Neural Implicit Fourier Transform for Remote Sensing Shadow Removal

Kaichen Chi, Junjie Li, Wei Jing, Qiang Li, *Member, IEEE*, and Qi Wang, *Senior Member, IEEE*

***Abstract*—Remote sensing shadow removal is an open issue. Previous studies focus on working in the spatial dimension, ignoring the potential of the Fourier dimension, while illumination degradation typically exists in the amplitude component. To address this limitation, our insight is a fresh dual-stage Fourier-based network (NeFour), which explores the best of both worlds between frequency and spatial information. In the frequency stage, we investigate the positive correlation between amplitude and brightness from channel and spatial statistics. Coupled with implicitly defined normalization, a controllable fitting amplitude transform map recreates the illumination. In the spatial stage, the inverted dark channel prior with 3D coordinates serves as modulation matrices that naturally reveal the spatial distribution of shadows, thus elegantly eliminating shadow remnants. With ingenious design, NeFour achieves nontrivial performance against state-of-the-art shadow removal methods in terms of both visual perception and quantitative evaluation. The code is publicly available at https://github.com/chi-kaichen/NeFour.**

***Index Terms*—Remote sensing imagery, shadow removal, Fourier transform, neural implicit.**

I. Introduction

SHADOWS are common natural phenomena, typically formed when light is partially or completely blocked. Such undesired illumination degradation not only fails to satisfy human perception [1], but also increases the difficulty of downstream tasks, such as object detection [2], segmentation [3], and localization [4]. Thus, it is crucial to apply shadow removal to recover contaminated surface information [5], [6].

Early methods analyze the statistics of illumination to detect and remove shadows through handcrafted priors (*e.g.*, region [7], morphology [8], and user interaction [9]). However, traditional methods typically fail when prior assumptions do not hold or shadow boundaries are intricate. Benefiting from large-scale training data and high generalization capability of neural networks, deep learning-based methods dominate this field and achieve encouraging performance. Pioneering works have been well studied from diverse perspectives, such as multi-task learning [10]–[12], image decomposition [13], [14], image generation [15], [16]. Unfortunately, these state-of-the-art methods produce visually unsatisfactory results to some extent, such as shadow remnants and artifacts. The main reason is that the design of current deep networks ignores the domain knowledge of the Fourier dimension.

Recently, Fourier transform has proven its effectiveness in low-light image enhancement [17], [18]. This is credited to the unique characteristic in the Fourier space, *i.e.*, the illumination representation is concentrated on amplitudes while phases contain the illumination-dependent structure or noise [19]. Inspired by the Fourier frequency property, we contribute a new insight for performing remote sensing shadow removal in the Fourier dimension. The design motivation for NeFour is shown in Fig. 1. For image pairs with identical content but inconsistent illumination (*i.e.*, shadow and shadow-free), we swap their amplitude components, then recombine them with corresponding phase components in the Fourier space. The amplitude components are swapped while the illumination conditions are swapped. Such observation suggests that it is feasible to remove shadows by enhancing the magnitude of amplitude component while keeping phase.

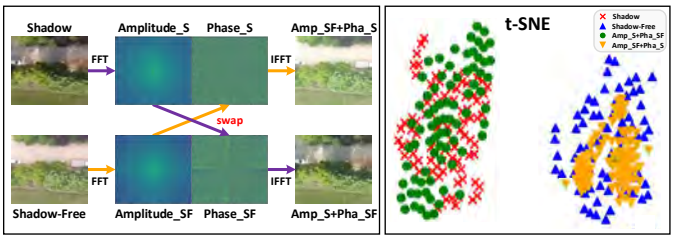

Fig. 1. Visualization of the relationship between amplitude and phase for shadow and shadow-free images. Shadow/Shadow-Free images are clustered with Amp_S+Pha_SF/Amp_SF+Pha_S versions by swapping the amplitude components, rendering similar distributions. Such observations are consistent with the t-SNE map.

Following the above rules, NeFour accomplishes shadow removal through the interaction and collaboration between frequency and spatial spaces. Notably, the spatial representation provides an understanding of global semantics, thus suggesting the presence of shadow remnants [11], [20], which is complementary to the frequency property. In the frequency stage, we propose to mine discriminative shadow representations by integrating spatial domain and channel domain Fourier transform, where the former modifies brightness through magnitude transform, and the latter serves brightness fit through channel discrepancy modeling. In particular, an implicit neural representation is incorporated into the spatial domain to normalize the diverse illumination distortions to be similar, resulting in better robustness. In the spatial stage, we embed the inverted dark channel prior into the 3D position encoder [21], which prompts the global spatial position of shadows from horizontal and vertical directions. Such a manner helps the deep model to understand "where the shadows are and what the intensity is". By and large, the main contributions of this paper are highlighted as follows:

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, lij55891@gmail.com, wei_adam@mail.nwpu.edu.cn, liqmges@gmail.com, crabwq@gmail.com).

- **Perspective contribution.** Shadow removal is innovatively treated as the learning of the positive correlation between amplitude and illumination. To the best knowledge, this is the first attempt to incorporate Fourier transform into the spatial dimension to accomplish remote sensing shadow removal.
- **Technical contribution.** A fresh dual-stage Fourier-based method capable of faithfully recovering color and texture of shadow appearance. The frequency stage focuses on enjoying the mutual benefits between Fourier transform and neural implicit for illumination recovery, while the spatial stage removes shadow remnants guided by scenario prior.
- **Practical contribution.** Extensive experiments on popular benchmarks demonstrate the superiority and scenario adaptability of NeFour.

II. Related Work

A. Shadow Removal

Classical shadow removal methods typically focus on exploring diverse physical shadow properties. Finlayson *et al.* [22] restored the chromaticity representation through dimensional expansion, and introduced shadow edge identification to relight pixels. Arbel and Hel-Or [23] used cubic smoothing splines to explore the per-pixel scale factor in shadows of varying width and profile, thus handling non-uniform shadows. Yang *et al.* [24] incorporated the details of the shadow image into RGB color space to recover the correct luminance values. Zhang *et al.* [25] employed texture similarity to establish the correspondence between shadow and shadow-free patches, then designed an illumination recovering operator to remove shadows. Movia *et al.* [26] computed specific parameters of RGB color space to perform shadow detection. Based on histogram matching, they further explored the probabilistic correspondence between sets of shadow and shadow-free pixels to remove shadows. Nevertheless, with only physical property constraints, scene-specific parameters of classical methods hardly cover real-world shadow detection and removal.

With the availability of large-scale benchmarks [10], [15], deep learning-based methods achieve great breakthroughs. Cun *et al.* [27] alleviated color inconsistency and artifacts on shadow boundaries by aggregating dilated multi-contexts. Zhu *et al.* [28] introduced a shadow illumination model, then progressively iterated the variational model to achieve fine-grained mapping between shadow and shadow-free pixels. Guo *et al.* [29] incorporated the strengths of Retinex and Transformer to leverage shadow-free regions to facilitate shadow region recovery. To address the visually disharmonious appearance, Wan *et al.* [30] estimated the style representation of shadow-free regions while designing learnable normalization to accomplish style-consistent de-shadowed results. Xu *et al.* [31] designed a shadow-aware dynamic convolution to decouple the interdependence between shadow and shadow-free pixels, thus strengthening the information flow from shadow-free regions to shadow regions. Liu *et al.* [32] regarded the structural information of shadow-free versions as guidance, then embedded it into the illumination recovery process with feature consistency regularization. Wang *et al.* [33] decomposed the mapping from the shadow domain to the shadow-free domain into pixel space and color space mappings, thus enabling shadow removal and color transfer progressively. Liu *et al.* [34] integrated the image reconstruction, shadow matte estimation, and shadow removal branches to generate an intensive information flow to recover light intensity. Fu *et al.* [35] formulated shadow removal as an exposure fusion issue, which removes shadow traces by fusing shadow and overexposed images based on the spatial-variant property. Built on a unified diffusion framework, Guo *et al.* [36] combined degradation and diffusive generative prior for highly effective illumination recovery. Zhu *et al.* [37] coupled the learning processes of shadow removal and shadow synthesis in a unified framework, then efficiently recovered colors and background contents through two-way constraints. Chen *et al.* [38] explored the physical property, spatial relation, and temporal coherence of video shadows, and then employed feature aggregation to strengthen spatio-temporal representations to recover illumination and texture. However, these methods do not deeply explore the potential power of the Fourier dimension, thus constraining their capability to learn luminance adjustment.

B. Fourier Transform

The amplitude component of Fourier space reflects the luminance representation of shadow and shadow-free regions. Inspired by this, Yu *et al.* [51] cascaded both amplitude recovery and phase recovery networks composed of residual blocks to reconstruct luminance and texture. In addition, the Fourier transform is likewise applied to the global illumination recovery task. Qiao *et al.* [18] employed the Fourier transform to simulate the degradation control factor in the image formation model, then enhanced the brightness through the bidirectional transfer of degradation rendering and physical restoration. Wang *et al.* [19] introduced the signal-to-noise-ratio map into the Fourier transform, thus recovering details while improving degradation of low-light images. Huang *et al.* [49] swapped the amplitude and phase components with various exposures to solve the mixed degradation of illumination and structure. Li *et al.* [52] refined amplitude and phase under the low-resolution regime, and then implemented minor adjustments at the high-resolution scale. Such a manner accomplishes low-light enhancement and Ultra-High-Definition compatibility. In summary, the Fourier transform is a promising direction to explore for illumination degradation.

III. Methodology

We present the overview of NeFour in Fig. 2. In what follows, we introduce the key components of NeFour, including the frequency stage, the spatial stage, and the loss function.

A. Frequency Stage

The frequency stage contributes a fresh perspective, *i.e.*, spatial domain and channel domain transform on amplitude and phase components for illumination recovery. The spatial domain Fourier transform leverages the global property of

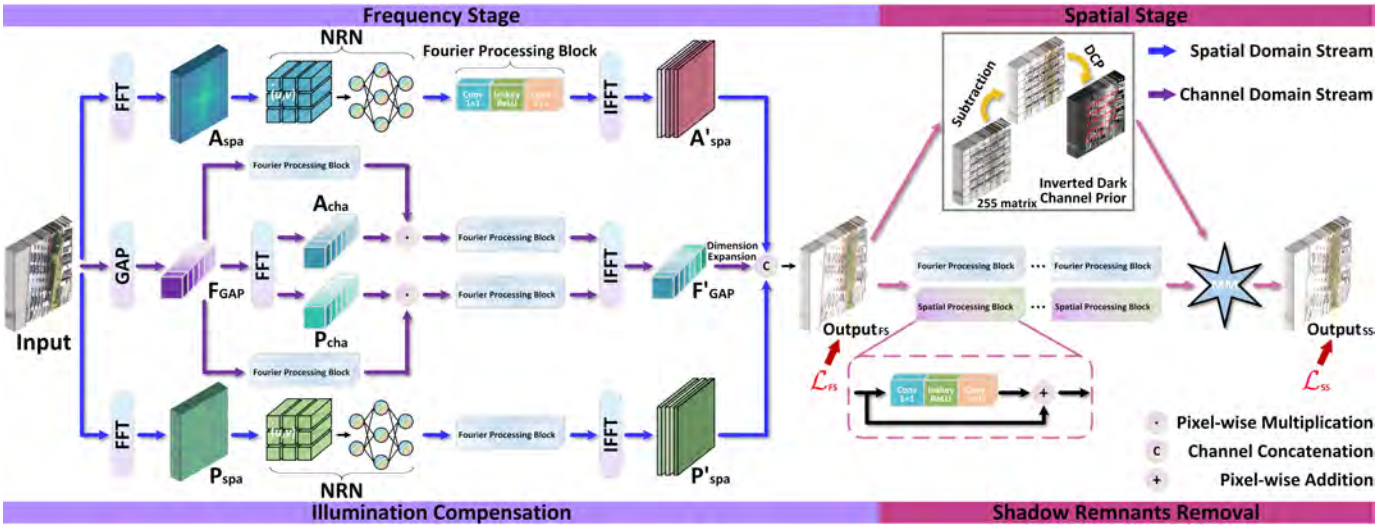

Fig. 2. Overview of proposed NeFour, consisting of a frequency stage and a spatial stage. The frequency stage exploits the global statistical property of the Fourier transform to capture representability in spatial and channel statistics to recover illumination. Notably, a neural representation normalization (NRN) is embedded into the spatial domain stream to suppress shadow boundary traces. The spatial stage considers the position embedding of shadows as a modulation matrix (MM) to remove shadow remnants, with the guidance of the inverted dark channel prior.

frequency information to relight shadow pixels. In contrast, as a valuable complement, the channel domain Fourier transform provides fitting brightness adjustment by enahncing the discriminability of global features. Both are in a win-win situation. Specifically, spatial domain and channel domain Fourier transform convert the shadow version x to the Fourier space, respectively:

$$\begin{aligned}\mathcal{F}_{spa}(x)(u,v) &= \frac{1}{\sqrt{HW}}\sum_{h=0}^{H-1}\sum_{w=0}^{W-1} x(h,w)e^{-j2\pi(\frac{h}{H}u+\frac{w}{W}v)},\\ \mathcal{F}_{cha}(y(c))(z) &= \frac{1}{C}\sum_{c=0}^{C-1} G(y(c))e^{-j2\pi\frac{c}{C}z},\end{aligned} \tag{1}$$

where h, w represent the coordinates in the spatial space, u, v represent the coordinates in the Fourier space, j represents the imaginary unit, and $G(\cdot)$ represents the global average pooling and is capable of encapsulating global representations. Correspondingly, amplitude components and phase components from the spatial domain and channel domain are expressed as:

$$\begin{aligned}\mathcal{A}_{spa}(x)(u,v) &= \sqrt{R^2(x)(u,v)+I^2(x)(u,v)},\\ \mathcal{P}_{spa}(x)(u,v) &= arctan[\frac{I(x)(u,v)}{R(x)(u,v)}],\\ \mathcal{A}_{cha}(y(c))(z) &= \sqrt{R^2(y(c))(z)+I^2(y(c))(z)},\\ \mathcal{P}_{cha}(y(c))(z) &= arctan[\frac{I(y(c))(z)}{R(y(c))(z)}],\end{aligned} \tag{2}$$

where $R(\cdot)$ and $I(\cdot)$ represent real and imaginary parts, respectively. For channel domain, we advocate a pooling layer as an identity connection since different channels exhibit different spectral properties:

$$\begin{aligned}\mathcal{A}^{'}_{cha} &= \mathcal{FP}(\mathcal{F}_{\mathcal{GAP}})\odot\mathcal{A}_{cha},\\ \mathcal{P}^{'}_{cha} &= \mathcal{FP}(\mathcal{F}_{\mathcal{GAP}})\odot\mathcal{P}_{cha},\\ \mathcal{F}^{'}_{\mathcal{GAP}} &= \mathcal{F}^{-1}(\mathcal{FP}(\mathcal{A}^{'}_{cha}),\mathcal{FP}(\mathcal{P}^{'}_{cha})),\end{aligned} \tag{3}$$

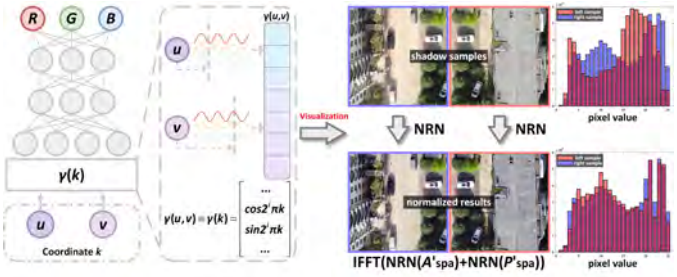

Fig. 3. Schematic illustration of the neural representation normalization, and data distribution between normalized samples. It can be found that NRN normalizes illumination to be similar.

where $\mathcal{FP}(\cdot)$ represents the Fourier processing block, consisting of convolutional layers with a Leaky ReLU activation to enrich global representations of channels. $\mathcal{F}^{-1}$ represents the inverse Fourier transform.

Unlike channel domain, the spatial domain discards some high-frequency components to some extent, such as adjacent pixels around shadow boundaries, during rendering [32]. Although their coordinates vary little, the pixel values vary a lot. Such inconsistent degradation levels are challenging for a well-trained network. Inspired by the implicit neural representation that parameterize coordinates [39], we design the neural representation normalization (NRN) to achieve more consistent degradation distribution. NRN aims to map the discrete grid of pixels $\mathbb{R}^2$ (*i.e.*, coordinate) to RGB space $\mathbb{R}^3$:

$$\mathcal{NRN}:\mathbb{R}^2\mapsto\mathbb{R}^3 \quad k\to\mathcal{NRN}(k)=[r,g,b], \tag{4}$$

In practice, 2D coordinates k are first mapped to high-dimension vectors through a high-frequency function $\gamma(\cdot)$:

$$\gamma(k)=(...,\sin(2^i\pi k),\cos(2^i\pi k),...), \quad i\in(0,8), \tag{5}$$

where i represents the dimension value and is capable of providing a more precise fit. Then, $\gamma(k)$ is passed to a multi-layer

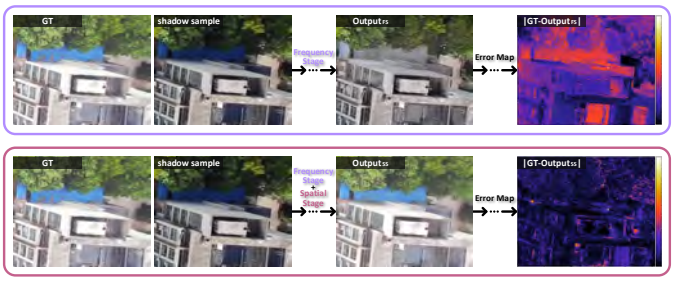

Fig. 4. Visualization of different stages of NeFour. Error maps indicate that the frequency stage improves brightness and the spatial stage removes shadow remnants.

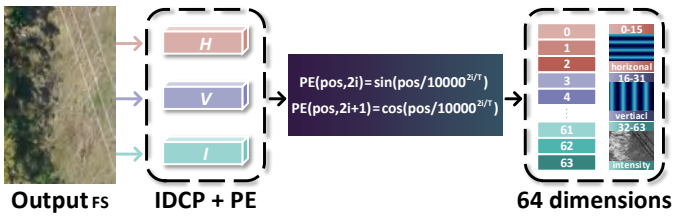

Fig. 5. Schematic illustration of the prior-guided modulation matrix. The dimensions from 0 to 15 stand for horizontal direction, from 16 to 31 stand for vertical direction, and from 32 to 63 stand for shadow intensity.

perceptron (MLP) that tends to learn an average frequency range of training sample. As depicted in Fig. 3, the normalized amplitude and phase components display similar histogram distributions, which decreases the difficulty of subsequent training procedure. In view of spatial and channel Fourier transform should have different contributions, we employ feature concatenation to implement dual-stream interaction:

$$Output_{\mathcal{FS}} = Cat[\mathcal{F}^{'}_{\mathcal{GAP}}, \mathcal{F}^{-1}(\mathcal{A}^{'}_{spa}, \mathcal{P}^{'}_{spa})], \tag{6}$$

Notably, we extend $\mathcal{F}^{'}_{\mathcal{GAP}}$ to the original resolution of $\mathbb{R}^{H\times W\times C}$ for size alignment.

B. Spatial Stage

The illumination of shadow images can be recovered well in the frequency stage, but still suffers from shadow remnants, as shown in Fig. 4. Simultaneously, based on the property of the dark channel prior, regions of low illumination intensity are naturally prompted. In this work, we prefer to provide the relative position and intensity of tiny shadows in the spatial space. Therefore, we embed the shadow-related prior knowledge into the position encoder, forming a modulation matrix consisting of horizontal dimension, vertical dimension, and shadow intensity, as shown in Fig. 5. Such a manner promotes that regions with similar shadow intensity can share similar nonlinear mapping relationships.

Specifically, we compute the inverted dark channel prior [40] of $Output_{\mathcal{FS}}$:

$$\mathcal{IDCP} = \mathbb{R}_{255} - \min_{y\in\Omega(x)}(\min_{c\in\{r,g,b\}} Output^{c}_{\mathcal{FS}}(y)), \tag{7}$$

where $\mathbb{R}_{255}$ represents a 255 matrix and $\Omega(x)$ represents a local patch centered at x. The main reason we choose the inverted dark channel prior based on its robust performance for low-light enhancement [18]. After that, we encode the position and shadow intensity information through sinusoidal position embedding:

$$\begin{aligned}\mathcal{PE}(pos, 2i) &= \sin(pos/10000^{2i/T}),\\ \mathcal{PE}(pos, 2i+1) &= \cos(pos/10000^{2i/T}),\end{aligned} \tag{8}$$

where pos represents the position of patch in horizontal direction, vertical direction, and shadow intensity. For horizontal direction, vertical direction, and intensity, we set T to 16, 16, and 32, respectively. Thus, our modulation matrix with 64 dimensions, where each row and column of patches share the same position embedding, provides relative positional relationships of shadow remnants. Moreover, the intensity embedding reveals shadow effects of diverse spatial regions.

With the prior-guided modulation matrix, the output item of NeFour is expressed as:

$$Output_{\mathcal{SS}} = \mathcal{SP}(Output_{\mathcal{FS}}) + \mathcal{M}(\mathcal{FP}(Output_{\mathcal{FS}})). \tag{9}$$

where $\mathcal{M}(\cdot)$ represents the modulation matrix and $\mathcal{SP}(\cdot)$ represents the spatial processing block with residual learning for preserving spatial information fidelity. We show that naively integrating frequency information and spatial information yields suboptimal versions, while the modulation matrix as a feature selector to efficiently implement information exchange and avoid any loss of information.

C. Loss Function

To achieve a good balance between visually pleasing appearance and quantitative scores, we employ the linear combination of frequency stage loss $\mathcal{L}_{\mathcal{FS}}$ and spatial stage loss $\mathcal{L}_{\mathcal{SS}}$, and the final loss $\mathcal{L}_{total}$ for training NeFour is expressed as:

$$\mathcal{L}_{total} = \mathcal{L}_{\mathcal{SS}} + \lambda\mathcal{L}_{\mathcal{FS}}, \tag{10}$$

where λ is set to 0.01 [49], [50] and it is employed to balance the frequency component and the spatial component. Specifically, the $\mathcal{L}_{\mathcal{FS}}$ loss explores the positive correlation between amplitude and illumination by constraining the amplitude gain:

$$\mathcal{L}_{\mathcal{FS}} = \|\mathcal{A}_{\mathcal{FS}} - \mathcal{A}_{\mathcal{GT}}\|_2, \tag{11}$$

where $\mathcal{A}_{\mathcal{FS}}$ and $\mathcal{A}_{\mathcal{GT}}$ represent the amplitude components of $Output_{\mathcal{FS}}$ and ground truth image, respectively. In [41], Zhao *et al.* suggested that the combination of ℓ_1 loss and SSIM loss is an effective driver of perceptual enhancement. A large effort has been devoted to demonstrating the effectiveness of this combinatorial loss function for super-resolution, artifact removal, denoising, and demosaicing tasks [41]. Therefore, the $\mathcal{L}_{\mathcal{SS}}$ loss likewise employs a weighted sum of ℓ_1 loss $\mathcal{L}_{\ell_1}$ and SSIM loss $\mathcal{L}_{\text{SSIM}}$ to balance illumination, color, and texture recovery:

$$\mathcal{L}_{\mathcal{SS}} = \|Output_{\mathcal{SS}} - \mathcal{GT}\|_1 + \zeta(1 - \text{SSIM}(p)). \tag{12}$$

where ζ represents the weight factor, following [41], ζ is set to 0.02. Besides, p represents the center pixel of patch.

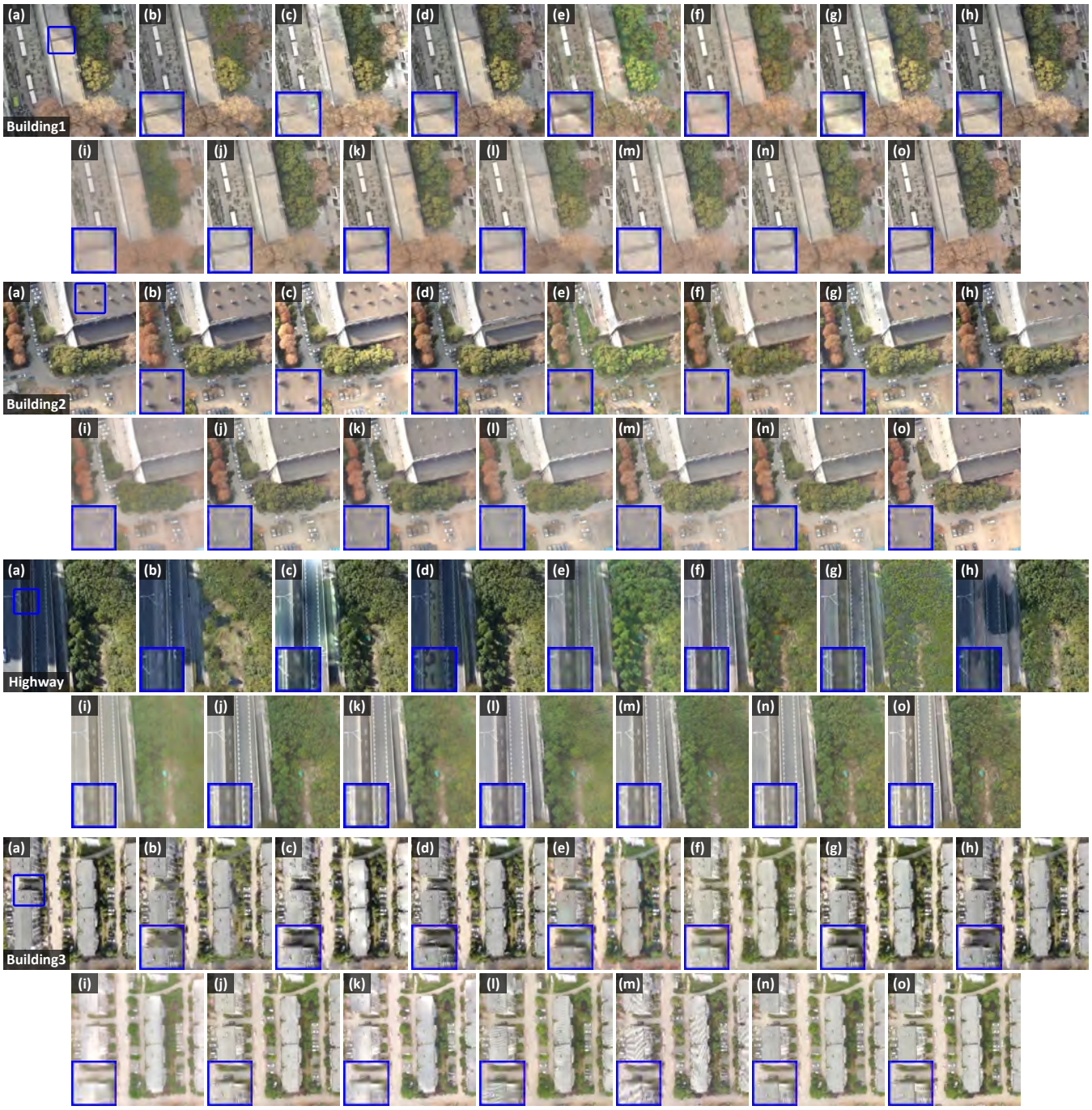

Fig. 6. Visual comparisons on shadow images with illumination variations and shadow overlaps sampled from **UAV-SC** [42]. (a) input. (b) Guo [7]. (c) Gong [9]. (d) Silva [8]. (e) Mask-ShadowGAN [15]. (f) DC-ShadowNet [44]. (g) LG-ShadowNet [45]. (h) G2R-ShadowNet [46]. (i) ST-CGAN [10]. (j) DHAN [27]. (k) ShadowFormer [29]. (l) DMTN [12]. (m) TBRNet [34]. (n) NeHour. (o) ground truth.

IV. Experiment

A. Experimental Settings

Implementation Details. The implementation of NeFour is done with PyTorch on an NVIDIA RTX 3090 GPU. The training epoch is set as 1000. A batch-mode learning strategy with a batch size of 2 is employed. ADAM is applied for network optimization and the learning rate is fixed to $1e^{-4}$. Notably, pre-training is not required for NeFour.

Benchmarks. We conduct experiments on UAV-SC [42] and AISD [43] benchmarks. **UAV-SC** consists of 6954 shadow images with corresponding shadow-free images, where 6924 image pairs are used for training and the rest 30 image pairs are used for qualitative and quantitative comparisons. In particular, UAV-SC maintains the consistency of image pairs by keeping the flight path and flight altitude of UAV constant. Besides, a standard processing procedure (*i.e.*, radiometric correction, geometric correction, and data standardization) is utilized to ensure the radiometric and geometric consistency of image pairs. **AISD**, a benchmark for shadow detection, contains 514

Fig. 7. Visual comparisons on shadow images with complex cast-surface geometric shapes and boundaries sampled from **AISD** [43]. (a) input. (b) Guo [7]. (c) Gong [9]. (d) Silva [8]. (e) Mask-ShadowGAN [15]. (f) DC-ShadowNet [44]. (g) LG-ShadowNet [45]. (h) G2R-ShadowNet [46]. (i) ST-CGAN [10]. (j) DHAN [27]. (k) ShadowFormer [29]. (l) DMTN [12]. (m) TBRNet [34]. (n) NeHour.

aerial images with shadows taken from Austin, Chicago, Tyrol, Vienna, and Innsbruck, which range from dense metropolitan financial districts to forested alpine resorts. Therefore, our training set and testing set cover diverse scenes, different illumination conditions, and a broad range of content.

Compared Methods. We compare NeFour with a series of representative shadow removal methods, including **traditional-based methods** (Guo [7], Gong [9], and Silva [8]) and **deep learning-based methods** (Mask-ShadowGAN [15], DC-ShadowNet [44], LG-ShadowNet [45], G2R-ShadowNet [46], ST-CGAN [10], DHAN [27], ShadowFormer [29], DMTN [12], and TBRNet [34]).

Evaluation Metrics. For UAV-SC, we perform full-reference evaluations by computing the root mean square error (RMSE) between the shadow-removal image and the corresponding ground truth image in the Lab color space. Moreover, we report the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) to measure the performance of competing methods in the RGB color space. A lower RMSE score suggests better performance, while the opposite is true for PSNR and SSIM scores. Notably, the above evaluation metrics are computed in the shadow region (S.), non-shadow region (N.S.), and whole image (All), respectively. For AISD without ground truths, we employ visual neuron matrix (VNM) [47] and Entropy to perform no-reference evaluations. VNM extracts visual features from shadow removal images 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. A higher VNM or Entropy score indicates a more attractive visual perception.

TABLE I
QUANTITATIVE COMPARISONS ON **UAV-SC** [42]. "↑" REPRESENTS THAT LARGER SCORES ARE BETTER, WHILE "↓" REPRESENTS THAT LOWER SCORES ARE BETTER. BEST AND SECOND-BEST SCORES ARE **HIGHLIGHTED** AND UNDERLINED. †, ‡, AND ¶ REPRESENT UNSUPERVISED, WEAKLY-SUPERVISED, AND FULLY-SUPERVISED METHODS, RESPECTIVELY.

Methods	RMSE(↓)			PSNR(↑)			SSIM(↑)		
	S.	N.S.	All	S.	N.S.	All	S.	N.S.	All
Guo [7] (TPAMI'13)	29.8381	17.9412	20.3919	23.5991	18.3971	16.6687	0.9042	0.8093	0.6945
Gong [9] (BMVC'14)	26.0850	18.8162	20.3155	25.1448	18.1944	16.8290	0.9179	0.7926	0.6984
Silva [8] (ISPRS'18)	32.0696	18.5089	21.3024	22.8902	18.1034	16.2621	0.8899	0.8035	0.6898
Mask-ShadowGAN [15] (ICCV'19)†	19.7786	17.5036	17.9722	27.5729	20.4507	19.1357	0.9450	0.8142	0.7374
DC-ShadowNet [44] (ICCV'21)†	16.1495	13.6091	14.1324	29.2519	22.4207	21.0712	0.9543	0.8479	0.7852
LG-ShadowNet [45] (TIP'21)†	23.0353	16.6409	17.9581	25.4526	19.6434	18.0556	0.9352	0.8352	0.7629
G2R-ShadowNet [46] (CVPR'21+Mask)‡	22.0473	18.6220	19.3276	24.6951	17.9899	16.8325	0.9015	0.8189	0.7183
ST-CGAN [10] (CVPR'18+Mask)¶	15.0976	13.0098	13.4399	27.4801	21.5584	20.2338	0.9525	0.8176	0.7486
DHAN [27] (AAAI'20+Mask)¶	9.1524	9.0602	9.1086	33.0654	25.1760	24.1399	0.9752	0.9010	0.8620
ShadowFormer [29] (AAAI'23+Mask)¶	9.6703	9.2937	9.5927	32.7140	24.3702	23.4432	0.9711	0.8804	0.8350
DMTN [12] (TMM'23+Mask)¶	11.5195	9.9282	10.2560	31.6098	24.1577	23.0287	0.9664	0.8696	0.8195
TBRNet [34] (TNNLS'23+Mask)¶	11.4115	10.8361	10.9546	31.2298	23.6601	22.5798	0.9619	0.8313	0.7722
NeFour	**9.0488**	**8.8464**	**8.8881**	**33.1123**	**25.3258**	**24.2568**	**0.9758**	**0.9091**	**0.8723**

B. Visual Comparisons

We first show the comparisons on challenging shadow images sampled from UAV-SC in Fig. 6. All traditional-based methods fail to recover the desired illumination, *e.g.*, Guo [7] fails to remove shadows obviously, Gong [9] introduces extra overexposure, and Silva [8] produces greenish artifacts. The main reason is that handcrafted priors of traditional-based methods are typically not satisfied for shadow scenes with texture complexity. Besides, Mask-ShadowGAN [15], DC-ShadowNet [44], and LG-ShadowNet [45] remove weak shadows to some extent, such as highway images. Nevertheless, shadows cast by diverse objects are still a challenge. G2R-ShadowNet [46] not only preserves shadow traces but even introduces severe color deviations in the highway image. Although ST-CGAN [10], DHAN [27], and ShadowFormer [29] recover satisfactory illumination, they either hide structural details or preserve shadow remnants, as depicted in zoomed-in regions of building3 images. DMTN [12] and TBRNet [34] introduce undesired streaks in building3 images. In contrast, our results are closest to the corresponding ground truth images, which is credited to the complementary nature of Fourier dimension and spatial dimension.

We then show the results of a series of methods on aerial images with obvious cast shadows in Fig. 7. All competing methods fail to achieve satisfactory visual perception. Some of them even introduce color deviations, such as Guo [7] and ST-CGAN [10]. Specifically, Guo [7] and ST-CGAN [10] change shadow regions of building2 images to greenish and grayish, respectively. In addition, Gong [9] tends to produce overexposure in shadow-free regions, such as the building3 image. Although Silva [8] is able to accurately detect shadow regions and improve illumination and color, it preserves obvious shadow boundaries. The rest compared methods either fail to cope with shadows cast by diverse objects or preserve shadow remnants. In contrast, NeFour effectively removes shadows and maintains color consistency without obvious shadow remnants and detail contamination, which demonstrates the convincing scenario adaptability of our well-design network structure.

TABLE II
QUANTITATIVE COMPARISONS ON **AISD** [43]. BEST AND SECOND-BEST SCORES ARE **HIGHLIGHTED** AND UNDERLINED. †, ‡, AND ¶ REPRESENT UNSUPERVISED, WEAKLY-SUPERVISED, AND FULLY-SUPERVISED METHODS, RESPECTIVELY.

Methods	VNM(↑)	Entropy(↑)
Guo [7] (TPAMI'13)	0.3659	6.6893
Gong [9] (BMVC'14)	0.3612	6.6056
Silva [8] (ISPRS'18)	0.3568	6.6352
Mask-ShadowGAN [15] (ICCV'19)†	0.4036	6.6548
DC-ShadowNet [44] (ICCV'21)†	0.4147	6.6232
LG-ShadowNet [45] (TIP'21)†	0.4445	6.7345
G2R-ShadowNet [46] (CVPR'21+Mask)‡	0.3386	6.6356
ST-CGAN [10] (CVPR'18+Mask)¶	0.5133	6.9166
DHAN [27] (AAAI'20+Mask)¶	0.5019	6.9464
ShadowFormer [29] (AAAI'23+Mask)¶	0.5399	6.8801
DMTN [12] (TMM'23+Mask)¶	0.5228	6.8391
TBRNet [34] (TNNLS'23+Mask)¶	0.5493	**6.9728**
NeFour	**0.5502**	6.9537

TABLE III
QUANTITATIVE SCORES OF THE ABLATION STUDY IN TERMS OF RMSE, PSNR, AND SSIM SCORES. THE BEST SCORE IS **HIGHLIGHTED**.

Baselines	UAV-SC		
	RMSE(↓)	PSNR(↑)	SSIM(↑)
w/o FS	10.5616	22.8012	0.8389
w/o SS	9.7501	23.0037	0.8507
w/o SDS	9.0082	24.0088	0.8542
w/o CDS	9.1945	23.7979	0.8523
w/o NRN	11.2553	21.0197	0.8073
w/o MM	9.2834	23.4044	0.8574
w/o $\mathcal{L}_{FS}$	10.5685	22.9411	0.8402
w/o $\mathcal{L}_{SS}$	9.4617	23.3019	0.8515
full model	**8.8881**	**24.2568**	**0.8723**

C. 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. Since weakly-supervised and fully-supervised compared methods require shadow, shadow mask, and shadow-free triplets for training, we employ BDRAR

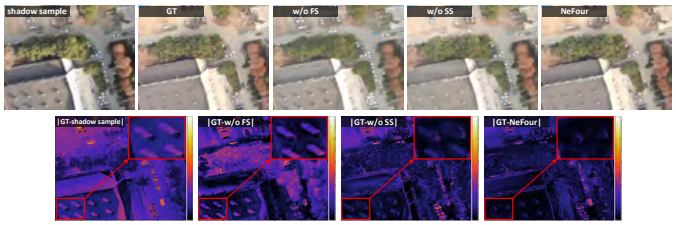

Fig. 8. Ablation study of the contributions of dual-stage framework. Compared with ablated models, NeFour improves illumination while removing shadow remnants.

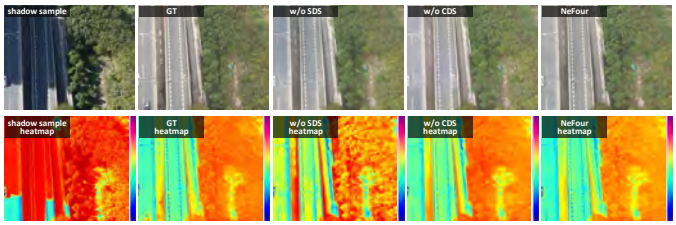

Fig. 9. Ablation study of the contributions of spatial and channel domain streams. Visualization is displayed through heatmaps, where bluish to reddish colors denote small to large values.

[48] to supplement shadow masks for UAV-SC. We first report the average scores of RMSE, PSNR, and SSIM of different methods in Table I. NeFour outperforms all compared methods on UAV-SC. Compared with the top-performing method DHAN [27], NeFour achieves a percentage gain of 2.4%/0.5%/1.2% in terms of RMSE/PSNR/SSIM, respectively. Next, we conduct non-reference evaluations on AISD. Table II reports the average VNM and Entropy scores of the results by different methods. NeFour achieves the highest VNM score while ranking the second-best in Entropy. To the best knowledge, no metric is overwhelmingly superior in state-of-the-art shadow removal quality evaluation. Thus, this still quantitatively demonstrates the superiority of NeFour.

D. Ablation Studies

We conduct extensive ablation studies to analyze the core components of NeFour, including the frequency stage, the spatial stage, the spatial domain stream, the channel domain stream, the neural representation normalization (NRN), and the modulation matrix (MM). In addition, we analyze the linear combination of the frequency stage loss $\mathcal{L}_{\mathcal{FS}}$ and the spatial stage loss $\mathcal{L}_{\mathcal{SS}}$. More specifically,

- w/o FS and w/o SS refer to NeFour without the frequency stage and spatial stage, respectively.
- w/o SDS and w/o CDS refer to the frequency stage without the spatial domain stream and channel domain stream, respectively.
- w/o NRN and w/o MM refer to NeFour without the neural representation normalization and modulation matrix, respectively.
- w/o $\mathcal{L}_{\mathcal{FS}}$ and w/o $\mathcal{L}_{\mathcal{SS}}$ mean that NeFour is trained only with the constraint of a single-stage loss.

The quantitative comparison of ablated models is depicted in Table III. The visual comparisons of the contributions

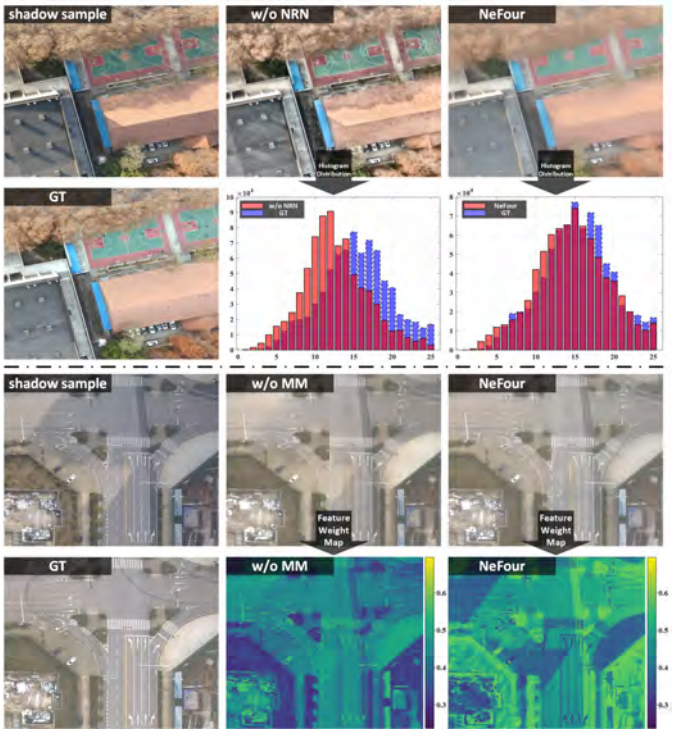

Fig. 10. Ablation study of the contributions of neural representation normalization and modulation matrix. On the one hand, NRN helps the deep network to learn pixel distributions that are more consistent with ground truths. On the other hand, MM encourages the deep network to focus more attention on shadow regions, thus removing shadow remnants.

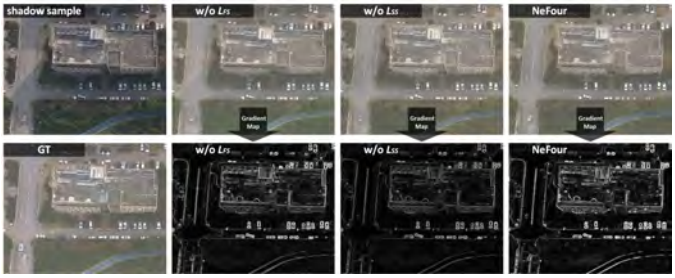

Fig. 11. Ablation study of the loss function. Compared with ablated models, NeFour renders more structural details while removing shadow traces.

of dual-stage framework, the effects of spatial and channel domain streams, the effectiveness of neural representation normalization and modulation matrix, and the effect of loss function are depicted in Fig. 8, 9, 10, and 11, respectively. The conclusions drawn from the ablation study are as follows.

- As shown in Table III, the full NeFour achieves the best quantitative scores on UAV-SC when compared with ablated models, implying the effectiveness of the combination of core components.
- Our well-design dual-stage framework promotes collaboration between illumination recovery task and shadow remnant removal task. As depicted in Fig. 8, the ablated model w/o FS presents an obvious error for shadows cast by the building, suggesting that the removal of the Fourier stage fails to recover satisfactory illumination. Besides, the ablated model w/o SS retains shadow remnants on

the error map. In contrast, NeFour naturally inherits the advantages of dual-stage with minimal error.

- In Fig. 9, the heatmap of NeFour is closer to the one of the ground truth image when compared with ablated models. Such visual comparisons indicate the illumination recovery capability from the spatial stream and the illumination fit capability from the channel stream, either should be neglected.
- As shown in Fig. 10, NRN works to guide NeFour to learn a more reasonable pixel distribution, while MM promotes NeFour to pay more attention to shadow regions through position embedding with prior knowledge. Besides, the quantitative scores of the full model are better than ablated models w/o NRN and w/o MM, which indicates the importance of NRN and MM.
- The quantitative scores in Table III show that removing the loss constraint at either stage degrades the shadow removal performance. In addition, the ablated model w/o $\mathcal{L}_{\mathcal{FS}}$ fails to alleviate obvious shadow boundaries, as shown in Fig. 11. Although the ablated model w/o $\mathcal{L}_{\mathcal{SS}}$ effectively removes shadows, the visualization of the gradient map suggests that structural details are contaminated to some extent. Therefore, a linear combination of dual-stage loss functions is crucial.

V. Conclusion

In this paper, we propose a dual-stage shadow removal network to explore the mutual benefits between the illumination recovery task and the shadow remnant removal task. The core design of the Fourier stage is the collaboration between spatial and channel streams to improve illumination by enlarging the magnitude of the amplitude component. Besides, the spatial stage employs position embedding to introduce prior knowledge, which naturally suggests the spatial location of shadow remnants. Extensive experiments demonstrate that NeFour achieves convincing shadow removal performance on multiple benchmarks. The effectiveness of core components of NeFour has been demonstrated in ablation studies.

As for our future focus, we would like to upgrade NeFour to video-level shadow removal, and plan to explore its applicability in other vision applications such as semantic segmentation. Besides, contributing a large-scale labeling benchmark with diverse shadow intensities, shapes, and positions to the community is also part of our future work.

References

[1] 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.

[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] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[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] 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, Nov. 2021.

[15] 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.

[16] 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.

[17] R. He, M. Guan, and C. Wen, "SCENS: Simultaneous contrast enhancement and noise suppression for low-light images," *IEEE Trans. Ind. Electron.*, vol. 68, no. 9, pp. 8687–8697, Sep. 2021.

[18] Y. Qiao, M. Shao, L. Wang, and W. Zuo, "Learning depth-density priors for fourier-based unpaired image restoration," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 4, pp. 2604–2618, Apr. 2024.

[19] C. Wang, H. Wu, and Z. Jin, "FourLLIE: Boosting low-light image enhancement by fourier frequency information," in *Proc. ACM Int. Conf. Multimedia*, Oct. 2023, pp. 7459–7469.

[20] Z. Wu, W. Liu, J. Li, C. Xu, and D. Huang, "SFHN: Spatial-frequency domain hybrid network for image super-resolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 11, pp. 6459–6473, Nov. 2023.

[21] C. Guo, Q. Yan, S. Anwar, R. Cong, W. Ren, and C. Li, "Image dehazing transformer with transmission-aware 3d position embedding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 5802–5810.

[22] 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.

[23] 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.

[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] 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.

[26] A. Movia, A. Beinat, and F. Crosilla, "Shadow detection and removal in RGB VHR images for land use unsupervised classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 119, pp. 485–495, Sep. 2016.

[27] 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.

[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] 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.

[30] J. Wan, H. Yin, Z. Wu, X. Wu, Y. Liu, and S. Wang, "Style-guided shadow removal," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Nov. 2022, pp. 361–378.
[31] 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.
[32] Y. Liu *et al.*, "Structure-informed shadow removal networks," *IEEE Trans. Image Process.*, vol. 32, pp. 5823–5836, Oct. 2023.
[33] Q. Wang, K. Chi, W. Jing, and Y. Yuan, "Recreating brightness from remote sensing shadow appearance," *IEEE Trans. Geosci. Remote Sens.*, early access, May. 09, 2024, doi: 10.1109/TGRS.2024.3398576.
[34] 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.
[35] 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.
[36] 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.
[37] 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.
[38] 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.*, early access, Feb. 26, 2024, doi: 10.1109/TCSVT.2024.3369910.
[39] L. Ma and F. Liu, "Motion-adjustable neural implicit video representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10728–10737.
[40] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.
[41] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imaging*, vol. 3, no. 1, pp. 47–57, Mar. 2017.
[42] 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.
[43] 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.
[44] 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.
[45] 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.
[46] 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.
[47] 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.
[48] 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.
[49] J. Huang *et al.*, "Deep fourier-based exposure correction network with spatial-frequency interaction," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Nov. 2022, pp. 163–180.
[50] M. Zhou *et al.*, "Adaptively learning low-high frequency information integration for pan-sharpening," in *Proc. ACM Int. Conf. Multimedia*, Oct. 2022, pp. 3375–3384.
[51] 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.
[52] C. Li *et al.*, "Embedding fourier for ultra-high-definition low-light image enhancement," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Feb. 2023, pp. 1–12.

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

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

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.

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.