Light Field Spatial Super-resolution via Multi-level Perception and View Reorganization

Yifan Mao[1]
School of Computer and Information,
Anqing Normal University
Anqing, 246000, China

Zaidong Tong[2]
School of Computer and Information,
Anqing Normal University
Anqing, 246000, China

Xin Zheng[3]
School of Computer and Information,
Anqing Normal University
Anqing, 246000, China

Xiaofei Zhou[4]
School of Automation,
Hangzhou Dianzi University,
Hangzhou 310018, China

Youzhi Zhang[5]
School of Computer and Information,
Anqing Normal University
Anqing, 246000, China

Deyang Liu[6]*
School of Computer and Information,
Anqing Normal University
Anqing, 246000, China

***Abstract*—Light field (LF) imaging can obtain spatial and angular information of three-dimensional (3D) scene through a single shot, which enables a wide range of applications in the fields of 3D reconstruction, refocusing, virtual reality, *etc*. However, due to the inherent trade-off problem, the spatial resolution of acquired LF images is low, which hinders the widespread application of LF imaging technique. In order to relieve this issue, an end-to-end LF spatial super-resolution network is proposed by considering the multi-level perception and view reorganization. This method can fully explore the highly interwoven LF spatial and angular structure information. Specifically, a multi-feature fusion enhancement block is introduced that can fully perceive LF spatial, angular, and EPI information for LF spatial super-resolution. Furthermore, the angular coherence between LF views is exploited by reorganizing the LF sub-aperture images and constructing a multi-angular stack structure. Compared with other state-of-the-art methods, the proposed method achieves superior performance in both visual and quantitative terms.**

Keywords*—*Light field image; spatial super-resolution; multi-level perception; view reorganization

I. Introduction

Light Field (LF) image with four-dimensional structure not only contains the intensities of light ray, but also records the directions of light ray. Compared with traditional 2D imaging which can only capture the spatial information of light ray, LF imaging technique has great potential in many fields, such as image refocusing [1], 3D reconstruction [2], and virtual reality [3], *etc*. However, due to the inherent trade-off problem between spatial and angular resolution in the imaging plane, low spatial resolution hinders the application of LF imaging. High-efficiency spatial super-resolution methods for LF imaging play a crucial role in advancing technological development and have wide-ranging applications in medical treatment, security monitoring, and related fields. The significance of these methods lies in their ability to enhance the resolution of LF data, enabling the reconstruction of high-quality images with greater detail and precision. Therefore, it is imperative to investigate and develop efficient LF super-resolution techniques to address the challenges posed by low spatial resolution of LF data and improve their usability in various applications.

LF image has several representations, such as lenslet image, Sub-Aperture Image (SAI) array and Pseudo Video Sequence (PVS), *et al*. For SAI (also called view) array representation, the adjacent SAIs records the same 3D scene information with a small disparity. This means that the SAI array is highly correlated, which benefits in enhancing the LF spatial super-resolution performance. SAI array representation is always adopted for LF spatial super-resolution task. Especially with the development of deep learning, the convolutional neural Network (CNN) has been widely used in LF image processing tasks with SAI array representation. However, due to the complex LF structure, and the interweaving of spatial and angular information in LF images, there are great challenges to further improve the super-resolution performance by using CNNs under SAI array representation. To solve this problem, most existing methods usually consider exploring the structure information of LF image or reducing the dimensionality of the LF image. Although these methods can reconstruct high spatial resolution LF images, their performance is limited. The reason lies in two aspects. One is that the LF structure information is under-explored. The other is that the rich angular information contained in LF image is under-used. Fully exploring LF structure and angular information is more conducive to improving LF super-resolution performance.

In order to mitigate these issues, in this paper, we propose a LF spatial super-resolution network via multi-level perception and view reorganization. By introducing the multi-feature fusion enhancement block, our network can adequately explore and fuse LF structure information, including spatial, angular, and Epipolar Plane Image (EPI) information, so as to recover more details, especially for some occlusion regions. In addition, in order to better mine the abundant angular information, we reorganize the LF SAIs in different angular directions. Specifically, we arrange the horizontal and vertical SAIs in the LF image array with the same angular coordinate element into a stack, and construct a Multi-Angular Stack (MAS) structure. The MAS structure can provide rich angular and spatial information for LF image spatial super-resolution. The main contributions of this paper are as follows:

- We propose a multi-feature fusion enhancement block to fully perceive LF spatial, angular, and EPI information for LF spatial super-resolution.

- We construct a multi-angular stack structure to adequately explore LF angular information to enhance LF spatial super-resolution performance.
- Comprehensive experiments demonstrate the superiority of the proposed method than the other state-of-the-art approaches.

The rest of this paper will be organized in the following way. A brief review of related work will be provided in Section II. In Sections III, we present our approach. Section IV discusses the simulation results. Finally, the paper is concluded in Section V.

II. Related Works

LF spatial super-resolution aims to generate high spatial resolution LF images from densely sampled low spatial resolution one. To achieve this goal, two approaches can be used. One is to apply a single image super-resolution method [4] to super-resolve each SAI separately. The other is to build a mathematical model based on prior information to directly reconstruct high spatial resolution LF image. With the development of deep learning, researchers are more inclined to use CNN to realize the spatial resolution reconstruction of LF images, which can take full use of LF abundant structure information and improve the LF reconstruction performance. A brief reviews of single image super-resolution and LF image super-resolution are given in this section.

A. Single Image Super-resolution

Single image super-resolution does not involve multi-view tasks, for which the goal is only to generate a high-resolution 2D image from a low-resolution 2D image. Shi *et al.* [5] constructed a structure-aware single image super-resolution network to further generate structure and details of images. Song *et al.* [6] developed a criss-cross network to reduce the computation complexity for single image super-resolution task. In their method, few feature points were used to compute long-range dependencies. Hsu *et al.* [7] proposed a detail-enhanced wavelet residual network for single image super-resolution to resolve the details over smooth problem. Wang *et al.* [8] developed an end-to-end joint framework to super-resolve single image by considering the issue of no ground truth high resolution images and degradation models are available. Lan *et al.* [9] put forward a lightweight network for single image super-resolution, which can decrease computational burden by expressing multiscale feature and learning feature correlation.

Single image super-resolution method can reconstruct high spatial resolution LF image by super-resolve each SAI. However, the inherent structure information is under-explored in this kind of method, which limits the LF super-resolution performance.

B. Light Filed Super-resolution

Different from 2D image super-resolution, the pixel information required for LF super-resolution actually exists in each SAI. The four-dimensional information of the LF image can be decomposed into many SAIs recording the scene, and there is a certain disparity between different SAIs, which has a strong correlation. Therefore, the SAIs of LF images are highly correlated, and the utilization of single view spatial information and angular correlation between different views is the key factor to improve the performance of LF image super-resolution.

Early studies followed the traditional paradigm by developing different models to achieve super-resolution in LF image space. Among them, LFBM5D [10] extends the BM3D [11] filtering to 5D to provide more prior information and thus improve the super-resolution performance. Mitra *et al.* [12] proposed a Gaussian mixture model for encoding the spatial structure of the LF to cope with noise and super-resolution issues. Farrugia *et al.* [13] used multivariate ridge regression to approximate the subspace linear projection method of the adjacent SAIs to the middle SAI. Rossi *et al.* [14] utilized the complementary information between different views to achieve spatial super-resolution through graph optimization based on regularized coupling of graphs. Although these models can encode the structure of the LF by establishing a mathematical model, they rely too much on the prior information of the image, resulting in limited super-resolution performance.

With the development of deep learning, researchers are more inclined to build different super-resolution networks to learn the mapping relationship between low-resolution and high-resolution LF images. For example, Yoon *et al.* [15] proposed a model for LF image super-resolution based on deep convolutional networks. Zhang *et al.* [16] divided the views into four groups, and used the residual information between adjacent views to cope with super-resolution tasks. They explored the correspondences between different viewpoints and divided the SAIs into multiple image stacks with a consistent sub-pixel offset. However, the complementary information between all views was not fully utilized, and the disparity consistency was not well maintained. In order to make full use of the high-dimensional features of LF data, Yeung *et al.* [17] alternately used convolutions to characterize the relationship between pixels in the 4D structural information of spatial domain and angular domain. However, the inherent disparity structure of LF images is ignored. Jin *et al.* [18] proposed an all-to-one light-field super-resolution strategy to strengthen the disparity structure. They explored the complementary information between the views to perform individual super-resolution for each SAI of LF image. Wang *et al.* [19] proposed a spatial-angular interaction strategy and designed different networks to extract spatial and angular features respectively. Wang *et al.* [20] further used separable convolutional networks to explore the spatial-angular information of LF. Although they explore the spatial-angular information to a certain extent, they do not effectively mine the high-dimensional information of the LF image. As a result, the LF super-resolution performance is affected to a certain extent. Liu *et al.* [21] extracted global view information and simultaneously modeled the correlation within each view to achieve better super-resolution performance. Although these methods have shown remarkable performance, there are still some problems that are not well addressed. One is that the high-dimensional information of LF images is not fully utilized, especially the complementary information between spatial angulars and the geometric consistency information of LF EPIs. The other is that the angular structure between the LF views is not broken, and the angular correlation between the LF images is not explored enough.

Focusing on the above problems, we propose a novel network for LF spatial SR based on the content characteristics of LFs. We introduce two strategies to mitigate the above problems. Since the high-dimensional features of the LF image contain rich information, we fully explore the LF spatial information, LF angular information, and the geometric information of the LF EPIs, respectively. The information is interacted and the channel attention is increased to obtain the enhanced information after interaction. In order to further explore the angular correlation between the LF views, we break the angular structure of the LF array and rearrange it into multiple stacks, and super-resolve each horizontal view stack separately. The network makes full use of the content characteristics of LF images and further improves the performance of LF spatial resolution reconstruction by fully exploring the information of different dimensions of the LF and designing the cross-arrangement of views to mine the angular correlation between views. Experimental results on both real and synthetic datasets demonstrate the superiority of the proposed method.

III. Proposed Method

In the proposed method, the spatial information of the LF image, the angular information and the geometric information of the LF EPIs were used to interact with the multi-dimensional features of the LF and reorganize the array structure based on the content characteristics of the LF. The method makes full use of multi-dimensional information and angular correlation of LF images, and is composed of two main modules: multi-stream feature fusion enhancement module and structure reorganization module. The overall network structure is shown in Fig. 1. We formulate LF in terms of a four-dimensional tensor $L(u, v, x, y) \in \mathbb{R}^{U \times V \times X \times Y}$, where U and V denote the angular dimensions, and H and W denote the spatial dimensions. Specifically, the SAI of a $U \times V$ array represents the LF, and the resolution of each SAI is $H \times W$. The high-resolution LF image $L \in \mathbb{R}^{U \times V \times \alpha X \times \alpha Y}$ is reconstructed from the low-resolution LF image $L \in \mathbb{R}^{U \times V \times X \times Y}$, where α is the magnification factor. Following [16-19], we perform SR only on the Y channel to reduce the computational complexity. The Cb and Cr channels are upsampled using bicubic interpolation algorithms. Then the super-resolved Y, Cb and Cr channels are converted into an RGB image. The proposed reconstruction network can be written as

$$\begin{aligned} L_{HR}(u, v, \alpha x, \alpha y) &= f(L_{LR}(u, v, x, y), \Theta), \\ \Theta^* = \arg \underset{\Theta}{min} ||L_{GT}(u, v, \alpha x, \alpha y) &- L_{HR}(u, v, \alpha x, \alpha y)|| \end{aligned} \tag{1}$$

where $L_{HR}(u, v, \alpha x, \alpha y)$ is the reconstructed dense LF , $L_{GT}(u, v, \alpha x, \alpha y)$ is the ground truth, $f(\cdot)$ represents the mapping from low resolution LF image to high resolution LF image, Θ is the network parameter.

To achieve a high-quality dense LF reconstruction and obtain optimal network parameter Θ, we propose a multi-stream reconstruction network. To effectively extract distinctive information from various view images, we propose a novel approach that combines multiple features to enhance the representation of LF spatial, angular, and EPI information. Specifically, we present a multi-feature fusion enhancement block that can accurately capture spatial and angular details contained in the LF data (See Sec. III-A). Moreover, we design a Structure-based Super-Resolution Module that utilizes the angular information present in the subaperture view array to perform super-resolution reconstruction, thus optimizing the quality of the reconstructed views (see Sec. III-B). To further enhance the geometric consistency between the reconstructed views and maintain the valuable disparity structure of the LF data, we propose a mixed loss function that incorporates both reconstruction loss and EPI gradient loss (see Sec. III-C). The network architecture is elaborated in the following subsections.

A. Multi-stream Feature Fusion Enhancement Module

To enhance the characteristics of decoupling, this paper proposes the addition of a channel information, $L \in \mathbb{R}^{U \times V \times X \times Y \times C}$, to the 5D data. Multiple representation methods of the LF image in various dimensions were utilized to explore its content characteristics and extract feature information. The fusion process of multi-stream feature, denoted as L_{MFFE}, is expressed as follows:

$$\begin{aligned} L_{MFFE} = f_{MFFE}(&\mathbf{CA}(\mathbf{SFE} + \mathbf{AFE}) \\ &+\mathbf{CA}(\mathbf{EPI^H} + \mathbf{EPI^W})) \end{aligned} \tag{2}$$

Here, $\mathbf{SFE}$ stands for the spatial feature extraction module of subaperture view, $\mathbf{AFE}$ indicates the angular feature extraction module of subaperture view, $\mathbf{EPI^H}$ and $\mathbf{EPI^V}$ denote the feature extraction modules of the EPI in the vertical and horizontal directions (EPI-H and EPI-V), respectively, and CA represents the attention module. In what follows, we expound on each module of the Multi-stream Feature Fusion Enhancement Module.

Spatial Feature Extraction(SFE): We focus on the information of SAI in the dimension and reshape the 5D LF data with increased channel information $S^{lr} \in \mathbb{R}^{UV \times C \times X \times Y}$. The SFE module is used to extract the spatial features of the SAI. Specifically, SFE is a module composed of three convolutions with a kernel size of 3×3, a step size of 1, a dilation rate of 2, and a Relu activation layer after each convolution layer. Since we focus on $H \times W$, the dimension information, SFE only includes the pixel information of the context in each SAI, which has a good refinement of the global features of each SAI and has rich texture information.

Angular Feature Extraction(AFE): In view of the multi-angular characteristics of LF, we centers on the $U \times V$ angular information. Similarly, we reshape the original 5D light field data $A^{lr} \in \mathbb{R}^{HW \times C \times U \times V}$. The AFE module is used to extract the angular feature from the pixel information of the same angular position of the SAI. Specifically, AFE is a module composed of three convolutions with a kernel size of 3×3 and a step size of 1. Each convolution layer is followed by a Relu activation layer. Different from SFE, AFE pays more attention to the correlation in angular, and different SAIs have strong correlation in the same pixel position, which can provide more pixel information for the occlusion area.

EPI Feature Extraction(EPI-H, EPI-V): EPI is the horizontal $E_H^{lr} \in \mathbb{R}^{VW \times C \times U \times H}$ or $E_W^{lr} \in \mathbb{R}^{UH \times C \times V \times W}$ vertical two-dimensional slice information of SAI by sampling angular coordinates and corresponding spatial coordinates in multi-dimensional data of LF. Acknowledging the effectiveness

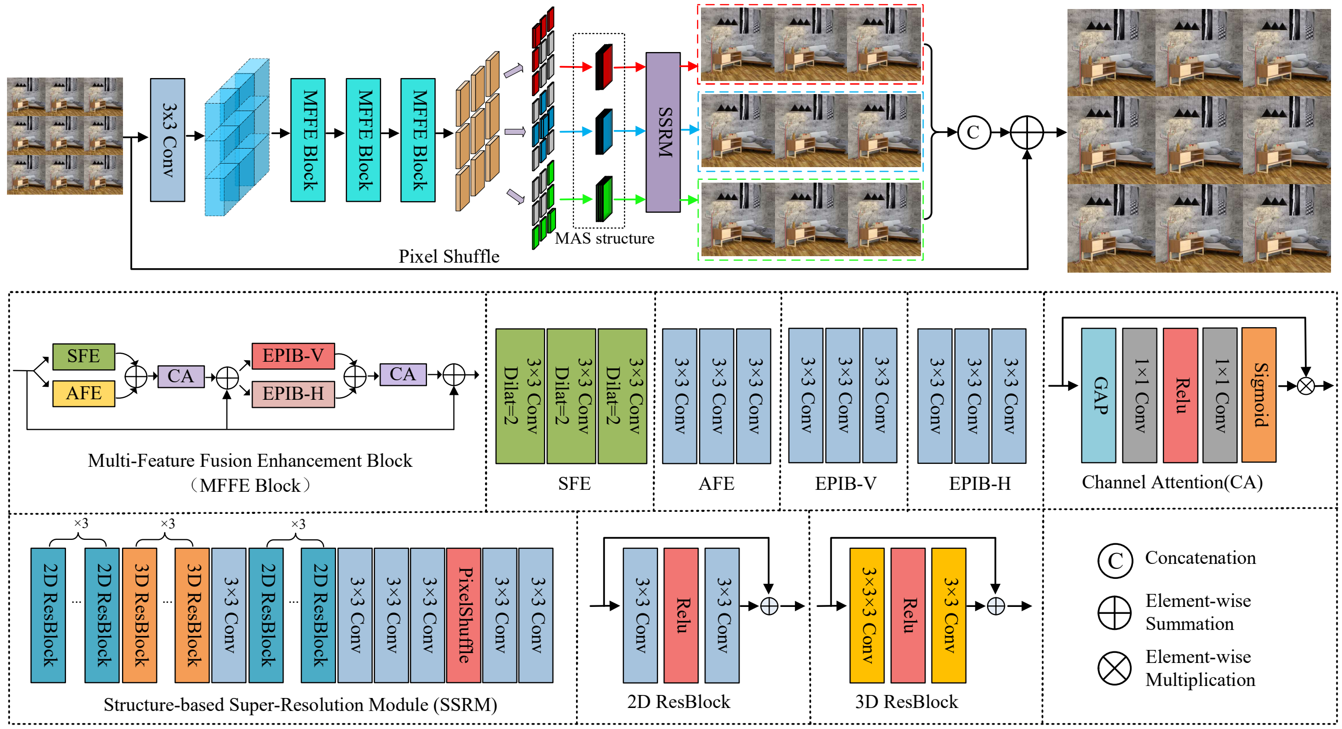

Fig. 1. The architecture of our light field spital reconstruction network.

of EPI in reflecting the geometrical consistency of LF, we delves into the analysis of LF geometric information, both horizontally and vertically. Specifically, EPI-H and EPI-V are modules composed of three convolutions with a kernel size of 3×3 and a step size of 1. Each convolution layer is followed by a relu activation layer. The EPI slices have a simple linear structure, which is basically a slanted straight line composed of homogeneous regions, and perform well for the analysis of features in the scene slices even for the rather complex shape and intensity variations in SAI.

Channel Attention(CA): Due to the local nature of convolutional operations, obtaining sufficient information to extract inter-channel relationships in LF imager can be challenging. To address this limitation, we integrate CA into our proposed architecture after the SFE and AFE fusion, as well as after the EPI-H and EPI-V fusion. The CA module compresses the feature map into a feature vector via global average pooling (GAP) to obtain a global description feature. Nonlinear relationships between channels are then learned by compressing the channel count through 1×1 convolutions, using the rule activation layer, and subsequently amplifying the channel count via another 1×1 convolution layer. Finally, the weighting coefficients assigned to each channel by the sigmoid function enable effective cross-channel interaction and enhancement of fusion interaction amongst the information.

B. Structure-base Super-resolution Module

This paper proposes a novel approach to leverage the angular information contained in the subaperture view array for super-resolution reconstruction. Specifically, a cross-arrangement structure of the angular view and a reorganized parallax structure of the view are proposed to enhance the utilization of angular information. Then, a multi-stream feature fusion module is introduced to extract rich and high-dimensional features, which are subsequently fed into the structure-based super-resolution module. The network structure can be expressed as:

$$L_{HR} = \textbf{Concat}(f^1_{SSRM}(MFFE^1), \cdots, f^i_{SSRM}(MFFE^i)) \tag{3}$$

where $f^i_{SSRM}(\cdot)$ represents the Structure-base Super-resolution module, $MFFE^i$ represents the input information of the first row in the cross-arrangement structure of angulars and views, and i represents the number of rows, where the main scenario in this paper is $i = 5$. The designed super-resolution network comprises three 2D ResBlock convolutions, three 3D ResBlock convolutions, one 3×3 convolution, three 2D ResBlock convolutions, three 3×3 convolutions, one pixel shuffle layer, and two 3×3 convolutions. Different ResBlock convolutions are utilized to fuse rich information in both the spatial and angular domains, while the pixel shuffle layer achieves spatial super-resolution via upsampling.

C. Training Details

In this method, we adopt the $L1$ loss function to measure the reconstructed target LF $L_{HR}(u, v, \alpha x, \alpha y)$, and supervise our network by its ground truth value $L_{GT}(u, v, \alpha x, \alpha y)$, which is defined as:

$$loss_1 = \sum_{u,v,x,y} (|L_{GT}(u, v, \alpha x, \alpha y) - L_{HR}(u, v, \alpha x, \alpha y)|) \tag{4}$$

To further preserve the valuable disparity structure of the

TABLE I. QUANTITATIVE COMPARISON RESULTS OF DIFFERENT METHODS FOR TASK 2 × SR AND 4 × SR (PSNR/SSIM)

Task	2 ×					4 ×				
Method	EPFL	HCI new	HCI old	INRIA	STFgantry	EPFL	HCI new	HCI old	INRIA	STFgantry
Bicubic	29.50/0.935	31.69/0.934	37.46/0.978	31.10/0.956	30.82/0.947	25.14/0.831	27.61/0.851	32.42/0.984	26.82/0.886	25.93/0.843
VDSR	32.50/0.960	34.37/0.956	40.61/0.987	34.43/0.974	35.54/0.979	27.25/0.878	29.31/0.883	34.81/0.952	29.19/0.921	28.51/0.901
EDSR	33.09/0.963	34.83/0.960	41.01/0.988	34.97/0.977	36.29/0.982	27.84/0.886	29.60/0.887	35.18/0.954	29.66/0.926	28.70/0.908
RCAN	33.16/0.964	34.98/0.960	41.05/0.988	35.01/0.977	36.33/0.983	27.88/0.886	29.63/0.888	35.20/0.954	29.76/0.927	28.90/0.921
resLF	32.75/0.967	36.07/0.972	42.61/0.992	34.57/0.978	36.89/0.987	27.46/0.890	29.92/0.901	36.12/0.965	29.64/0.934	28.99/0.921
LFSSR	33.69/0.975	36.86/0.975	43.75/0.994	35.27/0.983	38.07/0.990	28.27/0.908	30.72/0.912	36.70/0.969	30.31/0.945	30.15/0.939
LF-ATO	34.27/0.976	37.24/0.977	44.20/0.994	36.15/0.984	39.64/0.993	28.52/0.912	30.88/0.914	37.00/0.970	30.71/0.949	30.61/0.943
LF-InterNet	34.14/0.976	37.28/0.977	44.45/0.995	35.80/0.985	38.72/0.992	28.67/0.914	30.98/0.917	37.11/0.972	30.64/0.949	30.53/0.943
LF-DFnet	34.44/0.977	37.44/0.979	44.23/0.994	36.36/0.984	39.61/**0.994**	28.77/0.917	**31.23/0.920**	37.32/0.972	30.83/0.950	**31.15/0.949**
MEG-Net	34.31/0.977	37.42/0.978	44.10/0.994	36.10/0.985	38.77/0.992	28.75/0.916	31.10/0.918	37.29/0.972	30.67/0.949	30.77/0.945
DPT	34.49/0.976	37.36/0.977	44.30/0.994	**36.41**/0.984	39.42/0.993	**28.94/0.917**	31.20/0.919	**37.41/0.972**	30.96/0.950	**31.15/0.949**
Proposed	**34.56/0.977**	**37.64/0.979**	**44.55/0.995**	36.36/**0.985**	**39.64**/0.993	28.88/0.915	**31.23**/0.919	37.32/**0.972**	30.87/**0.950**	31.08/0.948

LF and promote the geometric consistency between the reconstructed views, this paper refers to the EPI gradient loss function proposed by [22], which is defined as follows

$$\begin{aligned} loss_2 = & \sum_{y,v}(|E^x_{GT}(x,u) - E^x_{HR}(x,u)| \\ & + |E^u_{GT}(x,u) - E^u_{HR}(x,u)|) \\ & + \sum_{x,u}(|E^y_{GT}(y,v) - E^y_{HR}(y,v)| \\ & + |E^v_{GT}(y,v) - E^v_{HR}(y,v)|) \end{aligned} \tag{5}$$

The training objective of our method is to minimize these two losses: min $loss_1 + loss_2$.

IV. EXPERIMENTS

To confirm the efficacy of the proposed approach, a range of detailed experimental results have been presented, comprising ablation experiments and comparisons with the existing methods. Specifically, we follow [20] which utilize five publicly available LF datasets (namely EPFL , HCInew , HCIold , INRIA , STFgantry) during both the training and testing phases. The training and test sets follow the same partitioning as provided in [20]. The LFs within these datasets possess an angular resolution of 9×9. During the training procedure, we downsample the SAI into LF patches of size 32×32 via bicubic downscaling. The optimization of our network utilizes both $L1$ and EPI loss functions and the Adam method. Our network is implemented in PyTorch, leveraging an RTX 5000 GPU. The learning rate is initially configured to 2×10^{-4} and subsequently reduced by a factor of 0.5 every 15 epochs. The performance of our proposed method is evaluated using objective measures, including Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM), while simultaneously conducting a subjective comparison of detail texture regions after SR.

A. Comparison With State-of-the-Art Methods

The proposed method is compared with several state-of-the-art methods, comprising three single-image SR techniques [7-9] and seven LF image SR methods [16,17,18,19,20,23,24]. To ensure a uniform training process, we retrained all these methods using the same dataset. **Quantitative Results:** Table I presents quantitative results for 2×SR and 4×SR. The proposed method significantly outperforms three single image super-resolution methods, VDSR[7], EDSR[8], and RCAN[9]. This improvement is mainly attributed to the complex texture details present in the comprehensive scene, which renders the reconstruction method of single image unsuitable for LF image reconstruction. Moreover, our approach attains the best overall performance compared to resLF[16], LFSSR[17], LF-ATO[18], LF-Internet[19], LF-DFnet[20], MEG-Net[23], and DPT[24]. Our proposed method outperforms the comparative methods in all five datasets for two primary reasons. Firstly, the comparative methods are less effective in fully exploiting the LF's rich angular information and handling complex scenes. resLF[16] constructs view stacks to explore LF information in five directions: horizontal, vertical, left, right, and tilt, fails to fully use complementary information among all views while maintaining disparity consistency. Similarly, though the LF-ATO[18] method proposes an all-to-one architecture that explores complementary information between views, the feature information between spatial and angular domains is not entirely fused. This deficiency affects the spatial super-resolution performance. Second, the comparative methods fail to exploit the LF's geometric structure information to its full potential. LF-Internet[19] utilized the spatial and angular interaction strategy and different networks to extract spatial and angular features while making use of the spatial and anglular correlations. However, they neglected the EPI structure information and angular geometric information, which deteriorated the quality of LF spatial reconstruction. Similarly, although LFSSR[17] utilized convolution to characterize the relationship between pixels in a 4D information space in the spatial and angular domains, it ignored the inherent geometric structure of the LF and failed to fully utilize the angular geometric structure. Observing the evaluation metrics presented in Table I, we note that the proposed method outperforms the comparative methods significantly.

Qualitative Results: The qualitative results of the Bedroom in the HCInew scene and ISO_Chart_1_Decoded in EPFL scene reconstructed by different methods under task 2×SR and 4×SR is presented in Fig.2 and Fig.3, respectively.

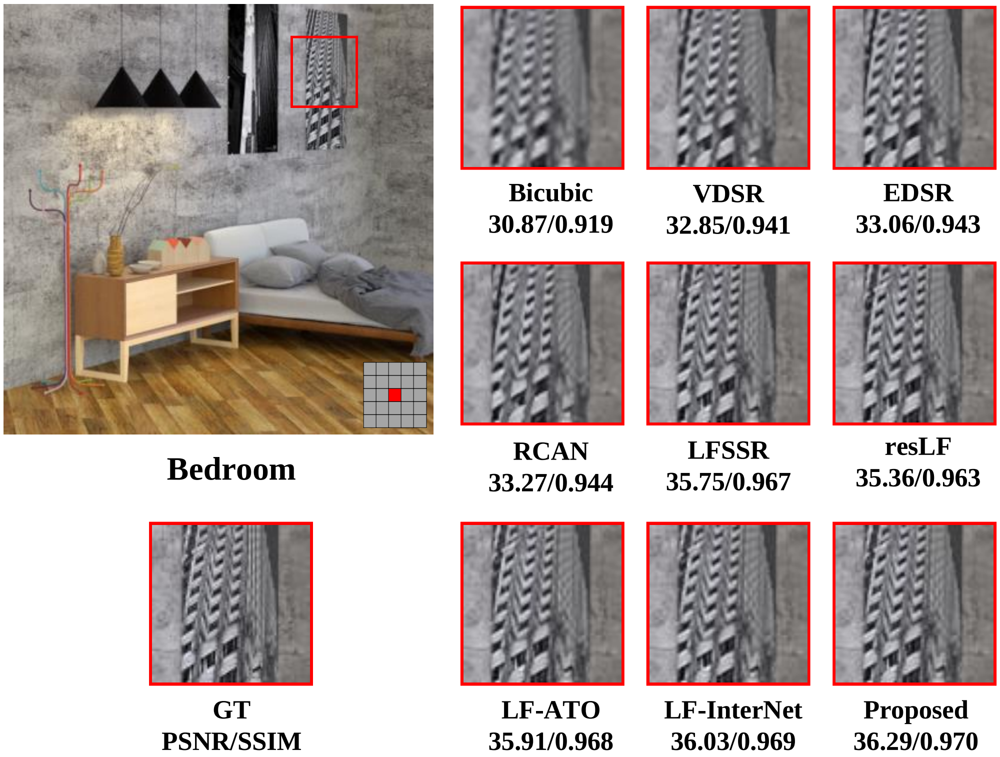

Fig. 2. Visual comparisons for 2×SR. The super-resolved center view images are shown. The PSNR and SSIM scores achieved by different methods on the presented scenes are reported below the zoom-in regions.

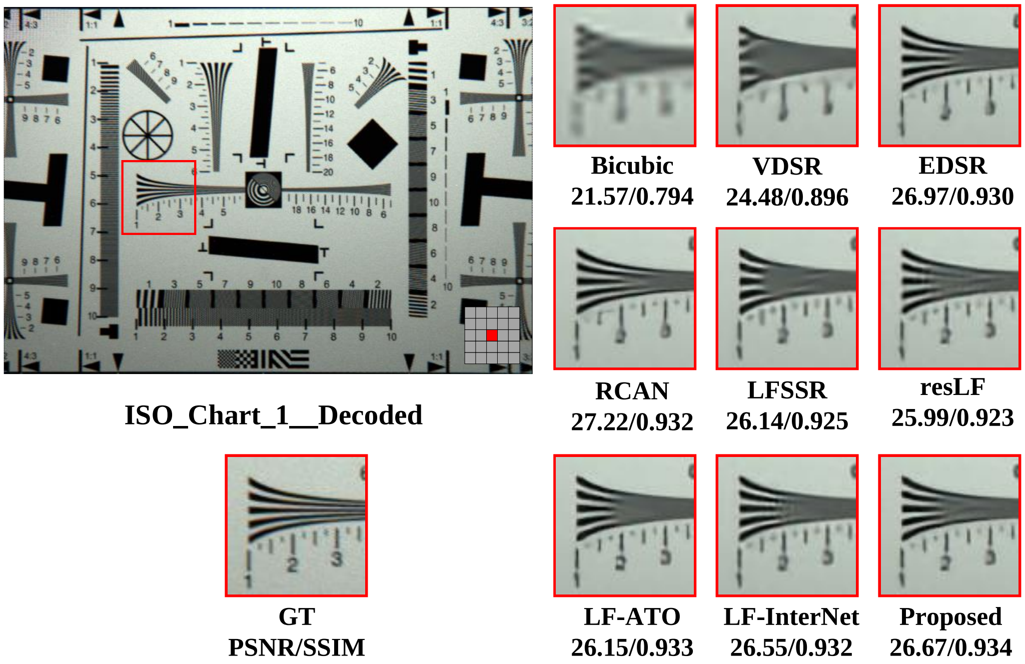

Fig. 3. Visual comparisons for 4×SR. The super-resolved center view images are shown. The PSNR and SSIM scores achieved by different methods on the presented scenes are reported below the zoom-in regions.

TABLE II. Task 2× Quantitative Comparison Results of Different Variants of the Proposed Method (PSNR/SSIM)

Method	EPFL	HCI new	HCI old	INRIA	STFgantry
w/o CA	34.43/0.977	37.58/0.979	44.56/0.995	36.22/0.985	39.55/0.993
w/o EPI-H	34.34/0.976	37.46/0.978	44.36/0.994	36.13/0.984	39.13/0.992
w/o EPI-V	34.33/0.976	37.39/0.978	44.25/0.994	36.09/0.984	39.09/0.992
w/o AFE	34.28/0.976	37.33/0.977	44.40/0.994	36.18/0.984	39.15/0.992
w/o SFE	34.38/0.977	37.65/0.979	44.54/0.995	36.22/0.985	39.67/0.993
w/o SSRM	34.21/0.975	37.11/0.976	44.11/0.994	36.09/0.984	38.77/0.992
Proposed	**34.56/0.977**	**37.64/0.979**	**44.55/0.995**	**36.36/0.985**	**39.64/0.993**

TABLE III. Task 2× SR Quantitative Comparison Results of Different Angular Resolutions of the Proposed Method (PSNR/SSIM)

Method	EPFL	HCI new	HCI old	INRIA	STFgantry
3 × 3	33.94/0.972	37.10/0.976	43.74/0.994	35.85/0.982	38.94/0.992
5 × 5	34.56/0.977	37.64/0.979	44.55/0.995	36.36/0.985	39.64/0.993
7 × 7	34.69/0.978	37.80/0.980	44.73/0.995	36.39/0.985	39.65/0.993

The magnification of the local view of the reconstructed sub-aperture is shown in the red box. In Fig.2, although the Bedroom scene contains complex textures, which makes reconstruction challenging. Our method leverages high-dimensional features of the LF and combines spatial and angular domain with EPI information to recover more detailed information of the scene. It can be seen from the Fig.3, the scene is composed of numerous lines and gaps that are difficult to reconstruct. While the LF reconstruction method can capture more information, it still has limitations in such complex line scenes with small gaps. Instead, EDSR[8] and RCAN[9], two single image super-resolution methods, show better reconstruction performance on these scenes. This is because the real pixel information is insufficient at a higher super-resolution size, and the LF reconstruction method synthesizes more new pixel information, which is intertwined with each other and blocks the gaps between lines. Compared with current state-of-the-art SISR and LF image SR methods, our method produces images with more accurate details and fewer artifacts.

B. Ablation Experiments

To gain a deeper understanding of the proposed network's properties, an ablation study was performed to demonstrate the efficacy of the feature fusion and angular view intersection arrangement structure for high-dimensional data in the LF context. The study involved removing various components from the network, including the channel attention module, EPI feature extraction module (EPI-H, EPI-V), angular feature extraction module, spatial feature extraction module, and structure-based super-resolution module. These were identified as the variants of the proposed network for the purposes of the study and are respectively denoted as "w/o CA", "w/o EPI-H", "w/o EPI-V", "w/o AFE", "w/o SFE" and "w/o the SSRM". The comparison results (PSNR/SSIM) of the different variants of the proposed method for task 2×SR on five public datasets are presented in Table II. The results indicate that the proposed method significantly outperforms the other variants with the removal of any module leading to an adverse effect on the reconstruction performance.

Specifically, compared with "w/o CA", the proposed method has obvious advantages in PSNR. This can be attributed to the channel attention module, which analyzes the weight of each channel by fusing spatial and angular with horizontal and vertical information of the polar plane. It strengthens the channel weight coefficient that has a greater impact on reconstruction. In comparison to "w/o EPI-H" and "w/o EPI-V", the proposed method attains higher PSNR and SSIM scores due to the EPI module's ability to analyze the section information of the LF geometrically, resulting in better recovery of the structural information. The proposed method outperforms "w/o AFE" by achieving a 0.28 dB PSNR gain by using the angular information to improve the LF reconstruction performance significantly. The multi-stream feature fusion module extracts diverse structural information by analyzing multiple dimensions of the high-dimensional data of the LF, thereby enhancing spatial angular correlations. Thus, all modules in multi-stream feature fusion contribute positively to the reconstruction performance. Significantly, the proposed method achieves the best gain compared to "w/o SSRM", with the PSNR value increasing from 34.21 dB to 34.56 dB for 2×SR. This is because the cross-arrangement of angular viewpoints offers geometric structure analysis of the angular correlation of the LF, leading to an improvement in reconstruction quality.

C. Extended Experiments

In this paper, we investigate the impact of angular resolution on the performance of our proposed method. We evaluate the super-resolution performance under different angular resolutions by extracting $A \times A$ sub-aperture views of the center from the input LF image, where A represents the number of views (A = 3, 5, 7). We train separate models for the 2× super-resolution task with each angular resolution setting. Our results, as shown in Table III, reveal that increasing the angular resolution from 3×3 to 7×7 improves the PSNR values. This improvement can be attributed to the richer angular information provided by additional views, which enhances the spatial super-resolution. However, we observe that the performance saturates for angular resolutions greater than 5×5. This is because the information obtained from the 7×7 subaperture views is already sufficient, and further increasing the angular resolution yields only marginal improvements in performance.

D. Discussions

This paper proposes a new method for learning LF spatial SR by interwovening LF spatial and angular structure information. Here, some discussions are presented. (1) Similar to previous literature, we adopt publicly available LF data to conduct detailed experiments. The qualitative and quantitative comparisons with the state-of-the-art SR methods demonstrate the superior performance of the proposed method. (2) Our ablation study highlights the effectiveness of multi-stream feature fusion by means of the integration and interlacing of high-dimensional data from diverse sources during the multi-stream feature fusion phase. This approach facilitates the extraction of comprehensive information, thereby enhancing reconstruction performance. Furthermore, the ablation experimental outcomes validate the effectiveness of both the proposed MFFE and SSRM. (4) Considering the quantitative results presented in Subsection IV , the performance of the proposed method for the narrow-baseline LF images is significantly better than that for the widebaseline LF images, mainly because the latter has a larger parallax range, posing greater challenges to feature extraction.

V. Conclusion

In this paper, we present a multi-stream feature fusion spatial reconstruction network with cross-arranged viewpoints. The network consists of two stages: multi-stream feature fusion and reconstruction based on cross-permutation of angular viewpoints. In the multi-stream feature fusion stage, we combine and interweave high-dimensional data from different sources to extract rich information that can be used to improve the reconstruction performance. Additionally, this stage allows us to fully explore the high-dimensional data of the LF and fuse different dimensional data. Then the rich information obtained in the multi-stream feature fusion stage is used to mine the LF information from the geometric structure level to improve the reconstruction performance. Through experiments on five public datasets, we demonstrate that our proposed method produces high-quality spatial reconstructions of LF images under both 2×SR and 4×SR reconstruction tasks. Furthermore, we analyze the influence of the input angular resolution on reconstruction performance. Our results show that our method significantly outperforms state-of-the-art approaches.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 62171002, 62271180, and in part by STCSM, China under Grant SKLSFO2021-05, in part by University Discipline Top Talent Program of Anhui under Grant gxbjZD2022034, in part by Anhui Outstanding Youth Fund by Colleges and Universities under Grant 2022AH030106, in part by Natural Science Foundation of Anhui under Grant 1908085MF194, in part by Jingfu Yuying Innovation and Entrepreneurship Leading Program of Anqing Normal University under Grant 2022JFYY031, and in part by the Open Research Fund of National Engineering Technology Research Center for RFID Systems under Grant RFID2021KF03.

References

[1] Y. Wang, J. Yang, Y. Guo, C. Xiao, and W. An, “Selective light field refocusing for camera arrays using bokeh rendering and super resolution,” IEEE Signal Processing Letters, vol. 26, no. 1, pp. 204-208, 2018.

[2] Z. Wang, L. Zhu, H. Zhang, *et al*, “Real-time volumetric reconstruction of biological dynamics with light-field microscopy and deep learning,” Nature Methods, vol. 18, pp. 551-556, 2021.

[3] J. Yu. “A light-field journey to virtual reality,” IEEE Multi Media, vol. 24, no. 2, pp. 104-112, 2017.

[4] W. Yang, X. Zhang, Y. Tian, W. Wang, J. -H. Xue and Q. Liao, “Deep Learning for Single Image Super-Resolution: A Brief Review,” IEEE Transactions on Multimedia, vol. 21, no. 12, pp. 3106-3121, 2019.

[5] W. Shi, F. Tao and Y. Wen, “Structure-Aware Deep Networks and Pixel-Level Generative Adversarial Training for Single Image Super-Resolution,” IEEE Transactions on Instrumentation and Measurement, vol. 72, pp. 1-14, 2023.

[6] Z. Song, B. Zhong, J. Ji and K. -K. Ma, “A Direction-Decoupled Non-Local Attention Network for Single Image Super-Resolution,” IEEE Signal Processing Letters, vol. 29, pp. 2218-2222, 2022.

[7] W. -Y. Hsu and P. -W. Jian, “Detail-Enhanced Wavelet Residual Network for Single Image Super-Resolution,” IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-13, 2022.

[8] L. Wang, T. -K. Kim and K. -J. Yoon, “Joint Framework for Single Image Reconstruction and Super-Resolution With an Event Camera,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 11, pp. 7657-7673, 2022.

[9] R. Lan, L. Sun, Z. Liu, H. Lu, C. Pang and X. Luo, “MADNet: A Fast and Lightweight Network for Single-Image Super Resolution,” in IEEE Transactions on Cybernetics, vol. 51, no. 3, pp. 1443-1453, 2021.

[10] M. Alain and A. Smolic, “Light field super-resolution via lfbm5d sparse coding,” IEEE International Conference on Image Processing (ICIP), 2018, pp. 2501-2505.

[11] K. Egiazarian and V. Katkovnik, “Single image super-resolution via bm3d sparse coding,” European Signal Processing Conference (EUSIPCO), 2015, pp. 2849-2853.

[12] K. Mitra and A. Veeraraghavan, “Light field denoising, light field super resolution and stereo camera based refocussing using a gmm light field patch prior,” IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2012, pp. 22-28.

[13] R. A. Farrugia, C. Galea, and C. Guillemot, “Super resolution of light field images using linear subspace projection of patch-volumes,” IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 7, pp.1058-1071, 2017.

[14] M. Rossi and P. Frossard. “Geometry-consistent light field super-resolution via graph-based regularization,” IEEE Transactions on Image Processing, vol. 27, no. 9, pp. 4207-4218, 2018.

[15] Y. Yoon, H.-G. Jeon, D. Yoo, *et al.*, “Learning a deep convolutional network for light-field image super-resolution,” IEEE International Conference on Computer Vision Workshop (ICCVW), 2015, pp. 24-32.

[16] S. Zhang, Y. Lin, H. Sheng, “Residual networks for light field image super-resolution,” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 11046-11055.

[17] H. W. F. Yeung,*et al.*, “Light field spatial super-resolution using deep efficient spatial-angular separable convolution,” IEEE Transactions Image Processing, vol. 28, no. 5, pp. 2319-2330, 2018.

[18] J. Jin, *et al.*, “Light field spatial super-resolution via deep combinatorial geometry embedding and structural consistency regularization,” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 2260-2269.

[19] Y. Wang, *et al.*, “Spatial-angular interaction for light field image super-resolution,” In Proceedings of the European Conference on Computer Vision, 2020, pp. 290-308.

[20] Y. Wang, J. Yang, L. Wang, X. Ying, T. Wu, W. An, and Y. Guo, “Light field image super-resolution using deformable convolution,” IEEE Transactions on Image Processing, vol. 30, pp. 1057-1071, 2020.

[21] G. Liu, H. Yue, J. Wu, *et al.*, "Intra-Inter View Interaction Network for Light Field Image Super-Resolution," IEEE Transactions on Multimedia, vol. 25, pp. 256-266, 2023.

[22] J. Jin, J. Hou, H. Yuan, and S. Kwong, "Learning light field angular super-resolution via a geometry-aware network," In Proceedings of AAAI Conference on Artificial Intelligence (AAAI), 2020, pp. 11141-11148.

[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] S. Wang, T. Zhou, Y. Lu, H. Di, "Detail-preserving transformer for light field image super-resolution," In Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, No. 3, pp. 2522-2530, 2022.