Real-World Light Field Image Super-Resolution Via Degradation Modulation

Yingqian Wang, Zhengyu Liang, Longguang Wang, Jungang Yang, Wei An, and Yulan Guo, Senior Member, IEEE

Abstract—Recent years have witnessed the great advances of deep neural networks (DNNs) in light field (LF) image super-resolution (SR). However, existing DNN-based LF image SR methods are developed on a single fixed degradation (e.g., bicubic downsampling), and thus cannot be applied to super-resolve real LF images with diverse degradation. In this article, we propose a simple yet effective method for real-world LF image SR. In our method, a practical LF degradation model is developed to formulate the degradation process of real LF images. Then, a convolutional neural network is designed to incorporate the degradation prior into the SR process. By training on LF images using our formulated degradation, our network can learn to modulate different degradation while incorporating both spatial and angular information in LF images. Extensive experiments on both synthetically degraded and real-world LF images demonstrate the effectiveness of our method. Compared with existing state-of-the-art single and LF image SR methods, our method achieves superior SR performance under a wide range of degradation, and generalizes better to real LF images. Codes and models are available at https://yingqianwang.github.io/LF-DMnet/.

Index Terms—Degradation modulation, dynamic convolution, image super-resolution (SR), light field (LF).

I. INTRODUCTION

Light field (LF) cameras record both intensity and direction of light rays, and enable many applications such as refocusing [1], depth estimation [2], [3], [4], and view rendering [5], [6], [7]. Since high-resolution (HR) LF images are beneficial to various applications but are generally obtained at an expensive cost, it is necessary to reconstruct HR LF images from low-resolution (LR) LF images, i.e., to achieve LF image super-resolution (SR).

In the past decade, deep neural networks (DNNs) have been successfully applied to LF image SR and achieved significant progress [8], [9], [10], [11], [12]. In the area of LF image SR, many networks [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30] were developed to improve SR accuracy. However, real-world LF image SR has remained under investigated due to the following two reasons. First, it is challenging to develop an LF image SR model that can handle real-world degradation. Real-world LF images suffer from diverse degradation which varies with both imaging devices (e.g., Lytro or RayTrix cameras) and shot conditions (e.g., scene depth, focal length, and illuminance). However, existing LF image SR methods focus on the design of network architecture, and develop models on the simple bicubic downsampling degradation. Consequently, these methods suffer a notable performance drop when applied to real LF images. Second, it is challenging to simultaneously utilize the degradation information while incorporating the complementary angular information. Existing methods generally achieve real-world SR on single images (i.e., ignore the view-wise correlation), and thus cannot achieve satisfactory performance on LF image SR.

In this article, we propose a simple yet effective method for real-world LF image SR. In our method, we first formulate...
a practical degradation model to approximate the degradation process of real LF images, and then develop a convolutional neural network to super-resolve LF images with diverse and real degradation. To incorporate the degradation prior into the SR process, we design a degradation-modulating convolution (DM-Conv) whose weights are dynamically generated according to the degradation representation. By integrating the proposed DM-Conv with the disentangling mechanism [27], our network (namely, LF-DMnet) can well incorporate spatial and angular information under diverse degradation. As shown in Fig. 1, compared with DistgSSR [27] and DASR [31], our method achieves better performance on real LF images and generates images with more clear details and fewer artifacts.

The contributions of this work are summarized as follows.

1) We propose a practical LF degradation model to handle the real-world LF image SR problem. Different from existing works which focus on the advanced network designs, we first address the importance of degradation formulation and modulation in LF image SR.

2) We propose a degradation-modulating network (i.e., LF-DMnet) to incorporate the degradation prior into the SR process. Extensive ablation studies and model analyses validate the effectiveness of our degradation modulation mechanism.

3) Our method achieves state-of-the-art SR performance on both synthetic and real-world degradation, which not only provides a simple yet strong baseline, but also takes a step toward practical real-world LF image SR.

The rest of this article is organized as follows. In Section II, we briefly review the related works. In Section III, we describe our degradation model for LF image SR. In Section IV, we introduce the details and design thoughts of our LF-DMnet. Experimental results are presented in Section V. Finally, we conclude this article in Section VI.

II. RELATED WORK

In this section, we briefly review several major works for DNN-based single image SR and LF image SR.

A. Single Image Super-Resolution

The goal of single image SR is to reconstruct an HR image from its LR version. According to different degradation settings, existing single image SR methods can be roughly categorized to single degradation-based methods and multidegradation-based methods.

Early works on DNN-based single image SR are generally developed on a single and fixed degradation (e.g., bicubic downsampling). Dong et al. [34] first applied convolution neural networks to image SR and developed a three-layer network named SRCNN. Although SRCNN is shallow and lightweight, it outperforms many traditional SR methods [35], [36], [37], [38]. Since then, deep networks have dominated the SR area and achieved continuously improved accuracy with large models and complex architectures. Kim et al. [39] applied global residual learning strategy to image SR and developed a 20-layer network called VDSR. Lim et al. [40] proposed an enhanced deep SR (EDSR) network by using both global and local residual connections. Zhang et al. [41] combined residual learning with dense connection to build a residual dense network with more than 100 layers. Subsequently, Zhang et al. [42] developed a very deep network in a residual-in-residual architecture to achieve competitive SR accuracy. More recently, attention mechanism [43], [44], [45] and Transformer architectures [46], [47] have been extensively studied to achieve state-of-the-art SR performance.

Although the aforementioned methods have achieved continuously improved SR performance, they are designed for a single fixed degradation (e.g., bicubic downsampling) and will suffer from a significant performance drop when the degradation differs from the assumed one. Consequently, many methods have been proposed to achieve image SR with multiple various degradation [48]. Zhang et al. [49] proposed an SRMD network where the degradation map was concatenated with the LR image as the input of the DNN. Subsequently, Xu et al. [50] applied dynamic convolutions to achieve better SR performance than SRMD. In [51], an unfolding SR network was developed to handle different degradation by alternately solving a data subproblem and a prior subproblem. Gu et al. [52] proposed an iterative kernel correction method (namely, IKC) to correct the estimated degradation by observing previous SR results. More recently, Wang et al. [31] achieved degradation representation learning in a contrastive manner and developed a degradation-aware SR network named DASR for real-world single image SR.

B. LF Image Super-Resolution

The goal of LF image SR is to super-resolve each subaperture image (SAI) of an LF. A straightforward scheme to achieve LF image SR is applying single image SR methods to each SAI independently. However, this scheme cannot achieve a good performance since the complementary angular information among different views is not considered. Consequently, existing LF image SR methods focus on designing advanced network architectures to fully use both spatial and angular information.

Yoon et al. [13] proposed the first DNN-based method called LFCNN to enhance both spatial and angular resolution of an LF. In their method, SAIs are first super-resolved using SRCNN [34], and then finetuned in pairs or quads to incorporate angular information. Wang et al. [15] proposed a bidirectional recurrent network for LF image SR, in which the angular information in adjacent horizontal and vertical views was incorporated in a recurrent manner. Zhang et al. [19] proposed a multibranch residual network to incorporate the multidirectional epipolar geometry prior for LF image SR. In their subsequent work MEG-Net [23], the SR performance was further improved by applying 3-D convolutions to SAI stacks of different angular directions. Jin et al. [20] developed an all-to-one method for LF image SR, and performed structural consistency regularization to preserve the LF parallax structure. Wang et al. [21] developed an LF-InterNet to repetitively interact spatial and angular information for LF image SR, and then generalized the spatial-angular interaction...
mechanism to the disentangling mechanism [27] to achieve state-of-the-art SR accuracy.

More recently, Wang et al. [22] used deformable convolutions [53], [54] to address the disparity problem in LF image SR. Cheng et al. [24] proposed a zero-shot learning scheme to handle the domain gap among different LF datasets. Liang et al. [29] proposed a Transformer-based LF image SR network, in which a spatial Transformer and an angular Transformer were designed to model long range spatial dependencies and angular correlation, respectively. Wang et al. [28] proposed a detail-preserving Transformer to exploit nonlocal context information and preserve details for LF image SR. Heber et al. [30] investigated the nonlocal spatial-angular correlations in LF image SR, and developed a Transformer-based network called EPIT to achieve state-of-the-art SR performance.

Although remarkable progress have been achieved in LF image SR, existing methods only focus on the advanced network design but ignored the generalization capability to real-world degradation. In this article, we handle the real-world LF image SR problem by formulating a practical LF degradation model and designing a degradation-modulating network.

III. LF IMAGE DEGRADATION FORMULATION

In this section, we formulate a general and practical degradation model for real-world LF image SR. In Section III-A, we analyze the camera imaging process and derive the image degradation model. In Section III-B, we extend the degradation model to 4-D LF images to build the LF image degradation model, and discuss its key components. In Section III-C, we compare the differences between our method and existing SR methods.

A. Degradation Formulation

In this section, we first formulate the camera imaging process considering three key factors including point spread function (PSF), sensor sampling, and additional noise. Then, we derive the image degradation model based on the formulated camera imaging process.

Fig. 2 shows a toy example of the camera imaging process, in which the light rays are first projected onto the sensor plane [as shown in Fig. 2(a)], and then sampled by the sensor units [as shown in Fig. 2(b)]. Let $I_{\text{real}} : (x, y) \rightarrow \mathbb{R}$ be the real image (a 2-D continuous function) on the sensor plane, $k_{psf}$ be the PSF of the camera imaging system, $I_{\text{ideal}} : (x, y) \rightarrow \mathbb{R}$ be the “ideal” image (a 2-D continuous function) without considering the point spread process. According to the camera imaging process, the “real” image is obtained by convolving the “ideal” image with the PSF, i.e.,

$$I_{\text{real}}(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} k_{psf}(u, v) \cdot I_{\text{ideal}}(x - u, y - v) \, du \, dv$$

which can be denoted as

$$I_{\text{real}} = I_{\text{ideal}} \otimes k_{psf}$$

where $\otimes$ represents the convolution operation. Assume that the size of each sensor unit is $\epsilon \times \epsilon$, the sampling process on the sensor unit $(h, w)$ can be formulated as

$$I_{LR}(h, w) = \int_{h-\frac{1}{2}}^{h+\frac{1}{2}} \int_{w-\frac{1}{2}}^{w+\frac{1}{2}} I_{\text{real}}(x, y) \, dx \, dy + \mathcal{N}(h, w)$$

where $I_{LR} \in \mathbb{R}^{H\times W}$ is the output of the sensor (i.e., a digital image). Here, we introduce $\lfloor \cdot \rceil_{\epsilon}$ to denote the sampling process with a sampling grid of $\epsilon \times \epsilon$. Then, (3) can be rewritten as

$$I_{LR} = \lfloor I_{\text{real}} \rceil_{\epsilon} + \mathcal{N}$$

where $\mathcal{N} \in \mathbb{R}^{H\times W}$ represents random noise in the imaging process.

In image SR task, it is expected to reconstruct (or estimate) the ideal image function $I_{\text{ideal}}$ from the observed LR image $I_{LR}$. Since continuous 2-D image function needs to be presented via a digital image, we further introduce HR image $I_{HR} \in \mathbb{R}^\alpha H\times \alpha W$ to quantize $I_{\text{ideal}}$, i.e.,

$$I_{HR} = \lfloor I_{\text{ideal}} \rceil_{\frac{1}{\alpha}}$$

where $\alpha$ is defined as the upsampling factor. From (5), we can consider that the HR image is obtained by sampling the
ideal image $I_{\text{ideal}}$ with smaller interval $(\epsilon/\alpha) \times (\epsilon/\alpha)$. Here, we introduce a downsampling operator $(\cdot)_{(\epsilon)}$, to build the relationship between LR image and its HR version, i.e.,

$$\left[I_k\right]_{(\epsilon)} = \left[I\right]_{(\epsilon)} \odot k_{psf}, \quad I = I_{\text{ideal}} \odot I_{\text{real}}. \quad (6)$$

Substitute (2) and (6) into (4), we can obtain

$$I_{LR} = \left[I_{\text{ideal}} \odot k_{psf}\right]_{(\epsilon)} + N. \quad (7)$$

According to the commutative law of convolution and sampling (see Appendix for prove), there is

$$\left[I_{\text{ideal}} \odot k_{psf}\right]_{(\epsilon)} = \left[I_{\text{ideal}}\right]_{(\epsilon)} \odot k_{psf}. \quad (8)$$

When we substitute (8) into (7), we can obtain the image degradation model as

$$I_{LR} = (I_{\text{HR}} \odot k)_{(\epsilon)} + N. \quad (9)$$

The above degradation model can be considered as a process in which the real observed LR image is obtained by blurring, downsampling and adding noise on the HR image. In Section III-B, we will apply the degradation model to 4-D LFs to formulate our LF image degradation model.

B. LF Image Degradation Model

We use the two-plane model [55] to parameterize 4-D LF as $L \in \mathbb{R}^{U \times V \times H \times W}$, where $U$ and $V$ represent angular dimensions, $H$ and $W$ represent spatial dimensions. Since this article focuses on enhancing the spatial resolution of LFs, we use the SAI representation in [27] to describe our method. That is, an LF can be considered as a $U \times V$ array of SAs, and each SAI has a spatial size of $H \times W$.

Here, we extend our degradation model [i.e., (9)] to 4-D LFs and build the LF image degradation model as

$$T_{u,v} = \left(T_{u,v}^r \odot k_{u,v}\right)_{(\epsilon)} + N_{u,v} \quad (10)$$

where $T_{u,v}^r \in \mathbb{R}^{H \times W \times 3}$ denotes the input LR SAI of view $(u, v)$, and $T_{u,v}^r \in \mathbb{R}^{H \times W \times 3}$ denotes the corresponding HR SAI. $k_{u,v} \in \mathbb{R}^{21 \times 21}$ and $N_{u,v} \in \mathbb{R}^{H \times W \times 3}$ represent the blur kernel and additional noise of view $(u, v)$, respectively. In the following text, we introduce the details of the three key components (i.e., blur kernel, noise, and downsampling) of our LF image degradation model.

1) Blur Kernel: We follow existing works [49], [52] to use the isotropic Gaussian kernel parameterized by kernel width to synthesize blurring LF images. Note that, although anisotropic kernels (e.g., anisotropic Gaussian blur and motion blur) are also used in recent single image SR methods [31], [31], [56], [57], [58] for degradation modeling, we do not consider these blur kernels in our method because under LF structures, the rotation angle of the anisotropic Gaussian kernel and the trajectory of the motion blur of each SAI should be different but correlated. The formulation of these anisotropic blur kernels depends on the 6-D pose changing of LF cameras, and belongs to the LF deblurring task [59], [60], [61]. As demonstrated in Section V-B2, based on the isotropic Gaussian blur assumption, our method can achieve promising SR performance on real LF images.

2) Noise: Real-world LF images (especially those captured by Lytro cameras) generally have large noise. Directly super-resolving noisy LF images without performing noise reduction can result in visually unpleasant artifacts (see Section V-D2). In this article, we consider the simple channel-independent additive white Gaussian noise in our degradation process. Each element in the noise tensor $N \in \mathbb{R}^{H \times W \times 3}$ is a random variable with a mean value of 0 and an adjustable standard deviation (i.e., noise level). It is demonstrated in Section V-D2 that, when the noise term is considered in our degradation model, the super-resolved images are more smooth and clean with less noise residual and ringing artifacts.

3) Downsampling: We adopt the widely used bicubic downsampling approach in our method. In this way, our degradation model can be degeneralized to a standard bicubic downsampling degradation when the kernel width and noise level equal to zero. Note that, different from blur kernel and noise level which can vary in the training phase, the downsampling approach is assumed to be fixed.

C. Comparison to Existing Works

1) Compared to Existing LF Image SR Methods: Compared to existing LF image SR methods [19], [20], [21], [22], [23], [27], [28], [29] which use the bicubic downsampling approach to produce LR LF images, our method adopts a more practical degradation model [i.e., (10)] since the blur kernel and noise level in our model can be adjusted in the training phase to enlarge the degradation space. It is shown in Section V-B2 that our LF-DMnet trained with this degradation model can achieve promising SR performance on real LF images, which demonstrates that our proposed degradation model can well cover the real-world degradation of LF images.

2) Compared to More Complex Synthetic Degradation: It is also worth noting that several recent works for single image SR [62], [63] designed very complex degradation models to train deep networks for real-world SR. In these methods, various kinds of blur, noise, and downsampling schemes were considered, and the order of these degradation elements (also including JPEG compression) were randomly shuffled to cover as much real-world degradation as possible. Although these methods [62], [63] achieve favorable visual performance on real-world images, we do not consider designing such a complex degradation model in this article because of the following three reasons. First, single images are generally captured by various cameras and transmitted multiple times on internet, and thus go through complex and high-order degradation [63]. In contrast, LF images are captured by a few kinds of imaging devices (e.g., Lytro or RayTrix), and saved to specific file formats that do not go through JPEG compression. Consequently, the degradation space of LF images is smaller than that of single images. Second, abundant high-quality HR images and diverse scenarios are required to train a network to fit such complex degradation. Networks in [62] and [63] were trained on multiple large-scale single image datasets [64], [65], [66], [67] with thousands of high-quality HR images. In contrast, publicly available high-quality LF datasets are limited in amount, spatial resolution, and scene diversity.
Consequently, it is difficult for an LF image SR network to learn such complex degradation with insufficient training samples. Third, as the first work to address LF image SR with multiple degradation, we aim to demonstrate the importance of degradation modulation to LF image SR, and propose a simple yet effective solution to this problem. Consequently, we do not make our degradation model over-complex.

IV. NETWORK ARCHITECTURE

A. Overview

Based on the degradation model in (10), we develop a degradation-modulating network (LF-DMnet) that can super-resolve LF images with various degradation. An overview of our LF-DMnet is shown in Fig. 3(a). Given an array of LR SAIs and their corresponding degradation (i.e., kernel width and noise level of each view), our LF-DMnet sequentially performs kernel prior embedding (KPE), degradation-modulated feature extraction, and upsampling. Following [27], we build our network by cascading four residual groups. In each residual group, a degradation-modulating block (DM-Block) is designed to process features according to the degradation, and four disentangling blocks (Distg-Blocks) are used to achieve spatial-angular information incorporation. The final output of our network is an array of HR LF images. Note that, since most LF image SR methods [16], [19], [20], [21], [22], [27], [29] use SAIs distributed in a square array as their inputs, in this article, we follow these methods and set $U = V = A$, where $A$ denotes the angular resolution. In the following sections, we will introduce the details of our network design.

B. Kernel Prior Embedding

Handling image SR with multidegradation is more challenging than handling that with bicubic downsampling only, since the solution space of the former one is much larger than the latter one. In such case, incorporating kernel priors into the SR process can constrain the solution space to a mainfold and thus reduce the ill-posedness of the SR process [56]. Since only isotropic Gaussian kernel (with different kernel widths) is considered in our method, we designed a KPE module to fully incorporate the kernel prior into the SR process.

In the KPE module, the isotropic Gaussian kernel $k \in \mathbb{R}^{21 \times 21}$ is first reconstructed according to the input kernel width (i.e., the only undetermined coefficient). The reconstructed kernel is then stretched into a 1-D tensor $v_k \in \mathbb{R}^{441 \times 1}$ and fed to a multilayer perceptron (MLP) unit with five fully connected (FC) layers to learn the internal characteristics. The output of the MLP is a compact blur representation with reduced dimensionality, i.e., $v_{\text{blur}} \in \mathbb{R}^{15 \times 1}$. Finally, the generated blur representation is concatenated with the noise level to produce the final degradation representation $v_{\text{dg}} \in \mathbb{R}^{16 \times 1}$. It is demonstrated in Section V-C that the proposed KPE module is beneficial to the SR performance.

C. Degradation-Modulating Block

DM-Block is designed to process image features based on the given degradation. To achieve this goal, a simple and straightforward scheme is to concatenate degradation representation with image features and fuse them via convolutions [49], [50]. However, as demonstrated in several recent works [31], [52], directly convolving image features with degradation representations can cause interference since there is a domain gap between these two kinds of representations. Motivated by the fact that images with different degradation are generated by convolving the original high-quality image using isotropic Gaussian kernel with different kernel widths, in this article, we design a DM-Conv whose kernels are dynamically generated according to the input degradation representation.

Specifically, in each DM-Block, the degradation representation $v_{\text{dg}}$ is first fed to two FC layers to produce a convolutional kernel $w$ (with a size of $3 \times 3 \times 64$ in this article). Then, the input feature $\mathcal{F}_{\text{input}}$ is processed with DM-Conv (using $w$) and another $1 \times 1$ convolution to generate $\mathcal{F}_{\text{mod}}$. Note...
that, we follow [31] to design our DM-Conv as a depthwise direct convolution, and use a channel attention layer to reweight the output features based on the statistics of both image feature (produced by performing average pooling on $F_{\text{mod}}^{\text{spa}}$) and degradation representation. Specifically, the degradation representation $v_{\text{deg}}$ is passed to another two FC layers and a Sigmoid activation layer to generate channel-wise modulation coefficients, which are then used to rescale different channels of $F_{\text{mod}}^{\text{spa}}$, resulting in $F_{\text{mod}}$. Finally, $F_{\text{mod}}$ is summed up with $F_{\text{input}}$ and $F_{\text{mod}}^{\text{spa}}$ to produce the output of our DM-Block. It is demonstrated in Section V-C that our method benefits from DM-Conv and degradation-modulating channel attention, and can well handle LF images with various degradation.

D. Disentangling Block

Although the proposed DM-Block can handle input images with various degradation, it processes the image features of different views separately without considering the interview correlation. Since information both within a single view and among different views is beneficial to the performance of LF image SR, in this article, we modify the Distg-Block [27] to incorporate multidimensional information for LF image SR.

Different from the Distg-Block in [27] where a series of specifically designed convolutions (i.e., spatial, angular, and epipolar feature extractors) are applied to a single macropixel image (MacPI) feature, in this article, we organize LF features into different shapes and apply plain convolutions to the reshaped features. Our modified approach is equivalent to the original design but is more simple and generic. Specifically, considering both batch and channel dimensions, the input feature of our Distg-Block can be denoted by $F_{\text{in}}^{C \times D} \in \mathbb{R}^{B \times U \times V \times C \times H \times W}$, where $B$, $C$, $H$, and $W$ represent batch, channel, height, and width, respectively, and $U = V = A$ represent angular resolution. As shown in Fig. 4, our Distg-Block has a spatial branch, an angular branch and two EPI branches (i.e., horizontal and vertical). In each branch, the input feature is reshaped into a 4-D feature and then convolved by several 2-D convolutions to achieve intra and interview information incorporation.

By adopting Distg-Blocks, our method can incorporate the beneficial spatial and angular information from the input LF to achieve state-of-the-art SR performance. The effectiveness of the Distg-Block for multidegraded LF image SR is validated in Section V-C.

E. Discussion on the Nonblind SR Setting

Recent single image SR methods [31], [52], [56], [57] generally adopt the blind SR settings, i.e., the "groundtruth" degradation is unknown for the SR networks. That is because, compared to nonblind SR methods [49], [50], [51], [58] where the degradation is also required as the input, blind SR is more practical since the real-world degradation is generally difficult to obtain.

However, in this article, we adopt the nonblind SR settings as in [49], and take both degraded LF images and their degradation (blur kernel width and noise level) as inputs of our network. Reasons are in three folds. First, performing nonblind SR helps us to better investigate the impact of input degradation to the SR performance, which has not been studied in LF image SR. Since the kernel width and noise level are independently fed to our network, performing nonblind SR can help us decouple different degradation elements and investigate their influence, respectively, as demonstrated in Section V-D. Second, performing nonblind SR helps us to explore the upper bound of blind SR because the groundtruth degradation information can be used as an accurate prior in nonblind SR. As the first work to achieve LF image SR with multidegradation, one of the major contributions of this article is to break the limitation of single fixed degradation and show the great potential and practical values of multidegraded LF image SR. To this end, nonblind SR is purer and more suitable than blind SR. Third, since the proposed degradation model has only two underdetermined coefficients, we can easily find a proper input degradation by observing the super-resolved images to correct the input degradation, or adopting a grid search strategy [49] to traverse kernel widths and noise levels in a reasonable range, as described in Section V-D2.

V. Experiments

In this section, we first introduce the datasets and implementation details, then compare our network to several state-of-the-art SR methods. Finally, we conduct ablation studies to investigate our design choices and further analyze the impact of the input kernel widths and noise levels.

A. Datasets and Implementation Details

Our method was trained and validated on synthetically degraded LFs generated according to (10), and further tested on real LFs captured by Lytro Illum and Raytrix cameras. For training and validation, three public LF datasets including HCInew [68], HClOld [69], and STFgantry [70] were adopted. The division of training and validation set was kept identical to that in [22], [27], [28], [29]. To test the generalization capability of our method to real-world degradation, three public LF datasets (i.e., EPFL [32],
TABLE I

| Method     | Kernel | HCInew | HCIold | STFlytro |
|------------|--------|--------|--------|----------|
| Noise level|        |        |        |          |
|            |        | 0      | 1.5    | 3        |
| Bicubic    | 27.71/82 | 25.07/70/79 | 19.45/49/2 | 14/82/27 |
| DistSGSR   | 33.08/22 | 33.08/22 | 15.98/28 | 20/11/0 |
| LPT        | 31.23/92 | 31.43/92 | 15.49/29 | 20/10/9 |
| SRMD       | 29.55/86 | 27.78/85 | 25.37/80 | 23/76/80 |
| DASR       | 29.31/86 | 27.88/85 | 24.78/78 | nan/nan |
| BLRNet     | 28.42/86 | 24.96/83 | 19.32/74 | 13/75/63 |
| Real-ESRNet| 27.16/82 | 26.98/83 | 23.68/74 | 15/77/72 |
| Ours       | 30.43/90 | 29.55/88 | 23.83/89 | 21/72/39 |

Note: 1) For the methods in [62], [63], we used the models trained with a pixel-wise 1-loss (i.e., BLRNet and Real-ESRNet) for comparison since they can achieve higher PSNR and SSIM values as compared to their GAN-based version (i.e., BSRGAN and Real-ESRGAN). 2) SRMD and our LF-DMnet are non-blind SR methods while DASR, BLRNet and Real-ESRNet are blind SR methods.

INRIA [71], and STFlytro [33]) developed with Lytro cameras and a dataset [72] developed with a Raytray camera were used as our test sets. Totally 39, 8, and 26 scenes were used for training, validation, and test in this article, respectively.

The LFs in the HCInew [68], HCIold [69], STFlytro [70], EPFL [32], INRIA [71], and STFlytro [33] datasets have an angular resolution of 9 × 9, and the LFs in the Raytray dataset [72] have an angular resolution of 5 × 5. For LFs with an angular resolution of 9 × 9, we followed the existing works [21], [22], [27], [28], [29] to use the central 5 × 5 SAIs in our experiments. In the training phase, we cropped HR SAIs into patches of size 152 × 152 with a stride of 32, and used the proposed degradation model to synthesize LR SAI patches of size 38 × 38.

We performed random horizontal flipping, vertical flipping, 90° rotation, and RGB channel shuffling to augment the training data by 48 ×. Note that, the spatial and angular dimension need to be flipped or rotated jointly to maintain LF structures.

Our network was trained using the L1 loss and optimized using the Adam method [73] with β1 = 0.9, β2 = 0.999 and a batch size of 8. Our LF-DMnet was implemented in PyTorch on a PC with two NVidia RTX 2080Ti GPUs. The learning rate was initially set to 2 × 10−4 and decreased by a factor of 0.5 for every 3 × 104 iterations. The training was stopped after 105 iterations.

Note: For the methods in [63], [64], [65], we used the models trained with a pixel-wise 1-loss (i.e., BLRNet and Real-ESRNet) for comparison since they can achieve higher PSNR and SSIM values as compared to their GAN-based version (i.e., BSRGAN and Real-ESRGAN). 2) SRMD and our LF-DMnet are non-blind SR methods while DASR, BLRNet and Real-ESRNet are blind SR methods.

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
1) **DistgSSR [27] and LFT [29]**: Two top-performing LF image SR methods developed on the bicubic downsampling degradation.

2) **SRMD [49]**: A popular nonblind single image SR method developed on isotropic Gaussian blur and Gaussian noise degradation.

3) **DASR [31]**: A state-of-the-art blind single image SR method developed on anisotropic Gaussian blur and Gaussian noise degradation.

4) **BSRGAN [62] and Real-ESRGAN [63]**: Two recent real-world single image SR methods developed on the complex synthetic degradation.

Besides the aforementioned compared methods, we also include bicubic upsampling method to produce baseline results.

1) **Results on Synthetically Degraded LFs**: Table I shows the quantitative PSNR and SSIM results achieved by different methods under synthetic degradation with different blur and noise levels. It can be observed that DistgSSR and LFT produce the top-2 highest PSNR and SSIM results under the bicubic downsampling degradation (i.e., kernel width = 0, noise level = 0), but suffer from significant performance drop when the kernel width and noise level are larger than zero. This demonstrates that existing LF image SR methods trained on the noise-free bicubic downsampling cannot generalize well to other degradation.

SRMD and DASR achieves much better performance than DistgSSR and LFT on blurry and noisy scenes since these two methods are designed for multidegraded image SR. Note that, SRMD is benefited from the input ground-truth degradation and thus slightly outperforms DASR. It can be also observed that the PSNR and SSIM values produced by BSRNet and Real-ESRNet are lower than SRMD and DASR. That is because, the degradation space in BSRNet and Real-ESRNet are much larger, so that the capability of these two methods in handling specific degradation is less powerful. It is worth noting that these single image SR methods only use spatial context information within single views for SR but overlook the correlations among different views, resulting in inferior SR performance and the angular inconsistency issue (see Section V-B3).

Compared to these state-of-the-art single and LF image SR methods, our LF-DMnet can simultaneously incorporate the complementary angular information and adapt to different degradation, and thus achieves the best PSNR and SSIM results on both in-distribution degradation and out-of-distribution (e.g., kernel width = 4.5 or noise level = 90) degradation except for the noise-free bicubic downsampling one. The benefits of angular information and degradation adaption are further analyzed in Section V-C.

Fig. 5 shows the visual results produced by different methods with blur kernel width and noise level being set to 1.5 and 15, respectively. It can be observed that our LF-DMnet can recover faithful details from the blurry and noisy input LFs.

2) **Results on Real LFs**: We test the practical values of different SR methods by directly applying them to LFs captured by Lytro and Raytrix cameras. Since the groundtruth HR images of the input LFs are unavailable, we compare the visual results produced by different methods in Figs. 6 and 7. It can be observed that the image quality of the input LFs is low since the bicubically upsampled images are blurry and noisy. DistgSSR and LFT augment the input noise and produce results with artifacts (see Fig. 6) or blurring details (see Fig. 7). This demonstrates that methods developed on the fixed bicubic downsampling degradation cannot handle real-world degradation and thus have limited practical values.

Although SRMD, DASR, BSRGAN, and Real-ESRGAN are specifically designed to handle image SR with multiple
degradation, these methods do not consider interview correlation and ignore the beneficial angular information. Consequently, these single image SR methods suffer from noise residual (e.g., see the results of DASR in Figs. 6 and 7), oversmoothness (e.g., see the results of SRMD in Figs. 6 and 7), and angular inconsistency (e.g., see the results of BSRGAN and Real-ESRGAN in Fig. 7) issues.

Compared to existing methods, our method achieves the best SR performance on real LFs, i.e., the results produced by our method have finer details (e.g., the words and characters in scene general_11) and less artifacts. This demonstrates that our network trained on the proposed degradation model can effectively handle real LF image SR problem. Readers are referred to the videos\(^4\) to view more visual SR results on real LFs.

3) Angular Consistency: Since LF image SR methods are required to preserve the LF parallax structure and generate angular-consistent HR LF images, we evaluate the angular consistency of different SR methods by visualizing their EPI slices. As shown below the zoom-in regions in Figs. 5–7,

\(^4\)https://github.com/YingqianWang/LF-DMnet/blob/main/demo_videos.md
our LF-DMnet can generate more straight and clear line patterns than other SR methods on both synthetic and real-world degradation, which demonstrates that the LF parallax structure is well preserved by our method. Readers can refer to this video\textsuperscript{5} for a visual comparison of angular consistency.

4) Efficiency: We compare our LF-DMnet to existing SR methods in terms of the number of parameters, FLOPs, and running time. As shown in Table II, our LF-DMnet has a moderate model size which is slightly larger than DistgSSR due to the additional KPE branch and the DM-Blocks. Note that, these additional 0.27M parameters only result in a 0.52G increase in FLOPs and running time, respectively. Compared to DASR, BSRNet (i.e., BSRGAN), and Real-ESRNet (i.e., Real-ESRGAN), our method has significantly smaller model size, lower FLOPs, and shorter running time. These results demonstrate the efficiency of our method.

C. Ablation Study

In this section, we investigate the effectiveness of our proposed modules and design choices by comparing our LF-DMnet with the following variants.

1) Model 1: We introduce a baseline model by removing the DM-Block and the angular and EPI branches in the Distg-Block. Consequently, this variant is equivalent to a plain single image SR network that neither performs degradation modulation nor incorporates angular information. Note that, we increase the number of convolution layers in this variant to make its model size not smaller than our LF-DMnet.

2) Model 2: We investigate the effectiveness of our DM-Conv by replacing it with a depth-wise $3 \times 3$ convolution and a vanilla $3 \times 3$ convolution. Distg-Block is maintained in this variant to incorporate angular information. Note that, the KPE module is also removed since vanilla convolutions do not take degradation as their input. This variant can be considered as an LF image SR method without degradation modulation (e.g., DistgSSR) retrained on our proposed degradation model.

3) Model 3: In this variant, we remove the angular and EPI branches in the Distg-Block and adopt the same strategy as in Model 1 to make the model size of this variant not smaller than our LF-DMnet. Since this model only incorporates intraview information to achieve degradation-modulated SR, it can be considered as a nonblind single image SR method, and the benefits of the angular information to real-world LF image SR can be validated.

4) Model 4: We modify the KPE module in this variant to investigate the effectiveness of KPE. Specifically, we do not perform isotropic Gaussian kernel reconstruction but directly fed the blur kernel width to a five-layer MLP to generate the blur degradation representation. Consequently, the isotropic Gaussian kernel prior cannot be incorporated by this variant.

5) Model 5: In this variant, we remove the degradation-modulating channel attention layer (i.e., DM-CA) from the DM-Block to investigate the benefits of channel-wise degradation modulation.

1) Degradation-Modulating Convolution: As the core components of our LF-DMnet, DM-Conv can adapt image features to the given degradation and thus enhances the capability to handle different degradation. As shown in Table III, without using DM-Conv, Model 2 suffers from a 1.61 dB decrease in average PSNR as compared to LF-DMnet. This is because, different degradation have different spatial characteristics (as analyzed in Section IV-C) and cannot be well handled via fixed convolution kernels. In contrast, our DM-Conv dynamically generates convolutional kernels conditioned on the input degradation to recover the degraded image features, and thus achieves higher PSNR values on a wide range of synthetic degradation. Moreover, we visualize the kernels of our DM-Conv (averaged along the channel dimension) with different input blur and noise levels. As shown in Fig. 8, all the four DM-Conv kernels can be validated.

2) Angular Information: The major difference between our LF-DMnet and nonblind single image SR methods (e.g., SRMD) is the incorporation of the angular information. As shown in Table III, when the angular information is not used (i.e., Model 3), the average PSNR value suffers a 2.25 dB drop. This performance gap is also consistent with the gap between SRMD and our method in Table II. This clearly demonstrates that the complementary interview correlation is crucial for real-world LF image SR.

3) Kernel Prior Embedding: It can be observed in Table III that Model 4 without KPE suffers a 0.22 dB decrease in PSNR as compared to our LF-DMnet, and the PSNR drop is more significant on noise-free scenes. That is because, without KPE, our network has to search for the best degradation kernel to recover the degraded image features. Since we adopt the isotropic Gaussian kernel as the blur kernel for synthetic degradation, KPE can help our network to reduce the searching space and thus facilitates our network to learn more accurate kernel representations.

4) Degradation-Modulating Channel Attention: As shown in Table III, when the degradation-modulating channel

### Table II

| Method   | #Param | FLOPs | Time (S) | PSNR (dB) | SSIM   |
|----------|--------|-------|----------|-----------|--------|
| SRMD [49]| 1.50M  | 39.76G| 0.070s   | 27.18/0.838| 31.16/0.931|
| DASR [31]| 5.80M  | 82.03G| 0.051s   | 26.74/0.831| 29.47/0.896|
| BSRNet [62]| 16.70M | 459.66G| 0.119s  | 24.03/0.807| 26.17/0.856|
| Real-ESRNet [63]| 16.70M | 459.66G| 0.119s  | 25.92/0.821| 28.52/0.888|
| DistgSSR [27]| 3.53M  | 65.41G| 0.037s  | 22.79/0.611| 24.97/0.642|
| LPT [50]| 1.11M  | 28.45G| 0.070s  | 22.92/0.612| 25.23/0.647|
| LF-DMnet (ours)| 3.80M  | 65.93G| 0.039s  | 28.75/0.869| 33.69/0.939|

Note that, these additional 0.27M parameters only result in a 0.52G increase in FLOPs and running time, respectively.
attention is removed, Model 5 suffers a 0.20 dB decrease in average PSNR as compared to LF-DMNet. This demonstrates the effectiveness of channel-wise degradation modulation. Since our DM-Conv can only adapt to different degradation in the spatial dimension, DM-CA can be used as a complementary part of DM-Conv to enhance its degradation adaptation capability. It is also worth noting that our DM-CA only introduces 0.01 M increase in model size, 2 ms increase in running time, and negligible increase in FLOPs. These results demonstrate the high efficiency of our model design.

D. Degradation Mismatch Analyses

In this section, we first analyze the performance variation of our method with mismatched input and groundtruth synthetic degradation. Then, we apply our LF-DMNet to real LF images and analyze its SR performance with various input blur kernel widths and noise levels.

1) Synthetic Degradation: Since our LF-DMNet is a nonblind SR method, it requires to take the blur kernel and noise level as its input. In the aforementioned experiments with synthetic degradation, we directly use the groundtruth degradation as the input degradation of our network. To investigate the performance of our method when the input degradation mismatches with the groundtruth one, we conduct the following experiments. First, we investigate the performance variation of our LF-DMNet with mismatched blur kernel widths by traversing the groundtruth kernel width $B_{\text{gt}}$ and the input kernel width $B_{\text{in}}$ from 0 to 3 with a step of 0.3. Fig. 9(a)–(d) visualizes the PSNR values achieved by our method (averaged on the validation scenes) under four different noise levels (i.e., $N_{\text{gt}} = 0, 15, 30, 50$). Second, we investigate the performance variation of our LF-DMNet with mismatched noise levels by traversing the groundtruth noise level $N_{\text{gt}}$ and the input noise level $N_{\text{in}}$ from 0 to 50 with a step of 5. Fig. 9(e)–(h) visualizes the PSNR values achieved by our method (averaged on the validation scenes) under four different blur kernel widths (i.e., $B_{\text{gt}} = 0, 1, 2, 3$). Third, we investigate the performance variation of our LF-DMNet with simultaneously mismatched blur kernel and noise level by traversing the input blur kernel width $B_{\text{in}}$ (from 0 to 3 with a step of 0.3) and the input noise level $N_{\text{in}}$ (from 0 to 50 with a step of 5). Fig. 9(i)–(l) visualizes the PSNR values achieved by our method (averaged on the validation scenes) under four representative degradation settings including $(B_{\text{gt}}, N_{\text{gt}}) = (0, 0), (3, 0), (1.5, 15)$, and $(0, 30)$.

From Fig. 9, we can draw the following conclusions.

1) Best SR performance can be achieved when the input degradation matches the groundtruth one.
2) The performance variation caused by blur mismatch is more significant than that caused by noise mismatch.
3) When $B_{\text{in}} \neq B_{\text{gt}}$, $B_{\text{in}} > B_{\text{gt}}$ leads to much more significant performance degradation than $B_{\text{in}} < B_{\text{gt}}$.
4) As the noise level increases, the PSNR variation caused by the blur kernel mismatch is reduced.

2) Real-World Degradation: To investigate the influence of the input kernel widths and noise levels to the SR performance under real-world degradation, we directly apply our LF-DMNet to the LFs captured by Lytro and Raytrix cameras, and traverse the input blur kernel width (from 0 to 3 with a step of 1) and the input noise level (from 0 to 60 with a step of 15). Since both the groundtruth HR images and their degradation are unavailable, we evaluate the performance of our method by visually comparing its SR results. Fig. 10 shows the $4 \times$ SR results achieved by our method with varied input degradation, from which we can obtain the following conclusions.

1) A large input kernel width can enhance the local contrast and sharpens edges and textures, but an over-large

| Model | DM-Conv | DM-CA | KPE | Ang | #Params. | FLOPs | Time | Noise = 0 | Noise = 15 | Noise = 50 | Average |
|-------|---------|-------|-----|-----|--------|-------|------|-----------|-----------|-----------|---------|
| Model 1 | 3.94M | 100.9G | 0.029s | 27.87 | 27.72 | 26.97 | 26.40 | 25.71 | 24.21 | 23.40 | 22.85 | 21.87 | 25.22 |
| Model 2 | ✓ | ✓ | ✓ | 3.77M | 69.1G | 0.038s | 28.94 | 28.28 | 27.21 | 25.14 | 27.31 | 25.82 | 26.82 | 25.99 | 24.48 | 26.00 |
| Model 3 | ✓ | ✓ | ✓ | 4.01M | 97.6G | 0.030s | 27.95 | 28.00 | 27.49 | 26.43 | 25.79 | 24.35 | 23.43 | 22.92 | 21.80 | 25.36 |
| Model 4 | ✓ | ✓ | ✓ | 3.77M | 65.9G | 0.038s | 29.36 | 28.91 | 27.90 | 26.56 | 27.77 | 26.20 | 26.95 | 26.14 | 24.70 | 27.39 |
| Model 5 | ✓ | ✓ | ✓ | 3.79M | 65.9G | 0.037s | 29.15 | 29.29 | 28.25 | 25.33 | 27.85 | 26.19 | 25.84 | 26.16 | 24.63 | 27.41 |
| LF-DMNet | ✓ | ✓ | ✓ | 3.80M | 65.9G | 0.039s | 29.77 | 29.47 | 28.51 | 26.82 | 27.91 | 26.22 | 26.99 | 26.25 | 24.72 | 27.61 |

Table III

PSNR VALUES ACHIEVED BY LF-DMNET AND ITS VARIANTS FOR 4 × SR. HERE, WE REPORT THE NUMBER OF PARAMETERS (#PARAMS.), FLOPS, AND RUNNING TIME OF EACH MODEL FOR EFFICIENCY EVALUATION.

Fig. 8. Kernel visualization of our DM-Conv with different input blur and noise levels. (a) DA-Conv 1. (b) DA-Conv 2. (c) DA-Conv 3. (d) DA-Conv 4.
Fig. 9. Visualization of the performance variation of our method with mismatched degradation. (a)–(d) PSNR values achieved with mismatched blur kernel widths under different noise levels. (e)–(h) PSNR values achieved with mismatched noise level under different blurs. (i)–(l) PSNR values achieved with simultaneously mismatched blurs and noise levels under four representative degradation settings (marked by white cross).

Fig. 10. Visual results achieved by our method on real LF images with different blur kernel width and noise levels. Three scenes from the EPFL [32], STFlytro [33], and Raytrix [72] datasets are used as examples for illustration. (a) EPFL_ISO_Chart. (b) STFlytro_general_11. (c) Raytrix_boxer.

Kernel width introduces ringing artifacts to the result images. 2) A large input noise level can enhance the local smoothness and helps to alleviate the artifacts, but an
over-large input noise level makes the result images blurring.

3) Our LF-DMnet can achieve better SR performance on Lytro LFs by setting kernel width and noise level to 2 and 30, respectively, and can achieve better SR performance on Raytrix LFs by setting kernel width and noise level to 4 and 50, respectively. Readers can further refer to our interactive online demo6 to view the influence of input degradation to the SR results.

VI. CONCLUSION AND DISCUSSION

In this article, we achieve real-world LF image SR via degradation modulation. We developed an LF degradation model based on the camera imaging process, and proposed an LF-DMnet that can modulate degradation priors into the SR process. Experimental results show that our method can produce visually pleasant and angular consistent SR results on real-world LF images. Through extensive ablation studies and model analyses, we validated the effectiveness of our designs and obtained a series of insightful observations.

It is worth noting that, although our LF-DMnet achieves significantly improved performance than existing methods on real-world LF image SR, it is sensitive to the input degradation and requires accurate degradation estimation. When the input blur kernel widths and noise levels mismatch with the real ones, our method will produce images with artifacts or oversmoothness. Moreover, due to the nonblind setting in our method, when applying our method to a novel LF camera with unknown degradation, we need to first “measure” the PSF and the noise level of this camera, which is user-unfriendly and not practical enough. In the future, we will study the more challenging blind LF image SR problem, and try to design a more practical method for real-world LF image SR. We believe that our LF-DMnet will serve as a fundamental work and can inspire more researchers to focus on real-world LF image SR.

REFERENCES

[1] S. S. Jayaweera, C. U. S. Edussooriya, C. Wijenayake, P. Agathoklis, and L. T. Bruton, “Multi-volumetric refocusing of light fields,” IEEE Signal Process. Lett., vol. 28, pp. 31–35, 2021.

[2] Y. Wang, L. Wang, Z. Liang, J. Yang, W. An, and Y. Guo, “Occlusion-aware cost constructor for light field depth estimation,” in Proc. IEEE/ICCV Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2022, pp. 19809–19818.

[3] C. Shin, H.-G. Jeon, Y. Yoon, I. S. Kweon, and S. J. Kim, “EPINET: A fully-convolutional neural network using epipolar geometry for depth from light field images,” in Proc. IEEE/ICCV Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2022.

[4] W. Zhou, L. Lin, Y. Hong, Q. Li, X. Shen, and E. E. Kurnioglu, “Beyond photometric consistency: Geometry-based occlusion-aware unsupervised light field disparity estimation,” IEEE Trans. Neural Netw. Learn. Syst., early access, Jul. 4, 2023, doi: 10.1109/TNNLS.2023.3289056.

[5] C. Zhu, H. Zhang, W. Chen, M. Tan, and Q. Liu, “An occlusion compensation learning framework for improving the rendering quality of light field,” IEEE Trans. Neural Netw. Learn. Syst., vol. 32, no. 12, pp. 5738–5752, Dec. 2021.

[6] G. Wu, Y. Liu, L. Fang, and T. Chai, “Revisiting light field rendering with deep anti-aliasing neural network,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 9, pp. 5430–5444, Sep. 2022.

[7] G. Wu, Y. Wang, Y. Liu, L. Fang, and T. Chai, “Spatial-angular attention network for light field reconstruction,” IEEE Trans. Image Process., vol. 30, pp. 8999–9013, 2021.

[8] M. Zhang, Q. Wu, J. Guo, Y. Li, and X. Gao, “Heat transfer-inspired network for image super-resolution reconstruction,” IEEE Trans. Neural Netw. Learn. Syst., vol. 35, no. 2, pp. 1810–1820, Feb. 2024.

[9] C. Tian, Y. Zhang, W. Zuo, C.-W. Lin, D. Zhang, and Y. Yuan, “A heterogeneous group CNN for image super-resolution,” IEEE Trans. Neural Netw. Learn. Syst., early access, Oct. 13, 2022, doi: 10.1109/TNNLS.2022.3210433.

[10] X. Hu, Z. Zhang, C. Shan, Z. Wang, L. Wang, and T. Tan, “Meta-USR: A unified super-resolution network for multiple degradation parameters,” IEEE Trans. Neural Netw. Learn. Syst., vol. 32, no. 9, pp. 4151–4165, Sep. 2021.

[11] K. Li, “Local means binary networks for image super-resolution,” IEEE Trans. Neural Netw. Learn. Syst., early access, Oct. 21, 2022, doi: 10.1109/TNNLS.2022.3212827.

[12] G. Wu, J. Jiang, and X. Liu, “A practical contrastive learning framework for single-image super-resolution,” IEEE Trans. Neural Netw. Learn. Syst., early access, Jul. 10, 2023, doi: 10.1109/TNNLS.2023.3290038.

[13] Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, and I. S. Kweon, “Learning a deep convolutional network for light-field image super-resolution,” in Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW), Dec. 2015, pp. 24–32.

[14] Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, and I. S. Kweon, “Light-field image super-resolution using convolutional neural network,” IEEE Signal Process. Lett., vol. 24, no. 6, pp. 848–852, Jun. 2017.

[15] Y. Wang, F. Liu, K. Zhang, G. Hou, Z. Sun, and T. Tan, “LFNet: A novel bidirectional recurrent convolutional neural network for light-field image super-resolution,” IEEE Trans. Image Process., vol. 27, pp. 4274–4286, 2018.

[16] H. W. F. Yeung, J. Hou, X. Chen, J. Chen, Z. Chen, and Y. Y. Chung, “Light field spatial super-resolution using deep efficient spatial-angular separable convolution,” IEEE Trans. Image Process., vol. 28, no. 5, pp. 2319–2330, May 2019.

[17] N. Meng, H. K.-H. So, X. Sun, and E. Y. Lam, “High-dimensional dense residual convolutional neural network for light field reconstruction,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 3, pp. 873–886, Mar. 2021.

[18] Z. Cheng, Z. Xiong, and D. Liu, “Light-field super-resolution by jointly exploiting internal and external similarities,” IEEE Trans. Circuits Syst. Video Technol., vol. 30, no. 8, pp. 2604–2616, Aug. 2020.

[19] S. Zhang, Y. Lin, and H. Sheng, “Residual networks for light field image super-resolution,” in Proc. IEEE/AAAI Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 11046–11055.

[20] J. Jin, J. Hou, J. Chen, and S. Kwong, “Light-field spatial super-resolution via deep combinatorial geometry embedding and structural consistency regularization,” in Proc. IEEE/ICCV Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 2260–2269.

[21] Y. Wang, L. Wang, J. Yang, W. An, J. Yu, and Y. Guo, “Spatial-angular interaction for light field image super-resolution,” in Proc. Eur. Conf. Comput. Vis. (ECCV), 2020, pp. 290–308.

[22] Y. Wang et al., “Light field image super-resolution using deformable convolution,” IEEE Trans. Image Process., vol. 30, pp. 1057–1071, 2021.

[23] S. Zhang, S. Chang, and Y. Lin, “End-to-end light field spatial super-resolution network using multiple epipolar geometry,” IEEE Trans. Image Process., vol. 30, pp. 5956–5968, 2021.

[24] Z. Cheng, Z. Xiong, C. Chen, D. Liu, and Z.-J. Zha, “Light field super-resolution with zero-shot learning,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2021, pp. 1131–1144.

[25] Z. Cheng, Y. Liu, and Z. Xiong, “Spatial-angular versatile convolution for light field reconstruction,” IEEE Trans. Comput. Imag., vol. 8, pp. 1131–1144, 2022.

[26] K. Ko, Y. J. Koh, S. Chang, and C. Kim, “Light field super-resolution via adaptive feature remixing,” IEEE Trans. Image Process., vol. 30, pp. 4114–4128, 2021.

[27] Y. Wang et al., “Disentangling light fields for super-resolution and disparity estimation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 1, pp. 425–443, Jan. 2023.

[28] S. Wang, T. Zhou, Y. Lu, and H. Di, “Detail-preserving transformer for light field image super-resolution,” in Proc. AAAI Conf. Artif. Intell. (AAAI), 2022, pp. 2522–2530.

[29] Z. Liang, Y. Wang, L. Wang, J. Yang, and S. Zhou, “Light field image super-resolution with transformers,” IEEE Signal Process. Lett., vol. 29, pp. 563–567, 2022.
Zhengyu Liang received the B.E. degree from Xidian University, Xi’an, China, in 2019, and the M.E. degree in information and communication engineering from the National University of Defense Technology (NUDT), Changsha, China, in 2021, where he is currently pursuing the Ph.D. degree with the College of Electronic Science and Technology. His current research interests mainly focus on low-level vision, particularly on light-field image processing and image super-resolution.

Longguang Wang received the B.E. degree in electrical engineering from Shandong University (SDU), Jinan, China, in 2015, and the Ph.D. degree in information and communication engineering from the National University of Defense Technology (NUDT), Changsha, China, in 2022. His current research interests include low-level vision and 3-D vision.

Jungang Yang received the B.E. and Ph.D. degrees from the National University of Defense Technology (NUDT), Changsha, China, in 2007 and 2013, respectively. He was a Visiting Ph.D. Student with the University of Edinburgh, Edinburgh, U.K., from 2011 to 2012. He is currently an Associate Professor with the College of Electronic Science, NUDT. His research interests include computational imaging, image processing, compressive sensing, and sparse representation.

Wei An received the Ph.D. degree from the National University of Defense Technology (NUDT), Changsha, China, in 1999. She was a Senior Visiting Scholar with the University of Southampton, Southampton, U.K., in 2016. She is currently a Professor with the College of Electronic Science and Technology, NUDT. She has authored or coauthored more than 100 journal and conference publications. Her current research interests include signal processing and image processing.

Yulan Guo (Senior Member, IEEE) received the B.E. and Ph.D. degrees from the National University of Defense Technology (NUDT), Changsha, China, in 2008 and 2015, respectively. He is currently an Associate Professor. He has authored more than 100 articles at highly referred journals and conferences. His current research interests focus on 3-D vision, particularly on 3-D feature learning, 3-D modeling, 3-D object recognition, and scene understanding.

Dr. Guo served as an Associate Editor for IEEE TRANSACTIONS ON IMAGE PROCESSING, a Guest Editor for IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, and an Area Chair for CVPR 2021, ICCV 2021, and ACM Multimedia 2021. He organized several tutorials and workshops in prestigious conferences, such as CVPR 2016, CVPR 2019, ICCV 2021, and 3DV 2021. He is a Senior Member of ACM.