LOCAL-GLOBAL FEATURE AGGREGATION FOR LIGHT FIELD IMAGE SUPER-RESOLUTION

Yan Wang[1], *Yao Lu*[1*], *Shunzhou Wang*[1], *Wenyao Zhang*[1] *and Zijian Wang*[1]

[1]Beijing Key Laboratory of Intelligent Information Technology,
School of Computer Science and Technology, Beijing Institute of Technology, China

ABSTRACT

Deep convolutional neural networks (CNNs) have been widely explored in light field (LF) image super-resolution (SR) to achieve remarkable progress. However, most of the existing CNNs-based methods ignore the similarity of local neighbor views in the 4D LF data. Besides, due to the limitations of CNNs, these methods can't fully model the global spatial properties of the whole LF images. In this paper, we propose a network with Local-Global Feature Aggregation (LF-LGFA) to handle these problems for LF image SR. Specifically, the Local Aggregation Module is designed to incorporate the local angular information by utilizing the similarity of the local neighbor views' features in LF images. Moreover, the Global Aggregation Module is designed to capture long-range spatial information via row-wise and column-wise self-attention. Extensive experimental results on five public LF datasets demonstrate that our method achieves comparable results against state-of-the-art techniques.

Index Terms— Light field, image super-resolution, feature aggregation, self-attention

1. INTRODUCTION

Light Field (LF) cameras capture scenes from multiple viewpoints and provide a richer interpretation of the scenes, enabling many attractive applications such as depth sensing [1, 2], de-occlusion [3, 4], etc. The commercialized light field cameras have an essential trade-off between the angular and spatial resolutions [5]. The limited spatial resolution restricts the development of LF imaging [6]. Since high-resolution (HR) images are required in various LF applications, light field image super-resolution (LFSR), which aims to reconstruct HR sub-aperture images (SAIs) from their corresponding low-resolution (LR) SAIs, has become an essential and popular research topic in recent years.

Due to the prosperity of deep learning techniques, convolutional neural networks (CNNs) based methods have demonstrated promising performance for LFSR. Yoon et al. [7] proposed the first CNN-based method LFCNN for LFSR. They used SRCNN to super-resolve SAIs separately and fine-tuned them in pairs to incorporate angular information. After that, many LFSR networks were proposed to exploit the spatial-angular information of LF images, achieving state-of-the-art performance. Wang et al. [8] proposed a spatial-angular interactive network named LF-InterNet to incorporate spatial and angular information. To address the angular disparity problem, Wang et al. [9] proposed an angular deformable alignment module and a collect-and-distribute approach to achieve the high-quality reconstruction of each LF image.

However, existing CNN-based LFSR methods have the following two limitations. First, they ignore the similarity of local neighbor SAIs in the 4D LF data. These methods integrate angular information without considering view position and the magnitude of parallax. Second, these methods only rely on the local receptive field of the CNNs and can't model global spatial features from the 4D LF data. These two limitations lead to insufficient utilization of angular and spatial information, resulting in sub-optimal results in LFSR.

For each SAI in LF images, its neighbor SAIs have smaller angular parallax with itself, and the texture details of neighbor SAIs are more similar to it. Based on these characteristics of the LF images, we design a Local Aggregation Module to incorporate local angular information by taking advantage of the similarity of the local neighbor view's features. Specifically, we regard each SAI as the central view and enhance its feature by aligning and aggregating its local neighbor SAIs' features. Recently, self-attention (SA) mechanisms showing promising results for computer vision tasks [10–20]. They consume fewer parameters and efficiently capture non-local context information for image restoration. Inspired by the recent advances of SA, we design a Global Aggregation Module via row-wise and column-wise self-attention to capture long-range spatial dependencies for each SAI.

The main contributions of this paper are summarized as follows: 1) We design a novel Local Aggregation Module to incorporate the local angular information and a Global Aggregation Module to capture global spatial information from the 4D LF data. 2) Based on these two modules, we propose a network named LF-LGFA, which achieves comparable

This work is supported by the National Key Research and Development Plan (No.2017YFC0112001), and by China Central Television (JG2018-0247). ∗Corresponding author: Yao Lu (vis_yl@bit.edu.cn).

results against state-of-the-art LFSR methods through Local-Global feature aggregation.

2. PROPOSED METHOD

2.1. Overview

LFSR aims to reconstruct high-resolution LF images from their low-resolution counterparts. Following [8, 9, 21, 22], we convert LF images to YCbCr color space and only super-resolve the Y channel. Thus, without considering the channel dimension, the input of our network is a low-resolution LF image with a 4D tensor $L_{\text{IN}} \in \mathbb{R}^{U\times V\times H\times W}$, where U and V represent angular dimensions, H and W represent spatial dimensions. Specifically, we achieve LFSR using SAIs distributed in a square array (i.e., $U = V = A$).

As illustrated in Fig.1(a), given the $L_{\text{IN}} \in \mathbb{R}^{U\times V\times H\times W}$, the Feature Extraction Module firstly extracts the shallow SAI features $\mathcal{F}_S \in \mathbb{R}^{A^2\times H\times W\times C}$. Then, the extracted features are sequentially processed by a Local Aggregation Module and a Global Aggregation Module to generate neighbor-aware SAI features $\mathcal{F}_L \in \mathbb{R}^{A^2\times H\times W\times C}$ and non-local SAI features $\mathcal{F}_G \in \mathbb{R}^{A^2\times H\times W\times 2C}$, respectively. Finally, the enhanced features $\mathcal{F}_{LG} = [\mathcal{F}_S, \mathcal{F}_L, \mathcal{F}_G] \in \mathbb{R}^{A^2\times H\times W\times 4C}$ are fed to the Reconstruction Module to generate the super-resolved results $L_{\text{OUT}} \in \mathbb{R}^{U\times V\times \alpha H\times \alpha W}$. We implement the Feature Extraction module with the residual atrous spatial pyramid pooling module (ResASPP) [23] and the Reconstruction Module with the information multi-distillation blocks (IMDB) [9] combined with the pixel shuffling operation [24]. The implementation details of the Local Aggregation Module and Global Aggregation Module are introduced as follows.

2.2. Local Aggregation

The Local Aggregation Module is designed to obtain the complementary information from each SAI's local neighborhood to enrich each SAI's high-frequency details. Specifically, as shown in Fig.1(b), we divide an LF image into nine groups according to their angular views (i.e.,$L = \{V_i\}_{i=1}^{9}$), and index them from top-left to bottom-right. We regard each SAI as the central view and enhance its feature by exploiting the complementary information from the neighbor SAIs. Without loss of generality, we take local aggregation on V_1 as an example, which can be readily generalized to other SAIs. When V_1 is regarded as the central view (i.e.,$V_c = V_1$), its neighbor views are denoted as $V_s = \{V_2, V_4, V_5\}$. The deformable convolution $\mathcal{DCN}$ is used to align neighbor views features $\mathcal{F}_{V_s}$ with their central-view feature $\mathcal{F}_{V_c}$ according to their corresponding offsets set $\Delta P_{V_s\to V_c}$ as:

$$\mathcal{F}_{V_s\to V_c}=\mathcal{DCN}(\mathcal{F}_{V_s}, \Delta P_{V_s\to V_c}), \tag{1}$$

where $\mathcal{F}_{V_s\to V_c}$ denotes the neighbor view features after aligning with their central view feature. The offsets set is obtained by a network with $conv1\times 1 \to ReLU \to ResASPP \to conv1\times 1$. The details of the Feature Alignment module are shown in Fig.1(d).

Then, we concatenate the aligned neighbor view features with the central view feature along the channel dimension and use a feature aggregation network $\mathcal{C}_{Agg}$ to incorporate local angular information. Next, the original central view feature $\mathcal{F}_{V_c}$ is fused as:

$$\mathcal{F}_L^i=\mathcal{C}_{Agg}([\mathcal{F}_{V_s\to V_c}, \mathcal{F}_{V_c}]) \oplus \mathcal{F}_{V_c}, \tag{2}$$

where $\oplus$ denotes the element-wise summation, $\mathcal{F}_L^i$ denotes the local-aggregation enhanced feature of the central view V_i.

Finally, each SAI's enhanced features are concatenated to generate the Local Aggregation Module output as:

$$\mathcal{F}_L = [\mathcal{F}_L^1, \mathcal{F}_L^2, ..., \mathcal{F}_L^n], \tag{3}$$

where n denotes the total number of the SAIs.

2.3. Global Aggregation

The Global Aggregation Module is proposed to aggregate long-range context information of different SAIs. Consequently, we design a multi-view aggregation module via a row-wise self-attention and a column-wise self-attention, which can discriminately model global spatial relations in the horizontal or vertical direction, as shown in Fig.1(c).

For simplicity, we take the row-wise self-attention as an example, which can be generalized to the column-wise self-attention. Specifically, given the j-th row of local-aggregation enhanced features $\mathcal{F}_{L,row}^j \in \mathbb{R}^{A\times H\times W\times C}$, three independent 1×1 convolutions $\mathcal{W}_q, \mathcal{W}_k$ and $\mathcal{W}_v$ are applied to extract the feature embeddings. Following [15], we use the $unfold$ operation to extract sliding local patches with stride s and patch size of $H_p \times W_p$. After that, we obtain three groups of 3D patches, and each group has $N = AHW/H_pW_p$ patches with the dimension of $d = C \times H_p \times W_p$. Finally, we generate the query, key, and value $\mathcal{Q}, \mathcal{K}, \mathcal{V} \in \mathbb{R}^{AN\times C\times H_p\times W_p}$ as:

$$\mathcal{Q} = unfold(\mathcal{W}_q(\mathcal{F}_{L,row}^j)), \tag{4}$$

$$\mathcal{K} = unfold(\mathcal{W}_k(\mathcal{F}_{L,row}^j)), \tag{5}$$

$$\mathcal{V} = unfold(\mathcal{W}_v(\mathcal{F}_{L,row}^j)), \tag{6}$$

Next, we reshape each 3D patches group into a new query matrix $\bar{\mathcal{Q}}$ and a key matrix $\bar{\mathcal{K}}$. Then, we calculate the similarity matrix between them and aggregate it with the value $\mathcal{V}$. Note that the similarity matrix is related to all embedding tokens of adjacent row SAIs. Thus, we can aggregate the non-local spatial information horizontally. Then, we use the fold operation $fold$ to combine those updating sliding local patches into a feature map with the size of $A\times H\times W\times C$. After that, we obtain the j-th row's aggregated feature $\mathcal{F}_{G,row}^j$ by using a 1×1 convolution layer $\mathcal{W}$ and adding the original feature $\mathcal{F}_{L,row}^j$. This process can be formulated as:

$$\mathcal{F}_{G,row}^j = \mathcal{F}_{L,row}^j + \mathcal{W}((fold(\mathcal{V}\cdot softmax(\bar{\mathcal{Q}}^T\cdot\bar{\mathcal{K}}))), \tag{7}$$

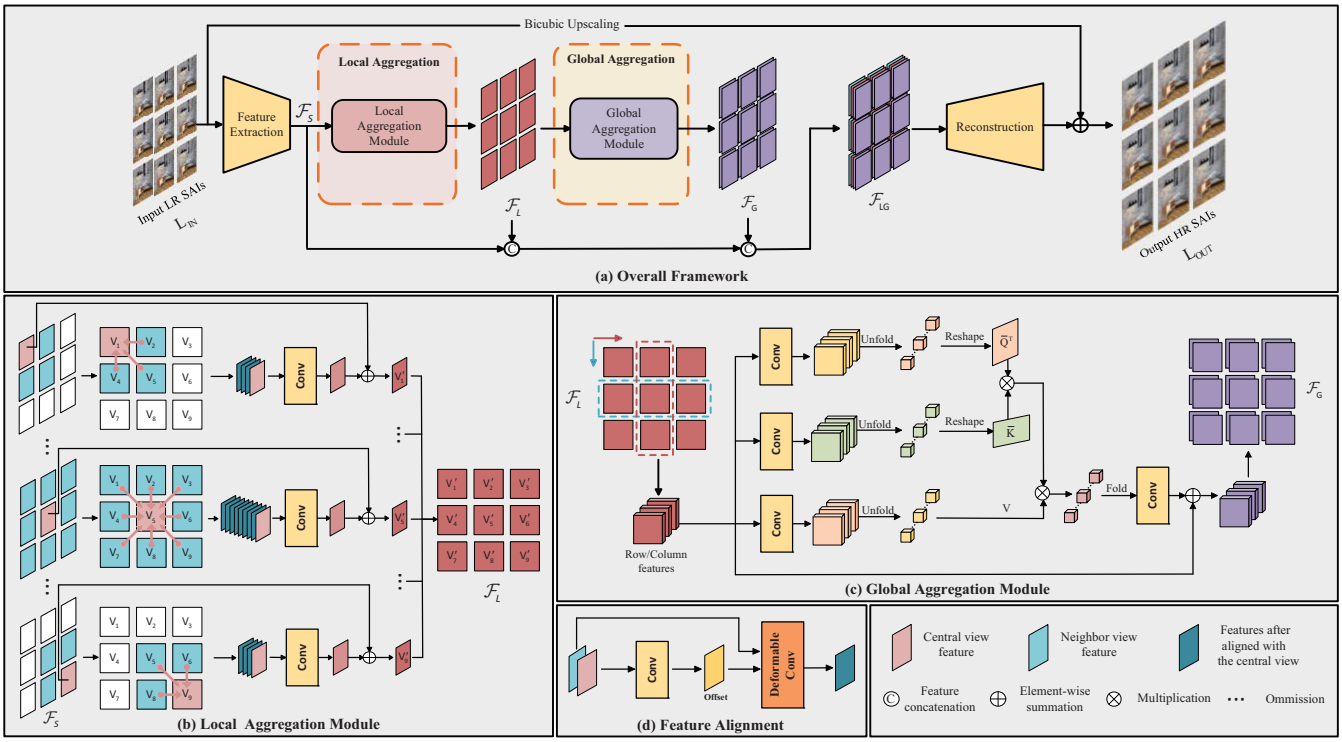

Fig. 1. An overview of our LF-LGFA.

For each row, we obtain the corresponding aggregated features $\mathcal{F}^{j}_{G,row}$ by performing the above operations. Finally, every row's aggregated features are concatenated to generate the row-wise global-aggregation enhanced features $\mathcal{F}_{G,row} = [\mathcal{F}^{1}_{G,row}, \mathcal{F}^{2}_{G,row}, ..., \mathcal{F}^{j}_{G,row}]$, where j denotes the total number of the rows.

In the same way, we can also get the column-wise global-aggregation enhanced features $\mathcal{F}_{G,col}$. Finally, we concatenate the two parts to generate the Global Aggregation stage's output $\mathcal{F}_G = [\mathcal{F}_{G,row}, \mathcal{F}_{G,col}]$ for the following light filed image reconstruction.

3. EXPERIMENTS

3.1. Implementation Details

We use five public LF datasets (i.e., EPFL [25], HCInew [26], HCIold [27], INRIA [28], STFgantry [29]) provided by [9], which contain 144 training images and 23 test images in total. All the LF images in the training and test sets have an angular resolution of 5×5. In the training stage, we crop each SAI into patches of size 64×64 and use the bicubic downsampling approach to generate LR patches of size 16×16.

Our network is trained with the L1 loss function and optimized using the Adam optimizer. The weights of our network are initialized using the Kaiming method [30]. All experiments are implemented in the PyTorch framework with one Nvidia GTX 2080Ti GPU card. The batch size is set to 8. The learning rate is initially set to 2×10^{-4} and decreases by a factor of 0.5 for every 15 epochs. We train LF-LGFA with a total of 50 epochs.

We calculate PSNR and SSIM [31] on the Y channel as quantitative metrics for performance evaluation. To obtain the score for a dataset with M scenes, we calculate the metrics on the $A \times A$ SAIs of each scene separately, then obtain the score for this dataset by averaging all the $M \times A^2$ scores.

3.2. Comparisons with state-of-the-art methods

We compare our LF-LGFA to several state-of-the-art methods, including two single image SR methods [32,33] and four LF image SR methods [8,9,21,22]. We also select the bicubic interpolation as the baseline method for comparisons.

1) *Quantitative Results:* Table 1 shows the quantitative comparisons between our method and other state-of-the-art methods for 4× SR. Our LF-LGFA achieves the highest PSNR results on all five datasets and comparable SSIM results. Note that, the PSNR improvements of our LF-LGFA are very significant on the EPFL dataset [25] (i.e., 0.24 dB higher than the second top-performing method [9]). That is because, SAIs in the EPFL dataset have more complex structures and larger disparity variations. By using Local-Global feature aggregation, our LF-LGFA can handle these complex scenes with outstanding performance.

2) *Qualitative Results:* Fig.2 shows the qualitative result comparisons achieved by different methods for 4× SR. Our

Table 1. Performance comparisons of different methods for 4× SR. The best and the second best results are marked with red and blue colors.

Methods	*EPFL*	*HCInew*	*HCIold*	*INRIA*	*STFgantry*
Bicubic	25.14/0.8311	27.61/0.8507	32.42/0.9335	26.82/0.8860	25.93/0.8431
EDSR [32]	27.84/0.8858	29.60/0.8874	35.18/0.9538	29.66/0.9259	28.70/0.9075
RCAN [33]	27.88/0.8863	29.63/0.8880	35.20/0.9540	29.76/0.9273	28.90/0.9110
resLF [21]	27.46/0.8899	29.92/0.9011	36.12/0.9651	29.64/0.9339	28.99/0.9214
LFSSR [22]	28.27/0.9080	30.72/0.9124	36.70/0.9690	30.31/0.9446	30.15/0.9385
LF-InterNet [8]	28.67/0.9143	30.98/0.9165	37.11/0.9715	30.64/0.9486	30.53/0.9426
LF-DFnet [9]	28.77/0.9165	31.23/0.9196	37.32/0.9718	30.83/0.9503	31.15/0.9494
LF-LGFA (Ours)	29.01/0.9159	31.24/0.9187	37.40/0.9721	30.96/0.9487	31.17/0.9481

Table 2. Parameters (Params), FLOPs, and reconstruction accuracy achieved by different methods for 4×SR. Note that, FLOPs is calculated on an input LF with a size of $5 \times 5 \times 32 \times 32$. We use PSNR and SSIM values averaged over five datasets [25]- [29] to represent their reconstruction accuracy.

Methods	Params	FLOPs	PSNR/SSIM
EDSR [32]	38.89M	40.66×25G	30.20/0.9121
RCAN [33]	15.36M	15.65×25G	30.27/0.9133
resLF [21]	6.79M	39.70G	30.43/0.9223
LFSSR [22]	1.61M	128.44G	31.23/0.9345
LF-InterNet [8]	5.23M	50.10G	31.59/0.9387
LF-DFnet [9]	3.99M	57.31G	31.86/0.9415
LF-LGFA (Ours)	3.78M	113.34G	31.96/0.9403

Table 3. Ablation Studies of LF-LGFA for 4 × SR. We use PSNR and SSIM values averaged over five datasets to represent their reconstruction accuracy. Note that, LA represents models using Local Aggregation. Row-GA and Column-GA represent models using Row-wise Global Aggregation and using Column-wise Global Aggregation, respectively.

#	LA	Row-GA	Column-GA	Params	Average
1				3.74M	30.22/0.9123
2	√			3.77M	31.76/0.9378
3	√	√		3.78M	31.88/0.9395
4	√		√	3.78M	31.89/0.9393
5		√	√	3.73M	30.64/0.9197
6	√	√	√	3.78M	31.96/0.9407

LF-LGFA can well preserve the details of the textures in the SR images and achieves better visual performance.

3) *Efficiency:* We compare our LF-LGFA to several competitive methods in terms of the number of parameters (i.e., Params) and FLOPs. As shown in Table 2, our method achieves the highest PSNR scores with a small number of parameters. Note that, by adding the complicated Global Aggregation module, our network has a notable performance improvement at the cost of a reasonable increase of FLOPs.

3.3. Ablation Study

We introduce several variants with different architectures to investigate the effectiveness of our network modules. Meanwhile, we replace the removed modules with convolutions to make these variant's parameters approximately the same. Table 3 shows the comparative results achieved by LF-LGFA and its variants.

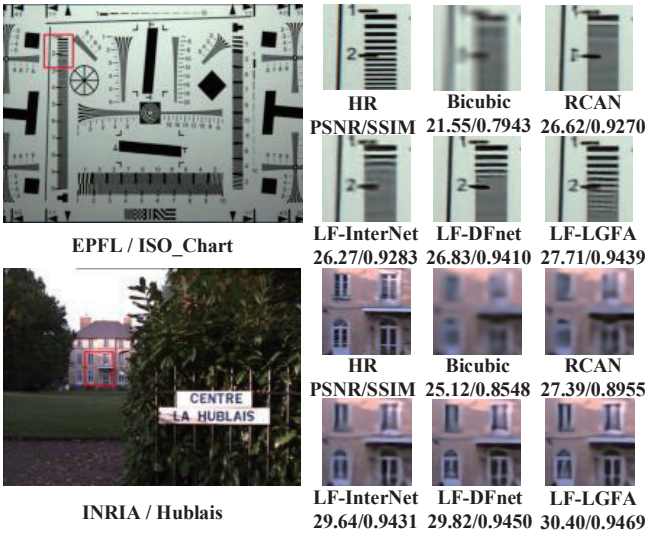

Fig. 2. Visual comparisons of different methods for 4× SR.

1) *Local Aggregation:* We compare the performance of *model-2* to *model-1* and *model-6* to *model-5* to validate the effectiveness of the Local Aggregation. By using the Local Aggregation module, *model-2* achieves 1.54 dB average PSNR improvement over *model-1*. Meanwhile, it can be observed in Table 3 that the average PSNR value of *model-5* suffers a decrease of 1.32 dB compared to *model-6*. This demonstrates the effectiveness of Local Aggregation, which can make use of the similarity of local neighbor features and learn the local neighbor complementary information.

2) *Global Aggregation:* We demonstrate the effectiveness of the Global Aggregation by comparing the performance of *model-5* to *model-1* and *model-6* to *model-2*. By adding Row-wise Global Aggregation and Column-wise Global Aggregation, *model-6* achieves 0.2 dB improvements over *model-2*, which proves that the combination of these two parts is effective in aggregating global spatial information. Moreover, we compare the performance of *model-3* to *model-2* and *model-4* to *model-2* to validate the effectiveness of the Row-wise Global Aggregation and Column-wise Global Aggregation, respectively.

4. CONCLUSION

In this paper, we propose a network with Local-Global feature aggregation (LF-LGFA) for LFSR. To make good use of the similarity of the local neighbor features in LF images, we design the Local Aggregation Module to incorporate the local angular information. Moreover, we design the Global Aggregation Module to capture the long-range spatial information from all SAIs via row-wise and column-wise self-attention. By using Local-Global feature Aggregation, our LF-LGFA achieves comparable results against state-of-the-art methods.

5. REFERENCES

[1] K. Mishiba, "Fast depth estimation for light field cameras," *IEEE TIP*, vol. 29, pp. 4232–4242, 2020.

[2] W. Wang, Y. Lin, and S. Zhang, "Enhanced spinning parallelogram operator combining color constraint and histogram integration for robust light field depth estimation," *IEEE SPL*, vol. 28, pp. 1080–1084, 2021.

[3] Y. Wang, T. Wu, J. Yang, L. Wang, W. An, and Y. Guo, "Deoccnet: Learning to see through foreground occlusions in light fields," in *WACV*, 2020, pp. 118–127.

[4] S. Zhang, Z. Shen, and Y. Lin, "Removing foreground occlusions in light field using micro-lens dynamic filter," in *IJCAI*, 2021, pp. 1302–1308.

[5] A. Levin, W. T. Freeman, and F. Durand, "Understanding camera trade-offs through a bayesian analysis of light field projections," in *ECCV*, 2008.

[6] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field image processing: An overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 926–954, 2017.

[7] Y. Yoon, H. Jeon, D. Yoo, J. Lee, and I. Kweon, "Light-field image super-resolution using convolutional neural network," *IEEE SPL*, vol. 24, no. 6, pp. 848–852, 2017.

[8] Y. Wang, L. Wang, J. Yang, W. An, J. Yu, and Y. Guo, "Spatial-angular interaction for light field image super-resolution," in *ECCV*, 2020, pp. 290–308.

[9] Y. Wang, J. Yang, L. Wang, X. Ying, T. Wu, W. An, and Y. Guo, "Light field image super-resolution using deformable convolution," *IEEE TIP*, vol. 30, pp. 1057–1071, 2021.

[10] T. Zhou, J. Li, S. Wang, R. Tao, and J. Shen, "Matnet: Motion-attentive transition network for zero-shot video object segmentation," *IEEE TIP*, vol. 29, pp. 8326–8338, 2020.

[11] T. Zhou, S. Wang, Y. Zhou, Y. Yao, J. Li, and L. Shao, "Motion-attentive transition for zero-shot video object segmentation," in *AAAI*, 2020, vol. 34, pp. 13066–13073.

[12] J. Yan, S. Chen, Y. Zhang, and X. Li, "Neural architecture search for compressed sensing magnetic resonance image reconstruction," *Computerized Medical Imaging and Graphics*, vol. 85, pp. 101784, 2020.

[13] H. Wang, Q. Xie, Q. Zhao, and D. Meng, "A model-driven deep neural network for single image rain removal," in *CVPR*, 2020, pp. 3103–3112.

[14] J. Cao, Y. Li, K. Zhang, and L. Van Gool, "Video super-resolution transformer," *arXiv preprint arXiv:2106.06847*, 2021.

[15] C. Mou, J. Zhang, X. Fan, H. Liu, and R. Wang, "Cola-net: Collaborative attention network for image restoration," *arXiv preprint arXiv:2103.05961*, 2021.

[16] S. Wang, T. Zhou, Y. Lu, and H. Di, "Contextual transformation network for lightweight remote sensing image super-resolution," *IEEE TGRS*, 2021, doi:10.1109/TGRS.2021.3132093.

[17] T. Zhou, L. Li, X. Li, C. Feng, J. Li, and L. Shao, "Group-wise learning for weakly supervised semantic segmentation," *IEEE TIP*, vol. 31, pp. 799–811, 2021.

[18] W. Wang, T. Zhou, F. Yu, J. Dai, E. Konukoglu, and L. Van Gool, "Exploring cross-image pixel contrast for semantic segmentation," in *ICCV*, 2021, pp. 7303–7313.

[19] W. Wang, T. Zhou, S. Qi, J. Shen, and S. Zhu, "Hierarchical human semantic parsing with comprehensive part-relation modeling," *IEEE TPAMI*, 2021, doi:10.1109/TPAMI.2021.3055780.

[20] T. Zhou, S. Qi, W. Wang, J. Shen, and S. Zhu, "Cascaded parsing of human-object interaction recognition," *IEEE TPAMI*, 2021, doi:10.1109/TPAMI.2021.3049156.

[21] S. Zhang, Y. Lin, and H. Sheng, "Residual networks for light field image super-resolution," in *CVPR*, 2019, pp. 11038–11047.

[22] H. Yeung, J. Hou, X. Chen, J. Chen, Z. Chen, and Y. Chung, "Light field spatial super-resolution using deep efficient spatial-angular separable convolution," *IEEE TIP*, vol. 28, no. 5, pp. 2319–2330, 2019.

[23] L. Wang, Y. Wang, Z. Liang, Z. Lin, J. Yang, W. An, and Y. Guo, "Learning parallax attention for stereo image super-resolution," in *CVPR*, 2019, pp. 12250–12259.

[24] W. Shi, J. Caballero, F. Huszár, J. Totz, A. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *CVPR*, 2016, pp. 1874–1883.

[25] M. Rerabek and T. Ebrahimi, "New light field image dataset," in *8th International Conference on Quality of Multimedia Experience (QoMEX)*, 2016, number CONF.

[26] K. Honauer, O. Johannsen, D. Kondermann, and B. Goldluecke, "A dataset and evaluation methodology for depth estimation on 4d light fields," in *ACCV*, 2016, pp. 19–34.

[27] S. Wanner, S. Meister, and B. Goldluecke, "Datasets and benchmarks for densely sampled 4d light fields.," in *VMV*, 2013, vol. 13, pp. 225–226.

[28] M. Le Pendu, X. Jiang, and C. Guillemot, "Light field inpainting propagation via low rank matrix completion," *IEEE TIP*, vol. 27, no. 4, pp. 1981–1993, 2018.

[29] V. Vaish and A. Adams, "The (new) stanford light field archive," *Computer Graphics Laboratory, Stanford University*, vol. 6, no. 7, 2008.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *ICCV*, 2015, pp. 1026–1034.

[31] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE TIP*, vol. 13, no. 4, pp. 600–612, 2004.

[32] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *CVPRW*, 2017, pp. 136–144.

[33] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *ECCV*, 2018, pp. 286–301.