NTIRE 2023 Challenge on Light Field Image Super-Resolution: Dataset, Methods and Results

Yingqian Wang*, Longguang Wang*, Zhengyu Liang*, Jungang Yang*†, Radu Timofte*, Yulan Guo*, Kai Jin, Zeqiang Wei, Angulia Yang, Sha Guo, Mingzhi Gao, Xiuzhuang Zhou, Vinh Van Duong, Thuc Nguyen Huu, Jonghoon Yim, Byeungwoo Jeon, Yutong Liu, Zhen Cheng, Zeyu Xiao, Ruikang Xu, Zhiwei Xiong, Gaosheng Liu, Manchang Jin, Huanjing Yue, Jingyu Yang, Chen Gao, Shuo Zhang, Song Chang, Youfang Lin, Wentao Chao, Xuechun Wang, Guanghui Wang, Fuqing Duan, Wang Xia, Yan Wang, Peiqi Xia, Shunzhou Wang, Yao Lu, Ruixuan Cong, Hao Sheng, Da Yang, Rongshan Chen, Sizhe Wang, Zhenglong Cui, Yilei Chen, Yongjie Lu, Dongjun Cai, Ping An, Ahmed Salem, Hatem Ibrahem, Bilel Yagoub, Hyun-Soo Kang, Zekai Zeng, Heng Wu

Abstract

In this report, we summarize the first NTIRE challenge on light field (LF) image super-resolution (SR), which aims at super-resolving LF images under the standard bicubic degradation with a magnification factor of 4. This challenge develops a new LF dataset called NTIRE-2023 for validation and test, and provides a toolbox called BasicLFSR to facilitate model development. Compared with single image SR, the major challenge of LF image SR lies in how to exploit complementary angular information from plenty of views with varying disparities. In total, 148 participants have registered the challenge, and 11 teams have successfully submitted results with PSNR scores higher than the baseline method LF-InterNet [1]. These newly developed methods have set new state-of-the-art in LF image SR, e.g., the winning method achieves around 1 dB PSNR improvement over the existing state-of-the-art method DistgSSR [2]. We report the solutions proposed by the participants, and summarize their common trends and useful tricks. We hope this challenge can stimulate future research and inspire new ideas in LF image SR.

*Yingqian Wang, Longguang Wang, Zhengyu Liang, Jungang Yang, Radu Timofte and Yulan Guo are the NTIRE 2023 challenge organizers, while the other authors participated in this challenge.
†Corresponding author: Jungang Yang
Section 6 provides the authors and affiliations of each team.
NTIRE 2023 webpage: https://cvlai.net/ntire/2023/
Challenge webpage: https://codalab.lisn.upsaclay.fr/competitions/9201
Leaderboard: https://codalab.lisn.upsaclay.fr/competitions/9201#results
Github: https://github.com/The-Learning-And-Vision-Atelier-LAVA/LF-Image-SR/tree/NTIRE2023
BasicLFSR toolbox: https://github.com/ZhengyuLiang24/BasicLFSR

1. Introduction

Light field (LF) cameras can capture both intensity and directions of light rays, and record 3D geometry in a convenient and efficient manner. By encoding 3D scene cues into 4D LF images (i.e., 2D for spatial dimension and 2D for angular dimension), LF cameras enable many attractive applications such as post-capture refocusing [3, 4], depth sensing [5–12], virtual reality [13, 14] and view rendering [15–18].

In many applications, high-resolution (HR) LF images are highly demanded to achieve higher perceptual quality and benefit downstream applications. However, HR LF images are generally obtained at an expensive cost due to the spatial-angular trade-off issue in LF imaging [19]. Consequently, it is highly necessary to reconstruct HR LF images from their low-resolution (LR) counterparts, i.e., to achieve LF image super-resolution (SR).

In recent years, remarkable progress has been achieved in image SR with deep learning techniques. However, most approaches focus on super-resolving single images [20–25], stereo images [26–30] or videos [31–34], and cannot be directly extended to the task of LF image SR. For LF images, how to effectively incorporate both spatial and angular information is important but challenging.

To develop and benchmark LF image SR methods, we host the first LF image SR challenge on the NTIRE 2023 workshop. This challenge employs the widely used and publicly available LF datasets [35–39] as training set, and proposes a new LF dataset called NTIRE-2023 for both validation (model development) and test (final ranking). The popular bicubic degradation is used to generate LR LF images, and the objective of this challenge is to make the super-resolved LF images as faithful as the groundtruth HR

ones. Besides, this challenge provides an open-source and easy-to-use toolbox named BasicLFSR to facilitate participants to quickly get access to LF image SR and develop their own models. In summary, this challenge aims at establishing a new benchmark for LF image SR, and aspires to highlight specific challenges and research problems. We hope that this challenge can inspire the community to explore the cross area of low-level vision and 3D vision, and stimulate future research in LF image processing.

This challenge is one of the NTIRE 2023 Workshop series of challenges on: night photography rendering [40], HR depth from images of specular and transparent surfaces [41], image denoising [42], video colorization [43], shadow removal [44], quality assessment of video enhancement [45], stereo super-resolution [46], light field image super-resolution [47], image super-resolution ($\times 4$) [48], 360° omnidirectional image and video super-resolution [49], lens-to-lens bokeh effect transformation [50], real-time 4K super-resolution [51], HR nonhomogenous dehazing [52], efficient super-resolution [53].

2. Related Work

In this section, we briefly review several major works in LF image SR. We divide existing LF image SR methods into traditional non-learning methods, CNN-based methods and Transformer-based methods. Note that, we only focus on the plain-lens based methods, and do not discuss those hybird-lens based LF image SR methods [54–58].

2.1. Traditional Methods

Light field image SR is a long-standing problem and has been investigated for decades. Bishop et al. [59] proposed a Bayesian deconvolution approach to super-resolve LF images based on the estimated disparities. Wanner et al. [60] first estimated disparity maps using structure tensor, and then developed a variational framework for LF image SR. Farrugia et al. [61] constructed a patch-volume dictionary of HR-LR LF image pairs, and proposed a multivariate ridge regression method to learn the linear mapping from LR patch volumes to their HR counterparts. In [62], Alain et al. considered the ill-posed LF image SR problem as an optimization problem based on the sparsity prior. Rossi et al. [63] combined the inter-view information using graph regularization, and formulated LF image SR as a quadratic problem which can be solved efficiently with standard convex optimization.

2.2. CNN-based Methods

In the past decade, convolutional neural networks (CNNs) have been extensively studied and achieved remarkable performance in LF image SR. Yoon et al. [64] proposed the first CNN-based LF image SR method (i.e., LFCNN). In their method, input LF images were grouped into pairs or quads, and fed to a three layer CNN to integrate complementary information from adjacent views. As the pioneering work, LFCNN [64] shows great potential of CNNs in LF image SR. Afterwards, many deeper CNNs with various angular information incorporation mechanisms were developed to achieve improved SR performance.

Wang et al. [65] proposed a bidirectional recurrent CNN (i.e., LFNet) to incorporate angular information from the sub-aperture images (SAIs) along the horizontal or vertical angular direction. Zhang et al. [66] stacked SAIs along four different angular directions, and developed a four-branch residual network to implicitly learn the epipolar geometry from stacked SAIs for LF image SR. In their subsequent work, Zhang et al. [67] improved the SR performance by performing 3D convolutions on SAI stacks of different angular directions. Cheng et al. [68] developed a framework to exploit both internal and external similarities for LF image SR. Meng et al. [69] applied 4D convolutions to simultaneously incorporate spatial and angular information from 4D LF data, and developed a high-dimensional dense residual network (HDDRNet) for LF image SR. Jin et al. [70] proposed an all-to-one method for LF image SR, and performed structural consistency regularization to preserve the parallax structure. Wang et al. [71] applied deformable convolution to LF spatial SR, and designed a collect-and-distribute scheme to incorporate the complementary information among different views. Mo et al. [72] proposed a dense dual-attention network (DDAN) for LF image SR, in which a view attention module and a channel attention module were designed to adaptively capture discriminative information from different views and channels, respectively.

Instead of directly processing 4D LF data or image stacks, some methods disentangled 4D LFs into different subspace for LF image SR. Yeung et al. [73] alternately reshaped LF images between SAI pattern and macro-pixel pattern, and designed spatial-angular separable convolutions for LF image SR. In [1], Wang et al. proposed spatial and angular feature extractors to extract corresponding information from macro-pixel images (MacPIs), and developed an LF-InterNet to repetitively interact the spatial and angular information for LF image SR. In their subsequent work, Wang et al. [2] further generalized the interaction mechanism into LF disentangling mechanism, and developed three CNNs (i.e., DistgSSR, DistgASR and DistgDisp) for spatial SR, angular SR and disparity estimation, respectively. Following [1], Liu et al. [74] proposed an intra-inter view interaction network (LF-IINet) with two parallel branches to extract global inter-view information and model the correlations among all intra-view features, respectively. These two branches are mutually interacted to fuse angular and spatial information for LF image SR.

https://cvlai.net/ntire/2023/

Besides the aforementioned works that design advanced network structures to pursuit superior SR accuracy, several works also studied some special yet important issue in LF image SR. Cheng et al. [75] addressed the domain gap issue by proposing a "zero-shot" learning framework, in which the network learns to achieve spatial SR without using external training data except the given input LR LF. Wang et al. [76] addressed the degradation formulation issue in LF image SR, and proposed a method to handle LF image SR with multiple degradation. Xiao et al. [77] proposed a data augmentation approach tailored for LF image SR, which can be applied to existing LF image SR networks to further improve their SR performance

2.3. Transformer-based Methods

Transformer networks, which were originally developed for natural language processing [78], have recently gained much attention in computer vision community. Recently, Transformers have been successfully applied to many low-level vision tasks such as image restoration [25, 79, 80] and video SR [81–83], and achieved superior performance than CNN-based methods.

In the past two years, researchers have explored Transformers for LF image SR. Wang et al. [84] proposed a detail-preserving Transformer (DPT) for LF image SR, in which SAIs of each vertical and horizontal views are considered as a sequence, and the long-range geometric dependency is learned via a spatial-angular locally enhanced self-attention layer. Liang et al. [85] proposed a simple yet effective Transformer network (i.e., LFT) for LF image SR. In their method, an angular Transformer is designed to incorporate complementary information among different views, and a spatial Transformer is developed to capture both local and long-range dependencies within each SAI. Guo et al. [86] develop a raw LF data generation pipeline to utilize the rich information from the raw LF data to enhance their spatial resolution. They introduced a volume Transformer to aggregate information of all views into center view, and designed a cross-view Transformer to align the center view feature to all views for non-local information utilization. Wang et al. [87] proposed a Multi-granularity Aggregation Transformer (MAT) for LF image SR, in which the LF feature representation was learned via three designed granularity aggregation units. More recently, Liang et al. [88] investigated the non-local spatial-angular correlations in LF image SR, and developed a Transformer-based network called EPIT to achieve state-of-the-art SR performance. The proposed EPIT achieves a global receptive field along the epipolar line, and is robust to disparity variations.

3. NTIRE 2023 Challenge

In this section, we introduce the NTIRE 2023 LF image SR Challenge. We first introduce the datasets used in this challenge, and then briefly describe the BasicLFSR toolbox. Afterwards, we review the two phases of this challenge, and finally summarize the common trends in the submitted solutions.

3.1. Dataset

Training Set. This challenge follows the existing LF image SR works [2, 71, 74, 84, 85, 88], and uses the EPFL [35], HCInew [36], HCIold [37], INRIA [38] and STFgantry [39] datasets for training. All the 144 LFs in the training set have an angular resolution of 9×9. The participants are required to use these LF images as HR groundtruth to train their models. External training data or models pretrained on other datasets are not allowed in this challenge.

Validation Set. In this challenge, we develop a new LF dataset (namely, NTIRE-2023) for both validation and test, as shown in Fig. 1. The validation set contains 16 synthetic scenes rendered by the 3DS MAX software[1] and 16 real-world images captured by Lytro Illum cameras. For synthetic scenes, all virtual cameras in the camera array have identical internal parameters and are co-planar with the parallel optical axes. All scenes in the validation set have an angular resolution of 5×5. The spatial resolutions of synthetic LFs and real-world LFs are 500×500 and 624×432, respectively. All the LF images in the validation set are bicubicly downsampled by a factor of 4, and only the LR versions are released to the participants. Challenge participants are required to apply their developed models to the LR LF images, and submit the super-resolved LF images to the CodaLab server for validation.

Test Set. To rank the submitted models, a new test set consisting of 16 synthetic LFs (rendered in the same way as in the validation set) and 16 real-world LFs (captured by Lytro Illum cameras) are provided, as shown in Fig. 1. Same as the validation set, only $4\times$ downsampled LR LF images with an angular resolution of 5×5 are released to the participants.

3.2. The BasicLFSR Toolbox

This challenge provides a PyTorch-based, open-source, and easy-to-use toolbox named BasicLFSR to facilitate participants to quickly get access to LF image SR and develop their own models. The BasicLFSR toolbox has the following three characteristics: (1) It provides a complete pipeline to develop novel LF image SR methods. (2) It integrates a number of LF image SR methods, and retrains them on unified LF datasets. The codes and checkpoints of each model are publicly available. (3) It provides a fair and comprehensive benchmark for LF image SR. The quantitative results of each method are listed, and their super-resolved LF images are available for download.

[1] https://www.autodesk.eu/products/3ds-max/overview

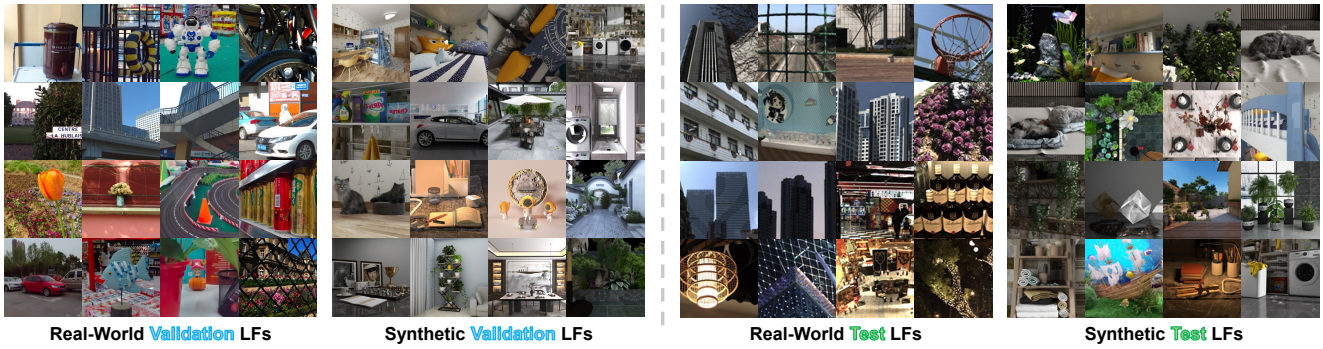

Figure 1. An illustration of the center-view images in the developed NTIRE-2023 LF dataset. Both validation and test sets contain 16 real-world and 16 synthetic LFs, respectively.

3.3. Challenge Phases

Development Phase. The participants can download the LR validation set and apply their developed models to the LR LF images to generate their SR versions. A validation leaderboard is available online, and the participants can compare their scores with the ones achieved by the baseline models (provided by the challenge organizers) or models developed by other participants.

Test phase. The participants are required to apply their models to the released LR test set, and submit their super-resolved LF images to the test server. The test server is available online during this phase, and will be closed after the test deadline. The participants are asked to submit the SR results, codes and a fact sheet of their methods before the given deadlines. After this challenge, the final rank is released to the participants, and the test server will be re-open to facilitate the development of novel LF image SR methods in the future.

Evaluation Metrics. Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are used as metrics for performance evaluation. The implementation details of PSNR and SSIM can be found in the BasicLFSR toolbox. The submitted results are ranked by the average PSNR values on the test set (both real-world and synthetic scenes).

3.4. Challenge Results

Among the 148 registered participants, 12 teams have successfully participated the final test phase and submitted their results, codes, and factsheets. The top 11 of them produced PSNR scores higher than the baseline method LF-InterNet [1]. Table 1 reports the PSNR and SSIM scores achieved by these methods on both test and validation sets, together with their major details. We briefly describe these solutions in Section 4, and introduce the corresponding team members in Appendix 6.

It can be observed from Table 1 that all these methods surpass the state-of-the-art method DistgSSR [2], and 9 of them surpass the recent top-performing method EPIT [88]. Note that, the winner solution proposed by the OpenMeow achieves around 1 dB improvement in PSNR over DistgSSR [2] on both test and validation sets, which significantly push the state-of-the-art of LF image SR to a new height. Moreover, the accuracy of the top 2 methods are very close with a minor PSNR difference of 0.02 dB on the test set. In addition, although the second runner-up solution proposed by the VIDAR team produces slightly inferior PSNR results than the winner solution and the runner-up solution, it achieves the highest SSIM score of 0.9323 on the test set.

Architectures and main ideas. All the proposed methods are based on deep learning techniques. Transformers are used as the basic architecture in 6 solutions, while other models are purely based on CNNs. The idea of LF disentangling [2] was adopted in most solutions, and the recently developed method EPIT [88] was used as the backbone by the OpenMeow team (winner) and the BNU-AI-TRY team.

Subspace division. Since an LF has a complex structure and its spatial and angular information is highly coupled with varying disparities, it is challenging for deep neural networks to exploit informative cues from such a high-dimensional tensor. Consequently, 7 teams adopted the disentangling mechanism in [2] to divide the 4D LFs into four 2D subspaces including spatial subspace (i.e., SAIs), angular subspace (i.e., macro-pixels), horizontal EPI subspace, and vertical EPI subspace. Three teams performed feature extraction and incorporation in spatial and EPI subspaces, while one team learned LF image SR in spatial and angular subspaces.

Data Augmentation. The participants commonly performed random flipping and rotation for training data augmentation. In addition, two teams randomly sampled 5×5 LFs from 9×9 LFs to further augment the training set. However, some advanced data augmentation approaches such as CutBlur [89] and RGB channel shuffling have not been adopted in this challenge.

Table 1. NTIRE 2023 LF Image SR Challenge results, final rankings, and the main characteristics of the solutions. Note that, the average PSNR value achieved on the test set is used for final ranking. The best results are in **red**, the second best results are in **blue**, and the third best results are in **green**.

Rank	Team	Test Set			Validation Set			#Params	Architec*	Subspace	Ensemble
		Average	Lytro	Synthetic	Average	Lytro	Synthetic				
1	OpenMeow ★	**30.66**/**.9314**	**30.82**/**.9475**	**30.51**/**.9152**	**32.71**/**.9496**	**33.36**/**.9562**	**32.07**/**.9430**	20.34M	Hybrid	Spa & Ang & EPI	Data & Model
2	DMLab ★	**30.64**/**.9318**	**30.92**/**.9489**	**30.35**/**.9146**	**32.43**/**.9485**	**33.24**/**.9559**	31.62/**.9410**	28.99M	CNN	Spa & Ang & EPI	Data
3	VIDAR ★	**30.56**/**.9323**	**30.67**/**.9491**	**30.45**/**.9154**	**32.54**/**.9494**	**33.24**/**.9568**	**31.85**/**.9419**	10.52M	Transf	Spa & Ang & EPI	Data & Model
4	IIR-Lab	30.38/.9285	30.56/.9450	30.20/.9119	32.24/.9465	32.84/.9529	**31.64**/.9402	2.63M	Transf	Spa & Ang & EPI	-
5	INSIS	30.35/.9287	30.56/.9458	30.15/.9117	32.12/.9455	32.86/.9526	31.39/.9383	5.43M	CNN	Spa & Ang & EPI	Data
6	BNU-AI-TRY	30.13/.9290	29.97/.9453	30.29/.9126	32.29/.9468	32.96/.9539	31.63/.9396	8.83M	Transf	Spa & EPI	Data & Model
7	BIT912	30.11/.9293	30.10/.9465	30.13/.9120	32.05/.9449	32.76/.9528	31.35/.9371	4.08M	Transf	Spa & Ang & EPI	-
8	HawkeyeGroup	30.06/.9285	29.99/.9447	30.13/.9124	32.13/.9463	32.86/.9543	31.40/.9383	3.35M	Transf	Spa & Ang	-
9	SHU-IVIPLab	29.90/.9265	29.78/.9433	30.01/.9096	32.01/.9442	32.69/.9517	31.32/.9366	7.79M	CNN	Spa & Ang & EPI	Data
10	CBNU-MIP-Lab	29.85/.9279	29.64/.9447	30.06/.9111	32.13/.9464	32.70/.9533	31.55/.9395	14.82M	CNN	Spa & EPI	-
11	LFSR-gdut-team	29.83/.9262	29.64/.9422	30.01/.9103	31.83/.9431	32.53/.9508	31.13/.9354	7.28M	CNN	Spa & EPI	-
-	EPIT [88]	29.87/.9259	29.72/.9420	30.03/.9097	32.04/.9447	32.54/.9507	31.53/.9387	1.47M	Transf	Spa & EPI	✗
-	LFT [85]	29.77/.9252	29.66/.9420	29.88/.9084	31.75/.9423	32.42/.9501	31.08/.9344	1.16M	Transf	Spa & Ang	✗
-	DistgSSR [2]	29.64/.9244	29.39/.9403	29.88/.9084	31.75/.9424	32.26/.9490	31.23/.9357	3.58M	CNN	Spa & Ang & EPI	✗
-	LF-InterNet [1]	29.45/.9198	29.23/.9369	29.45/.9028	31.33/.9381	32.06/.9468	30.61/.9295	5.48M	CNN	Spa & Ang	✗
-	Bicubic	25.79/.8378	25.11/.8404	26.46/.8352	27.51/.8714	27.49/.8719	27.53/.8710	-	✗	Spa	✗

Note: "Transf" denotes that the model adopts Transformer as a basic component, "CNN" denotes that the model was developed based on convolutions only. "Hybird" denotes that the model contains sub-models which are developed based on CNNs and Transformers, respectively.

Ensemble Strategy. Both data ensemble (a.k.a. test-time augmentation) and model ensemble were adopted in several solutions to boost the SR performance. For data ensemble [90], the inputs were flipped and rotated, and the resultant SR images were aligned and averaged for enhanced prediction. Note that, the INSIS team proposed a shear ensemble approach tailored with LF image SR for performance enhancement. The OpenMeow, VIDAR, and BNU-AI-TRY teams adopted model ensemble, and averaged the results produced by multiple models for better results.

Conclusions. By analyzing the settings, the proposed methods and their results, we can conclude that:

- The proposed solutions significantly improve the state-of-the-art in LF image SR.
- Transformers are increasingly popular in LF image SR, but the well-designed CNNs (e.g., the solution proposed by the DMLab team) can also achieve competitive SR performance.
- Most methods exploring multi-dimensional information from spatial, angular and EPI subspaces. Spatial and EPI subspaces are quite important for achieving competitive SR performance.
- There seems to be a considerable room of further performance improvement, because ensemble strategy and some advanced data augmentation approaches have not been widely used.

4. Challenge Teams and Methods

4.1. OpenMeow: DistgEPIT ★

The OpenMeow team proposed a hybrid network called DistgEPIT for LF image SR. Readers can refer to [91] for more details of their method. The proposed DistgEPIT contains a DistgSSR-based branch [2] and an EPIT-based branch [88], which can learn the spatial-angular relationship from the MacPI representation while handling the large disparity issue by adopting the EPI representation. As shown in Fig. 2, the DistgEPIT network adopts the non-local cascading block (i.e., Basic-Transformer unit in EPIT [88]) to exploit information from sub-aperture images (SAIs) along the horizontal and vertical angular directions. The long-range modeling ability of the non-local cascading block benefits the learning of pixel-wise correlations from remote views. After extracting deep features via several non-local cascading blocks, the OpenMeow team uses several DistgBlocks [2] for refinement. The final SR results are generated by fusing the bicubicly upsampled image, the output of the EPIT branch, and the output of the DistgSSR branch.

Moreover, this team proposed a position-sensitive post-processing method to eliminate the margin of LF patches introduced by the commonly used zero-padding in the *LF divide-and-integrate operation*[2]. Specifically, they adopted a sliding window approach to crop the chop in an overlapping manner without introducing any padding operations. In cases where the last row or last column is cropped, the window backtracks to make up the entire chop.

[2]Please refer to the BasicLFSR toolbox for the implementation details.

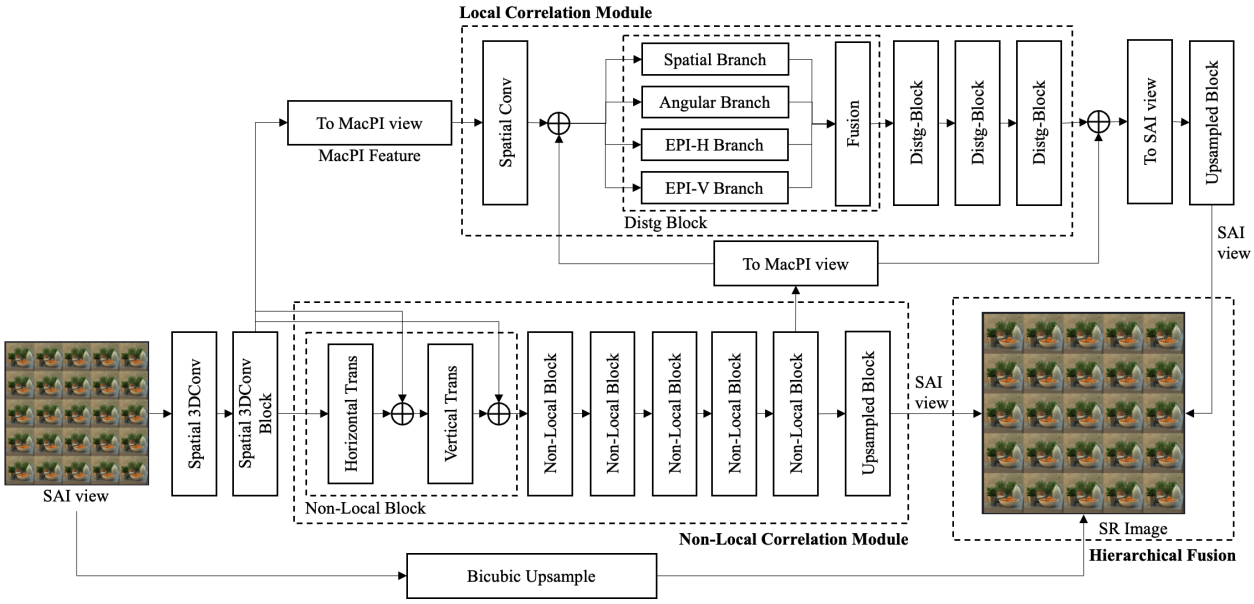

Figure 2. The OpenMeow Team: The network architecture of the proposed DistgEPIT.

Ensemble Strategy: The OpenMeow team performed model ensemble by using three different configurations of DistgEPIT and two different configurations of DistgSSR. Specifically, in the first DistgEPIT model (i.e., DistgEPIT_wider), each local correlation module has 128 channels and includes 4 Distg-Groups (each Distg-Group has 4 Distg-Blocks). The second configuration of DistgEPIT, called DistgEPIT_deeper, has 64 channels but increases the number of non-local correlation blocks from 5 to 8. Moreover, the number of Distg-Groups in the local correlation module is increased from 4 to 8. The third configuration of DistgEPIT, called DistgEPIT_Parallel, extracts features from both local and non-local correlation modules in parallel, and fuses MacPIs at the top level using two cascaded Distg-Groups (each Distg-Group has two Distg-Blocks). The two configurations of DistgSSR have 64 and 128 channels, respectively, and the convolution kernels in the original upsampling layer of DistgSSR are modified from 1×1 to 3×3. In total, 12 groups of model parameters were obtained from different training phases. For data augmentation, horizontal flip, vertical flip, and 90-degree rotation were used, and the final results were obtained by aligning and averaging the results of all models and data.

4.2. DMLab: RR-HLFSR

This team presented a residual in residual learning based hybrid LF image SR network (namely, RR-HLFSR), which is an enhanced version of their recently published method HLFSR [92]. The main improvement of RR-HLFSR as compared to HLFSR is that the local residual learning and global residual learning are introduced to the basic hybrid feature extraction, as shown in Fig. 3. Thanks to the residual learning mechanism, the RR-HLFSR network can be developed deeper than HLFSR, and achieves considerable improvements in SR performance.

The proposed RR-HLFSR network contains three types of 2D feature extractors that work in different sub-spaces of 4D LFs: Inter-Intra Spatial Feature Extractor (II-SFE), Inter-Intra Angular Feature Extractor (II-AFE), and Multi-Orientation Epipolar Feature Extractor (MO-EFE). Specifically, the II-SFE and II-AFE are designed to explore the correlation among pixels within each SAI and each macro-pixel, respectively. The MO-EFE is designed to handle multiple stacks of SAIs with different epipolar geometry orientations to extract abundant sub-pixel information.

Moreover, since diverse information can be extracted from multiple sub-spaces, how to effectively fuse various features from different feature extractors is crucial in further improving the quality of recovered LF images. This method designed an attention fusion module (AFM) that handles fused information from different branches. By using the simple but effective modules, the SR performance is enhanced.

4.3. VIDAR: SAVformer

This method is mainly inspired by their published work LFSSR-SAV [93] and the mile-stone single image restoration method Swin-Transformer [79]. In LFSSR-SAV [93], the authors proposed a novel spatial-angular correlated convolution (SAC-conv) and adopted the spatial-angular separable convolution (SAS-conv) [73] for efficient LF feature extraction, and verified that both SAS-conv and SAC-

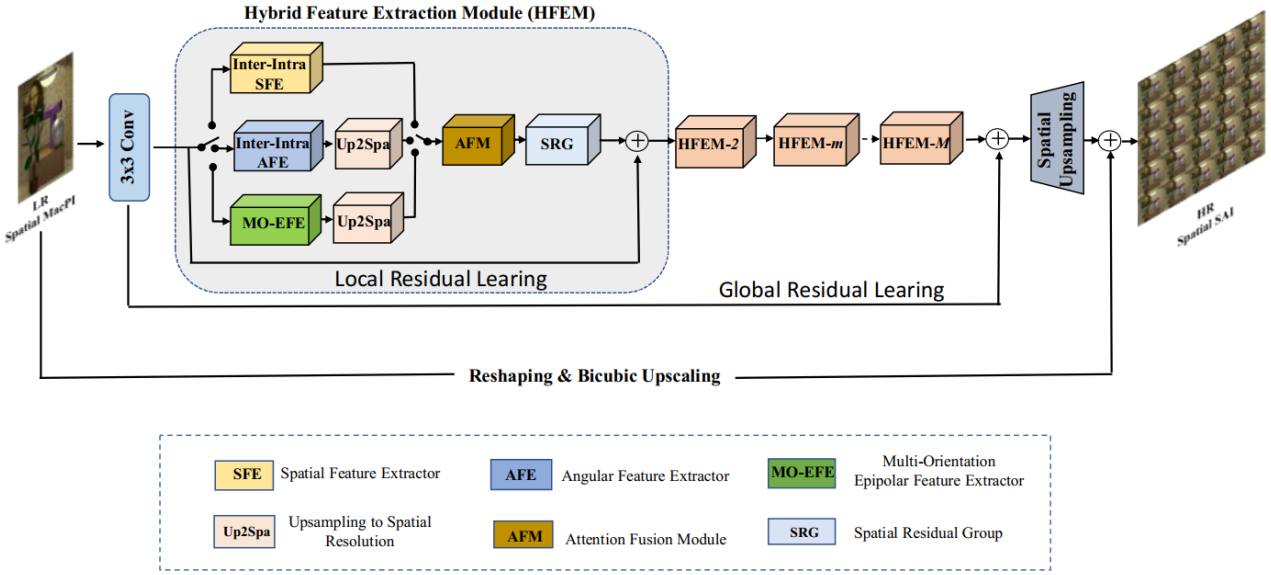

Figure 3. The DMLab Team: The network architecture of the proposed RR-HLFSR.

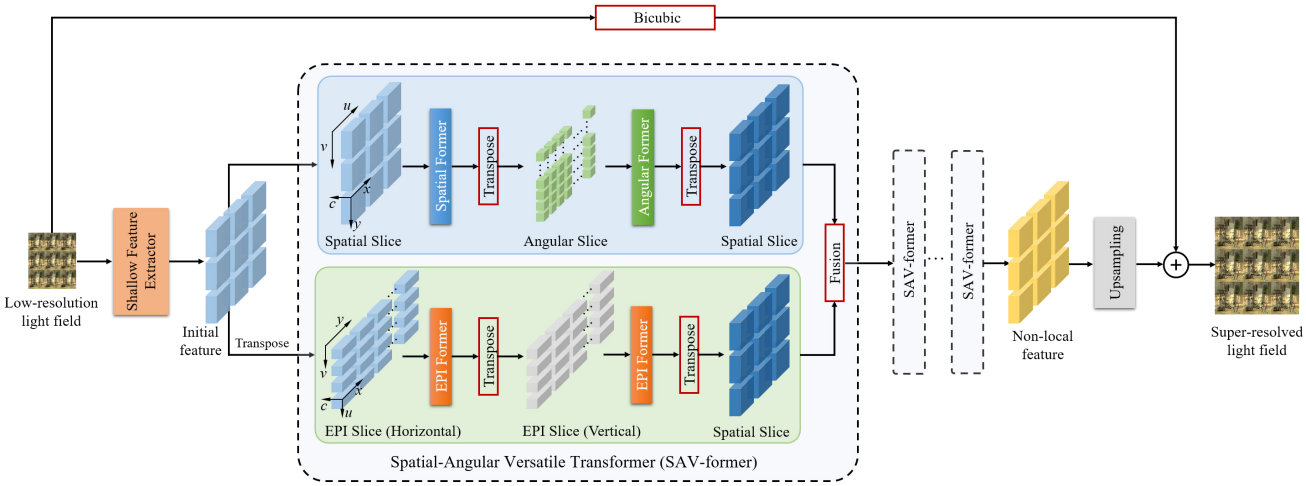

Figure 4. The VIDAR Team: The network architecture of the proposed SAVformer.

conv are complementary at different aspects of 4D LF feature embedding. However, LFSSR-SAV is a CNN-based method, and the limited receptive field of convolutions hinders the utilization of the non-local self-similarity information, especially the inter-view correspondence. Therefore, this team introduced the Swin-Transformer to LFSSR-SAV, and designed the spatial-angular versatile Transformer network (namely, SAVformer) for LF image SR. Figure 4 shows the architecture of their SAVformer, which contains Spatial-Former, Angular-Former and EPI-Former.

Loss Function: To better preserve the geometric consistency, this team followed LF-ATO [70] to use the EPI gradient loss $\mathcal{L}_e$ and the $\mathcal{L}_1$ loss for network training, i.e.,

$$\mathcal{L}_{total} = \mathcal{L}_1 + \alpha \mathcal{L}_e, \tag{1}$$

where α denotes the weighting factor which is set to 0.1.

Training Strategies: They trained their network in four stages: 1) They first trained SAVformer with a batch size of 4, a patch size of 32×32, and loss $\mathcal{L}_{total}$ for 16000 epochs (144 iterations per epoch). The learning rate was initially set to 2×10^{-4} and decreased by a factor of 0.5 for every 5000 epochs. 2) They finetuned SAVformer with a batch size of 4, a patch size of 48×48, and loss $\mathcal{L}_{total}$ for 5000 epochs (144 iterations per epoch). The learning rate

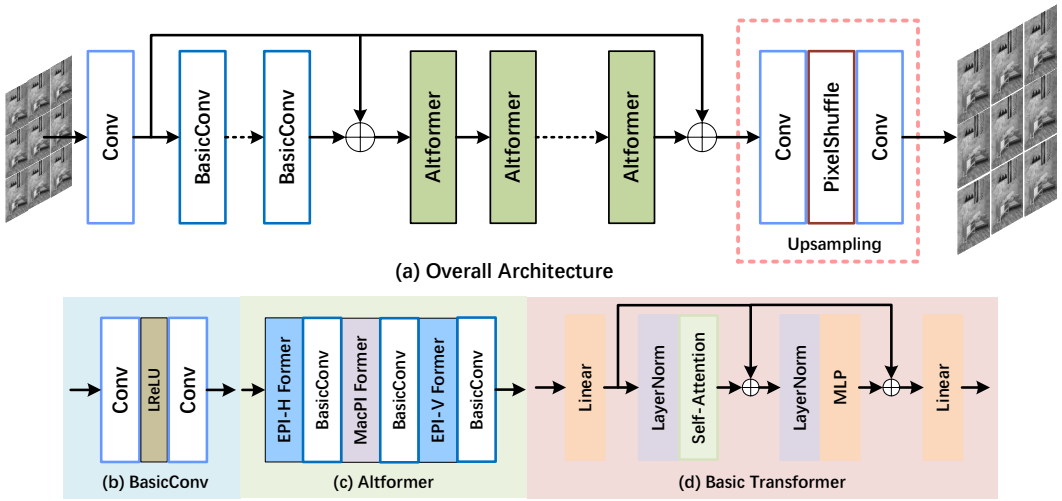

Figure 5. The IIR-Lab Team: The network architecture of the proposed SA-Altfromer.

was initially set to 1×10^{-4} and decreased by a factor of 0.5 for every 1000 epochs. 3) They finetuned SAVformer with a batch size of 8, a patch size of 32×32, and loss $\mathcal{L}_{total}$ for 10 epochs (9039 iterations per epoch). The learning rate was initially set to 2×10^{-5} and decreased by a factor of 0.5 for every 15 epochs. 4) They finetuned SAVformer with a batch size of 4, a patch size of 32×32, and loss $\mathcal{L}_1$ for 50 epochs (9039 iterations per epoch). The learning rate was initially set to 2×10^{-5} and decreased by a factor of 0.5 for every 15 epochs.

4.4. IIR-Lab: SA-Altfromer

Considering the EPIs and MacPIs are the two typical LF representations that reflect the angular correlations, this team applies Transformers on these two representations to exploit the spatial and angular correlations. An overview of the proposed SA-Altfromer is shown in Fig. 5(a).

The proposed method first cascades several BasicConv blocks (as shown in Fig. 5(b)) to gradually extract the intra-view features (i.e., spatial correlations). Then, this method adopts 6 Altformer modules (see Fig. 5(c)) to alternately perform multi-head self-attention (MHSA) operations on EPI and MacPI subspace. In each Altformer, the horizontal EPI features, vertical EPI features and MacPI features are sequentially fed into the EPI-H, EPI-V, and MacPI Formers. As shown in Fig. 5(d), the EPI-H, EPI-V, and MacPI Formers are developed on the Basic Transformer modules. After the Altformers, the enhanced feature by local connection is fed into an upsampling block to generate the final super-resolved results.

4.5. INSIS: SAMSSR

As shown in Fig. 6, this team proposed a spatial-angular multi-scale spatial SR network (namely, SAMSSR) to cover the long-range disparity range and explicitly exploit the sub-pixel correspondence in LF images. Readers can refer to [94] for more details of their method. This team first designed a Multi-Dimension Interaction Block (MDIB) consisting of four branches to separately extract the spatial information, angular information, and horizontal and vertical spatial-angular coupling information. To decouple the spatial-angular information along the epipolar line, they designed a Spatial-Angular Multi-Scale Process Module (MSPB) based on horizontal or vertical EPI structures, and adopted dilated convolutions to fully incorporate the long-range disparity information. In addition, to better integrate the multi-dimension and multi-scale characteristics, this team adopted the channel attention mechanism at the end of both MDIB and MSPB to fuse information from different branches.

Refinement with Shear Operation. To ensure that the proposed SAMSSR performs well under large disparities, this team additionally introduced the LF Shear Attention network [95] as a second-stage model to improve the accuracy of the final result. Specifically, they first applied the pre-trained SAMSSR model to the sheared LF images with different disparity values {-1, -0.5, 0, 0.5, 1}, and obtained a set of SR results which were then sheared back with the $4\times$ disparity values {4, 2, 0, -2, -4} to restore the original disparity. Afterward, they trained the LF Shear Attention network [95] to distinguish the relevant information from different sheared levels, and fused them to generate the final SR result.

4.6. BNU-AI-TRY: EPITv2_max

This method is mainly inspired by the recent EPIT method [88], and aims to improve the capability of the spatial-angular correlation modeling. Specifically, this team

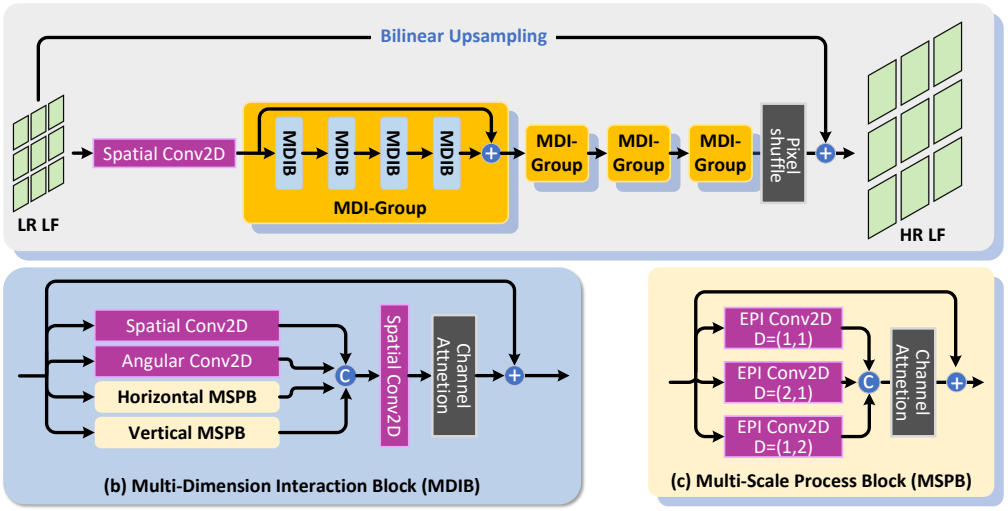

Figure 6. The INSIS Team: The network architecture of the proposed SAMSSR.

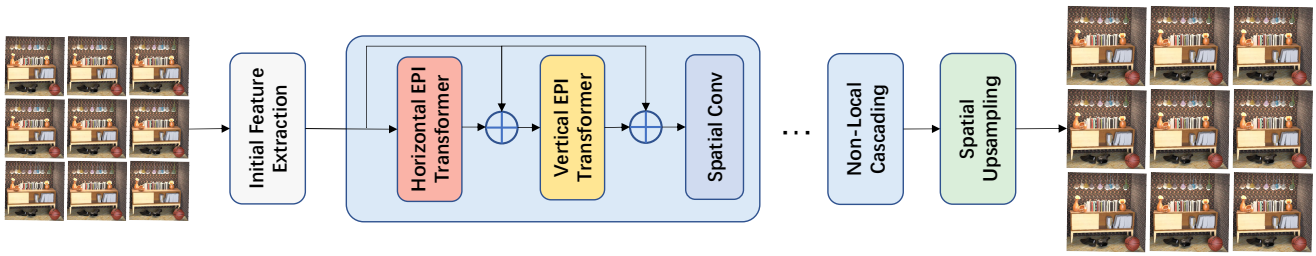

Figure 7. The BNU-AI-TRY Team: The network architecture of the proposed EPITv2_max.

increased the channels of feature maps in EPIT (64→128) and re-designed the non-local cascading block in EPIT by sequentially cascading the horizontal Basic-Transformer unit, the vertical Basic-Transformer unit, and the spatial convolution. An overview of their EPITv2_max is shown in Fig. 7.

Data Augmentation. During training, this team cropped each SAI into patches of size 128×128 with a smaller stride than EPIT (32 v.s. 64) to generate more LF training patches to alleviate the over-fitting issue.

4.7. BIT912: CSWinLFSR

This team observed that the Transformer-based method LFT [85] requires a large number of computational resources during LF feature extraction, and thus aimed to reduce the computation cost of LFT and increase the network layers for stronger modeling capability. Inspired by the novel CSwin Transformer [96], this team replaced the global self-attention operation in LFT with the criss-cross shifted window self-attention in CSwin Transformer, and proposed CSwinSpa, CSwinAng, and CSwinEPI modules to extract spatial, angular, and EPI information, respectively. Figure 8 shows the overview of the CSwinLFSR network, which consists of three stages: shallow feature extraction, spatial-angular feature learning module, and LF reconstruction.

4.8. HawkeyeGroup: LF-DET

This team proposed a deep efficient Transformers (i.e., LF-DET) for LF image SR.

4.9. SHU-IVIPLab: SA-VSNet

Inspired by LFSSR-SAV [93], this team designed a spatial-angular separable convolution (SAS-conv) module and a spatial-angular correlated convolution (SAC-conv) module for LF image processing. This team further introduced 3D convolutions on neighboring view sequences to explore the complementary benefits from the joint spatial context and specific directional views for LF image SR. An overview of proposed Spatial-Angular View-Sequence Network (SA-VSNet) is illustrated in Fig. 9.

Specifically, this method follows a "coarse-to-fine" strategy to obtain the SR results progressively. In the coarse stage, this method first extracts the spatial features from

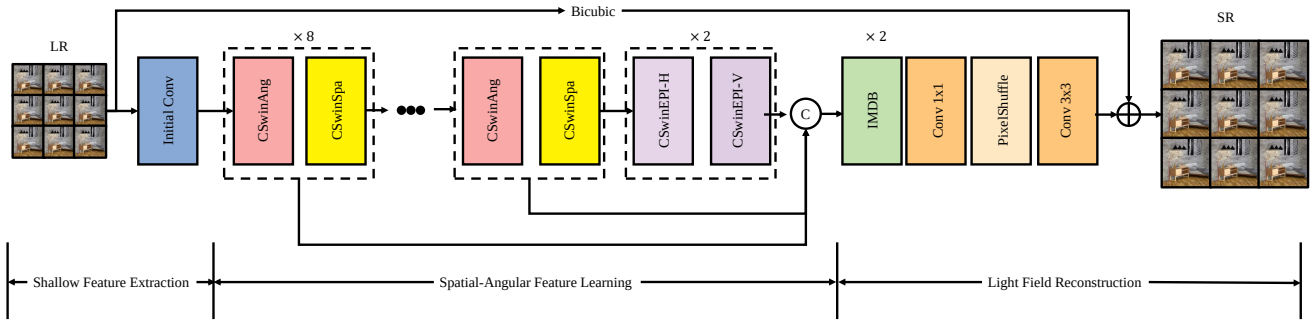

Figure 8. The BIT912 Team: The network architecture of the proposed CSWinLFSR.

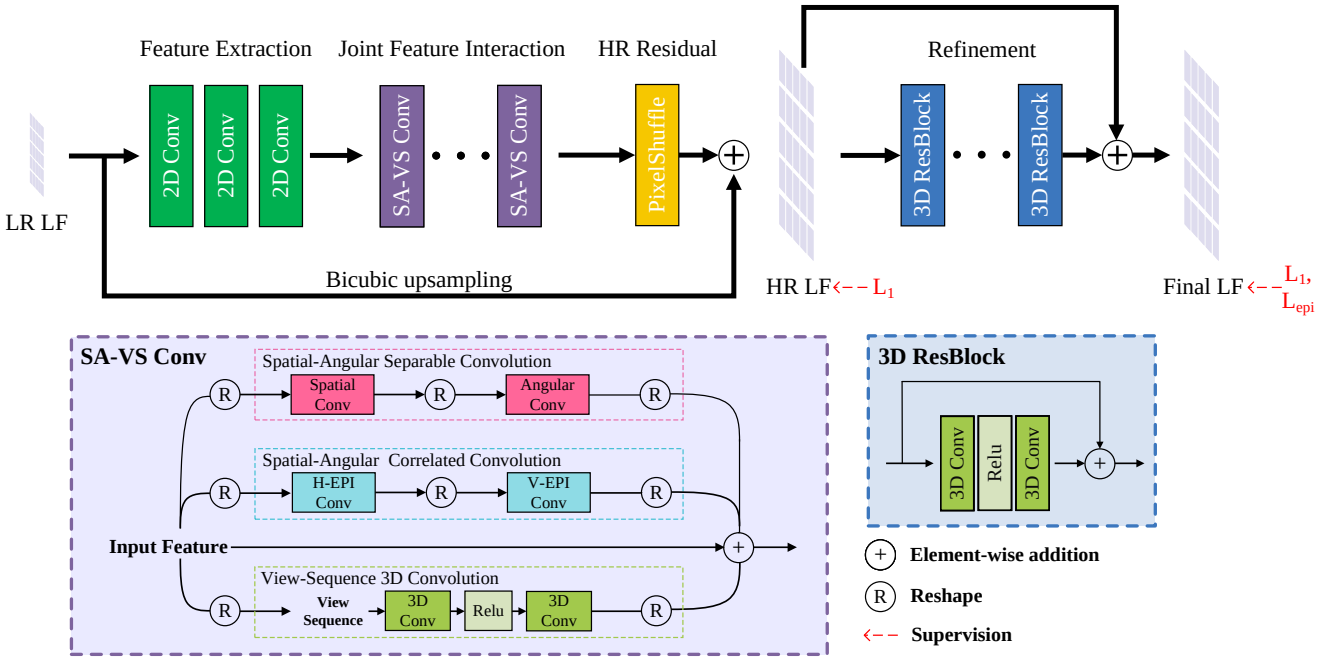

Figure 9. The SHU-IVIPLab Team: The network architecture of the proposed SA-VSNet.

each view of the input LR LF, which are then fed to 16 SA-VS Conv blocks for joint feature interaction. Each SA-VS Conv block consists of an SAS-Conv module, an SAC-Conv module, and a View-Sequence 3D Convolution module. The generated features are then processed by a PixelShuffle layer to predict the initial HR LF images which are supervised by the $\mathcal{L}_1$ loss. In the fine stage, this method adopts four 3D residual blocks to further refine the initial super-resolved results, and employs a hybrid loss function consisting of the EPI gradient loss [70] and the $\mathcal{L}_1$ loss to enhance details.

4.10. CBNU-MIP-Lab: EPIS-LFSR

Following the pipeline of LF-InterNet [1], this team rearranged the input LF images into MacPI pattern, and carefully designed a series of 2D convolutions for MacPIs. An overview of the proposed network is shown in Fig. 10, and readers can refer to [97] for more details of their method. The input LR LF is first processed by a spatial convolution to extract shallow features. Then, the shallow features are processed by 8 Extract-Groups (each group consists of 8 cascaded Extract-Blocks) to generate the deep features. The proposed network is built in a residual-in-residual manner for better SR performance.

4.11. LFSR-gdut-team: MAFNetSR

This team followed DistgSSR [2] to develop a series of 2D convolutions with channel attention for LF image SR. This method was trained using the default training setting in the BasicLFSR toolbox, and achieved better performance

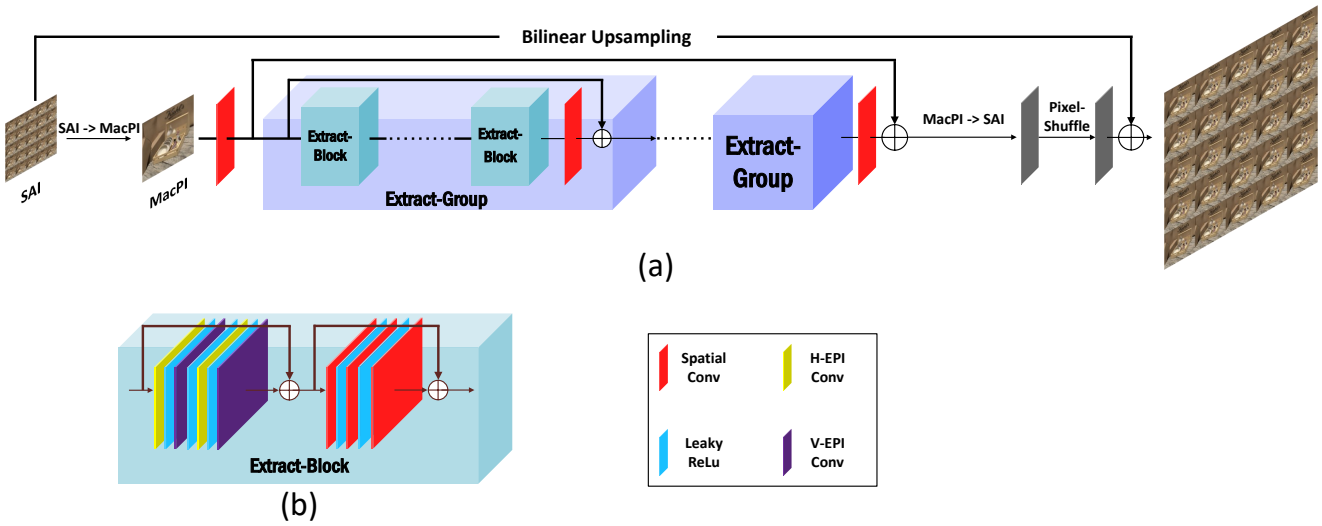

Figure 10. The CBNU-MIP-Lab Team: The network architecture of the proposed EPIS-LFSR.

than most baselines.

5. Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (No. 61921001, U20A20185, 61972435), the Young Talents Project of Hunan (2020RC3026), the Guangdong Basic and Applied Basic Research Foundation (2022B1515020103), and the Shenzhen Science and Technology Program (No. RCYX20200714114641140, JCYJ20190807152209394).

We thank the NTIRE 2023 sponsors: Sony Interactive Entertainment, Meta Reality Labs, ModelScope, ETH Zürich (Computer Vision Lab) and University of Würzburg (Computer Vision Lab).

6. Teams and Affiliations

Challenge Organizers

Members:
Yingqian Wang[1] (wangyingqian16@nudt.edu.cn),
Longguang Wang[2] (wanglongguang15@nudt.edu.cn),
Zhengyu Liang[1] (zyliang@nudt.edu.cn),
Jungang Yang[1] (yangjungang@nudt.edu.cn),
Radu Timofte[3,4] (timofter@vision.ee.ethz.ch),
Yulan Guo[5,1] (guoyulan@sysu.edu.cn).

Affiliations:
[1]National University of Defense Technology
[2]Aviation University of Air Force
[3]University of Würzburg
[4]ETH Zürich
[5]The Shenzhen Campus of Sun Yat-sen University, Sun Yat-sen University

(1) The OpenMeow Team

Members:
Kai Jin[1] (jinkai@bigo.sg), Zeqiang Wei[2,3], Angulia Yang[1], Sha Guo[4], Mingzhi Gao[1], Xiuzhuang Zhou[5]

Affiliations:
[1]Bigo Technology Pte. Ltd.
[2]Smart Medical Innovation Lab, Beijing University of Posts and Telecommunications
[3]Global Explorer Ltd., Suzhou China
[4]National Engineering Research Center of Visual Technology, School of Computer Science, Peking University
[5]School of Artificial Intelligence, Beijing University of Posts and Telecommunications

(2) The DMLab Team

Members:
Vinh Van Duong[1] (duongvinh@skku.edu), Thuc Nguyen Huu[1], Jonghoon Yim[1], Byeungwoo Jeon[1]

Affiliations:
[1]Department of Electrical and Computer Engineering, Sungkyunkwan University

(3) The VIDAR Team

Members:
Yutong Liu[1] (ustclyt@mail.ustc.edu.cn), Zhen Cheng[1], Zeyu Xiao[1], Ruikang Xu[1], Zhiwei Xiong[1]

Affiliations:
[1]University of Science and Technology of China

(4) The IIR-Lab Team

Members:
Gaosheng Liu[1] (gaoshengliu@tju.edu.cn), Manchang Jin[1], Huanjing Yue[1], Jingyu Yang[1]

Affiliations:
[1]School of Electrical and Information Engineering, Tianjin University

(5) The INSIS Team

Members:
Chen Gao[1] (gaochen@bjtu.edu.cn), Shuo Zhang[1], Song Chang[1], Youfang Lin[1]

Affiliations:
[1]Beijing Key Lab of Traffic Data Analysis and Mining, School of Computer and Information Technology, Beijing Jiaotong University

(6) The BNU-AI-TRY Team

Members:
Wentao Chao[1] (chaowentao@mail.bnu.edu.cn), Xuechun Wang[1], Guanghui Wang[2], Fuqing Duan[1]

Affiliations:
[1]Beijing Normal University
[2]Toronto Metropolitan University

(7) The BIT912 Team

Members:
Wang Xia[1] (3220221027@bit.edu.cn), Yan Wang[1], Peiqi Xia[1], Shunzhou Wang[1], Yao Lu[1,2]

Affiliations:
[1]Beijing Institute of Technology
[2]Shenzhen MSU-BIT University

(8) The HAWKEYE Group Team

Members:
Ruixuan Cong[1,2,3] (congrx@buaa.edu.cn), Hao Sheng[1,2,3], Da Yang[1,2,3], Rongshan Chen[1,2,3], Sizhe Wang[1,2,3], Zhenglong Cui[1,2,3]

Affiliations:
[1]State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University
[2]Beihang Hangzhou Innovation Institute Yuhang
[3]Faculty of Applied Sciences, Macao Polytechnic University

(9) The SHU-IVIPLab Team

Members:
Yilei Chen[1] (yileichen@shu.edu.cn), Yongjie Lu[1], Dongjun Cai[1], Ping An[1]

Affiliations:
[1]School of Communication and Information Engineering, Shanghai University

(10) The CBNU-MIP-Lab Team

Members:
Ahmed Salem[1] (ahmeddiefy@cbnu.ac.kr), Hatem Ibrahem[1], Bilel Yagoub[1], Hyun-Soo Kang[1]

Affiliations:
[1]School of Information and Communication Engineering, Chungbuk National University

(11) The LFSR-gdut-team Team

Members:
Zekai Zeng[1] (2112204431@mail2.gdut.edu.cn), Heng Wu[1]

Affiliations:
[1]Guangdong University of Technology

References

[1] Yingqian Wang, Longguang Wang, Jungang Yang, Wei An, Jingyi Yu, and Yulan Guo. Spatial-angular interaction for light field image super-resolution. In *European Conference on Computer Vision (ECCV)*, 2020. 1, 2, 4, 5, 10

[2] Yingqian Wang, Longguang Wang, Gaochang Wu, Jungang Yang, Wei An, Jingyi Yu, and Yulan Guo. Disentangling light fields for super-resolution and disparity estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 1, 2, 3, 4, 5, 10

[3] Vaibhav Vaish, Bennett Wilburn, Neel Joshi, and Marc Levoy. Using plane+ parallax for calibrating dense camera arrays. In *CVPR*, 2004. 1

[4] Yingqian Wang, Jungang Yang, Yulan Guo, Chao Xiao, and Wei An. Selective light field refocusing for camera arrays using bokeh rendering and superresolution. *IEEE Signal Processing Letters*, 26(1):204–208, 2018. 1

[5] Shuo Zhang, Hao Sheng, Chao Li, Jun Zhang, and Zhang Xiong. Robust depth estimation for light field via spinning parallelogram operator. *Computer Vision and Image Understanding*, 145:148–159, 2016. 1

[6] Williem, In Kyu Park, and Kyoung Mu Lee. Robust light field depth estimation using occlusion-noise aware data costs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(10):2484–2497, 2018. 1

[7] Changha Shin, Hae-Gon Jeon, Youngjin Yoon, In So Kweon, and Seon Joo Kim. Epinet: A fully-convolutional neural network using epipolar geometry for depth from light field images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4748–4757, 2018. 1

[8] Yu-Ju Tsai, Yu-Lun Liu, Ming Ouhyoung, and Yung-Yu Chuang. Attention-based view selection networks for light-field disparity estimation. In *AAAI Conference on Artificial Intelligence (AAAI)*, volume 34, pages 12095–12103, 2020. 1

[9] Jiaxin Chen, Shuo Zhang, and Youfang Lin. Attention-based multi-level fusion network for light field depth estimation. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2021. 1

[10] Yingqian Wang, Longguang Wang, Zhengyu Liang, Jungang Yang, Wei An, and Yulan Guo. Occlusion-aware cost constructor for light field depth estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1

[11] Wentao Chao, Xuechun Wang, Yingqian Wang, Liang Chang, and Fuqing Duan. Learning sub-pixel disparity distribution for light field depth estimation. *arXiv preprint arXiv:2208.09688*, 2022. 1

[12] Hao Sheng, Yebin Liu, Jingyi Yu, Gaochang Wu, Ruixuan Cong, Rongshan Chen, et al. Lfnat 2023 challenge on light field depth estimation: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023. 1

[13] Ryan S Overbeck, Daniel Erickson, Daniel Evangelakos, Matt Pharr, and Paul Debevec. A system for acquiring, processing, and rendering panoramic light field stills for virtual reality. *ACM Transactions on Graphics*, 37(6):1–15, 2018. 1

[14] Jingyi Yu. A light-field journey to virtual reality. *IEEE MultiMedia*, 24(2):104–112, 2017. 1

[15] Gaochang Wu, Yebin Liu, Lu Fang, and Tianyou Chai. Revisiting light field rendering with deep anti-aliasing neural network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 1

[16] Vincent Sitzmann, Semon Rezchikov, Bill Freeman, Josh Tenenbaum, and Fredo Durand. Light field networks: Neural scene representations with single-evaluation rendering. *Advances in Neural Information Processing Systems (NeurIPS)*, 34, 2021. 1

[17] Huan Wang, Jian Ren, Zeng Huang, Kyle Olszewski, Menglei Chai, Yun Fu, and Sergey Tulyakov. R2l: Distilling neural radiance field to neural light field for efficient novel view synthesis. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1

[18] Benjamin Attal, Jia-Bin Huang, Michael Zollhoefer, Johannes Kopf, and Changil Kim. Learning neural light fields with ray-space embedding networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1

[19] H Zhu, M Guo, H Li, Q Wang, and A Robles-Kelly. Revisiting spatio-angular trade-off in light field cameras and extended applications in super-resolution. *IEEE Transactions on Visualization and Computer Graphics*, 2019. 1

[20] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision (ECCV)*, pages 184–199, 2014. 1

[21] Longguang Wang, Yingqian Wang, Xiaoyu Dong, Qingyu Xu, Jungang Yang, Wei An, and Yulan Guo. Unsupervised degradation representation learning for blind super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10581–10590, 2021. 1

[22] Longguang Wang, Xiaoyu Dong, Yingqian Wang, Xinyi Ying, Zaiping Lin, Wei An, and Yulan Guo. Exploring sparsity in image super-resolution for efficient inference. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4917–4926, 2021. 1

[23] Juncheng Li, Faming Fang, Kangfu Mei, and Guixu Zhang. Multi-scale residual network for image super-resolution. In *ECCV*, pages 517–532, 2018. 1

[24] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *European Conference on Computer Vision (ECCV)*, pages 286–301, 2018. 1

[25] Zhisheng Lu, Juncheng Li, Hong Liu, Chaoyan Huang, Linlin Zhang, and Tieyong Zeng. Transformer for single image super-resolution. In *CVPRW*, pages 457–466, 2022. 1, 3

[26] Longguang Wang, Yingqian Wang, Zhengfa Liang, Zaiping Lin, Jungang Yang, Wei An, and Yulan Guo. Learning parallax attention for stereo image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1

[27] Yingqian Wang, Xinyi Ying, Longguang Wang, Jungang Yang, Wei An, and Yulan Guo. Symmetric parallax attention for stereo image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 766–775, 2021. 1

[28] Qinyan Dai, Juncheng Li, Qiaosi Yi, Faming Fang, and Guixu Zhang. Feedback network for mutually boosted stereo image super-resolution and disparity estimation. In *ACM MM*, 2021. 1

[29] Xiaojie Chu, Liangyu Chen, and Wenqing Yu. Nafssr: Stereo image super-resolution using nafnet. In *CVPRW*, 2022. 1

[30] Hansheng Guo, Juncheng Li, Guangwei Gao, Zhi Li, and Tieyong Zeng. Pft-ssr: Parallax fusion transformer for stereo image super-resolution. 2023. 1

[31] Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 0–0, 2019. 1

[32] Xinyi Ying, Longguang Wang, Yingqian Wang, Weidong Sheng, Wei An, and Yulan Guo. Deformable 3d convolution for video super-resolution. *IEEE Signal Processing Letters*, 27:1500–1504. 1

[33] Kelvin CK Chan, Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Basicvsr: The search for essential components in video super-resolution and beyond. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4947–4956, 2021. 1

[34] Longguang Wang, Yulan Guo, Li Liu, Zaiping Lin, Xinpu Deng, and Wei An. Deep video super-resolution using HR optical flow estimation. *IEEE Transactions on Image Processing*, 2020. 1

[35] Martin Rerabek and Touradj Ebrahimi. New light field image dataset. In *International Conference on Quality of Multimedia Experience (QoMEX)*, 2016. 1, 3

[36] Katrin Honauer, Ole Johannsen, Daniel Kondermann, and Bastian Goldluecke. A dataset and evaluation methodology for depth estimation on 4d light fields. In *Asian Conference on Computer Vision (ACCV)*, pages 19–34, 2016. 1, 3

[37] Sven Wanner, Stephan Meister, and Bastian Goldluecke. Datasets and benchmarks for densely sampled 4d light fields. In *Vision, Modelling and Visualization (VMV)*, volume 13, pages 225–226, 2013. 1, 3

[38] Mikael Le Pendu, Xiaoran Jiang, and Christine Guillemot. Light field inpainting propagation via low rank matrix completion. *IEEE Transactions on Image Processing*, 27(4):1981–1993, 2018. 1, 3

[39] Vaibhav Vaish and Andrew Adams. The (new) stanford light field archive. *Computer Graphics Laboratory, Stanford University*, 6(7), 2008. 1, 3

[40] Alina Shutova, Egor Ershov, Georgy Perevozchikov, Ivan A Ermakov, Nikola Banic, Radu Timofte, Richard Collins, Maria Efimova, Arseniy Terekhin, et al. NTIRE 2023 challenge on night photography rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[41] Pierluigi Zama Ramirez, Fabio Tosi, Luigi Di Stefano, Radu Timofte, et al. NTIRE 2023 challenge on hr depth from images of specular and transparent surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[42] Yawei Li, Yulun Zhang, Luc Van Gool, Radu Timofte, et al. NTIRE 2023 challenge on image denoising: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[43] Xiaoyang Kang, Xianhui Lin, Kai Zhang, Zheng Hui, Wangmeng Xiang, Jun-Yan He, Xiaoming Li, Peiran Ren, Xuansong Xie, Radu Timofte, et al. NTIRE 2023 video colorization challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[44] Florin-Alexandru Vasluianu, Tim Seizinger, Radu Timofte, et al. NTIRE 2023 image shadow removal challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[45] Xiaohong Liu, Xiongkuo Min, Wei Sun, Yulun Zhang, Kai Zhang, Radu Timofte, Guangtao Zhai, Yixuan Gao, Yuqin Cao, Tengchuan Kou, Yunlong Dong, Ziheng Jia, et al. NTIRE 2023 quality assessment of video enhancement challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[46] Longguang Wang, Yulan Guo, Yingqian Wang, Juncheng Li, Shuhang Gu, Radu Timofte, et al. NTIRE 2023 challenge on stereo image super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[47] Yingqian Wang, Longguang Wang, Zhengyu Liang, Jungang Yang, Radu Timofte, Yulan Guo, et al. NTIRE 2023 challenge on light field image super-resolution: Dataset, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[48] Yulun Zhang, Kai Zhang, Zheng Chen, Yawei Li, Radu Timofte, et al. NTIRE 2023 challenge on image super-resolution (x4): Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[49] Mingdeng Cao, Chong Mou, Fanghua Yu, Xintao Wang, Yinqiang Zheng, Jian Zhang, Chao Dong, Ying Shan, Gen Li, Radu Timofte, et al. NTIRE 2023 challenge on 360° omnidirectional image and video super-resolution: Datasets, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[50] Marcos V Conde, Manuel Kolmet, Tim Seizinger, Thomas E. Bishop, Radu Timofte, et al. Lens-to-lens bokeh effect transformation. NTIRE 2023 challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[51] Marcos V Conde, Eduard Zamfir, Radu Timofte, et al. Efficient deep models for real-time 4k image super-resolution. NTIRE 2023 benchmark and report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[52] Codruta O Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluianu, Radu Timofte, et al. NTIRE 2023 challenge on nonhomogeneous dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[53] Yawei Li, Yulun Zhang, Luc Van Gool, Radu Timofte, et al. NTIRE 2023 challenge on efficient super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[54] Song Chang, Youfang Lin, and Shuo Zhang. Flexible hybrid lenses light field super-resolution using layered refinement. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 5584–5592, 2022. 2

[55] Jing Jin, Junhui Hou, Jie Chen, Sam Kwong, and Jingyi Yu. Light field super-resolution via attention-guided fusion of hybrid lenses. In *ACM International Conference on Multimedia (ACM MM)*, pages 193–201, 2020. 2

[56] Haitian Zheng, Minghao Guo, Haoqian Wang, Yebin Liu, and Lu Fang. Combining exemplar-based approach and learning-based approach for light field super-resolution using a hybrid imaging system. In *International Conference on Computer Vision Workshops (ICCVW)*, pages 2481–2486, 2017. 2

[57] Yuwang Wang, Yebin Liu, Wolfgang Heidrich, and Qionghai Dai. The light field attachment: Turning a dslr into a light field camera using a low budget camera ring. *IEEE Transactions on Visualization and Computer Graphics*, 23(10):2357–2364, 2016. 2

[58] Yeyao Chen, Gangyi Jiang, Mei Yu, Haiyong Xu, and Yo-Sung Ho. Deep light field spatial super-resolution using heterogeneous imaging. *IEEE Transactions on Visualization and Computer Graphics*, 2022. 2

[59] Tom E Bishop and Paolo Favaro. The light field camera: Extended depth of field, aliasing, and superresolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5):972–986, 2011. 2

[60] Sven Wanner and Bastian Goldluecke. Variational light field analysis for disparity estimation and super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):606–619, 2013. 2

[61] Reuben A Farrugia, Christian Galea, and Christine Guillemot. Super resolution of light field images using linear subspace projection of patch-volumes. *IEEE Journal of Selected Topics in Signal Processing*, 11(7):1058–1071, 2017. 2

[62] Martin Alain and Aljosa Smolic. Light field denoising by sparse 5d transform domain collaborative filtering. In *International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6, 2017. 2

[63] Mattia Rossi and Pascal Frossard. Graph-based light field super-resolution. In *International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6, 2017. 2

[64] Youngjin Yoon, Hae-Gon Jeon, Donggeun Yoo, Joon-Young Lee, and In So Kweon. Learning a deep convolutional network for light-field image super-resolution. In *IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 24–32, 2015. 2

[65] Yunlong Wang, Fei Liu, Kunbo Zhang, Guangqi Hou, Zhenan Sun, and Tieniu Tan. Lfnet: A novel bidirectional recurrent convolutional neural network for light-field image super-resolution. *IEEE Transactions on Image Processing*, 27(9):4274–4286, 2018. 2

[66] Shuo Zhang, Youfang Lin, and Hao Sheng. Residual networks for light field image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11046–11055, 2019. 2

[67] Shuo Zhang, Song Chang, and Youfang Lin. End-to-end light field spatial super-resolution network using multiple epipolar geometry. *IEEE Transactions on Image Processing*, 2021. 2

[68] Zhen Cheng, Zhiwei Xiong, and Dong Liu. Light field super-resolution by jointly exploiting internal and external similarities. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019. 2

[69] Nan Meng, Hayden Kwok-Hay So, Xing Sun, and Edmund Lam. High-dimensional dense residual convolutional neural network for light field reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 2

[70] Jing Jin, Junhui Hou, Jie Chen, and Sam Kwong. Light field spatial super-resolution via deep combinatorial geometry embedding and structural consistency regularization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2260–2269, 2020. 2, 7, 10

[71] Yingqian Wang, Jungang Yang, Longguang Wang, Xinyi Ying, Tianhao Wu, Wei An, and Yulan Guo. Light field image super-resolution using deformable convolution. *IEEE Transactions on Image Processing*, 2020. 2, 3

[72] Yu Mo, Yingqian Wang, Chao Xiao, Jungang Yang, and Wei An. Dense dual-attention network for light field image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(7):4431–4443, 2021. 2

[73] Henry Wing Fung Yeung, Junhui Hou, Xiaoming Chen, Jie Chen, Zhibo Chen, and Yuk Ying Chung. Light field spatial super-resolution using deep efficient spatial-angular separable convolution. *IEEE Transactions on Image Processing*, 28(5):2319–2330, 2018. 2, 6

[74] Gaosheng Liu, Huanjing Yue, Jiamin Wu, and Jingyu Yang. Intra-inter view interaction network for light field image super-resolution. *IEEE Transactions on Multimedia*, 2021. 2, 3

[75] Zhen Cheng, Zhiwei Xiong, Chang Chen, Dong Liu, and Zheng-Jun Zha. Light field super-resolution with zero-shot learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10010–10019, 2021. 3

[76] Yingqian Wang, Zhengyu Liang, Longguang Wang, Jungang Yang, Wei An, and Yulan Guo. Learning a degradation-adaptive network for light field image super-resolution. *arXiv preprint arXiv:2206.06214*, 2022. 3

[77] Zeyu Xiao, Yutong Liu, Ruisheng Gao, and Zhiwei Xiong. Cutmib: Boosting light field super-resolution via multi-view image blending. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 3

[78] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, pages 6000–6010, 2017. 3

[79] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *International Conference on Computer Vision Workshops (ICCVW)*, pages 1833–1844, 2021. 3, 6

[80] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12299–12310, 2021. 3

[81] Jingyun Liang, Yuchen Fan, Xiaoyu Xiang, Rakesh Ranjan, Eddy Ilg, Simon Green, Jiezhang Cao, Kai Zhang, Radu Timofte, and Luc V Gool. Recurrent video restoration transformer with guided deformable attention. *Advances in Neural Information Processing Systems*, 35:378–393, 2022. 3

[82] Jingyun Liang, Jiezhang Cao, Yuchen Fan, Kai Zhang, Rakesh Ranjan, Yawei Li, Radu Timofte, and Luc Van Gool. Vrt: A video restoration transformer. *arXiv preprint arXiv:2201.12288*, 2022. 3

[83] Jiezhang Cao, Yawei Li, Kai Zhang, and Luc Van Gool. Video super-resolution transformer. *arXiv preprint arXiv:2106.06847*, 2021. 3

[84] Shunzhou Wang, Tianfei Zhou, Yao Lu, and Huijun Di. Detail-preserving transformer for light field image super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2022. 3

[85] Zhengyu Liang, Yingqian Wang, Longguang Wang, Jungang Yang, and Shilin Zhou. Light field image super-resolution with transformers. *IEEE Signal Processing Letters*, 2022. 3, 5, 9

[86] Xiao Guo, Xinzhu Sang, Binbin Yan, Duo Chen, and Peng Wang. Light field image super-resolution based on raw data with transformers. *JOSA A*, 39(12):2131–2141, 2022. 3

[87] Zijian Wang and Yao Lu. Multi-granularity aggregation transformer for light field image super-resolution. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 261–265, 2022. 3

[88] Zhengyu Liang, Yingqian Wang, Longguang Wang, Jungang Yang, Zhou Shilin, and Yulan Guo. Learning non-local spatial-angular correlation for light field image super-resolution. *arXiv preprint arXiv:2302.08058*, 2023. 3, 4, 5, 8

[89] Jaejun Yoo, Namhyuk Ahn, and Kyung-Ah Sohn. Rethinking data augmentation for image super-resolution: A comprehensive analysis and a new strategy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8375–8384, 2020. 4

[90] Radu Timofte, Rasmus Rothe, and Luc Van Gool. Seven ways to improve example-based single image super resolution. In *CVPR*, pages 1865–1873, 2016. 5

[91] Kai Jin, Angulia Yang, Zeqiang Wei, Sha Guo, Mingzhi Gao, and Xiuzhuang Zhou. Distgepit: Enhanced disparity learning for light field image super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023. 5

[92] Vinh Van Duong, Thuc Nguyen Huu, Jonghoon Yim, and Byeungwoo Jeon. Light field image super-resolution network via joint spatial-angular and epipolar information. *IEEE Transactions on Computational Imaging*, 2023. 6

[93] Zhen Cheng, Yutong Liu, and Zhiwei Xiong. Spatial-angular versatile convolution for light field reconstruction. *IEEE Transactions on Computational Imaging*, 8:1131–1144, 2022. 6, 9

[94] Chen Gao, Youfang Lin, Chang Song, and Shuo Zhang. Spatial-angular multi-scale mechanism for light field spatial super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023. 8

[95] Yangling Chen, Shuo Zhang, Song Chang, and Youfang Lin. Light field reconstruction using efficient pseudo 4d epipolar-aware structure. *IEEE Transactions on Computational Imaging*, 8:397–410, 2022. 8

[96] Xiaoyi Dong, Jianmin Bao, Dongdong Chen, Weiming Zhang, Nenghai Yu, Lu Yuan, Dong Chen, and Baining Guo. Cswin transformer: A general vision transformer backbone with cross-shaped windows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12124–12134, 2022. 9

[97] Ahmed Salem, Hatem Ibrahem, and Hyun-Soo Kang. Learning epipolar-spatial relationship for light field image super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023. 10