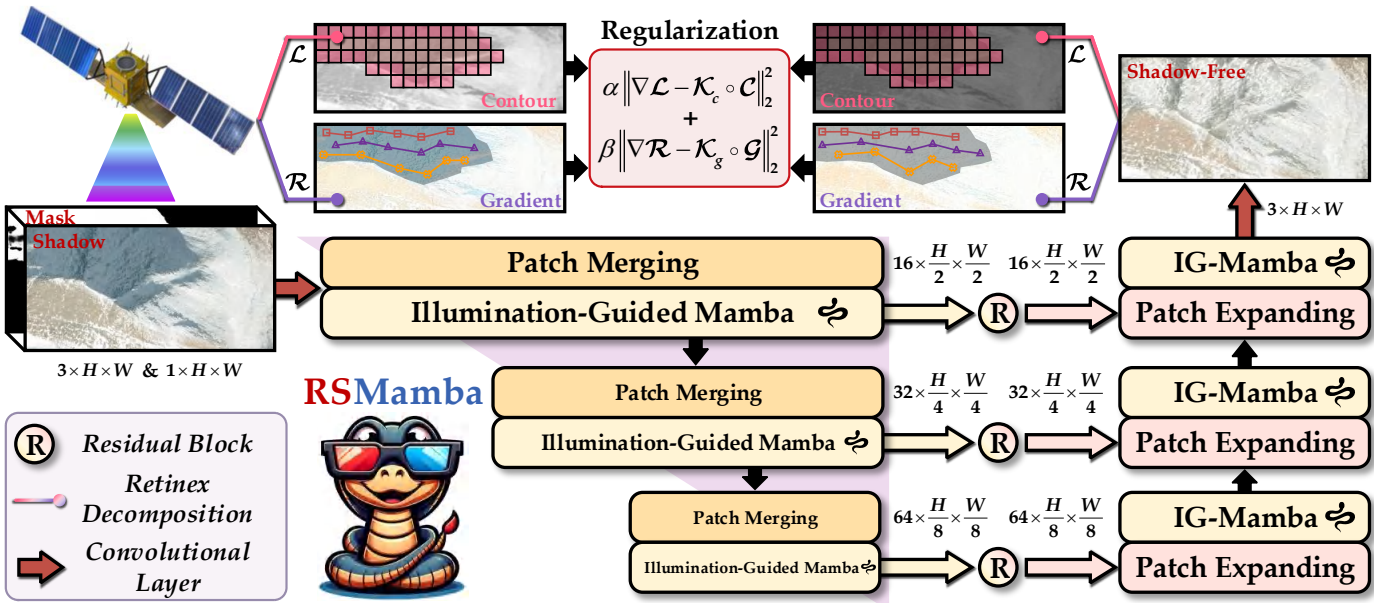

Fig. 2. Schematic illustration of RSMamba. RSMamba consists of multiple illumination-guided Mamba (IG-Mamba) modules and follows a U-shaped architecture to perform shadow removal. For the loss function, RSMamba simulates the dependency of the retina on gradient and contour perception, thus encoding discontinuities in illumination, similar to the brain.

process to smooth shadow boundaries. Liu *et al.* [29] proposed a bilateral network to progressively recover illumination and texture. Local illumination was corrected through conditional denoising, then local texture was enhanced through scale-adaptive consistency. Based on physical property, spatial relation, and temporal coherence, Chen *et al.* [30] designed a video shadow removal network to cope with complex illumination and textures. Guo *et al.* [31] proposed a diffusion-based shadow removal network, which recovers shadow boundary structure through underlying reflectance. Similarly, Guo *et al.* [32] integrated degradation prior and diffusive generative prior to progressively refine shadow boundary artifacts. Li *et al.* [33] formulated shadow removal as an adaptive fusion process, thus enjoying the collaboration between illumination recovery and image inpainting. They pre-trained the shadow removal network on image inpainting benchmarks to remove shadow remnants. Yu *et al.* [34] computed the mean and variance of shadow-free regions, and then achieved illumination and color consistency between regions through statistical constraints. Liu *et al.* [35] leveraged shadow-free structural information to regularize shadow pixels for removing shadows at the image-structure level. Xu *et al.* [59] designed a lightweight dynamic shadow-aware convolution bridge the color difference between shadow and shadow-free regions. Yu *et al.* [64] rethought shadow removal in terms of Fourier transform, and then employed frequency and spatial interaction strategy to repair luminance and structural information. Chang *et al.* [65] proposed a two-stage transformer-based shadow removal network, where the shadow removal stage for global illumination restoration while the content refinement stage for missing pixel compensation. Nevertheless, the above methods are limited by local bias or quadratic complexity, thus reaching a plateau in performance.

III. Methodology

In this section, we first describe an overview of State Space Models and introduce the details of RSMamba. Subsequently, the loss function to capture illumination discontinuities and reflectance variations is illustrated.

A. Preliminaries

State Space Models. SSMs are typically regarded as linear time-invariant systems, which project a one-dimensional stimulus $x(t) \in \mathbb{R}^L$ to $y(t) \in \mathbb{R}^L$ by the intermediary hidden state $h(t) \in \mathbb{R}^N$. Mathematically, SSMs are formulated by linear ordinary differential equations (ODEs):

$$\begin{aligned} h^{'}(t) &= \mathbf{A}h(t) + \mathbf{B}x(t), \\ y(t) &= \mathbf{C}h(t) + \mathbf{D}x(t), \end{aligned} \tag{2}$$

where $\mathbf{A} \in \mathbb{C}^{N\times N}$ represents the state projection matrix, $\mathbf{B} \in \mathbb{C}^N$ and $\mathbf{C} \in \mathbb{C}^N$ represent projection parameters, and $\mathbf{D} \in \mathbb{C}^1$ represents the skip connection.

In view of the continuous-time property, SSMs face significant challenges in integrating into deep learning. To this end, a zero-order hold (ZOH) rule with a timescale parameter Δ is employed to discretize projection parameters:

$$\begin{aligned} \bar{\mathbf{A}} &= \exp(\Delta\mathbf{A}), \\ \bar{\mathbf{B}} &= (\Delta\mathbf{A})^{-1}(\exp(\Delta\mathbf{A}) - \mathbf{I}) \cdot \Delta\mathbf{B}, \end{aligned} \tag{3}$$

In summary, the discrete SSMs can be expressed as:

$$\begin{aligned} h_t &= \bar{\mathbf{A}}h_{t-1} + \bar{\mathbf{B}}x_t, \\ y_t &= \mathbf{C}h_t + \mathbf{D}x_t, \end{aligned} \tag{4}$$

Nevertheless, the discrete version designed for one-dimensional data fails to capture two-dimensional spatial information. Fortunately, the 2D selective scan [36] provides crucial contextual sensitivity.

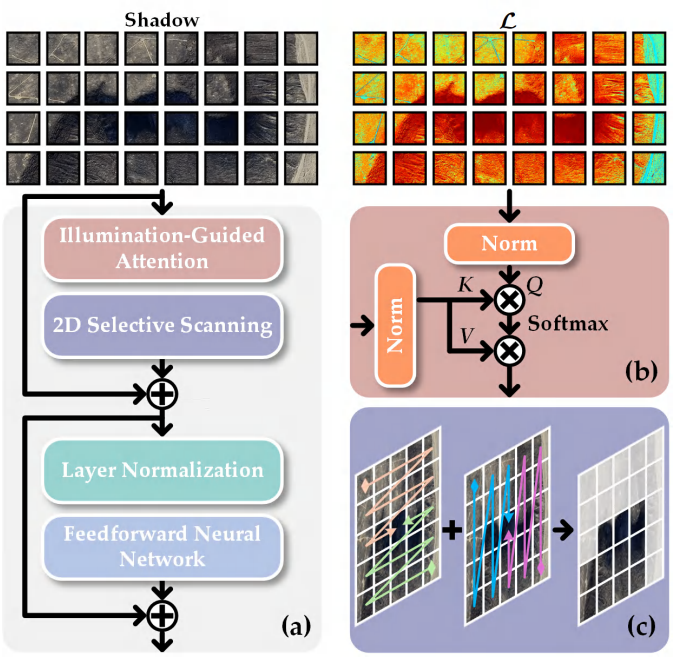

Fig. 3. (a) Schematic illustration of the illumination-guided Mamba module. (b) Schematic illustration of the illumination-guided attention. (c) Schematic illustration of the 2D selective scanning. Notably, the illumination prior (displayed through the heatmap) enforce the neural network to understand shadow intensity, thus incorporating the advantage of physical imaging into the data-driven paradigm.

2D Selective Scan. 2D selective scan (SS2D) arranges separate patch sequences along four directions, *i.e.*, top-left to bottom-right, bottom-right to top-left, top-right to bottom-left, and bottom-left to top-right. Such a location-aware scan strategy ensures that the discrete state-space equation captures multi-directional long-term dependencies, thus establishing a global receptive field. SS2D can be expressed as:

$$\begin{aligned} \mathbf{t}_v &= \text{expand}(t, v), \\ \bar{\mathbf{t}}_v &= \text{S6}(\mathbf{t}_v), \\ \bar{\mathbf{t}} &= \text{merge}(\bar{\mathbf{t}}_1, \bar{\mathbf{t}}_2, \bar{\mathbf{t}}_3, \bar{\mathbf{t}}_4), \end{aligned} \tag{5}$$

where $v = 1, 2, 3, 4$ represents four scanning directions. Besides, $\text{expand}(\cdot)$ and $\text{merge}(\cdot)$ represent scan expand and scan merge operations [36].

B. Overall Pipeline

The overview architecture of RSMamba is depicted in Fig. 2. The well-designed RSMamba takes shadow sample and auxiliary shadow mask as input, which contributes to the discrimination between shadow and shadow-free regions. Meanwhile, we employ a pre-trained Retinex decomposition network [37] to extract illumination and reflectance components. Subsequently, we concatenate the contamination representation $\mathbb{R}^{H\times W\times 4}$ and the illumination component $\mathbb{R}^{H\times W\times 1}$, then feed the five-channel cascade to the illumination-guided Mamba module, where the illumination component serves as a supervision signal revealing the quality degradation degree.

As shown in Fig. 3, IG-Mamba module consists of illumination-guided attention, SS2D, layer normalization (LN), and feedforward neural network (FFN). Similar to cross attention [38] and spatial reduced self-attention [69], illumination-guided attention employs the illumination prior as a query vector, thus enhancing the response to shadow traces. Such a manner explores the mutual benefits between optical imaging model and neural networks. More importantly, we employ average pooling for $\mathcal{K}$ and $\mathcal{V}$ to a fixed dimension. Such a manner solves the quadratic computational overhead of traditional self-attention [69]. The self-attention calculation process can be expressed as:

$$\begin{gathered} \mathcal{Q} = \mathcal{L}\mathcal{W}_{\mathcal{Q}},\ \mathcal{K} = \text{Cat}[\mathcal{S}, \mathcal{M}]\mathcal{W}_{\mathcal{K}},\ \mathcal{V} = \text{Cat}[\mathcal{S}, \mathcal{M}]\mathcal{W}_{\mathcal{V}}, \\ \mathcal{K}' = \text{AvgPool}(\mathcal{K}),\ \mathcal{V}' = \text{AvgPool}(\mathcal{V}), \end{gathered} \tag{6}$$

where $\mathcal{Q}$ represents Query, $\mathcal{K}$ represents Key, $\mathcal{V}$ represents Value, $\mathcal{S}$ represents the shadow sample, $\mathcal{M}$ represents the shadow mask, $\text{Cat}[,]$ represents the feature concatenation operation, $\text{AvgPool}(\cdot)$ represents the average pooling, $\mathcal{W}_{\mathcal{Q}}$, $\mathcal{W}_{\mathcal{K}}$, and $\mathcal{W}_{\mathcal{V}}$ represent learnable parameter matrices.

$$\text{Attn}(\mathcal{Q}, \mathcal{K}, \mathcal{V}) = \text{softmax}\left(\frac{\mathcal{Q}\mathcal{K}'^{T}}{\sqrt{d}}\right)\mathcal{V}'. \tag{7}$$

where $\sqrt{d}$ represents the normalization factor. After that, we leverage SS2D to enable per-element in the 1D array engage with the scanned samples, thus maintaining the integrity of 2D sample structures. After a LN layer and a FFN layer, a pixel-wise addition serves as an identity connection to preserve the data fidelity.

The IG-Mamba module with superior global dependency capture capability is integrated into the U-Net architecture. Skip connections between encoders and corresponding decoders contribute to the intermediate network learning irregular and inhomogeneous shadow intensity distributions. Finally, we employ a projection convolution to reshape decoded features back to $\mathbb{R}^{H\times W\times 3}$, leading to a shadow-free version $\hat{\mathcal{S}}$.

C. Loss Function

The inhibitory connection of retinal ganglion cells produces a center-surround opponency [39], which is attributed to gradient variations. This research inspired us to design a reflectance gradient loss for human perceivability:

$$\mathcal{L}_{\text{gradient}} = \|\nabla\mathcal{R}_{\text{GT}} - \mathcal{K}_g \circ \mathcal{G}\|_2^2, \tag{8}$$

where $\nabla\mathcal{R}_{\text{GT}}$ represents the reflectance gradient of ground truths, $\mathcal{G}$ represents the reflectance gradient of shadow-free versions, and $\mathcal{K}_g$ represents a weighting matrix of $\mathcal{G}$. Notably, we use the traditional sobel operator to compute the gradient instead of a pre-trained neural network, which helps to reduce the computational overhead. Besides, the strong excitatory stimuli of the visual cortex for illumination contours provide us the inspiration to design an illumination contour loss [40]:

$$\mathcal{L}_{\text{contour}} = \|\nabla\mathcal{L}_{\text{GT}} - \mathcal{K}_c \circ \mathcal{C}\|_2^2, \tag{9}$$

where $\nabla\mathcal{L}_{\text{GT}}$ represents the illumination contour of ground truths, $\mathcal{C}$ represents the illumination contour of shadow-free versions, and $\mathcal{K}_c$ represents a weighting matrix of $\mathcal{C}$. Unlike the reflectance gradient loss, the illumination contour loss use canny operator to capture contours.

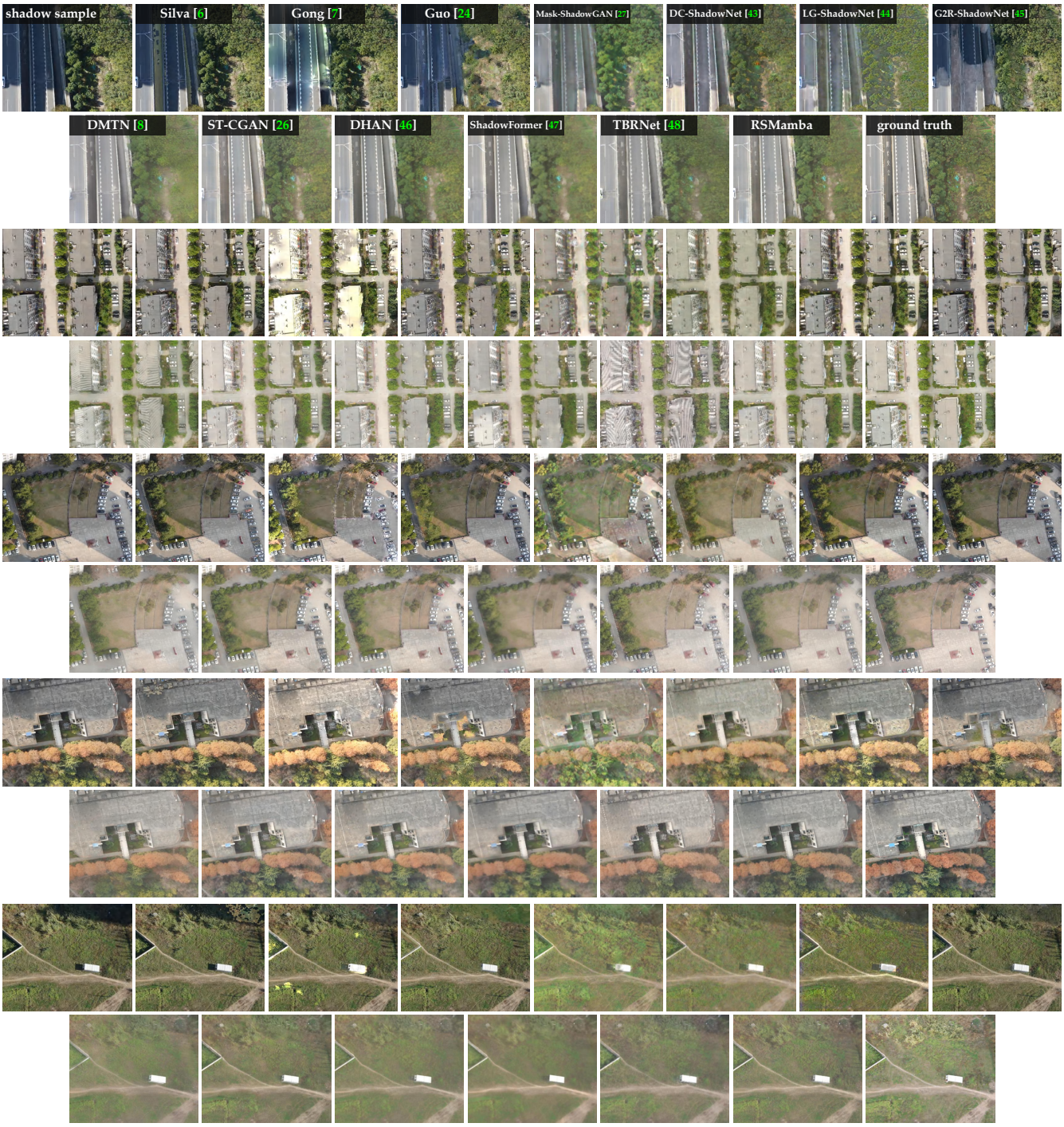

Fig. 4. Qualitative comparisons on the UAV-SC benchmark [41].

Afterward, we employ the ℓ_1 loss to maintain illumination and color consistency between shadow and shadow-free domains:

$$\mathcal{L}_{\ell_1} = \|\hat{\mathcal{S}} - \mathcal{S}_{\mathrm{GT}}\|_1, \tag{10}$$

In a nutshell, we exploit a linear combination of ℓ_1 loss, reflectance gradient loss $\mathcal{L}_{\mathrm{gradient}}$, and illumination contour loss $\mathcal{L}_{\mathrm{contour}}$ as the final loss:

$$\mathcal{L} = \mathcal{L}_{\ell_1} + \alpha\mathcal{L}_{\mathrm{contour}} + \beta\mathcal{L}_{\mathrm{gradient}}. \tag{11}$$

where α and β follow [15] and are set to 10^{-4} and 10^{-3} for balancing the scales of multiple losses.

IV. Experiment

A. Experimental Settings

Implementation Details. RSMamba is end-to-end trained on an NVIDIA RTX 3090 GPU by employing an ADAM optimizer with 500 epochs. Notably, except for the pre-trained Retinex decomposition model, the rest of RSMamba is learned from scratch. The initial learning rate is 1×10^{-4}, then

Fig. 5. Qualitative comparisons on the AISD benchmark [42].

progressively decreased to 1×10^{-6} through a cosine annealing strategy.

Benchmarks. We conduct qualitative and quantitative experiments on UAV-SC [41] and AISD [42] benchmarks. UAV-SC includes 6954 shadow samples and corresponding references, where 6924 sample pairs are divided into a training set and the rest into a testing set. UAV-SC captures street views of Wuhan through drone platforms while employing radiometric correction and geometric correction to maintain geometric consistency. AISD serves shadow detection, including 514 aerial samples of metropolitans and forests from Austin, Chicago, Tyrol, Vienna, and Innsbruck.

Competitor. We compare RSMamba with the following methods: Silva [6], Gong [7], Guo [24], Mask-ShadowGAN [27], DC-ShadowNet [43], LG-ShadowNet [44], G2R-ShadowNet [45], DMTN [8], ST-CGAN [26], DHAN [46], ShadowFormer [47], and TBRNet [48].

Quantitative Metrics. Toward the UAV-SC benchmark [41], root mean square error (RMSE), peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) metrics are employed for full-reference evaluations. Lower RMSE scores indicate better visual quality, which is the opposite of PSNR and SSIM metrics. Not only that, the full-reference evaluation is performed in shadow regions (S.), shadow-free regions (N.S.), and whole image (All). Toward the AISD benchmark (without reference) [42], we refer to [27], [28], [53] using a user study to evaluate visual quality in terms of both shadow removal (Des) and visually realistic (Real). We recruit 20 participants: 12 males and 8 females, aged 23-30 with experience in computer vision. We present the result of developed and compared methods to participants in random order, then ask participants to rate them from 1 (bad) to 10 (good).

B. Qualitative Comparison

We first display the visual comparisons on UAV images in Fig. 4. Traditional competitors either fail to alleviate shadow contamination or alter the brightness of shadow-free regions. In addition, G2R-ShadowNet [45], Mask-ShadowGAN [27], DC-ShadowNet [43], and LG-ShadowNet [44] repair illumination to some extent, but fail to cope with variable shadow intensities. This is because the randomness of shadows in terms of position, shape and size challenges unsupervised and weakly supervised strategies. DMTN [8] and TBRNet [48] introduce undesired streaky artifacts, especially on the roof. ST-CGAN [26] and DHAN [46] are sensitive and vulnerable towards cast shadows. Although ShadowFormer [47] removes shadow traces, it blurs details such as grasses and woods.

TABLE I
QUANTITATIVE COMPARISONS ON UAV-SC [41]. "↑" REPRESENTS THAT LARGER SCORES ARE BETTER, WHILE "↓" REPRESENTS THAT LOWER SCORES ARE BETTER. ⋆, ¶, †, AND ‡ REPRESENT TRADITIONAL, UNSUPERVISED, WEAKLY SUPERVISED, AND FULLY SUPERVISED METHODS, RESPECTIVELY. ∗ REPRESENTS THAT THE CODE IS NOT PUBLICLY AVAILABLE AND IS IMPLEMENTED BY OURSELVES. BEST AND SECOND-BEST SCORES ARE **HIGHLIGHTED** AND UNDERLINED.

Methods	RMSE(↓)			PSNR(↑)			SSIM(↑)		
	S.	N.S.	All	S.	N.S.	All	S.	N.S.	All
Silva [6] (ISPRS'18) ⋆	32.0696	18.5089	21.3024	22.8902	18.1034	16.2621	0.8899	0.8035	0.6898
Gong [7] (BMVC'14) ⋆	26.0850	18.8162	20.3155	25.1448	18.1944	16.8290	0.9179	0.7926	0.6984
Guo [24] (TPAMI'13) ⋆	29.8381	17.9412	20.3919	23.5991	18.3971	16.6687	0.9042	0.8093	0.6945
Mask-ShadowGAN [27] (ICCV'19) ¶	19.7786	17.5036	17.9722	27.5729	20.4507	19.1357	0.9450	0.8142	0.7374
DC-ShadowNet [43] (ICCV'21) ¶	16.1495	13.6091	14.1324	29.2519	22.4207	21.0712	0.9543	0.8479	0.7852
LG-ShadowNet [44] (TIP'21) ¶	23.0353	16.6409	17.9581	25.4526	19.6434	18.0556	0.9352	0.8352	0.7629
G2R-ShadowNet [45] (CVPR'21) †	22.0473	18.6220	19.3276	24.6951	17.9899	16.8325	0.9015	0.8189	0.7183
DMTN [8] (TMM'23) ‡	11.5195	9.9282	10.2560	31.6098	24.1577	23.0287	0.9664	0.8696	0.8195
ST-CGAN [26] (CVPR'18) ‡ ∗	9.4592	9.2458	9.2898	32.4301	24.8846	23.8219	0.9563	0.8679	0.8317
DHAN [46] (AAAI'20) ‡	9.1524	9.0602	9.1086	33.0654	25.1760	24.1399	0.9752	0.9010	0.8620
ShadowFormer [47] (AAAI'23) ‡	9.6703	9.2937	9.5927	32.7140	24.3702	23.4432	0.9711	0.8804	0.8350
TBRNet [48] (TNNLS'23) ‡	11.4115	10.8361	10.9546	31.2298	23.6601	22.5798	0.9619	0.8313	0.7722
RSMamba ‡	**9.1059**	**8.9469**	**9.0003**	**33.2074**	**25.4818**	**24.4199**	**0.9767**	**0.9102**	**0.8746**

TABLE II
QUANTITATIVE COMPARISONS ON AISD [42]. "↑" REPRESENTS THAT LARGER SCORES ARE BETTER. BEST AND SECOND-BEST SCORES ARE **HIGHLIGHTED** AND UNDERLINED.

Methods	Des(↑)	Real(↑)
Silva [6] (ISPRS'18) ⋆	5.2 ± 2.3	5.4 ± 1.8
Gong [7] (BMVC'14) ⋆	5.0 ± 2.0	5.1 ± 2.7
Guo [24] (TPAMI'13) ⋆	6.2 ± 1.8	6.0 ± 0.8
Mask-ShadowGAN [27] (ICCV'19) ¶	5.8 ± 2.1	5.2 ± 1.7
DC-ShadowNet [43] (ICCV'21) ¶	5.9 ± 1.5	5.3 ± 0.7
LG-ShadowNet [44] (TIP'21) ¶	5.8 ± 0.9	5.5 ± 1.5
G2R-ShadowNet [45] (CVPR'21) †	5.3 ± 2.2	5.2 ± 1.6
DMTN [8] (TMM'23) ‡	6.7 ± 1.7	6.4 ± 1.8
ST-CGAN [26] (CVPR'18) ‡ ∗	7.2 ± 1.2	6.9 ± 1.4
DHAN [46] (AAAI'20) ‡	6.9 ± 1.3	7.1 ± 1.5
ShadowFormer [47] (AAAI'23) ‡	7.0 ± 1.5	6.5 ± 1.8
TBRNet [48] (TNNLS'23) ‡	7.1 ± 2.2	7.0 ± 0.9
RSMamba ‡	**8.5 ± 1.3**	**7.4 ± 1.0**

TABLE III
QUANTITATIVE SCORES OF THE ABLATION STUDY. "↑" REPRESENTS THAT LARGER SCORES ARE BETTER, WHILE "↓" REPRESENTS THAT LOWER SCORES ARE BETTER. THE BEST SCORE IS **HIGHLIGHTED**.

Baselines	UAV-SC		
	RMSE(↓)	PSNR(↑)	SSIM(↑)
w/o IAM	10.9825	23.1598	0.8337
w/o SS2D	11.1149	22.7704	0.8189
w/o $\mathcal{L}_{\text{gradient}}$	12.7911	21.0881	0.7310
w/o $\mathcal{L}_{\text{contour}}$	12.0895	21.7792	0.7783
w/o $\mathcal{L}_{\ell_1}$	11.5768	22.3260	0.8051
Full Model	**9.0003**	**24.4199**	**0.8746**

In contrast, RSMamba effectively removes shadows without obvious over-saturation or color deviation.

We also display the visual comparisons on aerial images in Fig. 5. Competing methods either introduce yellowish or greenish color deviations, such as Gong [7] and Mask-ShadowGAN [27], or remain shadow remnants. Therefore, all compared methods do not have convincing scenario adaptability. Although RSMamba is able to remove large cast shadows and tiny shadow traces, it introduces a mild overexposure. This is because neither the developed nor the compared methods are trained on the AISD benchmark, which significantly tests the robustness of the methods. Nevertheless, our method demonstrates convincing visual quality, which is an advantage that other methods do not have.

C. Quantitative Comparison

We first present the full-reference evaluation scores of RSMamba and competitors in Table I. RSMamba outperforms all compared methods on the UAV-SC benchmark. Compared with the top-performing competitor [46], RSMamba achieves the percentage gain of 0.5∼1.3%/0.4∼1.2%/0.2∼1.5% in terms of RMSE/PSNR/SSIM, respectively. Subsequently, we report the user study scores in Table II. RSMamba achieves more votes with high scores, which suggests that our method is more attractive to participants in terms of de-shadowing performance and realism. Such quantitative scores verify the effectiveness of well-designed RSMamba.

D. Ablation Study

We perform extensive ablation experiments to analyze core components of RSMamba, consisting of illumination-guided attention mechanism and 2D selective scan. Besides, we explore the combination of the reflectance gradient loss, the illumination contour loss, and the ℓ_1 loss.

- w/o IAM refers to RSMamba without the illumination-guided attention mechanism, employing native Mamba block instead of cross attention, thus removing Retinex-aware guidence.
- w/o SS2D refers to RSMamba without the 2D selective scan, maintaining only the Transformer architecture.
- w/o $\mathcal{L}_{\text{gradient}}$ indicates that RSMamba is trained independently of the reflectance gradient loss.
- w/o $\mathcal{L}_{\text{contour}}$ indicates that RSMamba is trained independently of the illumination contour loss.

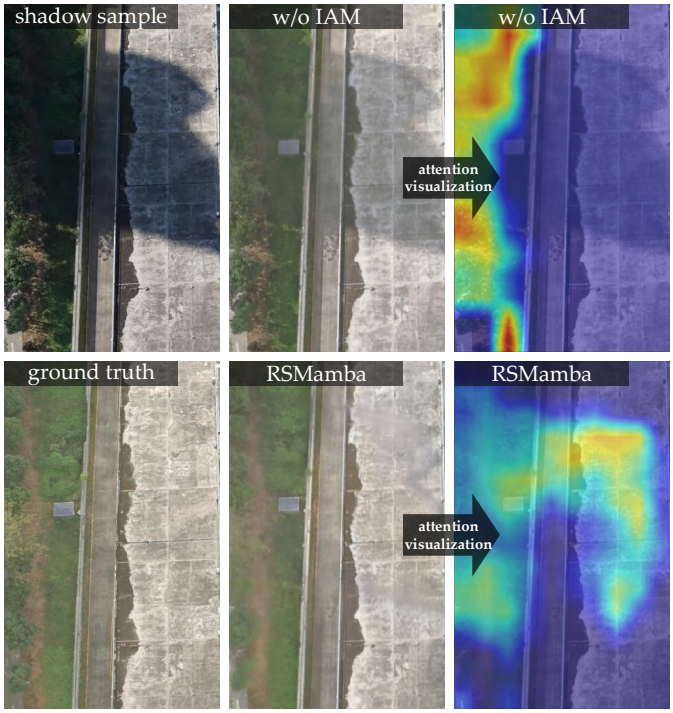

Fig. 6. Ablation study towards the illumination-guided attention mechanism. Attention visualization suggests that the cross-modal interaction contributes to forcing RSMamba to pay attention to quality-degraded regions.

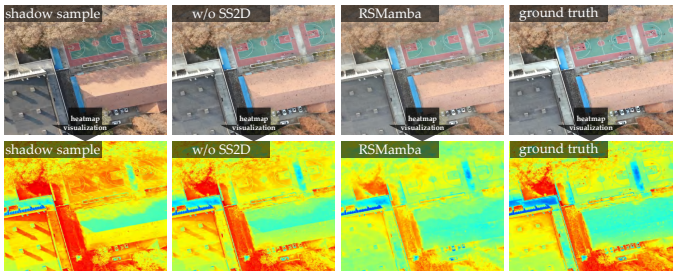

Fig. 7. Ablation study towards the 2D selective scan. Heatmap visualization highlights the favorable global modeling capability of SS2D, which contributes to removing tiny shadow remnants.

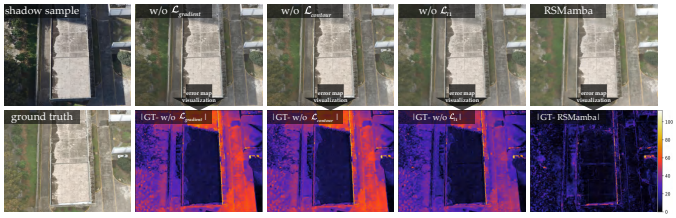

Fig. 8. Ablation study towards the loss function. RSMamba efficiently locates and removes shadow traces through gradient regularization and contour regularization.

- w/o $\mathcal{L}_{\ell_1}$ indicates that RSMamba is trained independently of the ℓ_1 loss.

The quantitative scores of ablated models are presented in Table III. Besides, the effects of illumination-guided attention mechanism, the effectiveness of 2D selective scan, and the contributions of loss function are shown in Figs. 6, 7, and 8, respectively. The conclusions drawn from the ablation study are as follows:

- As reported in Table III, RSMamba achieves the best RMSE, PSNR, and SSIM scores compared with ablated models, implying that the combination of IAM and SS2D modules is imperative.
- In Fig. 6, the ablated model w/o IAM focuses more attention on local shadows, leading to large shadow remnants. In contrast, RSMamba enhances the response toward quality-degraded regions under illumination guidance, thus maintaining luminance equalization.
- Tiny cast shadows conquer the ablated model w/o SS2D since it fails to capture long sequences with hidden layers, as shown in Fig. 7. In contrast, RSMamba exploits long-range dependency with linear complexity to remove shadows with diverse locations, intensities, and sizes.
- As depicted in Fig. 8, the ablated models w/o $\mathcal{L}_{\text{gradient}}$, w/o $\mathcal{L}_{\text{contour}}$, and w/o $\mathcal{L}_{\ell_1}$ suffer from shadow contamination. RSMamba resembles the human visual system in discriminating shadows from gradient and contour perspectives under the supervision of multiple regularizations, thus enjoying the mutual benefits between shadow detection and shadow removal.

V. Conclusion

In this paper, we design a retinex-based Mamba inspired by the visual cortex. RSMamba inherits the robustness of the human visual system through gradient and contour capture for shadow acquisition. Such a manner is physiologically plausible, since gradients and contours are perceptible to humans. Meanwhile, the illumination map serves as an intensity cue to remove shadows by assigning more attention weight to quality degradation pixels. Extensive experiments on multiple benchmarks have demonstrated the superiority of RSMamba. Moreover, the efficacy of core components of RSMamba has been proven in ablation studies.

References

[1] L. Jie and H. Zhang, "RMLANet: Random multi-level attention network for shadow detection and removal," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 12, pp. 7819–7831, Dec. 2023.

[2] J. Shen, C. Zhang, Y. Yuan, and Q. Wang, "Enhancing prospective consistency for semisupervised object detection in remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–12, Aug. 2023.

[3] Y. Jia, J. Gao, W. Huang, Y. Yuan, and Q. Wang, "Holistic mutual representation enhancement for few-shot remote sensing segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–13, Oct. 2023.

[4] 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.

[5] 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.

[6] 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.

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

[8] 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.

[9] N. Inoue and T. Yamasaki, "Learning from synthetic shadows for shadow detection and removal," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 11, pp. 4187–4197, Nov. 2021.

[10] L. Fu *et al.*, "Auto-exposure fusion for single-image shadow removal," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10566–10575.

[11] H. Le and D. Samaras, "Physics-based shadow image decomposition for shadow removal," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 9088–9101, Dec. 2022.

[12] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," 2023, *arXiv: 2312.00752*.

[13] S. Zhao, H. Chen, X. Zhang, P. Xiao, L. Bai, and W. Ouyang, "RS-Mamba for large remote sensing image dense prediction," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–14, Jul. 2024.

[14] B. Montrucchio, C. Celozzi, and P. Cerutti, "Thresholds of vision of the human visual system: visual adaptation for monocular and binocular vision," *IEEE Trans. Human-Machine Syst.*, vol. 45, no. 6, pp. 739–749, Dec. 2015.

[15] R. Cai and Z. Chen, "Brain-like retinex: A biologically plausible retinex algorithm for low light image enhancement," *Pattern Recognit.*, vol. 136, pp. 109195, Apr. 2023.

[16] E. H. Land and J. J. McCann, "Lightness and Retinex theory," *J. Opt. Soc. Amer.*, vol. 61, no. 1, pp. 1–11, Jan. 1971.

[17] Y. Li, Y. Luo, L. Zhang, Z. Wang, and B. Du, "MambaHSI: Spatial-spectral Mamba for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–16, Jul. 2024.

[18] H. Chen, J. Song, C. Han, J. Xia, and N. Yokoya, "ChangeMamba: Remote sensing change detection with spatiotemporal state space model," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–20, Jun. 2024.

[19] M. Ju, S. Xie, and F. Li, "Improving skip connection in U-Net through fusion perspective with Mamba for image dehazing," *IEEE Trans. Consum. Electron.*, vol. 70, no. 4, pp. 7505–7514, Nov. 2024.

[20] C. Ma and Z. Wang, "Semi-Mamba-UNet: Pixel-level contrastive and cross-supervised visual Mamba-based UNet for semi-supervised medical image segmentation," *Knowledge-based Syst.*, vol. 300, pp. 112203, Sep. 2024.

[21] Y. Ge, Z. Chen, M. Yu, Q. Yue, R. You, and L. Zhu, "MambaTSR: You only need 90k parameters for traffic sign recognition," *Neurocomputing*, vol. 599, pp. 128104, Sep. 2024.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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)*, Oct. 2019, pp. 2472–2481.

[28] Y. Jin, W. Ye, W. Yang, Y. Yuan, and R. Tan, "DeS3: Adaptive attention-driven self and soft shadow removal using ViT similarity," in *Proc. AAAI Conf. Artif. Intell.*, vol. 38, 2024, pp. 2634–2642.

[29] Y. Liu, Z. Ke, K. Xu, F. Liu, Z. Wang, and R. Lau, "Recasting regional lighting for shadow removal," in *Proc. AAAI Conf. Artif. Intell.*, vol. 38, 2024, pp. 3810–3818.

[30] 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.

[31] L. Guo, C. Wang, W. Yang, Y. Wang, and B. Wen, "Boundary-aware divide and conquer: A diffusion-based solution for unsupervised shadow removal," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 12999–13008.

[32] 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.

[33] X. Li *et al.*, "Leveraging inpainting for single-image shadow removal," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 13009–13018.

[34] Q. Yu, N. Zheng, J. Huang, and F. Zhao, "CNSNet: A cleanness-navigated-shadow network for shadow removal," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Nov. 2022, pp. 221–238.

[35] Y. Liu *et al.*, "Structure-informed shadow removal networks," *IEEE Trans. Image Process.*, vol. 32, pp. 5823–5836, Nov. 2023.

[36] Y. Liu *et al.*, "Vmamba: Visual state space model," 2024, *arXiv: 2401.10166*.

[37] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Sep. 2018, pp. 1–12.

[38] C.-F. R. Chen, Q. Fan, and R. Panda, "CrossViT: Cross-attention multi-scale vision transformer for image classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 347–356.

[39] D. Dacey, O. S. Packer, L. Diller, D. Brainard, B. Peterson, and B. Lee, "Center surround receptive field structure of cone bipolar cells in primate retina," *Vision Res.*, vol. 40, no. 14, pp. 1801–1811, Jun. 2000.

[40] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, no. 1, pp. 106–154, Jan. 1962.

[41] 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.

[42] S. Luo, H. Li, and H. Shen, "Deeply supervised convolutional neural network for shadow detection based on a novel aerial shadow imagery dataset," *ISPRS J. Photogramm. Remote Sens.*, vol. 167, pp. 443–457, Sep. 2020.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

[47] 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.

[48] 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.

[49] 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.

[50] 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.

[51] Q. Li, Y. Yuan, X. Jia, and Q. Wang, "Dual-stage approach toward hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 31, pp. 7252–7263, Nov. 2022.

[52] Q. Li, M. Gong, Y. Yuan, and Q. Wang, "RGB-induced feature modulation network for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–11, May. 2023.

[53] K. Niu, Y. Liu, E. Wu, and G. Xing, "A boundary-aware network for shadow removal," *IEEE Trans. Multimedia*, vol. 25, pp. 6782–6793, Oct. 2022.

[54] 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.

[55] 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.

[56] C. Yang, M. Chen, Y. Yuan, and Q. Wang, "Reinforcement shrink-mask for text detection," *IEEE Trans. Multimedia*, vol. 25, pp. 6458–6470, Sep. 2022.

[57] C. Yang, M. Chen, Z. Xiong, Y. Yuan, and Q. Wang, "CM-Net: Concentric mask based arbitrary-shaped text detection," *IEEE Trans. Image Process.*, vol. 31, pp. 2864–2877, Mar. 2022.

[58] D. Liao, Q. Wang, T. Lai, and H. Huang, "Joint classification of hyperspectral and LiDAR data Based on Mamba," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–15, Sep. 2024.
[59] Y. Xu, M. Lin, H. Yang, F. Chao, and R. Ji, "Shadow-aware dynamic convolution for shadow removal," *Pattern Recognit.*, vol. 146, pp. 109969, Feb. 2024.
[60] Q. Yan *et al.*, "Toward high-quality HDR deghosting with conditional diffusion models," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 5, pp. 4011–4026, May. 2024.
[61] Q. Yan *et al.*, "From dynamic to static: Stepwisely generate HDR image for ghost removal," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Sep. 25, 2024, doi: 10.1109/TCSVT.2024.3467259.
[62] Y. Zhang, H. Zhang, N. M. Nasrabadi, and T. S. Huang, "Multi-metric learning for multi-sensor fusion based classification," *Inf. Fusion*, vol. 14, no. 4, pp. 431–440, Oct. 2013.
[63] Q. Yan *et al.*, "Uncertainty estimation in HDR imaging with Bayesian neural networks," *Pattern Recognit.*, vol. 156, pp. 110802, Dec. 2024.
[64] J. Yu, P. He, and Z. Peng, "FSR-Net: Deep fourier network for shadow removal," in *Proc. ACM Int. Conf. Multimedia*, Oct. 2023, pp. 2335–2343.
[65] H.-E. Chang *et al.*, "TSRFormer: Transformer based two-stage refinement for single image shadow removal," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2023, pp. 1436–1446.
[66] W. Wu, J. Weng, P. Zhang, X. Wang, W. Yang, and J. Jiang, "URetinex-Net: Retinex-based deep unfolding network for low-light image enhancement," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5891–5900.
[67] Q. Jiang, Y. Mao, R. Cong, W. Ren, C. Huang, and F. Shao, "Unsupervised decomposition and correction network for low-light image enhancement," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 19440–19455, Oct. 2022.
[68] L. Ma, R. Liu, Y. Wang, X. Fan, and Z. Luo, "Low-light image enhancement via self-reinforced retinex projection model," *IEEE Trans. Multimedia*, vol. 25, pp. 3573–3586, Sep. 2023.
[69] S. Chen, A. Atapour-Abarghouei, H. Zhang, and H. P. H. Shum, "MxT: Mamba x transformer for image inpainting," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Nov. 2024, pp. 1–14.

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.

Sai Guo received the B.E. degree in internet of things engineering from Northwestern Polytechnical University, Xi'an, China, in 2023. He is currently pursuing the M.S. degree in School of Cybersecurity, Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision and pattern recognition.

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/).