

# Recreating Brightness from Remote Sensing Shadow Appearance

Qi Wang, *Senior Member, IEEE*, Kaichen Chi, Wei Jing, and Yuan Yuan, *Senior Member, IEEE*

**Abstract**—Shadow removal from remote sensing images is still an open issue. Recently, deep network training on unpaired data is preferable since corresponding ground truths of shadow images are not available in practice. Nevertheless, unsupervised shadow removal research for remote sensing imagery is limited by the scarcity of publicly available benchmarks. This paper proposes an unsupervised progressive network (UP-ShadowGAN) to jointly learn decoupled features for shadow removal and color transfer. UP-ShadowGAN explores the mapping between shadow and shadow-free domains through adversarial learning and cycle consistency constraint. In particular, we employ progressive learning to decompose the overall mapping process into more manageable shadow removal and color transfer steps. Specifically, the realistic illumination is restored by propagating spatial context between shadow and shadow-free nodes. Coupled with a multi-color space aggregation strategy, diverse color space representations alleviate color deviation caused by spatial inconsistency. More importantly, we contribute the first unpaired remote sensing shadow removal dataset (URSSR), which encourages future exploration. Extensive experiments demonstrate that UP-ShadowGAN competes favorably with state-of-the-art methods. The dataset and code are available at <https://github.com/chi-kaichen/UP-ShadowGAN>.

**Index Terms**—Shadow removal, graph reasoning, unpaired data, progressive learning.

## I. INTRODUCTION

**S**HADOW is formed when illumination is completely or partially blocked by an object, and is a ubiquitous natural phenomenon. Unfortunately, the presence of shadows typically deteriorates the performance of computer vision tasks, such as geo-localization [1], scene classification [2], [3], and object detection [4], [5]. An effective solution to restore illumination in shadow regions is of great significance for improving the visual quality of remote sensing images and accurately understanding surface information [48]–[50].

Early shadow removal methods detect and remove shadows by designing physics models to analyze the statistics of illumination [6]–[10]. However, physics model-based methods highly rely on handcrafted prior assumptions, and tend to produce unsatisfactory results in complex cases. The main reason is that prior assumptions introduce disturbances and uncertainties to some extent.

Recently, large-scale shadow removal benchmarks [11]–[13] have stimulated the emergence of deep learning-based

This work was supported in part by the National Natural Science Foundation of China under 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.)

Q. Wang, K. Chi, W. Jing, and Y. Yuan are with the School of Artificial Intelligence, Optics and Electronics, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: crabwq@gmail.com, chikaichen@mail.nwpu.edu.cn, wei\_adam@mail.nwpu.edu.cn, y.yuan.ieee@gmail.com).

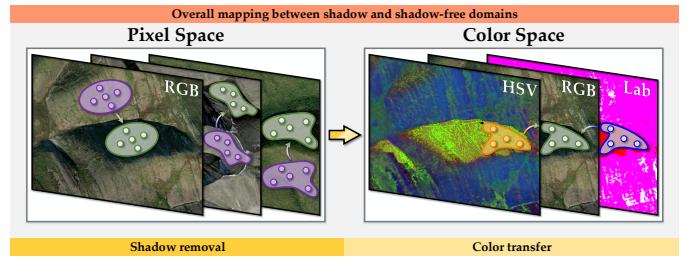


Fig. 1. The schematic illustration of progressive learning. In pixel space, shadow and shadow-free nodes contain similar and complementary scene priors. In color space, HSV and Lab color spaces are able to reflect stronger changes in luminance, hue, and saturation from shadow regions. These motivate us to learn the mapping between shadow and shadow-free domains through spatial context and color space representation interactions.

methods. These methods utilize paired data, *i.e.*, both shadow and shadow-free versions, to train deep networks [14]–[16], which achieve promising performance. However, collecting such sample pairs is both time-consuming and labor-intensive. More importantly, highly-controlled light sources and strictly static scenes are essential, which are not available for remote sensing data collection.

To solve these issues, Hu *et al.* [17] presented Mask-ShadowGAN trained on unpaired data, which is inspired by CycleGAN [18]. Mask-ShadowGAN trains the shadow generator to provide shadow masks, which contribute to guiding the generation of matched shadow samples. Such a manner solves the training collapse caused by highly under-constrained mappings. However, Mask-ShadowGAN is unstable and sensitive to changes in color, shadow, occlusion, and other factors [19]. This is because 1) the color transfer from shadow-free to shadow regions is ill-posed; 2) the high-level semantic knowledge between shadow and shadow-free regions is ignored.

In this work, we propose to explore a complex balance between spatial contextualized information and color space representation through progressive learning. Although the proposed method, CycleGAN [18], and Mask-ShadowGAN [17] all follow the cycle consistent property, CycleGAN [18] and Mask-ShadowGAN [17] directly employ simple bidirectional mappings to achieve a transformation from shadow to shadow-free versions, thus not applicable to soft shadow and inconsistent color [19], [30]. In contrast, our proposal is to embed a multi-stage process in the inter-domain transformation, decomposing the mapping between shadow and shadow-free domains into two subtasks, aiming to enjoy the mutual benefits between shadow removal and color transfer,

as depicted in Fig. 1. Since the collection of shadow pixels shares content-related information (*e.g.*, intensity and texture) with the collection of long-range well-perceived pixels, we design a domain graph reasoning block to model long-range dependency between shadow and shadow-free domains. The interaction of spatial context between domains brings a richer information compensation for non-homogeneous shadow regions. Considering that RGB components are highly correlated and inevitably affected by occlusion [20], we integrate the characteristics from multiple color spaces into a unified visual representation to enhance the response of color deviation. The aggregation of RGB, HSV, and Lab color spaces allows us to incorporate the advantage of physical prior into the deep network, which improves chroma correction performance. This is rarely explored in the context of shadow removal. In a nutshell, the main contributions of this work are as follows:

- We construct an unpaired remote sensing shadow removal dataset, which enables non-expensive unsupervised learning for shadow removal. To our knowledge, this is the first publicly available benchmark in the remote sensing field, alleviating the data scarcity dilemma.
- We propose a UP-ShadowGAN with multi-stage nature, which decomposes the challenging domain mapping task into easier sub-problems to progressively restore illumination. It explores the best of both worlds between shadow removal and color transfer through prior compensation in pixel and color spaces.
- With unpaired data training, UP-ShadowGAN achieves encouraging performance in terms of both perceptual quality and quantitative metrics.

## II. RELATED WORK

In this section, we briefly review existing shadow removal methods according to the mechanism of shadow formation, *i.e.*, surface object cast and cloud cover.

### A. Shadow Removal Toward Surface Object Cast

Shadow removal methods toward surface object cast can be classified into physics model-based and deep learning-based methods, aiming to recover the original intensity of cast shadow regions and make them compatible with shadow-free regions without degradation traces.

**Physics model-based methods** incorporate prior knowledge into illumination models to remove shadows. Finlayson *et al.* [7] restored the chromaticity representation through dimensional expansion, and introduced shadow edge identification to relight pixels. Guo *et al.* [21] predicted relative illumination conditions between shadow and shadow-free regions to remove shadows. Gong and Cosker [22] employed user interaction of shadow and lit regions to highlight shadow boundary intensity changes, leading to robust shadow removal. Based on texture similarity, Zhang *et al.* [23] designed an illumination recovery operator for removing shadows and restoring texture under shadow patches. Yang *et al.* [24] incorporated the details of the intrinsic image into RGB color space to recover the correct luminance values. Arbel and Hel-Or [25] used cubic smoothing splines to explore the per-pixel scale factor in

shadows of varying width and profile, thus handling non-uniform shadows. Although physics model-based methods can improve the de-shadowing performance to some extent, they tend to produce overexposure or underexposure results since assumed illumination models do not always hold.

**Deep learning-based methods** typically rely on paired samples for supervised training. Khan *et al.* [26] predicted shadow masks through a conditional random field model and established a Bayesian formulation to remove shadows. Vicente *et al.* [27] trained a support vector machine to distinguish between shadow and shadow-free regions. Then, illumination compensation was performed by histogram matching between shadow and lit regions. Zhu *et al.* [28] designed an illumination model-driven network, which performs finer-grained identity mapping through variational optimization. Hu *et al.* [29] analyzed the spatial context through a direction-aware mechanism to progressively achieve shadow detection and removal. However, such methods are limited by the consistency between data pair distributions, leading to unsatisfactory robustness and generalization capability. To eliminate the dependence on paired samples, Liu *et al.* [19] proposed an unsupervised network with the guidance of lightness to perform shadow removal. Jin *et al.* [30] incorporated a shadow/shadow-free domain classifier into an unsupervised network to handle hard and soft shadows. Nevertheless, the inevitable color deviation of unpaired samples challenges current unsupervised methods. In contrast, our method explores the interaction and collaboration between shadow removal and color transfer in a divide-and-conquer way.

### B. Shadow Removal Toward Cloud Cover

Shadow removal methods toward cloud cover can be classified into self-complementation-based, multispectral-complementation-based, and multitemporal-complementation-based methods, aiming to correct clouds and accompanying shadows due to climatic factors and atmospheric conditions.

**Self-complementation-based methods** recover the contamination of cloud/shadow regions through geometrical flow from cloud/shadow-free regions. Feng *et al.* [31] detected the cloud/shadow locations through high reflection characteristics, then analyzed the context relationship of contrast ratios between adjacent pixels to remove cloud and cloud shadow. Maalouf *et al.* [32] employed the Bandelet transform with multi-scale grouping to represent the geometry of cloud/shadow regions, then performed the integral line continuation to reconstruct the contaminated information. Unfortunately, self-complementation-based methods fail to cope with large-scale cloud/shadow regions, which are the actualities of remote sensing scenarios.

**Multispectral-complementation-based methods** use extra cloud/shadow-free bands to reconstruct quality degradation bands. Based on compressive sensing theory, Lorenzi *et al.* [33] designed a reconstruction method under an  $\ell_0$ -norm perspective, which relies on the assumption of spectral nonstationarity to reconstruct all spectral bands with just one cloud/shadow-free band. Li *et al.* [34] fused multiple cloud/shadow-free bands to obtain a reference image through a



Fig. 2. The geographical location of URSSR. Taking the Chersky Range as an example, we collect adjacent shadow and shadow-free regions as the training sample pairs.

spatial and temporal nonlocal filter-based fusion strategy, then employed a nonnegative matrix factorization to remove cloud cover and shadow contamination. Although multispectral-complementation-based methods are capable of reconstructing large-scale degraded regions, they fail when cloud/shadow regions appear in all bands.

**Multitemporal-complementation-based methods** remove cloud and cloud shadow by integrating spectral, spatial, and temporal coherence. Chen *et al.* [35] decomposed the cloud/shadow image into the surface-reflected component and the degraded component. Then, they employed low-rank regularization to model the spectral-temporal correlation of surface-reflected information while leveraging the  $\ell_1$ -norm to describe the sparse prior of cloud/shadow. Such a manner preserved the information in cloud/shadow-free regions while reconstructing contaminated information. Xu *et al.* [36] proposed a multitemporal dictionary learning strategy to recover contaminated pixels by mining dictionaries of degraded regions and cloud/shadow-free regions in the spectral domain. However, the temporal gap between multitemporal images may bring negative effects, such as spectral reflectance variation.

### III. PROPOSED DATASET

Existing paired shadow removal datasets are typically limited by light sources, occluding objects, and fixed image capturing systems. In such a controlled environment, the scenario adaptability of trained models is hardly convincing. More importantly, there is still a gap in publicly available unpaired shadow removal benchmarks in remote sensing. To this end, we contribute the first unpaired remote sensing shadow removal dataset (URSSR), which contains 1350 shadow samples and 1230 shadow-free samples, both with 4K resolution. As shown in Fig. 2, our sample pairs are collected at the same time, *i.e.*, midday with strong direct sunlight, to avoid interference from illumination change between samples. In addition, the collection locations are multiple mountains and national parks, thus covering comprehensive image content, complex shadow morphologies, and diverse color deviations. With URSSR, we perform a comprehensive analysis for remote sensing shadow removal methods both qualitatively and

quantitatively, and shed light on future research. It also makes unsupervised remote sensing image shadow removal without the constraints of specific scenarios possible.

### IV. PROPOSED METHOD

In this section, we first introduce the overview architecture of UP-ShadowGAN. Then, we introduce the details of shadow removal and color transfer stages accordingly.

**Overview.** Our UP-ShadowGAN (Fig. 3) consists of two parts, *i.e.*, forward cycle ( $I_s \rightarrow I_f \rightarrow I_s$ ) and backward cycle ( $I_f \rightarrow I_s \rightarrow I_f$ ). These two parts complement each other and achieve the cycle consistency constraint through inter-domain inverse mapping. Such a manner learns the underlying relationship between shadow and shadow-free domains. For forward cycle, we convert the shadow image  $I_s$  to the shadow-free image  $I_f$  through the generator  $G_{s2f}$ , then utilize the discriminator  $D_f$  to distinguish whether  $I_f$  is real or fake. For mapping  $I_s \rightarrow I_f$ , the adversarial loss encourages  $G_{s2f}$  to learn the data distribution of shadow-free samples:

$$\mathcal{L}_{adv}^{s2f}(G_{s2f}, D_f) = \mathbb{E}_{I_s, I_f} [\log(1 - D_f(G_{s2f}(I_s)))] + \mathbb{E}_{I_s, I_f} [\log D_f(I_f)], \quad (1)$$

Meanwhile, we restore the shadow-free image  $I_f$  back to shadow version through another generator  $G_{f2s}$ , which alleviates artifacts caused by adversarial learning. Notably, the input of  $G_{f2s}$  is the concatenation of  $I_f$  and shadow mask  $M$ . The shadow mask is the binarization of the difference between  $I_s$  and  $I_f$ , encouraging  $G_{s2f}$  and  $G_{f2s}$  to generate the same content in shadow regions. Besides, we achieve pixel-wise similarity from the loss function perspective by the cycle consistency loss between  $G_{s2f}$  and  $G_{f2s}$ :

$$\mathcal{L}_{cycle}(G_{s2f}, G_{f2s}) = \varepsilon_{I_s \sim Q(I_s)} \| G_{f2s}(I_f, M) - I_s \|_1, \quad (2)$$

where  $\varepsilon$  represents the error and  $Q(I_s)$  represents the data distribution of  $I_s$ . In comparison, the backward cycle randomly selects a shadow mask from the forward cycle to guide the creation of diverse shadow samples, thus improving the generalization capability. We employ the adversarial loss to optimize the shadow creation process:

$$\mathcal{L}_{adv}^{f2s}(G_{f2s}, D_s) = \mathbb{E}_{I_s, I_f} [\log(1 - D_s(G_{f2s}(I_f, M)))] + \mathbb{E}_{I_s, I_f} [\log D_s(I_s)], \quad (3)$$

For mapping  $I_s \rightarrow I_f$ , we implement the cycle consistency constraint utilizing the cycle consistency loss:

$$\mathcal{L}_{cycle}(G_{f2s}, G_{s2f}) = \varepsilon_{I_f \sim Q(I_f)} \| G_{s2f}(I_s) - I_f \|_1, \quad (4)$$

In summary, the final loss  $\mathcal{L}_{final}$  of UP-ShadowGAN is the weighted sum of the losses from both forward cycle and backward cycle:

$$\mathcal{L}_{final} = \alpha(\mathcal{L}_{adv}^{s2f}(G_{s2f}, D_f) + \mathcal{L}_{adv}^{f2s}(G_{f2s}, D_s)) + \beta(\mathcal{L}_{cycle}(G_{s2f}, G_{f2s}) + \mathcal{L}_{cycle}(G_{f2s}, G_{s2f})), \quad (5)$$

where  $\alpha$  and  $\beta$  are empirically set to 1 and 10 for balancing the scales of adversarial loss and cycle consistency loss.

**Shadow Removal.** In the forward cycle and backward cycle parts, we treat the mapping between shadow and shadow-free

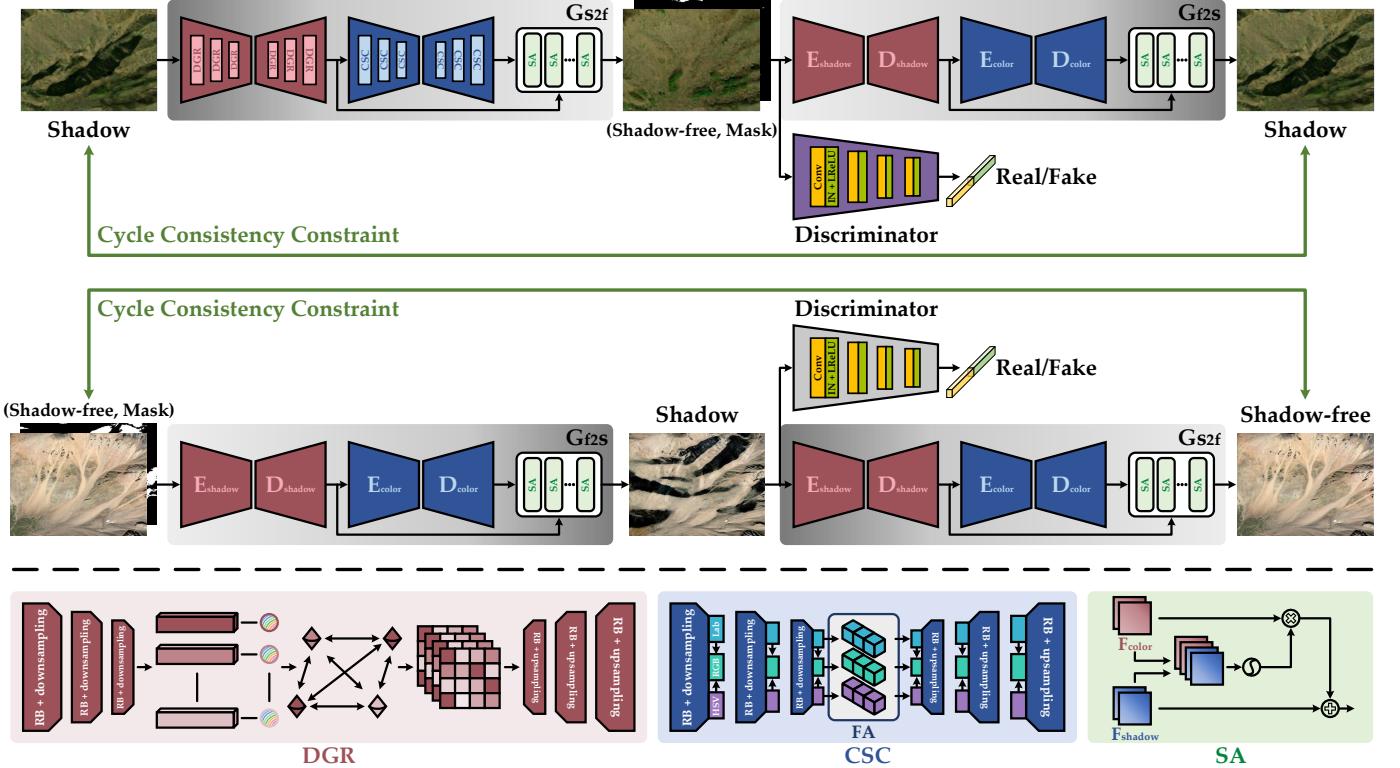


Fig. 3. The architecture of UP-ShadowGAN. UP-ShadowGAN simultaneously learns forward cycle and backward cycle to explore the potential relationship between shadow and shadow-free samples. Considering the balance between spatial details and color information, the mapping between shadow and shadow-free domains (*i.e.*,  $G_{s2f}$  and  $G_{f2s}$ ) is decoupled into more manageable steps, where  $E_{\text{shadow}}$ ,  $D_{\text{shadow}}$ ,  $E_{\text{color}}$ , and  $D_{\text{color}}$  consist of domain graph reasoning blocks (DGR) and color space collaboration blocks (SCS) with different scales, respectively. In addition, a stage aggregation block (SA) is designed to integrate the complementary merits between shadow removal and color transfer.

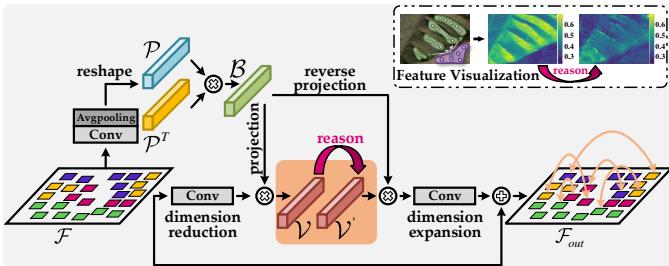


Fig. 4. The schematic illustration of the domain graph reasoning block. We project the relationship between shadow and shadow-free regions to nodes in latent interaction space, thus forming a fully connected graph. After propagating scene prior between shadow and shadow-free nodes, node features are projected back to pixel space.

domains as a joint shadow removal and color transfer task. As pointed out by [29] and [37], shadow removal requires spatial context. We analyze the spatial context of shadow and other distant regions in a graph reasoning manner, and propagate scene prior between shadow and shadow-free nodes through node-wise similarity. Specifically, we embed a well-designed domain graph reasoning block into the encoder-decoder structure. The domain graph reasoning block consists of three operations, *i.e.*, projection, reason, and reverse projection. The projection aims to project the feature  $\mathcal{F} \in \mathbb{R}^{H \times W \times C}$  to a set of spatial nodes in latent interaction space:

$$\mathcal{V} = \mathcal{P}_{ro}(\mathcal{F}) = \mathcal{B} \times \mathcal{F}, \quad (6)$$

where  $\mathcal{P}_{ro}(\cdot)$  represents the projection operation,  $\mathcal{B}$  represents the projection matrix,  $\mathcal{V} \in \mathbb{R}^{N \times C}$  represents the node feature, and  $N$  represents the node numbers. As shown in Fig. 4, we utilize the average pooling to gather the compact representation  $\mathcal{P}$ , which helps to reduce redundancy. Meanwhile, we transpose  $\mathcal{P}$  and obtain  $\mathcal{B} = \mathcal{P} \times \mathcal{P}^T$  through pixel-wise multiplication.

With the node feature  $\mathcal{V}$ , we propose to construct the edges between nodes through the adjacent matrix  $\mathcal{A} \in \mathbb{R}^{N \times N}$ . Notably,  $\mathcal{A}_{ij} = 1$  when regions  $i$  and  $j$  are spatially adjacent, otherwise  $\mathcal{A}_{ij} = 0$ . Considering that the relationships between regions are simplified to the interactions between corresponding nodes, we utilize graph convolution [38] to reason long-range dependency:

$$\mathcal{V}' = \mathcal{G}(\mathcal{V}) = \mathcal{A} \times \mathcal{V} \times \mathcal{W}, \quad (7)$$

where  $\mathcal{G}(\cdot)$  represents graph convolution and  $\mathcal{W}$  represents the weights of  $\mathcal{G}(\cdot)$ . Since the graph convolution covers the receptive field of all nodes, the spatial context of shadow and shadow-free regions are accurately captured. As depicted in the feature visualization, information from shadow-free regions is diffused to shadow regions with similar content. This is demonstrated by more consistent feature weights from shadow and shadow-free regions. After that, we reverse project  $\mathcal{V}'$  back to the pixel space, thus providing more powerful representations for subsequent color transfer:

$$\mathcal{F}_{shadow} = \mathcal{R}\mathcal{P}_{ro}(\mathcal{V}') + \mathcal{F} = \mathcal{B}^T \times \mathcal{V}' + \mathcal{F}. \quad (8)$$

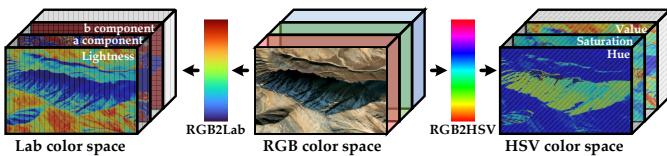


Fig. 5. Visual examples of channels extracted from RGB, HSV, and Lab color spaces. These channels are displayed by heatmaps, where the bluish to reddish colors denote the small to large values.

where  $\mathcal{RP}_{ro}(\cdot)$  represents the reverse projection operation. Besides, a residual loop is utilized as an identity connection. Without bells and whistles, such a three-step process gracefully implements inter-domain spatial context interaction.

**Color Transfer.** In view of human perception, visual content is typically stored and displayed in the RGB color space. Unfortunately, the RGB color space is vulnerable to changes in illumination and shadow. In comparison, the HSV color space is able to intuitively reflect the hue, saturation, and value of shadow regions, as depicted in Fig. 5. Besides, the Lab color space is perceptually uniform, which reflects the color difference of the visible spectrum. The complementary advantages of these color spaces motivate us to incorporate features from RGB, HSV, and Lab color spaces into a unified color representation to deal with color inconsistencies. Specifically, we establish triple information flow through color space transformation. In each information flow, the output item from the shadow removal stage is fed to the serial residual block (RB) [39] and downsampling operation to preserve data fidelity. Meanwhile, the multi-scale features of HSV and Lab information flow boost the corresponding RGB information flow through dense connections. The contributions of multiple color spaces are aggregated by the interdependence between features. Coupled with feature attention (FA) [40], the color deviation-related components (*e.g.*, hue, saturation, and lightness) from encoder features are highlighted adaptively. After two upsampling operations, the colorfully pleasing result is reconstructed. Such a color space collaboration strategy solves the color difference between unpaired data from a physical perspective with convincing interpretability.

In a multi-stage framework, the combination of synergistic design is imperative for the mapping between shadow and shadow-free domains. We show that naively cascading the shadow removal and color transfer stages produces suboptimal results [41]. Such sequential processing implies that subtasks are relatively independent, restricting information utilization. To this end, we design a stage aggregation block to implement information supplementation between different stages. Specifically, we amplify the valuable color information by learning feature weights. Meanwhile, pixel-wise addition serves as a tool for identity connection, propagating color information to shadow-free features to preserve the unique contribution of each subtask:

$$\mathcal{F}_{final} = \mathcal{F}_{shadow} + \mathcal{F}_{color} \times \mathcal{S}(\text{Cat}(\mathcal{F}_{shadow}, \mathcal{F}_{color})). \quad (9)$$

where  $\mathcal{S}(\cdot)$  represents the Sigmoid function and  $\text{Cat}(\cdot)$  represents the concatenation layer. More importantly, a simple yet effective stage aggregation block simplifies the information

flow among shadow removal and color transfer stages while stabilizing the multi-stage framework optimization.

## V. EXPERIMENTS

In this section, we describe the implementation details and experiment settings. Besides, we compare UP-ShadowGAN with state-of-the-art methods and analyze the contribution of each component through a series of ablation studies.

### A. Implementation Details

To train UP-ShadowGAN, we randomly selected 1300 shadow images from the URSSR dataset and utilized all 1230 shadow-free images as unpaired ground truths. To augment training data, we cropped image patches of size  $640 \times 360$ . For testing, we utilized the rest 50 shadow images of the URSSR dataset. We implemented UP-ShadowGAN utilizing PyTorch and trained on an NVIDIA RTX 3090 GPU. We employed ADAM with learning rate of  $2 \times 10^{-4}$  for deep model optimization. In addition, the batch size and epoch were set to 1 and 400, respectively.

### B. Experiment Settings

**Compared Methods.** We compared UP-ShadowGAN with the following representative methods.

- **Physics model-based methods:** Guo [21] based on relative illumination conditions between shadow and shadow-free regions, Gong [22] based on rough user strokes, and Silva [42] based on the illumination ratio between shaded and unshaded pixels.
- **Deep learning-based methods:** CycleGAN [18], an unsupervised method based on cycle consistency, Mask-ShadowGAN [17], an unsupervised method based on simultaneous learning of shadow mask supply and shadow removal, DC-ShadowNet [30], an unsupervised domain-classifier guided method, LG-ShadowNet [19], an unsupervised Lab channels guided method, and G2R-ShadowNet [43] an weakly-supervised method based on region cropping.

**Evaluation Metrics.** For the unpaired dataset URSSR, we conducted performance measures using several blind image quality evaluation metrics, *i.e.*, visual neuron matrix (VNM) [44], Entropy, and chroma-contrast (2C). VNM extracts visual features from shadow removal results by simulating visual neurons in the cerebral cortex. Then, the neural network is utilized to associate visual features with corresponding quality scores. Entropy serves as a statistical tool for visual features, which typically reflects the information richness of images. A higher VNM or Entropy score suggests a better visual perception. Although VNM and Entropy are applicable to low-level visual tasks, a metric dedicated to measuring the visual quality of shadow images is left unexplored. To bridge this gap, we propose a chroma-contrast metric. Inspired by [45], we employed the standard deviation <sup>1</sup>  $\sigma$  of chroma to reflect color-related perception. Considering that the difference between shadow and shadow-free regions significantly increases

<sup>1</sup><https://github.com/paulwong16/UCIQE>

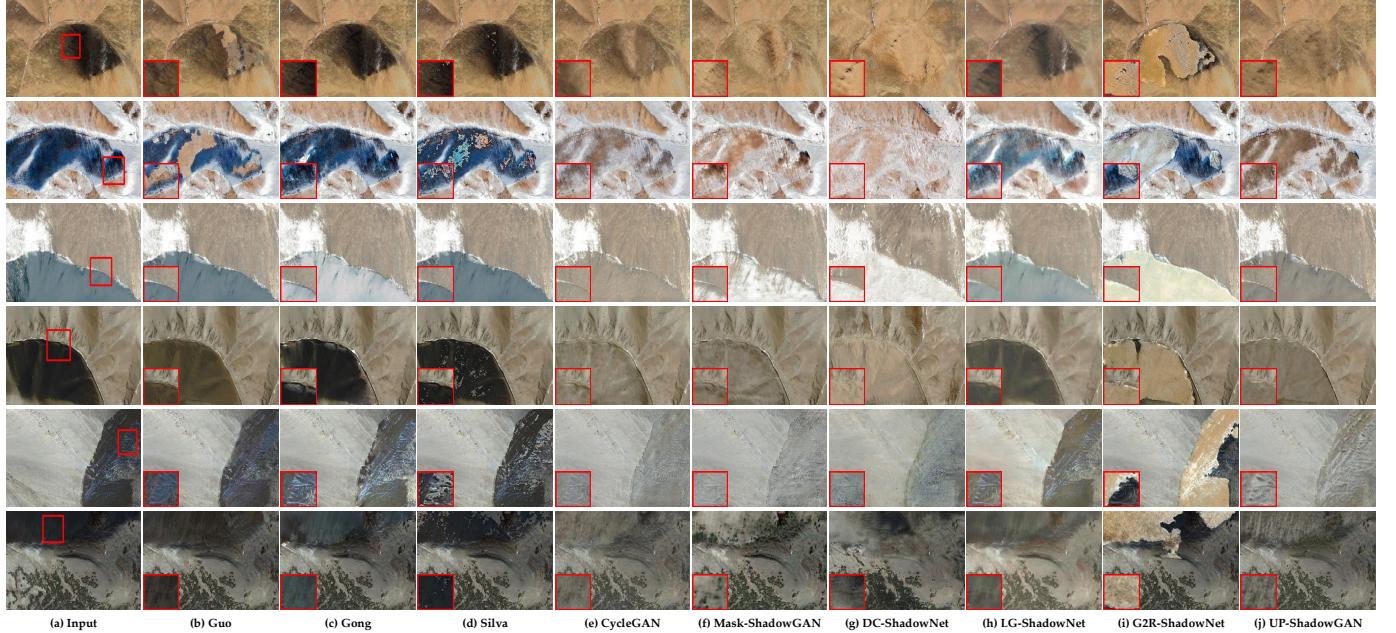


Fig. 6. Visual comparisons on shadow images sampled from **URSSR**. These images suffer from obvious and irregularly shaped shadows. Moreover, the diverse color appearance challenges current shadow removal methods. The red boxes of shadow removal methods are enlarged clips corresponding to the red box of the input image.

TABLE I

THE AVERAGE VNM, ENTROPY, AND 2C SCORES ON **URSSR**. THE BEST SCORE IS IN **RED** AND THE SECOND-BEST IS IN **BLUE**. THE REQUIRED MASK IS PROVIDED BY [46].

Methods	Publication	VNM $\uparrow$	Entropy $\uparrow$	2C $\uparrow$
Guo [21]	TPAMI'13	0.3089	6.6653	0.1175
Gong [22]	BMVC'14	0.3010	6.6002	0.1130
Silva [42]	ISPRS'18	0.2993	6.7209	0.1148
CycleGAN [18]	ICCV'17	0.3581	6.7909	0.1231
Mask-ShadowGAN [17]	ICCV'19	<b>0.3623</b>	<b>6.9372</b>	<b>0.1246</b>
DC-ShadowNet [30]	ICCV'21	0.3572	6.8337	0.1208
LG-ShadowNet [19]	TIP'21	0.3188	6.7473	0.1135
G2R-ShadowNet [43]	CVPR'21	0.3319	6.7123	0.1173
UP-ShadowGAN	–	<b>0.3682</b>	<b>7.0778</b>	<b>0.1254</b>

contrast, we employed contrast distortion <sup>2</sup>  $\omega$  as an attenuation item. The chroma-contrast metric is defined as:

$$2C = \sigma - \gamma \times \omega. \quad (10)$$

where  $\gamma$  is a weighted coefficient, set to 0.03. A higher 2C score denotes less distortion caused by shadow and color deviation.

### C. Visual Comparisons

We present visual comparisons on challenging shadow images sampled from URSSR in Fig. 6. For these shadow images, all physics model-based methods fail to achieve satisfactory results. For example, Guo [21] still retains shadow traces, which is due to inaccurate shadow detection. Gong [22] and Silva [42] cannot cope with shadow effects and even introduce color deviations. This is because priors or assumptions are inevitably violated when facing complex remote sensing scenarios. As

shown in regions framed in red, G2R-ShadowNet [43] and CycleGAN [18] either change surface information or color tone, which is undesirable. LG-ShadowNet [19] fail to completely remove shadow effects. Although Mask-ShadowGAN [17] and DC-ShadowNet [30] can remove shadows, they fail to restore clear mountain details. In contrast, UP-ShadowGAN effectively removes shadows and maintains color consistency without obvious loss of surface information. This is credited to the effective design of progressive learning.

### D. Quantitative Comparisons

For fair quantitative comparisons, we employ the source code provided by the authors, then retrain the compared methods using the consistent training set and achieve the best quantitative scores. The average VNM, Entropy, and 2C scores of competing methods are reported in Table I. For three blind image quality evaluation metrics, UP-ShadowGAN outperforms all compared methods. Compared with the top-performing method Mask-ShadowGAN [17], UP-ShadowGAN achieves a percentage gain of 1.63%/2.03%/0.64% in terms of VNM/Entropy/2C, respectively. Such a result suggests that our method is able to render visually pleasing shadow removal images. There is an interesting finding from the quantitative comparisons. Although both UP-ShadowGAN and Mask-ShadowGAN employ the cycle consistency constraint, the performance is significantly different. This demonstrates that the mutual benefits between shadow removal and color transfer are attractive.

### E. Ablation Study

We conduct extensive ablation studies to explore the core components of UP-ShadowGAN, including the shadow removal stage (SR stage), the color transfer stage (CT stage),

<sup>2</sup><https://github.com/imfing/CEIQ>

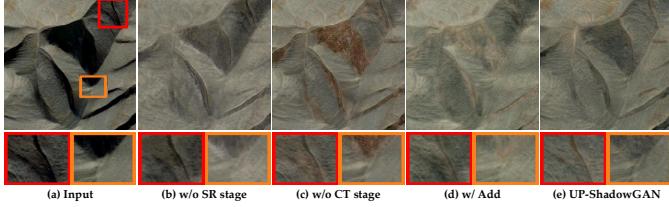


Fig. 7. Ablation study of the contributions of progressive learning. Compared with ablated models, UP-ShadowGAN combines the complementary advantages of shadow removal and color transfer, demonstrating progressive learning is an encouraging direction to explore.

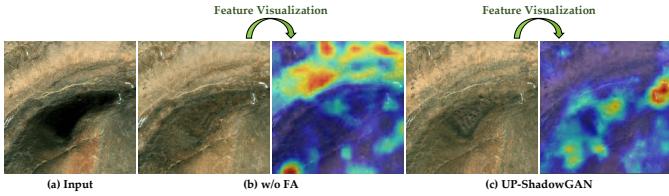


Fig. 8. Ablation study of the contributions of feature attention. More realistic mountain texture and color appearance are achieved by UP-ShadowGAN than the ablated model w/o FA.

the stage aggregation block, the feature attention block (FA), and color space encoder-decoder paths. More specifically,

- w/o SR stage represents UP-ShadowGAN without the shadow removal stage.
- w/o CT stage represents UP-ShadowGAN without the color transfer stage.
- w/ Add is the model by replacing the stage aggregation block with pixel-wise addition.
- w/o FA represents UP-ShadowGAN without the feature attention block.
- w/o HSV, w/o Lab, and w/o RGB represent UP-ShadowGAN without HSV encoder-decoder path, Lab encoder-decoder path, and RGB encoder-decoder path, respectively.

**Ablation Study of Progressive Learning.** In Fig. 7, the ablated models w/o SR stage and w/o CT stage either remain obvious shadows or introduce color artifacts, as depicted in regions framed in red and orange. In addition, the ablated model w/ Add produces visually incongruous shadow boundaries and colors. In contrast, our full model removes shadows while maintaining the consistency of restored illumination and surroundings. Such visual comparisons suggest the effectiveness of the combination of shadow removal stage and color transfer stage. More importantly, the well-designed stage aggregation block is promising for integrating merits between subtasks.

**Ablation Study of Feature Attention.** As shown in Fig. 8, our full model pay more attention to quality-degraded regions and highlights the most discriminative features extracted from RGB, HSV, and Lab color spaces. In contrast, the ablated model w/o FA fails to cope with shadows of irregular shapes.

**Ablation Study of Multiple Color Spaces.** As shown in Fig. 9, the ablated models w/o HSV, w/o Lab, and w/o RGB produce unsatisfactory results, e.g., the ablated model w/o HSV produces the ghost, and the ablated model w/o Lab introduces color deviation. This may be induced by

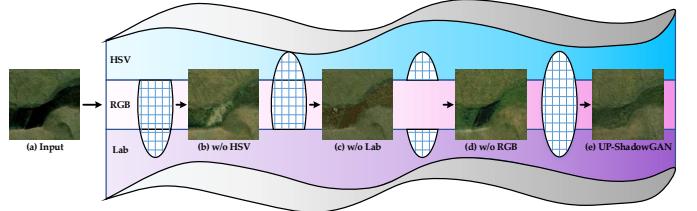


Fig. 9. Ablation study of the contributions of color spaces. Compared with ablated models, UP-ShadowGAN converges information flow from multiple color spaces, which helps to restore the vivid color of shadow regions.

removing the complementary information flow. In contrast, the aggregation of multiple color spaces contributes to the recovery of illumination-related content. Besides, removing any one of the information flow from HSV, Lab, and RGB color spaces will degrade the performance, as depicted in Fig. 10. More importantly, the nonlinear operation brought by color space transformation helps to improve the robustness.

#### F. Exploration of Scenario Adaptability

In this section, we generalize UP-ShadowGAN to shadow removal in natural scenes to demonstrate scenario adaptability. Similar to remote sensing shadow removal, each competing method is retrained using the USR dataset [17] and achieves the best quantitative results. As shown in Fig. 11, physics model-based methods either fail to remove shadows or remain undesired artifacts. This is because 1) standard and generalized user interaction paradigms are not available in practice; 2) the sequential shadow detection and removal process amplifies inter-task domain errors to some extent. CycleGAN [18], Mask-ShadowGAN [17], LG-ShadowNet [19], and G2R-ShadowGAN [43] do not have the capability to deal with obvious shadows. G2R-ShadowGAN [43] even introduces color casts. Although DC-ShadowNet [30] restores illumination in shadow regions, it also changes the hue of non-shadow regions. In contrast, UP-ShadowGAN remains color consistency while restoring illumination. Besides, the quantitative results in Table II demonstrate the superior trade-off between shadow removal and color transfer with UP-ShadowGAN.

We then generalize the proposed method to UAV imagery. The developed and compared methods are retrained on the UAV-SC dataset [47]. In Fig. 12, all compared physics model-based, unsupervised, and weakly-supervised methods fail to achieve satisfactory results, such as Guo [21], Gong [22], Silva [42], LG-ShadowNet [19], and G2R-ShadowNet [43] are ineffective for morphologically complex shadows, while CycleGAN [18], Mask-ShadowGAN [17], and DC-ShadowNet [30] change the original tones of UAV images. In contrast, UP-ShadowGAN and compared supervised methods [12], [14], [37], [56], [57] are closer to the ground-truth images. Such qualitative comparisons suggest the visually pleasing quality of our results. In terms of quantitative comparisons, our method is superior to compared physics model-based, unsupervised, and weakly-supervised methods, as shown in Table III. To the best of knowledge, supervised methods have overwhelming superiority over unsupervised methods in full-reference evaluation [17], [30], [43]. Nevertheless, our

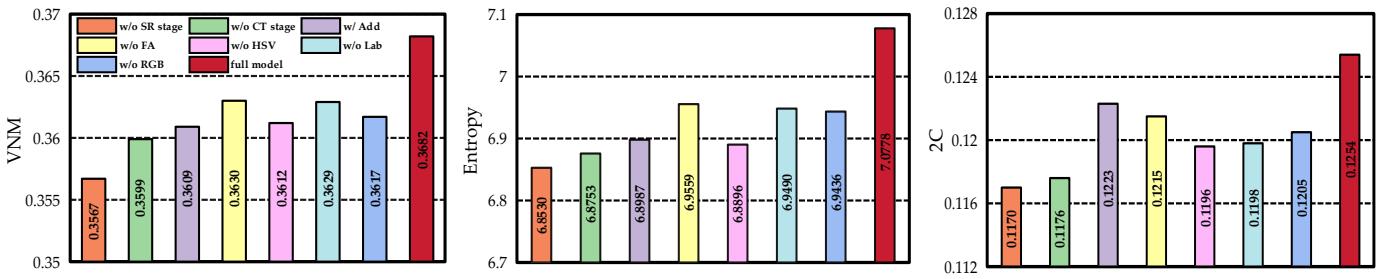


Fig. 10. Quantitative comparisons of the ablation study in terms of average VNM, Entropy, and 2C scores.

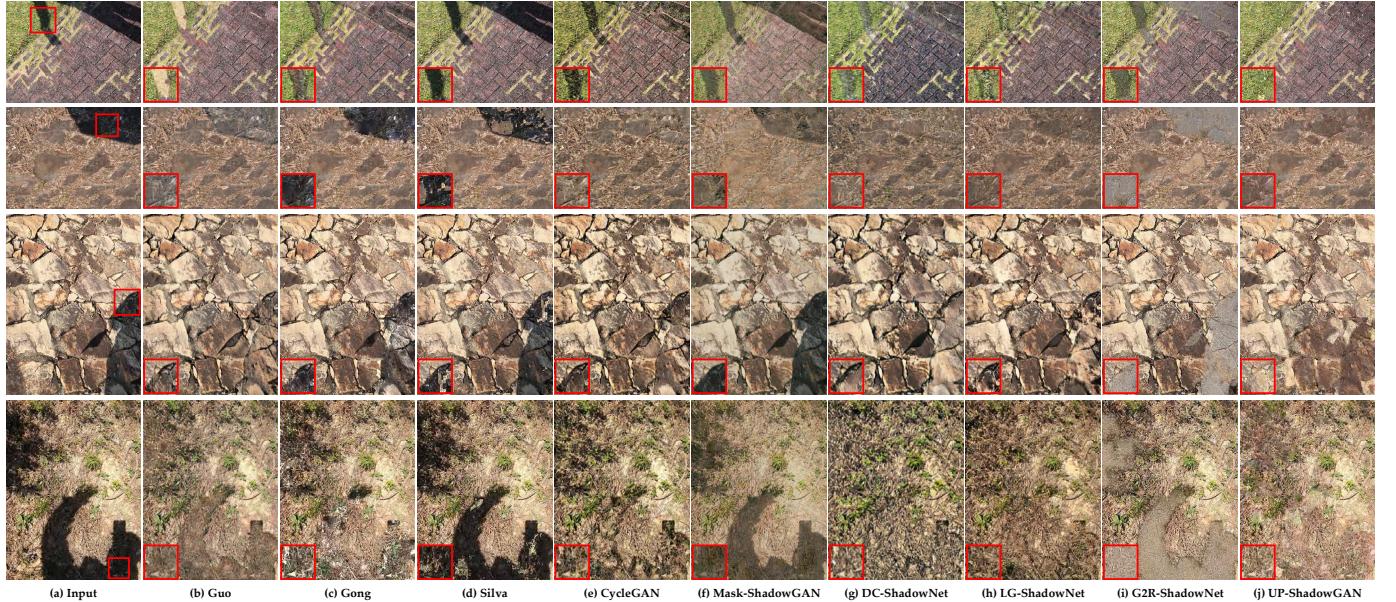


Fig. 11. Visual comparisons on shadow images sampled from **USR**. These images suffer from shadows cast by various objects, unsatisfactory colors and luminance. The red boxes of shadow removal methods are enlarged clips corresponding to the red box of the input image.



Fig. 12. Visual comparisons on shadow images sampled from **UAV-SC**. Shadows formed due to occlusion significantly hide structural details of UAV images. The red boxes of shadow removal methods are enlarged clips corresponding to the red box of the input image.

method outperforms some supervised methods, such as ST-CGAN [12]. Combined with the rigorous and expensive data collection of supervised learning, we can still conclude that UP-ShadowGAN is convincing and promising.

We also show comparisons on thick cloud images sampled from Rsipac<sup>3</sup> and WHU-Cloud [58] datasets in Figs. 13 and 14. Rsipac and WHU-Cloud both provide multitemporal cloud/shadow, cloud/shadow-free, and cloud mask triplets for training and testing. Inspired by [51], we employ WLR [52], HaLRTC [53], TMacTT [54], and TRLRF [55] as competing

methods. Notably, the cloud mask only serves WLR [52], and neither the proposed method nor the other compared methods utilize the cloud mask. Therefore, our method still follows the unsupervised learning paradigm. As shown in Figs. 13 and 14, HaLRTC [53], TMacTT [54], and TRLRF [55] fail to cope with thick clouds and cloud shadows. Some of them even introduce undesirable additive noise, such as HaLRTC [53] and TMacTT [54]. Although WLR [52] effectively removes small-scale clouds from WHU-Cloud samples, it cannot improve the visual quality of thick cloud images from Rsipac. Therefore, the scenario adaptability of WLR [52] is unsatisfactory. In

<sup>3</sup>[http://rsipac.whu.edu.cn/subject\\_one](http://rsipac.whu.edu.cn/subject_one)

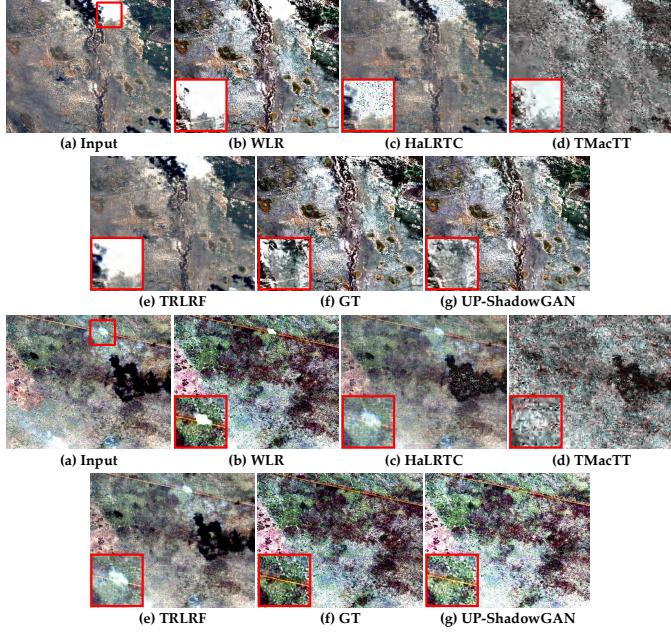


Fig. 13. Visual comparisons on cloud images sampled from **Rsipac**. Thick clouds significantly hides the structural details of surface objects. The red boxes of cloud removal methods are enlarged clips corresponding to the red box of the input image.

TABLE II  
THE AVERAGE VNM, ENTROPY, AND 2C SCORES ON **USR**. THE BEST SCORE IS IN **RED** AND THE SECOND-BEST IS IN **BLUE**. THE REQUIRED MASK IS PROVIDED BY [46].

Methods	Publication	VNM $\uparrow$	Entropy $\uparrow$	2C $\uparrow$
Guo [21]	TPAMI'13	0.3257	7.1435	0.1264
Gong [22]	BMVC'14	0.3263	7.0970	0.1259
Silva [42]	ISPRS'18	0.3115	7.0173	0.1304
CycleGAN [18]	ICCV'17	0.2791	<b>7.1876</b>	0.1192
Mask-ShadowGAN [17]	ICCV'19	<b>0.3339</b>	7.1358	<b>0.1369</b>
DC-ShadowNet [30]	ICCV'21	0.3292	6.9191	0.1063
LG-ShadowNet [19]	TIP'21	0.3003	7.0958	0.1033
G2R-ShadowNet [43]	CVPR'21	0.3322	7.0728	0.1181
UP-ShadowGAN	—	<b>0.3508</b>	<b>7.2069</b>	<b>0.1397</b>

TABLE III  
THE AVERAGE VNM, ENTROPY, 2C, PSNR (DB), AND SSIM SCORES ON **UAV-SC**. THE BEST SCORE IS IN **RED** AND THE SECOND-BEST IS IN **BLUE**. THE REQUIRED MASK IS PROVIDED BY [46].

Methods	Publication	VNM $\uparrow$	Entropy $\uparrow$	2C $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$
Guo [21]	TPAMI'13	0.4385	7.2522	0.1288	16.6687	0.6945
Gong [22]	BMVC'14	0.4399	7.2083	0.1265	16.8290	0.6984
Silva [42]	ISPRS'18	0.4234	7.2472	0.1260	16.2621	0.6898
CycleGAN [18]	ICCV'17	0.4552	7.1824	0.1143	19.7296	0.7410
Mask-ShadowGAN [17]	ICCV'19	0.4623	7.1792	0.1207	19.1357	0.7374
DC-ShadowNet [30]	ICCV'21	0.4634	7.2636	0.1302	21.0712	0.7852
LG-ShadowNet [19]	TIP'21	0.4613	<b>7.3009</b>	0.1338	18.0556	0.7629
G2R-ShadowNet [43]	CVPR'21	0.4367	7.2369	0.1253	16.8325	0.7183
ST-CGAN [12]	CVPR'18	0.5450	7.2474	0.1348	20.2338	0.7486
DHAN [14]	AAAI'20	<b>0.5817</b>	7.2629	<b>0.1395</b>	<b>24.1399</b>	<b>0.8620</b>
ShadowFormer [37]	AAAI'23	0.5512	7.2568	0.1389	<b>23.4432</b>	<b>0.8350</b>
DMTN [56]	TMM'23	<b>0.5529</b>	7.2401	0.1372	23.0287	0.8195
TBRNet [57]	TNNLS'23	0.5474	7.2427	0.1279	22.5798	0.7722
UP-ShadowGAN	—	0.4772	<b>7.2676</b>	<b>0.1393</b>	21.7753	0.7904

comparison, UP-ShadowGAN effectively remove cloud cover and shadow contamination, and achieve the best full-reference evaluation scores in Table IV.

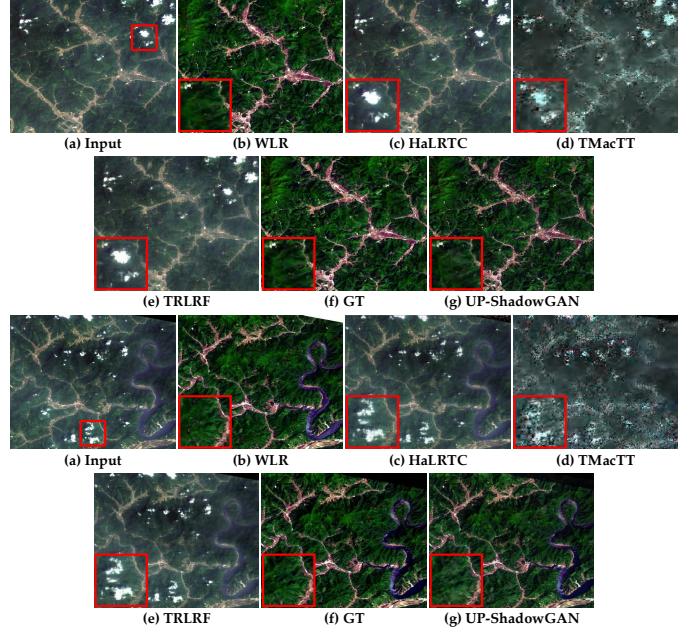


Fig. 14. Visual comparisons on cloud images sampled from **WHU-Cloud**. Thick clouds significantly hides the structural details of surface objects. The red boxes of cloud removal methods are enlarged clips corresponding to the red box of the input image.

TABLE IV  
THE AVERAGE PSNR (DB), SSIM, AND MSE ( $\times 10^3$ ) SCORES ON **RSIPAC** AND **WHU-CLOUD**. THE BEST SCORE IS IN **RED** AND THE SECOND-BEST IS IN **BLUE**.

Methods	Publication	Rsipac			WHU-Cloud		
		PSNR $\uparrow$	SSIM $\uparrow$	MSE $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	MSE $\downarrow$
WLR [52]	RSE'13	<b>24.8916</b>	<b>0.9469</b>	<b>0.2734</b>	<b>21.2475</b>	<b>0.8228</b>	<b>0.6260</b>
HaLRTC [53]	TPAMI'13	23.1564	0.9391	0.3146	15.0846	0.4881	2.0577
TMacTT [54]	TIP'17	11.2734	0.3124	4.8691	13.9210	0.3807	3.6440
TRLRF [55]	AAAI'19	15.6566	0.5048	2.6046	15.1093	0.4916	2.0618
UP-ShadowGAN	—	<b>27.5587</b>	<b>0.9601</b>	<b>0.1683</b>	<b>22.8239</b>	<b>0.8662</b>	<b>0.5083</b>

## VI. CONCLUSION

In this work, we proposed a progressive learning model, called UP-ShadowGAN, for shadow removal from unpaired remote sensing images. As the name suggests, the core insight of UP-ShadowGAN is to decouple the mapping between shadow and shadow-free domains into more manageable sub-tasks. Such a divide-and-conquer manner provides a win-win situation in terms of shadow removal and color transfer. Our method restores illumination by employing graph inference to establish long-range dependencies between shadow and shadow-free regions. Meanwhile, color attenuation is compensated by embedding visual representations from diverse color spaces. In addition, we contribute the first unpaired remote sensing shadow removal dataset. With this dataset, we comprehensively verify UP-ShadowGAN, both quantitatively and visually. In the future, we plan to provide masks for URSSR while collecting more remote sensing shadow samples (e.g., urban images). The mask as an auxiliary tool helps to alleviate color and luminosity inconsistency.

## REFERENCES

- [1] B. Sun, G. Liu, and Y. Yuan, “F3-Net: Multiview scene matching for drone-based geo-localization,” *IEEE Trans. Geosci. Remote Sens.*, vol.

- 61, pp. 1–11, May. 2023.
- [2] Q. Wang, W. Huang, Z. Xiong, and X. Li, “Looking closer at the scene: Multiscale representation learning for remote sensing image scene classification,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 33, no. 4, pp. 1414–1428, Dec. 2022.
- [3] Y. Yuan, J. Fang, X. Lu, and Y. Feng, “Remote sensing image scene classification using rearranged local features,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1779–1792, Oct. 2019.
- [4] Y. Yuan, Z. Li, and D. Ma, “Feature-aligned single-stage rotation object detection with continuous boundary,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, Sep. 2022.
- [5] W. Jing, Y. Yuan, and Q. Wang, “Dual-field-of-view context aggregation and boundary perception for airport runway extraction,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–12, May. 2023.
- [6] S. Nadimi and B. Bhanu, “Physical models for moving shadow and object detection in video,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 1079–1087, Aug. 2004.
- [7] 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.
- [8] G. D. Finlayson, M. S. Drew, and C. Lu, “Entropy minimization for shadow removal,” *Int. J. Comput. Vis.*, vol. 85, pp. 35–57, Jan. 2009.
- [9] H. Zhang, K. Sun, and W. Li, “Object-oriented shadow detection and removal from urban high-resolution remote sensing images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6972–6982, Nov. 2014.
- [10] H. Song, B. Huang, and K. Zhang, “Shadow detection and reconstruction in high-resolution satellite images via morphological filtering and example-based learning,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2545–2554, May. 2014.
- [11] L. Qu, J. Tian, S. He, Y. Tang, and R. W. H. Lau, “DeshadowNet: A multi-context embedding deep network for shadow removal,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2017, pp. 2308–2316.
- [12] 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.
- [13] H. Le and D. Samaras, “Shadow removal via shadow image decomposition,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2019, pp. 8577–8586.
- [14] 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.
- [15] H. Le and D. Samaras, “From shadow segmentation to shadow removal,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Aug. 2020, pp. 264–281.
- [16] Y. Zhu, H. Jie, 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. 2022, pp. 5617–5626.
- [17] 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.
- [18] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2017, pp. 2242–2251.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.
- [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] 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.
- [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] S. H. Khan, M. Bennamoun, F. Sohel, and R. Togneri, “Automatic shadow detection and removal from a single image,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 3, pp. 431–446, Mar. 2016.
- [27] T. F. Y. Vicente, M. Hoai, and D. Samaras, “Leave-one-out kernel optimization for shadow detection and removal,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 682–695, Mar. 2018.
- [28] Y. Zhu, Z. Xiao, Y. Fang, X. Fu, Z. Xiong, and Z.-J. Zha, “Efficient model-driven network for shadow removal,” in *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 3, 2022, pp. 3635–3643.
- [29] X. Hu, C.-W. Fu, L. Zhu, J. Qin, and P.-A. Heng, “Direction-aware spatial context features for shadow detection and removal,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 11, pp. 2795–2808, Nov. 2020.
- [30] 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.
- [31] J. Feng, L. Bai, H. Tang, S. Liu, Q. Zhou, and J. Liu, “A new context-based procedure for the detection and removal of cloud shadow from moderate-and-high resolution satellite data over land,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2010, pp. 1718–1721.
- [32] A. Maalouf, P. Carre, B. Augereau, and C. Fernandez-Maloigne, “A bandelet-based inpainting technique for clouds removal from remotely sensed images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2363–2371, Jul. 2009.
- [33] L. Lorenzi, F. Melgani, and G. Mercier, “Missing-area reconstruction in multispectral images under a compressive sensing perspective,” *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 3998–4008, Jul. 2013.
- [34] X. Li, L. Wang, Q. Cheng, P. Wu, W. Gan, and L. Fang, “Cloud removal in remote sensing images using nonnegative matrix factorization and error correction,” *ISPRS J. Photogramm. Remote Sens.*, vol. 148, pp. 103–113, Feb. 2019.
- [35] Y. Chen, W. He, N. Yokoya, and T.-Z. Huang, “Total variation regularized low-rank sparsity decomposition for blind cloud and cloud shadow removal from multitemporal imagery,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 1970–1973.
- [36] M. Xu, X. Jia, M. Pickering, and A. J. Plaza, “Cloud removal based on sparse representation via multitemporal dictionary learning,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2998–3006, May. 2016.
- [37] 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.
- [38] Y. Chen, M. Rohrbach, Z. Yan, S. Yan, J. Feng, and Y. Kalantidis, “Graph-based global reasoning networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 433–442.
- [39] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [40] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, “FFA-Net: Feature fusion attention network for single image dehazing,” in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 7, 2020, pp. 11908–11915.
- [41] S. W. Zamir *et al.*, “Multi-stage progressive image restoration,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14816–14826.
- [42] 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.
- [43] 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.
- [44] H.-W. Chang, X.-D. Bi, and C. Kai, “Blind image quality assessment by visual neuron matrix,” *IEEE Signal Process. Lett.*, vol. 28, pp. 1803–1807, Aug. 2021.
- [45] 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.
- [46] L. Zhu *et al.*, “Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Oct. 2018, pp. 122–137.
- [47] 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.
- [48] 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.

- [49] 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.
- [50] Q. Li, M. Gong, Y. Yuan, and Q. Wang, "Symmetrical feature propagation network for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, Sep. 2022.
- [51] Y. Chen, W. He, N. Yokoya, and T.-Z. Huang, "Blind cloud and cloud shadow removal of multitemporal images based on total variation regularized low-rank sparsity decomposition," *ISPRS J. Photogramm. Remote Sens.*, vol. 157, pp. 93–107, Nov. 2019.
- [52] C. Zeng, H. Shen, and L. Zhang, "Recovering missing pixels for Landsat ETM + SLC-off imagery using multi-temporal regression analysis and a regularization method," *Remote Sens. Environ.*, vol. 131, pp. 182–194, Apr. 2013.
- [53] J. Liu, P. Musalski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2013.
- [54] J. A. Bengua, H. N. Phien, H. D. Tuan, and M. N. Do, "Efficient tensor completion for color image and video recovery: Low-rank tensor train," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2466–2479, May. 2017.
- [55] L. Yuan, C. Li, D. Mandic, J. Cao, and Q. Zhao, "Tensor ring decomposition with rank minimization on latent space: An efficient approach for tensor completion," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 1, 2019, pp. 9151–9158.
- [56] 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.
- [57] 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.*, early access, Aug. 02, 2023, doi: [10.1109/TNNLS.2023.3290078](https://doi.org/10.1109/TNNLS.2023.3290078).
- [58] S. Ji, P. Dai, M. Lu, and Y. Zhang, "Simultaneous cloud detection and removal from bitemporal remote sensing images using cascade convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 732–748, Jan. 2021.



**Wei Jing** received the B.M. degree in e-commerce and the M.S. degree in computer software and theory from Shandong University of Science and Technology, Qingdao, China, in 2019 and 2022 respectively. He is currently working toward the Ph.D. degree in the National Elite Institute of Engineering and the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. His research interests include remote sensing image processing and deep learning.



**Yuan Yuan** (M'05-SM'09) is currently a Full Professor with the School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. She has authored or co-authored over 150 papers, including about 100 in reputable journals, such as the IEEE TRANSACTIONS AND PATTERN RECOGNITION, as well as the conference papers in CVPR, BMVC, ICIP, and ICASSP. Her current research interests include visual information processing and image/video content analysis.



**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.



**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.