LF-MAGNet: Learning Mutual Attention Guidance of Sub-Aperture Images for Light Field Image Super-Resolution

Zijian Wang[1], Yao Lu[1(✉)], Yani Zhang[2], Haowei Lu[1], Shunzhou Wang[1], and Binglu Wang[3]

[1] Beijing Laboratory of Intelligent Information Technology, School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China
vis_yl@bit.edu.cn

[2] School of Marine Electrical Engineering, Dalian Maritime University, Dalian 116026, China

[3] School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, China

Abstract. Many light field image super-resolution networks are proposed to directly aggregate the features of different low-resolution sub-aperture images (SAIs) to reconstruct high-resolution sub-aperture images. However, most of them ignore aligning different SAI's features before aggregation, which will generate sub-optimal light field image super-resolution results. To handle this limitation, we design a mutual attention mechanism to align the SAI's features and propose a Light Field Mutual Attention Guidance Network (LF-MAGNet) constructed by multiple Mutual Attention Guidance blocks (MAGs) in a cascade manner. MAG achieves the mutual attention mechanism between center SAI and any surrounding SAI with two modules: the center attention guidance module (CAG) and the surrounding attention guidance module (SAG). Specifically, CAG first aligns the center-SAI features and any surrounding SAI features with the attention mechanism and then guides the surrounding SAI feature to learn from the center-SAI features, generating refined-surrounding SAI features. SAG aligns the refined-surrounding SAI feature and the original surrounding SAI feature and guides the refined surrounding SAI feature to learn from the original surrounding SAI features, generating the final outputs of MAG. With the help of MAG, LF-MAGNet can efficiently utilize different SAI features and generate high-quality light field image super-resolution results. Experiments are performed on commonly-used light field image super-resolution benchmarks. Qualitative and quantitative results prove the effectiveness of our LF-MAGNet.

This is a student paper.

This work is supported by the National Natural Science Foundation of China (No. 61273273), by the National Key Research and Development Plan (No. 2017YFC0112001), and by China Central Television (JG2018-0247).

H. Ma et al. (Eds.): PRCV 2021, LNCS 13021, pp. 105–116, 2021.
https://doi.org/10.1007/978-3-030-88010-1_9

Keywords: Light-field image super-resolution · Visual attention mechanism · Feature alignment · Sub-aperture image

1 Introduction

Light field image super-resolution (LFSR) is a newly emerging computer vision task which aims to enlarge the spatial resolution of sub-aperture images of light field image. It is a basic technology for many applications, such as virtual reality [5,29], depth sensing [15,16], 3D reconstruction [37,38] and so on.

Benefited from the development of deep learning, LFSR has achieved significant progress recently. LFCNN [28] is the first LFSR network that employs the bicubic interpolation to enlarge the sub-aperture images' spatial size and applies the convolution neural network to learn the angular information of LF image. After that, many LFSR networks are proposed for efficiently utilizing the spatial-angular information of LF images. For example, Yeung *et al.* [27] propose a spatial-angular separable convolution to extract the spatial-angular features. Wang *et al.* [21] formulate the different sub-aperture images into a macro-pixel image and propose an LF-InterNet to process the generated macro-pixel image to learn spatial-angle information for LFSR. Although they obtain promising LFSR performance, the features of SAIs are not efficiently utilized due to lack of feature alignment, resulting in sub-optimal results. Thus, we should align different SAI's features before the aggregation operation to improve the LFSR performance.

Recently, visual attention mechanism has been successfully applied to many computer vision tasks [2,4,12,20,25,32]. It can help the model to focus on more task-relevant feature representations and suppress the irrelevant features. In this paper, we want to use the visual attention mechanism to highlight the similar feature representations of different SAIs for aligning features, and the aligned features are subsequently processed for learning the complementary information of different SAIs. Based on this motivation, we propose a Mutual Attention Guidance Network (namely LF-MAGNet). Specifically, we propose a Mutual Attention Guidance block (MAG) to align the features of different SAIs. Each MAG includes two modules: the Center Attention Guidance module (CAG) and the Surrounding Attention Guidance module (SAG). Given two SAIs (*i.e.,* the center SAI (c-SAI) and its any surrounding SAI (s-SAI)), CAG first uses the attention module to align the feature of c-SAI and s-SAI, and then extract the complementary information from them to guide the s-SAI feature learning and generate the refined s-SAI feature. SAG uses the attention module to align the refined s-SAI and s-SAI features and then extract the complementary information from them to guide refined s-SAI feature learning. With the help of MAG, LF-MAGNet can fully utilize all SAIs and suppress the irrelevant representations for final LFSR. Extensive experiments are performed on five commonly-used LFSR benchmarks. Compared with the other advanced LFSR models, LF-MAGNet achieves new state-of-the-art results, which demonstrate the effectiveness of our model.

To summarize, our contributions are two-fold:

- We propose a new Mutual Attention Guidance block (MAG) to efficiently learn the complementary representations of different SAIs for final LFSR.
- Based on the proposed MAG block, we build an LF-MAGNet which achieves new state-of-the-art results on five commonly-used LFSR benchmarks.

The rest of this paper is organized as follows. We will review the related works in Sect. 2. Section 3 illustrates the details of our LF-MAGNet. Experiment and implementation details are illustrated in Sect. 4. We give a conclusion of this paper in Sect. 5.

2 Related Work

2.1 Light Field Image Super-Resolution

With the renaissance of deep learning, many CNN-based LFSR networks are proposed in recent years. Early CNN-based methods can be mainly divided into two categories: two-stage methods and one-stage methods. Two stages methods usually enlarge the spatial size of sub-aperture images firstly and then learn the LF angular information. For example, Yoon *et al.* [28] applied the bicubic interpolation to enlarge the spatial size of sub-aperture Images and employ a convolution neural network to reconstruct final LFSR results. Yuan *et al.* [30] first applied the SISR network to super-resolved the sub-aperture images and then used the designed epipolar plane image network to learn the LF structure information. One-stage methods simultaneously learn the spatial and angular information of LF images for LFSR. For instance, Yeung *et al.* [27] proposed a spatial-angular separable convolution to learn the spatial-angular information of LF images for LFSR. Recently, the research interests of LFSR are mainly about how to utilize the different views of sub-aperture images for LFSR efficiently. They employ part [23,31] or all [6,22] of Sub-Aperture Images to learn the complementary information of each other for final LFSR. Although the above methods achieve satisfactory performance, the complementary information of different sub-aperture images is not explicitly modeled to improve super-resolution results. Different from the above methods, we propose a mutual attention guidance mechanism to efficiently learn the similar representations of different sub-aperture images to align features for improving the LFSR performance.

2.2 Visual Attention Mechanism

Visual attention mechanism aims to enhance the task-relevant feature representations of a network, and has been successfully applied to various computer vision tasks, such as image or video classification [4,12,20,25], video object segmentation [11,33,35], human parsing [19,34,36], temporal action localization [18,26], single image super-resolution [2,32] and so on. There are some representative attention blocks among them. For example, Hu *et al.* [4] proposed a Squeeze-and-Excitation block to enhance the feature representations of different channels.

Woo *et al.* [12,25] modeled channel attention and spatial attention respectively to highlight the task-relevant feature representations. Wang *et al.* [20] designed a non-local block to extract the long-range context information for efficiently suppressing the irrelevant features. There are few works to explore the visual attention mechanism for LFSR. To this end, we propose a Mutual Attention Guidance block to efficiently align the features of different sub-aperture images for generating high-quality LF images.

3 Proposed Method

Following [21,22,31], we convert the input LF image from the RGB space to the YCbCr space. Given an LF image, only the Y channel of the image is super-resolved, and the bicubic interpolation is used to process the Cb and Cr channel of the images. We don't consider the channel dimension and denote the LF image as a 4-D tensor $L \in \mathbb{R}^{U\times V\times H\times W}$, where U and V stand for the LF angular resolution, and H and W represent the spatial resolution of each SAI. Given a low-resolution LF image $L_{lr} \in \mathbb{R}^{U\times V\times H\times W}$, LFSR aims to generate high-resolution SAIs while maintaining the angular resolution of LF image unchanged. We denote the LFSR result as $L_{sr} \in \mathbb{R}^{U\times V\times sH\times sW}$, where s $(s > 1)$ represents the upsampling scale factor. In this paper, we only explore LFSR with the square array distributed (*i.e.,* $U = V = A$).

LF-MAGNet is illustrated in Fig. 1. It consists of three modules: Shallow Feature Extraction $\mathcal{F}_{\text{feat}}$, Hierarchical Feature Extraction $\mathcal{F}_{\text{sa}}$, and LF Image Reconstruction $\mathcal{F}_{\text{up}}$. Given a low resolution LF image $L_{lr} \in \mathbb{R}^{U\times V\times H\times W}$, the feature process of LF-MAGNet is as follows:

$$\text{Shallow Feature Extraction: } F = \mathcal{F}_{\text{feat}}(L_{lr}) \in \mathbb{R}^{A^2\times H\times W}, \tag{1}$$

$$\text{Hierarchical Feature Extraction: } S = \mathcal{F}_{\text{sa}}(F) \in \mathbb{R}^{A^2\times H\times W}, \tag{2}$$

$$\text{LF Image Reconstruction: } L_{sr} = \mathcal{F}_{\text{up}}(S) \in \mathbb{R}^{A\times A\times sH\times sW}. \tag{3}$$

In the rest of this paper, we will introduce the $\mathcal{F}_{\text{feat}}$, $\mathcal{F}_{\text{sa}}$ and $\mathcal{F}_{\text{up}}$ in detail.

3.1 Shallow Feature Extraction

Residual learning has been successfully applied to single image super-resolution and achieved promising progresses [9,32]. To obtain efficient feature representations of each SAI, we also construct $\mathcal{F}_{\text{feat}}$ with multiple residual blocks for learning efficient shallow feature representations. The structure of $\mathcal{F}_{\text{feat}}$ is shown in Fig. 1(b). It consists of four residual blocks in a cascaded way. Each residual block is constructed with two 3×3 convolutions in a residual manner. We denote the center SAI of the low-resolution LF image as $I_c \in \mathbb{R}^{H\times W}$, and the surrounding SAI of the low-resolution LF image as $I_s^i \in \mathbb{R}^{H\times W}$, where $i \in [1, \cdots, A^2-1]$. Thus, the low-resolution LF image can be denoted as $L_{lr} = \{I_c, I_s^1, \cdots, I_s^{A^2-1}\}$. We use the same $\mathcal{F}_{\text{feat}}$ to separately process I_c and I_s^i, and obtain shallow feature F_c and F_s^i as follows:

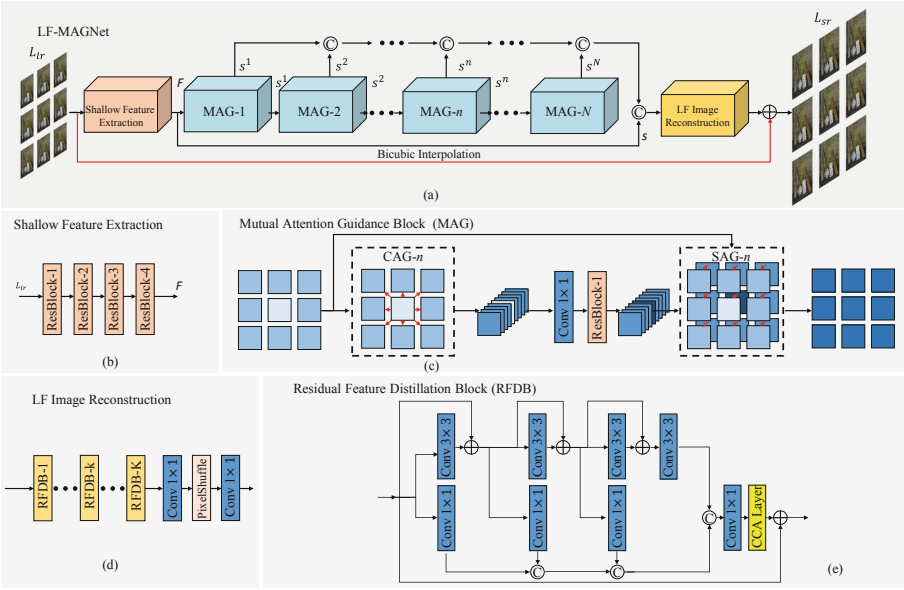

Fig. 1. Overview of LF-MAGNet. Given LR SAIs input, *Shallow Feature Extraction* module first processes the input. Then, multiple *MAG* blocks process the extracted feature to learn from each others. Finally, the output features are fed to *LF Image Reconstruction* module to generate the final LFSR results.

$$F_c = \mathcal{F}_{\text{feat}}(I_c) \in \mathbb{R}^{C\times H\times W}, F_s^i = \mathcal{F}_{\text{feat}}(I_s^i) \in \mathbb{R}^{C\times H\times W}, \tag{4}$$

where C denotes the feature channel number. The shallow features can be denoted as $F = [F_c, F_s^1, \cdots, F_s^{A^2-1}]$, and $[\cdot,\cdot]$ represents the feature concatenate operation.

3.2 Mutual Attention Guidance

Shallow Feature Extraction module $\mathcal{F}_{\text{feat}}$ extracts features $\boldsymbol{F}$ from input SAIs, but it does not efficiently utilize the complementary information between different SAIs, which is important for the network to reconstruct the high-quality LFSR image. The SAIs in an LF image share a similar appearance and vary slightly due to the view disparity of LF structure. Thus, to efficiently obtain the complementary information, we should first align different SAI's features and then aggregate them for LFSR. Inspired by the success of the visual attention mechanism, we propose a hierarchical feature extraction module $\mathcal{F}_{\text{sa}}$, which is constructed with N mutual attention guidance blocks (MAGs), to process different SAIs. The MAG block includes two modules: The Center Attention Guidance module (CAG) and the Surrounding Attention Guidance module (SAG). Given a center SAI feature F_c and the ith surrounding SAI feature F_s^i, CAG obtains the aligned feature A_{CAG}^i between F_c and F_s^i, and use the A_{CAG}^i to guide the F_s^i

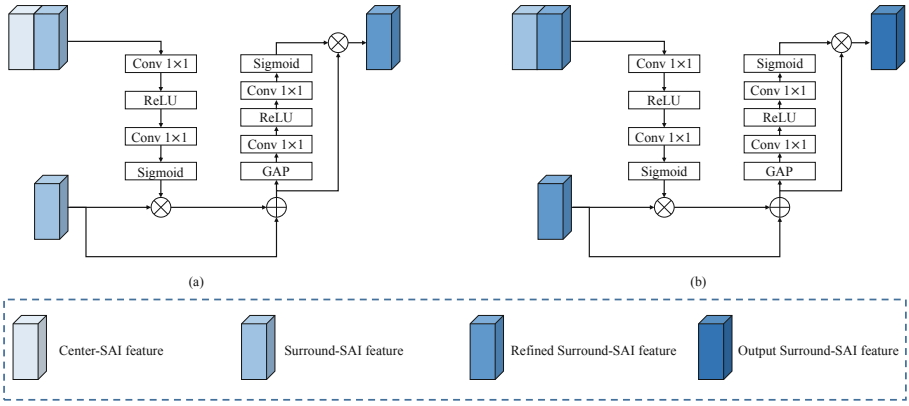

Fig. 2. illustration of MAG block. (a) Center attention guidance module. (b) Surrounding attention guidance module.

to learn the complementary information to generate the refined surrounding-SAI feature $\hat{F}_s^i$. While SAG learns the aligned feature A_{SAG}^i between the $\hat{F}_s^i$ and the F_s^i, and guide the $\hat{F}_s^i$ to learn the complementary representations from F_s^i. The above two attention guidance modules construct our proposed MAG, and the detailed illustrations of each module are as follows. For simplicity, we take the n-th MAG as an example.

Center Attention Guidance Module. The structure of CAG is illustrated in Fig. 2(a). Given the center-SAI feature F_c and the ith surrounding feature F_s^i, we first use the spatial attention $\mathcal{F}_{\mathcal{SA}}$, which is implemented with `conv` $1 \times 1 \rightarrow$`ReLU`$\rightarrow$`conv`$1 \times 1 \rightarrow$`Sigmoid`, to process the concatenated feature of F_c and F_s^i for extracting the complementary information between F_c and F_s^i, and obtain the aligned feature A_{CAG}^i as follows:

$$A_{CAG}^i = \mathcal{F}_{\mathcal{SA}}([F_c, F_s^i]). \tag{5}$$

Then, the surrounding-SAI feature F_s^i multiplies with A_{CAG}^i to receive the supplementary information from F_c, and adds with the original feature F_s^i to obtain the refined surrounding-SAI feature $\overline{F}_s^i$ as follows:

$$\overline{F}_s^i = (F_s^i \otimes A_{CAG}^i) \oplus F_s^i, \tag{6}$$

where $\otimes$ denotes the element-wise multiplication, $\oplus$ represents the element-wise summation. Finally, to further enhance the feature representations, the refined surrounding-SAI feature $\overline{F}_s^i$ is further processed by the channel attention $\mathcal{F}_{\mathcal{CA}}$ which is implemented with `Global Average Pooling (GAP)`$\rightarrow$`conv` $1 \times 1 \rightarrow$`ReLU`$\rightarrow$`conv`$1 \times 1 \rightarrow$`Sigmoid` as follows:

$$\hat{F}_s^i = \mathcal{F}_{\mathcal{CA}}(\overline{F}_s^i) \otimes \overline{F}_s^i. \tag{7}$$

Both $\mathcal{F}_{\mathcal{SA}}$ and $\mathcal{F}_{\mathcal{CA}}$ adopt the same channel dimension reduction ratio r to highlight the relevant representations. The effect of different settings of reduction ratio r for final performance are explored in Sect. 4.2.

Surrounding Attention Guidance Module. CAG helps F_s^i to align with F_c, learn the complementary information from F_c and the feature representations of the surrounding SAI F_s^i are refined. However, CAG only considers the center-SAI features while ignoring the surrounding SAI features, generating sub-optimal results. We also need to utilize the surrounding SAI features to improve the final LFSR performance. To this end, we propose a SAG module to guide the F_s^i to learn from the $\hat{F}_s^i$. The structure of SAG is illustrated in Fig. 2(b). The whole feature process is the same with CAG, and the major difference is that the input of $\mathcal{F}_{\mathcal{SA}}$. The detail processes of SAG are as follows:

$$A_{SAG}^i = \mathcal{F}_{\mathcal{SA}}([\hat{F}_s^i, F_s^i]), \tag{8}$$

$$\widetilde{F}_s^i = \hat{F}_s^i \otimes A_{SAG}^i \oplus \hat{F}_s^i, \tag{9}$$

$$\dot{F}_s^i = \mathcal{F}_{\mathcal{CA}}(\widetilde{F}_s^i) \otimes \widetilde{F}_s^i. \tag{10}$$

The refined center and surrounding SAI features construct the n-th MAG output $S^n = [\dot{F}_s^1, \cdots, \dot{F}_s^n]$, and different MAG outputs are concatenated to generate the hierarchical spatial-angular features $S = [S^1, \cdots, S^n]$ for LF image reconstruction.

3.3 LF Image Reconstruction

After getting the hierarchical features processed by the cascade MAGs, we need to fuse and upsample the extracted features for LFSR. Thus, we propose an LF Image Reconstruction module as illustrated in Fig. 1(d). It mainly consists of two components: the feature fusion part and the feature upsampling part. Feature fusion part is constructed by multiple lightweight residual feature distillation blocks (RFDBs) [10], which are illustrated in Fig. 1(e). With the help of RFDB, the feature fusion part can efficiently obtain the complementary information of different SAIs with fewer network parameters and computational resources. Afterward, the fused features are sent to 1×1 convolutions and PixelShuffle layer to enlarge the spatial size of each SAI for LFSR.

4 Experiment

4.1 Dataset and Implementation Details

Following [22], we select five commonly-used LFSR datasets (*i.e.,* EPFL [13], HCInew [3], HCIold [24], INRIA [8], and STFgantry [17]) to train and evaluate the performance of our LF-MAGNet on them. The angular resolution of LF images from the above datasets are all 9×9. For the training stage, we crop

Table 1. Performance comparisons of different numbers of MAG in LF-MAGNet on INRIA dataset for ×4 SR. MAG-n indicates the nth MAG block in LF-MAGNet.

MAG-1	MAG-2	MAG-3	MAG-4	MAG-5	PSNR	SSIM
✓					30.34	0.9418
✓	✓				30.48	0.9429
✓	✓	✓			30.88	0.9478
✓	✓	✓	✓		**30.94**	**0.9489**
✓	✓	✓	✓	✓	30.84	0.9480

64×64 patch from each SAI and use the bicubic interpolation to generate ×2 and ×4 LR patch. Random horizontal rotation, vertical rotation, and 90° rotation are employed to augment the training data. Spatial and angular resolution all needs to be processed simultaneously for maintaining the LF image structure.

We only process 5×5 angular resolution for ×2 and ×4 SR. LF-MAGNet is optimized with *L1* loss, and we select Adam to optimize the network. All experiments are performed on a single NVIDIA RTX 2080Ti GPU card with the Pytorch framework. The batch size is set to 8, and the initial learning rate is set to 2×10^{-4}. We train LF-MAGNet with a total of 50 epochs, and the learning rate is decreased to half after every 15 epochs.

Following [21,22], we choose PSNR and SSIM to evaluate the LFSR performance on the Y channel. To obtain the performance score of M scenes with angular resolution $A \times A$, we first calculate the performance score of each SAI. Then, we average the performance score of A^2 SAIs to get the performance score of one scene. Finally, The performance score of M scenes are averaged to obtain the final performance score.

4.2 Ablation Studies

Number of MAG. We explore the number of MAG from LF-MAGNet for final LF-SR performance. The results are displayed in Table 1. We can see that with the increase of the number of MAG, the LFSR performances are improved. LF-MAGNet achieves the best result when $N = 4$. After that, the performance is decreased when the number becomes large. The reason is that large network parameters hinder the network optimization. Thus, we set $N = 4$ in our LF-MAGNet.

Table 2. Performance comparisons of different attention guidance on INRIA dataset for ×4 SR.

Variants	PSNR	SSIM
LF-MAGNet w/o CAG and SAG	30.61	0.9480
LF-MAGNet w/o CAG	30.72	0.9486
LF-MAGNet w/o SAG	30.79	0.9484
LF-MAGNet	**30.94**	**0.9489**

Table 3. Performance comparisons of different reduction ratio r settings in MAG block on INRIA dataset for ×4 SR.

r	1	2	4	8
LF-MAGNet	30.69	30.80	**30.94**	30.88

Table 4. Performance comparison of different methods for ×2 and ×4 SR. The best and the second best results are marked with bold and italic.

Method	Scale	EPFL	HCInew	HCIold	INRIA	STFgantry
Bicubic	×2	29.50/0.9350	31.69/0.9335	37.46/0.9776	31.10/0.9563	30.82/0.9473
VDSR	×2	32.50/0.9599	34.37/0.9563	40.61/0.9867	34.43/0.9742	35.54/0.9790
EDSR	×2	33.09/0.9631	34.83/0.9594	41.01/0.9875	34.97/0.9765	36.29/0.9819
RCAN	×2	33.16/0.9635	34.98/0.9602	41.05/0.9875	35.01/0.9769	36.33/0.9825
LFBM5D	×2	31.15/0.9545	33.72/0.9548	39.62/0.9854	32.85/0.9695	33.55/0.9718
GB	×2	31.22/0.9591	35.25/0.9692	40.21/0.9879	32.76/0.9724	35.44/0.9835
resLF	×2	32.75/0.9672	36.07/0.9715	42.61/0.9922	34.57/0.9784	36.89/0.9873
LFSSR	×2	33.69/0.9748	36.86/0.9753	43.75/0.9939	35.27/0.9834	38.07/0.9902
LF-InterNet	×2	*34.14*/*0.9761*	37.28/0.9769	*44.45*/*0.9945*	*35.80*/**0.9846**	38.72/0.9916
LF-DFnet	×2	**34.44**/**0.9766**	*37.44*/**0.9786**	44.23/0.9943	**36.36**/*0.9841*	*39.61*/**0.9935**
LF-MAGNet	×2	**34.44**/0.9761	**37.62**/*0.9783*	**44.62**/**0.9946**	**36.36**/**0.9846**	**39.66**/*0.9931*
Bicubic	×4	25.14/0.8311	27.61/0.8507	32.42/0.9335	26.82/0.8860	25.93/0.8431
VDSR	×4	27.25/0.8782	29.31/0.8828	34.81/0.9518	29.19/0.9208	28.51/0.9012
EDSR	×4	27.84/0.8858	29.60/0.8874	35.18/0.9538	29.66/0.9259	28.70/0.9075
RCAN	×4	27.88/0.8863	29.63/0.8880	35.20/0.9540	29.76/0.9273	28.90/0.9110
LFBM5D	×4	26.61/0.8689	29.13/0.8823	34.23/0.9510	28.49/0.9137	28.30/0.9002
GB	×4	26.02/0.8628	28.92/0.8842	33.74/0.9497	27.73/0.9085	28.11/0.9014
resLF	×4	27.46/0.8899	29.92/0.9011	36.12/0.9651	29.64/0.9339	28.99/0.9214
LFSSR	×4	28.27/0.9080	30.72/0.9124	36.70/0.9690	30.31/0.9446	30.15/0.9385
LF-InterNet	×4	28.67/0.9143	30.98/0.9165	37.11/0.9715	30.64/0.9486	30.53/0.9426
LF-DFnet	×4	*28.77*/*0.9165*	**31.23**/**0.9196**	*37.32*/*0.9718*	*30.83*/**0.9503**	**31.15**/**0.9494**
LF-MAGNet	×4	**29.03**/**0.9170**	*31.09*/*0.9162*	**37.40**/**0.9721**	**30.94**/*0.9489*	*30.71*/*0.9428*

Ablation of LF-MAGNet Design. To explore the effectiveness of our MAG block, we design three variants of LF-MAGNet, which are LF-MAGNet w/o CAG and SAG, LF-MAGNet w/o CAG, and LF-MAGNet w/o SAG. The results are presented in Table 2. Without the assistance of MAG, LF-MAGNet achieves the lowest LFSR performance. Only CAG or SAG can improve the LFSR results, but they don't obtain the best results. LF-MAGNet with full implementation of MAG achieves the best performance, which demonstrates the effectiveness of our proposed mutual attention guidance mechanism. We also explore the reduction ratio settings in each MAG block. The performance comparisons are displayed in Table 3. We can see that LF-MAGNet achieves the best result when $r = 4$. Thus, we select $r = 4$ in our final model.

4.3 Comparisons with the State-of-The-Arts

We compare our LF-MAGNet with other image super-resolution (SISR) methods, including single image super-resolution methods (*i.e.,* VDSR [7], EDSR [9],

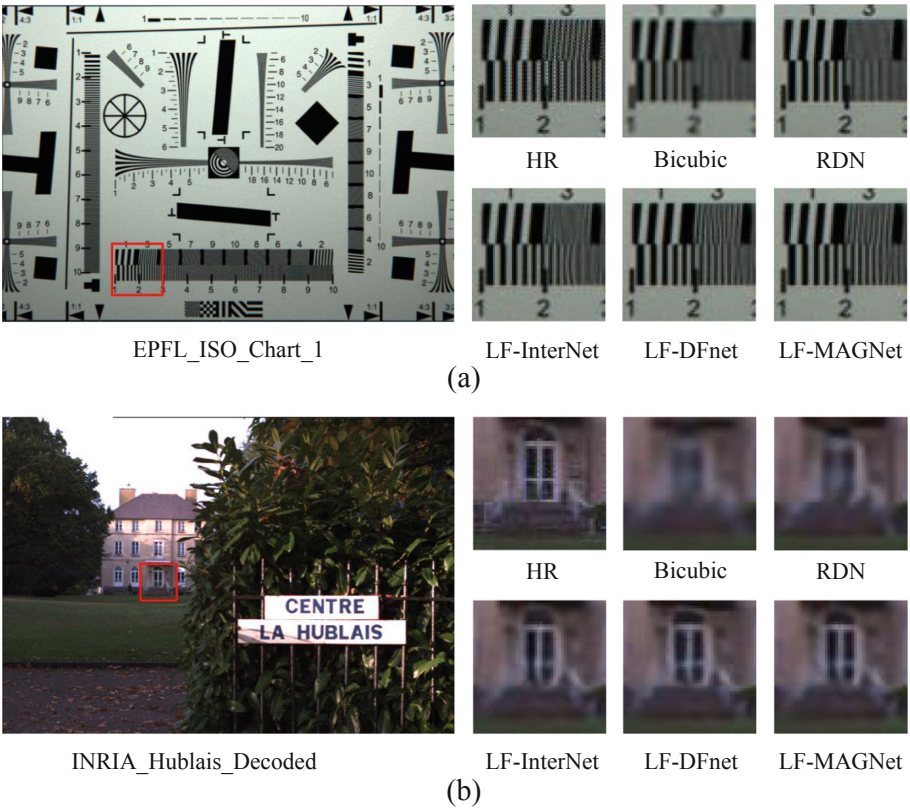

Fig. 3. Visual results of different methods for (a) ×2 SR and (b) ×4 SR.

and RCAN [32]), and LFSR methods (*i.e.,* LFBM5D [1] GB [14] LFSSR [27] resLF [31] LF-InterNet [21] LF-DFNet [22]). We also select bicubic interpolation as the baseline method for performance comparisons.

Table 4 reports the results on 5×5 LF images for ×2 and ×4 SR. We can see that LF-MAGNet achieves new state-of-the-art results on all LFSR benchmarks. Compared with SISR methods, LF-MAGNet achieves a significant performance improvement by efficiently utilizing the complementary information from different SAIs. Compared with the other LFSR methods, LF-MAGNet also outperforms them, demonstrating the effectiveness of our network.

Figure 3 present the qualitative results of different super-resolution methods. We can see that LF-MAGNet outputs more clear LF images with abundant textures and details compared with other methods, which further proves the superiority of LF-MAGNet.

5 Conclusion

In this paper, we propose a Mutual Attention Guidance Network (namely LF-MAGNet) for LFSR. LF-MAGNet is mainly constructed by multiple MAGs, which helps the center-SAI and surrounding SAIs of an LF image learn from each other efficiently. Extensive experiments are performed on commonly-used light filed image super-resolution benchmarks. Our LF-MAGNet achieves new state-of-the-art results compared with other advanced light filed image super-resolution networks.

References

1. Alain, M., Smolic, A.: Light field super-resolution via lfbm5d sparse coding. In: ICIP, pp. 2501–2505 (2018)
2. Dai, T., Cai, J., Zhang, Y., Xia, S.T., Zhang, L.: Second-order attention network for single image super-resolution. In: CVPR, pp. 11065–11074 (2019)
3. Honauer, K., Johannsen, O., Kondermann, D., Goldluecke, B.: A dataset and evaluation methodology for depth estimation on 4D light fields. In: ACCV, pp. 19–34 (2016)
4. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: CVPR, pp. 7132–7141 (2018)
5. Huang, F.C., et al.: The light field stereoscope-immersive computer graphics via factored near-eye field displays with focus cues. SIGGRAPH (2015)
6. Jin, J., Hou, J., Chen, J., Kwong, S.: Light field spatial super-resolution via deep combinatorial geometry embedding and structural consistency regularization. In: CVPR, pp. 2260–2269 (2020)
7. Kim, J., Lee, J.K., Lee, K.M.: Accurate image super-resolution using very deep convolutional networks. In: CVPR, pp. 1646–1654 (2016)
8. Le Pendu, M., Jiang, X., Guillemot, C.: Light field inpainting propagation via low rank matrix completion. IEEE TIP **27**(4), 1981–1993 (2018)
9. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: CVPRW, pp. 136–144 (2017)
10. Liu, J., Tang, J., Wu, G.: Residual feature distillation network for lightweight image super-resolution. arXiv preprint arXiv:2009.11551 (2020)
11. Lu, X., Wang, W., Danelljan, M., Zhou, T., Shen, J., Van Gool, L.: Video object segmentation with episodic graph memory networks. In: ECCV, pp. 661–679 (2020)
12. Park, J., Woo, S., Lee, J.Y., Kweon, I.S.: Bam: Bottleneck attention module. arXiv preprint arXiv:1807.06514 (2018)
13. Rerabek, M., Ebrahimi, T.: New light field image dataset. In: 8th International Conference on Quality of Multimedia Experience. No. CONF (2016)
14. Rossi, M., Frossard, P.: Geometry-consistent light field super-resolution via graph-based regularization. IEEE TIP **27**(9), 4207–4218 (2018)
15. Sheng, H., Zhang, S., Cao, X., Fang, Y., Xiong, Z.: Geometric occlusion analysis in depth estimation using integral guided filter for light-field image. IEEE TIP **26**(12), 5758–5771 (2017)
16. Shi, J., Jiang, X., Guillemot, C.: A framework for learning depth from a flexible subset of dense and sparse light field views. IEEE TIP **28**(12), 5867–5880 (2019)
17. Vaish, V., Adams, A.: The (new) stanford light field archive. Computer Graphics Laboratory, Stanford University 6(7) (2008)

18. Wang, B., Yang, L., Zhao, Y.: POLO: learning explicit cross-modality fusion for temporal action localization. IEEE Signal Process. Lett. **28**, 503–507 (2021)
19. Wang, W., Zhou, T., Qi, S., Shen, J., Zhu, S.C.: Hierarchical human semantic parsing with comprehensive part-relation modeling. IEEE TPAMI (2021)
20. Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: CVPR, pp. 7794–7803 (2018)
21. Wang, Y., Wang, L., Yang, J., An, W., Yu, J., Guo, Y.: Spatial-angular interaction for light field image super-resolution. In: ECCV, pp. 290–308 (2020)
22. Wang, Y., Yang, J., Wang, L., Ying, X., Wu, T., An, W., Guo, Y.: Light field image super-resolution using deformable convolution. IEEE TIP **30**, 1057–1071 (2020)
23. Wang, Y., Liu, F., Zhang, K., Hou, G., Sun, Z., Tan, T.: LFNet: a novel bidirectional recurrent convolutional neural network for light-field image super-resolution. IEEE TIP **27**(9), 4274–4286 (2018)
24. Wanner, S., Meister, S., Goldluecke, B.: Datasets and benchmarks for densely sampled 4D light fields. Vis. Modell. Visual. **13**, 225–226 (2013)
25. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: CBAM: convolutional block attention module. In: ECCV, pp. 3–19 (2018)
26. Yang, L., Peng, H., Zhang, D., Fu, J., Han, J.: Revisiting anchor mechanisms for temporal action localization. IEEE TIP **29**, 8535–8548 (2020)
27. Yeung, H.W.F., Hou, J., Chen, X., Chen, J., Chen, Z., Chung, Y.Y.: Light field spatial super-resolution using deep efficient spatial-angular separable convolution. IEEE TIP **28**(5), 2319–2330 (2018)
28. Yoon, Y., Jeon, H.G., Yoo, D., Lee, J.Y., Kweon, I.S.: Light-field image super-resolution using convolutional neural network. IEEE Signal Process. Lett. **24**(6), 848–852 (2017)
29. Yu, J.: A light-field journey to virtual reality. IEEE Multimedia **24**(2), 104–112 (2017)
30. Yuan, Y., Cao, Z., Su, L.: Light-field image superresolution using a combined deep CNN based on EPI. IEEE Signal Process. Lett. **25**(9), 1359–1363 (2018)
31. Zhang, S., Lin, Y., Sheng, H.: Residual networks for light field image super-resolution. In: CVPR, pp. 11046–11055 (2019)
32. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: ECCV, pp. 286–301 (2018)
33. Zhou, T., Li, J., Wang, S., Tao, R., Shen, J.: MATNet: motion-attentive transition network for zero-shot video object segmentation. IEEE TIP **29**, 8326–8338 (2020)
34. Zhou, T., Qi, S., Wang, W., Shen, J., Zhu, S.C.: Cascaded parsing of human-object interaction recognition. IEEE TPAMI (2021)
35. Zhou, T., Wang, S., Zhou, Y., Yao, Y., Li, J., Shao, L.: Motion-attentive transition for zero-shot video object segmentation. In: AAAI, pp. 13066–13073 (2020)
36. Zhou, T., Wang, W., Liu, S., Yang, Y., Van Gool, L.: Differentiable multi-granularity human representation learning for instance-aware human semantic parsing. In: CVPR, pp. 1622–1631 (2021)
37. Zhu, H., Wang, Q., Yu, J.: Occlusion-model guided antiocclusion depth estimation in light field. IEEE J. Selected Top. Signal Process. **11**(7), 965–978 (2017)
38. Zhu, H., Zhang, Q., Wang, Q.: 4D light field superpixel and segmentation. In: CVPR, pp. 6384–6392 (2017)