

NTIRE 2025 Challenge on Light Field Image Super-Resolution: Methods and Results

Yingqian Wang*, Zhengyu Liang*, Fengyuan Zhang*, Lvli Tian*, Longguang Wang*,
 Juncheng Li*, Jungang Yang*†, Radu Timofte*, Yulan Guo*, Kai Jin, Zeqiang Wei, Angulia Yang,
 Di Wu, Mingzhi Gao, Xiuzhuang Zhou, Yue Yan, Yuhao Wang, Shuang Chen, Zeping Tian, Yizhi Hu ,
 Yao Lu, Haosong Liu, Xiancheng Zhu, Huanqiang Zeng, Jianqing Zhu, Yifan Shi, Junhui Hou,
 Mingyang Yu, Zhijian Wu, Dingjiang Huang, Wenli Zheng, Zekai Xu, Huiyuan Fu, Heng Zhang,
 Zhijuan Huang, Hongyuan Yu, Zeke Zexi Hu, Haodong Chen, Vera Yuk Ying Chung,
 Xiaoming Chen, Zean Chen, Yeyao Chen, Gangyi Jiang, Haiyong Xu, Ting Luo, Guanglong Liao,
 Danhao Zhang, Siyu Zhang, Wendong Mao, Zhongfeng Wang, Sunita Arya, Abhishek Kumar Sinha,
 S Manthira Moorthi, Hao Zhang, Hao Sheng, Da Yang, Zhenglong Cui, Shuai Wang, Haotian Zhang,
 Xingzheng Wang, Yuanbo Huang, Jiahao Lin, Yuhang Lin, Ahmed Salem, Ebrahem Elkady,
 Hatem Ibrahim, Jae-Won Suh, Hyun-Soo Kang, Changguang Wu, Hao Hou, Pengpeng Li,
 Peng Huang, Jiangxin Dong, Jinhui Tang

Abstract

This report summarizes the 3rd NTIRE challenge on light field (LF) image super-resolution (SR), focusing on novel methods and their outcomes. This challenge aims to super-resolve LF images degraded by bicubic downsampling, and comprises three tracks: a classical track, an efficiency track, and a large model track. In total, 308 participants registered, and 13 teams submitted results that outperformed the baseline methods. The challenge has established a new state-of-the-art in LF image SR, e.g., the winning method in Track 1 achieves a 0.36 dB PSNR improvement over last year's champion on the test set. We present the submitted solutions, analyze their common trends, and highlight practical techniques. We hope this challenge will inspire further advancements in LF image SR.

* Yingqian Wang, Zhengyu Liang, Fengyuan Zhang, Lvli Tian, Longguang Wang, Juncheng Li, Jungang Yang, Radu Timofte and Yulan Guo are the NTIRE 2025 challenge organizers, while the other authors participated in this challenge.

† Corresponding author: Jungang Yang

Section 5 provides the authors and affiliations of each team.

NTIRE 2025 webpage: <https://cvlai.net/ntire/2025/>

Challenge webpage (Track 1): <https://codalab.lisn.upsaclay.fr/competitions/21276>

Challenge webpage (Track 2): <https://codalab.lisn.upsaclay.fr/competitions/21276>

Challenge webpage (Track 3): <https://codalab.lisn.upsaclay.fr/competitions/21278>

Github: <https://github.com/The-Learning-And-Vision-Atelier-LAVA/LF-Image-SR/tree/NTIRE2025>

1. Introduction

Light field (LF) cameras can capture the intensity and directions of light rays, and record 3D geometry in an effective way. Through encoding 3D scene cues into 4D LF images (2D for spatial dimension and 2D for angular dimension), LF cameras can facilitate numerous appealing applications, including post-capture refocusing [1–3], depth sensing [4–6], virtual reality [7, 8], and view rendering [9–12].

High-resolution (HR) light field imaging has become critically important across various applications, as it significantly enhances perceptual quality and facilitates advanced post-processing tasks. However, the acquisition of HR LF imagery faces substantial technical challenges stemming from the fundamental spatial-angular resolution trade-off inherent in LF imaging systems. This physical constraint necessitates the development of sophisticated reconstruction techniques to generate HR LF images from their low-resolution (LR) counterparts, which is a technical challenge formally referred to as LF image super-resolution (SR).

To develop and benchmark LF image SR methods, the NTIRE LF image SR challenge series [13, 14] were hosted every year since 2023. In the challenge, the popular bicubic downsampling degradation is used to generate LR LF images, and the objective is to make the super-resolved LF images as faithful as the groundtruth HR ones. Specifically,

BasicLFSR toolbox: <https://github.com/ZhengyuLiang24/BasicLFSR>

the NTIRE-LFSR 2023 challenge employed the widely used and publicly available LF datasets [15–19] as training set, and proposed a new LF dataset called NTIRE-LFSR [13] for both validation (model development) and test (final ranking). The NTIRE-LFSR 2024 challenge is inherited from NTIRE-LFSR 2023, with a special focus on the optimization of model size (*i.e.*, the number of parameters) and computational cost (*i.e.*, FLOPs).

Building upon the success of previous challenges, we hold the 3rd LF image SR challenge at NTIRE 2025. The NTIRE-LFSR 2025 challenge extends NTIRE-LFSR 2024 through the introduction of a novel “large model track”. Distinguished from conventional settings, this new track permits the use of external training data while removing the restrictions on model size and computational complexity. We aim to catalyze the exploration of the potential of foundation models in LF image SR.

This challenge is one of the NTIRE 2025 Workshop associated challenges on: ambient lighting normalization [20], reflection removal in the wild [21], shadow removal [22], event-based image deblurring [23], image denoising [24], XGC quality assessment [25], UGC video enhancement [26], night photography rendering [27], image super-resolution (x4) [28], real-world face restoration [29], efficient super-resolution [30], HR depth estimation [31], efficient burst HDR and restoration [32], cross-domain few-shot object detection [33], short-form UGC video quality assessment and enhancement [34, 35], text to image generation model quality assessment [36], day and night raindrop removal for dual-focused images [37], video quality assessment for video conferencing [38], low light image enhancement [39], light field super-resolution [40], restore any image model (RAIM) in the wild [41], raw restoration and super-resolution [42] and raw reconstruction from RGB on smartphones [43].

2. Related Work

In this section, we aim to offer a concise review of the major achievements in the field of LF image SR. Existing LF image SR methods can be broadly divided into two categories: traditional (*i.e.*, non-learning) methods and deep learning-based methods.

2.1. Traditional Methods

LF image SR has long been a challenging research topic that has garnered attention for decades. Wanner et al. [44] initially estimated disparity maps using a structure tensor and subsequently developed a variational framework for LF image SR. Farrugia et al. [45] constructed a patch-volume dictionary of HR and LR LF image pairs and introduced a multivariate ridge regression method to learn the lin-

ear mapping from LR patch volumes to their HR counterparts. Alain et al. [46] addressed the ill-posed LF image SR problem as an optimization problem based on the sparsity prior. Rossi et al. [47] integrated inter-view information using graph regularization and formulated LF image SR as a quadratic problem, which can be efficiently solved with standard convex optimization techniques.

2.2. Deep learning-based Methods

In the past decade, deep learning-based methods have revolutionized the field of LF image SR. Yoon et al. [48] introduced the first CNN-based LF image SR method (*i.e.*, LFCNN). This pioneering work demonstrated the potential of CNNs in LF image SR. Since then, numerous deeper CNN architectures with various mechanisms for incorporating angular information have been developed to achieve improved SR performance in LF image SR tasks.

Wang et al. [49] introduced a bidirectional recurrent CNN to integrate angular information from sub-aperture images (SAIs) along the horizontal or vertical angular direction. Zhang et al. [50] stacked SAIs along four different angular directions and developed a four-branch residual network to implicitly learn the epipolar geometry from the stacked SAIs for LF image SR. Meng et al. [51] applied 4D convolutions to simultaneously incorporate spatial and angular information from 4D LF data and developed the high-dimensional dense residual network (HDDRNet) for LF image SR. Jin et al. [52] proposed an all-to-one method for LF image SR and performed structural consistency regularization to preserve the parallax structure. Moreover, several methods have decomposed high-dimensional LF data into different subspaces for LF image SR. Wang et al. [53] proposed spatial and angular feature extractors to extract corresponding information from macro-pixel images and developed the LF-InterNet to repeatedly interact spatial and angular information for LF image SR. In their subsequent work [54], Wang et al. further generalized the interaction mechanism into an LF disentangling mechanism and developed three CNNs (DistgSSR, DistgASR, and DistgDisp) for spatial super-resolution, angular super-resolution, and disparity estimation, respectively. Following LF-InterNet, Liu et al. [55] proposed an intra-inter view interaction network (LF-IINet) with two parallel branches to extract global inter-view information and model correlations among all intra-view features.

Over the past three years, researchers have begun to explore the application of Transformers in LF image SR. Wang et al. [56] introduced the Detail-Preserving Transformer (DPT) for LF image SR. In this approach, the sub-aperture images (SAIs) of each vertical and horizontal view are treated as sequences, and long-range geometric dependencies are learned through a spatial-angular locally enhanced self-attention layer. Liang et al. [57] proposed a

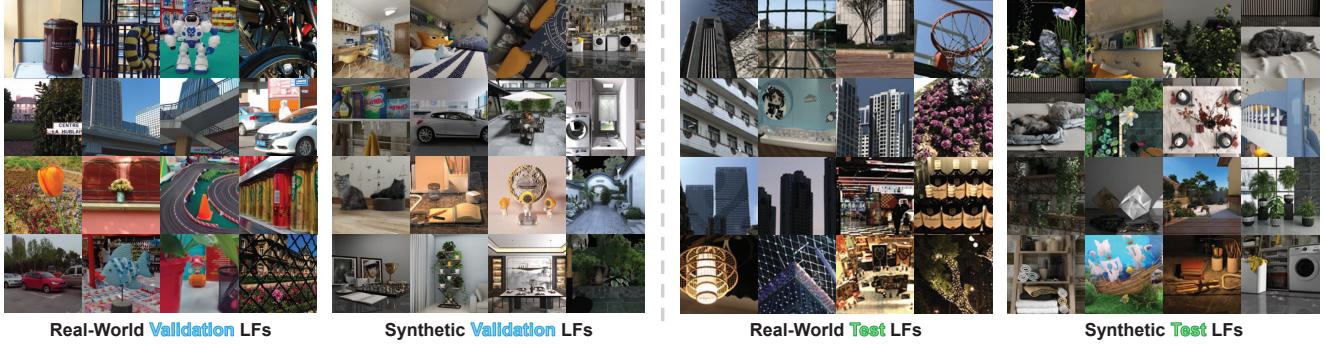


Figure 1. An illustration of the center-view images in the NTIRE-LFSR dataset [13]. Both validation and test sets contain 16 real-world and 16 synthetic LFs, respectively.

straightforward yet effective Transformer network, known as LFT, for LF image SR. Their method designs an angular Transformer to integrate complementary information across different views and a spatial Transformer to capture both local and long-range dependencies within each SAI. More recently, Liang et al. [58] investigated the non-local spatial-angular correlations in LF image SR and developed a Transformer-based network called EPIT, which achieves state-of-the-art SR performance. The proposed EPIT attains a global receptive field along the epipolar line and demonstrates robustness to disparity variations. Cong et al. [59] designed a deep transformer-based network called LF-DET for LF spatial SR. LF-DET leverages a spatial-angular separable transformer encoder with innovative strategies: sub-sampling spatial modeling helps manage computational costs when processing spatial information, while multi-scale angular modeling adapts to varying disparity ranges by focusing on multi-scale macro-pixel regions. Based on these established techniques, participants have proposed many solutions to the LF image SR challenge. Jin et al. [60] combined EPIT [58] and DistgSSR [54] to create the DistgEPIT network for LF image SR. This network won the NTIRE 2023 LF Image SR Challenge [13]. Most recently, BigEPIT improves upon the EPIT architecture by increasing the number of feature channels and cascading blocks to better capture spatial-angular correlations, achieving state-of-the-art performance. This network won the NTIRE 2024 LF Image SR Challenge [14].

Transformers have improved LFSR by capturing long-range dependencies but face efficiency issues. To overcome this, Mamba, a State Space Model (SSM) with linear complexity, has been introduced. LFMamba [61] uses Mamba on 2D slices of 4D light fields to effectively enhance SR performance.

3. NTIRE 2025 Challenge

In this section, we introduce the NTIRE 2025 LF image SR Challenge. We first introduce the official datasets and

toolbox of this challenge. Then, we review the three tracks and two phases of this challenge. Finally, we summarize the common trends in the submitted solutions.

3.1. Datasets, Toolbox and Evaluation

Training set. This challenge follows the common settings in [13, 14, 54, 58], and uses the EPFL [15], HCInew [16], HCIold [17], INRIA [18] and STFgantry [19] datasets for training. All the 144 LFs in the training set have an angular resolution of 9×9 . Challenge participants are required to use these LF images as HR groundtruth to train their models. External training data and pretrained models are not allowed in Tracks 1 and 2, but can be used in Track 3.

Validation and test set. We use the NTIRE-LFSR dataset developed in the 1st NTIRE LF image SR challenge [13] for validation and test, as shown in Fig. 1. Both validation and test sets contain 16 synthetic scenes (rendered by 3DS MAX) and 16 real-world scenes (captured by Lytro Illum cameras). Details of the NTIRE-LFSR dataset can be referred to [13]. All the LF images in the validation and test set are bicubically 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 and ranking.

Toolbox. We provide BasicLFSR, an open-source and easy-to-use toolbox to facilitate participants to quickly get access to LF image SR and develop their own models. The BasicLFSR toolbox is publicly available at <https://github.com/ZhengyuLiang24/BasicLFSR>.

Evaluation. 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).

Table 1. NTIRE 2025 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.	FLOPs	Architec* (C/T/M)
			Average	Lytro	Synthetic	Average	Lytro	Synthetic			
Track 1	1	OpenMeow*	31.16/.9366	31.43/.9531	30.89/.9201	33.15/.9534	33.92/.9606	32.38/.9462	9.10M	445.08G	T&M
	2	BITSMBU*	31.09/.9354	31.47/.9522	30.71/.9186	32.81/.9511	33.46/.9579	32.15/.9442	3.91M	115.75G	C&M
	3	SmartVIPLab*	30.86/.9336	31.08/.9503	30.64/.9170	32.81/.9511	33.53/.9578	32.10/.9443	11.20M	340.80G	T&M
	4	Only My Railgun	30.63/.9313	30.84/.9478	30.43/.9150	32.64/.9492	33.49/.9574	31.79/.9410	2.74M	79.13G	T
	5	BuptMM	30.62/.9312	30.79/.9475	30.45/.9148	32.40/.9478	33.12/.9544	31.68/.9411	15.53M	791.46G	T&M
	6	NBULFLab	30.60/.9313	30.86/.9483	30.35/.9143	32.28/.9474	33.11/.9547	31.46/.9400	6.16M	101.56G	C&Others
	7	Icais-AI-team	30.34/.9284	30.58/.9453	30.10/.9114	31.75/.9423	32.68/.9506	30.83/.9341	5.95M	146.86G	C&M
	8	SpaceVision	30.27/.9311	30.19/.9476	30.35/.9145	32.44/.9489	33.07/.9558	31.81/.9419	7.83M	166.76G	T
	9	HawkeyeGroup	30.15/.9295	30.04/.9457	30.26/.9132	32.27/.9473	32.87/.9541	31.66/.9405	19.65M	710.71G	T
	10	SZU-VS	30.12/.9297	30.02/.9462	30.22/.9131	32.24/.9471	32.93/.9546	31.56/.9396	4.08M	141.67G	T
Track 2	1	BITSMBU*	30.39/.9289	30.57/.9454	30.21/.9125	32.31/.9466	33.02/.9541	31.60/.9390	0.54M	17.03G	C&M
	2	OpenMeow*	30.34/.9280	30.50/.9445	30.18/.9116	32.30/.9466	33.02/.9543	31.58/.9388	0.45M	19.33G	T&M
	3	LFSR-DASE*	30.23/.9270	30.41/.9434	30.05/.9107	32.03/.9450	32.88/.9527	31.18/.9373	0.98M	19.87G	T
	4	CBNU-MIP&VC-Labs	30.13/.9255	30.26/.9419	30.00/.9092	32.04/.9441	32.58/.9504	31.50/.9379	0.66M	19.88G	C&T
	5	SmartVIPLab	30.13/.9258	30.29/.9424	29.97/.9092	32.00/.9434	32.68/.9505	31.32/.9363	0.58M	19.01G	T&M
	6	IMAG	30.09/.9275	30.05/.9433	30.12/.9116	32.13/.9458	32.80/.9534	31.46/.9382	0.97M	16.69G	T
	7	BuptMM	30.06/.9248	30.24/.9416	29.88/.9079	31.86/.9426	32.65/.9503	31.06/.9349	0.83M	19.59G	T
	8	Only My Railgun	30.02/.9245	30.17/.9411	29.87/.9078	32.21/.9450	33.05/.9538	31.36/.9363	0.69M	19.58G	T
Track 3	1	OpenMeow*	31.22/.9370	31.49/.9536	30.95/.9204	33.21/.9538	33.98/.9610	32.45/.9467	12.04M	590.99G	T&M
	2	BITSMBU*	31.09/.9354	31.47/.9522	30.71/.9186	32.81/.9511	33.46/.9579	32.15/.9442	3.91M	115.75G	C&M
	3	BuptMM*	30.62/.9312	30.79/.9475	30.45/.9148	32.53/.9499	33.27/.9570	31.79/.9428	15.53M	791.46G	T&M
	4	SpaceVision	30.27/.9311	30.19/.9476	30.35/.9145	32.44/.9489	33.07/.9558	31.81/.9419	7.83M	166.76G	T
	5	SZU-VS	30.12/.9297	30.02/.9462	30.22/.9131	32.24/.9471	32.93/.9546	31.56/.9396	4.08M	141.67G	T
Baselines	-	BigEPIT [62]	30.80/.9332	31.00/.9496	30.60/.9167	32.74/.9508	33.46/.9576	32.01/.9441	11.04M	569.30G	T
	-	DistgEPIT [60]	30.66/.9314	30.82/.9475	30.51/.9152	32.71/.9496	33.36/.9562	32.07/.9430	20.34M	566.48G	C&T
	-	EPIT [58]	29.87/.9259	29.72/.9420	30.03/.9097	32.04/.9447	32.54/.9507	31.53/.9387	1.47M	76.39G	T
	-	DistgSSR [54]	29.64/.9244	29.39/.9403	29.88/.9084	31.75/.9424	32.26/.9490	31.23/.9357	3.58M	65.27G	C
	-	Bicubic	25.79/.8378	25.11/.8404	26.46/.8352	27.51/.8714	27.49/.8719	27.53/.8710	-	-	-

Note: “C” denotes that the model is developed based on convolutions, “T” denotes that the model adopts Transformer as basic components, and “M” denotes that the model takes Mamba as basic components.

3.2. Tracks

Track 1: Classical. This track aims to encourage participants to explore the precision upper bound of LF image SR. In this track, the rankings are determined by the average PSNR value on the test set only. DistgSSR [54] is set as the baseline method in this track. The solutions with PSNR values lower than DistgSSR will not be ranked in the final leaderboard.

Track 2: Efficiency. In this track, the model size (*i.e.*, number of parameters) is restricted to 1 MB, and the FLOPs is restricted to 20 G (with an input LF of size $5 \times 5 \times 32 \times 32$). The rankings are determined by the average PSNR value on the test set, but the solutions with model size larger than 1M or FLOPs larger than 20G will not be ranked in the final leaderboard. Bicubic interpolation is set as the baseline method in this track. The solutions with PSNR values lower than the bicubic interpolation will not be ranked in the final leaderboard.

Track 3: Large Model. In this track, the participants are allowed to use external training data and pretrained models for model development, and there is no efficiency limita-

tion. The rankings are determined by the average PSNR value on the test set only. DistgSSR [54] is set as the baseline method in this track, and the solutions with PSNR values lower than DistgSSR will not be ranked in the final leaderboard.

3.3. Challenge Phases

Development Phase. The participants can download the validation set and apply their developed models to the LR LF images to generate their SR versions. A validation leaderboard is available during this phase. The participants can compare their scores with the ones achieved by the baseline models or models developed by other participants.

Test phase. The participants are required to apply their models to the released 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 deadline.

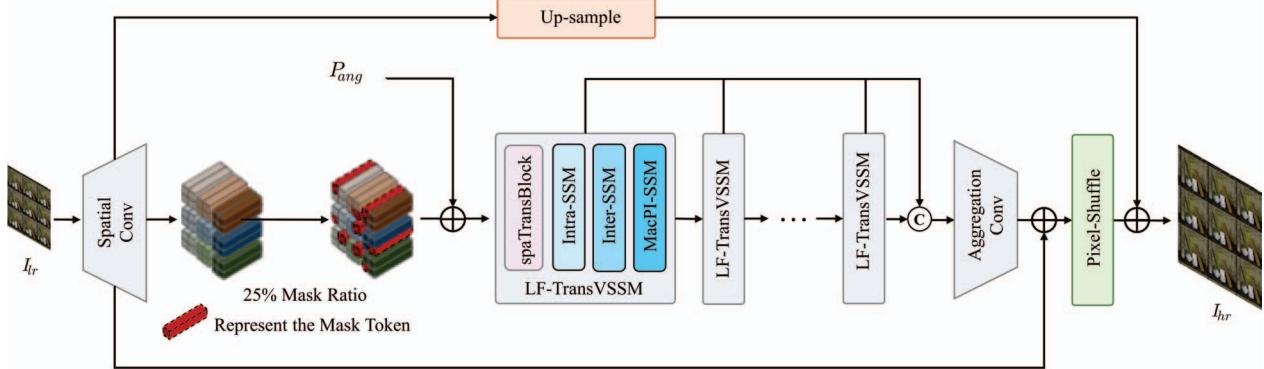


Figure 2. An overview of the LFTransMamba network.

3.4. Challenge Results

Among the 308 registered participants, 13 teams have participated in the final test phase of the NTIRE 2025 LF Image SR Challenge and submitted their results, codes, and factsheets. This year’s challenge consists of three tracks. Tracks 1 and 2 continue with the settings established in the 2024 competition, attracting participation from all teams, and the newly introduced Track 3 attracted 5 teams choosing to participate, reflecting its growing appeal and the increasing interest in tackling large-model challenges.

Table 1 reports the PSNR and SSIM scores achieved by these methods on both test and validation sets, together with their major details. Notably, all of the top three teams in both Tracks 1 and 2 surpassed NTIRE-LFSR 2024’s highest scores in their corresponding tracks, **setting new performance benchmarks in LF image SR**. The Track 1 winner, team OpenMeow, achieved a 0.36dB improvement over the winning method of the 2024 competition, *i.e.*, BigEPIT [62]. In Track 2, team BITSMBU, which also won the championship of Track 2 in 2024, recorded PSNR improvements of 0.23dB and 0.21dB on the test and validation sets, respectively.

Notably, Track 3 introduced new challenges by no longer restricting the training datasets and parameter computation. The winning solution from team OpenMeow, leveraged additional synthetic LF datasets, and achieved the best performance over all Tracks while maintaining a small parameter count, thereby showing the effectiveness of external data in boosting LF image SR performance.

Across all three Tracks, a significant observation is **the widespread adoption of Mamba architecture**. Among the 23 submitted competition solutions, 10 implemented a combination of Transformer and Mamba, one utilized a fusion of CNN and Mamba, while the remaining solutions followed the mainstream Transformer-based approach established in NTIRE-LFSR 2024. This trend highlights the generalizability of Mamba architecture and its effectiveness in enhancing LF image SR.

We briefly describe these solutions in Section 4, and introduce the corresponding team members in Appendix 5.

4. Challenge Teams and Methods

4.1. OpenMeow: LFTransMamba (Tracks 1 \star , 2 \star , 3 \star)

This team participated in three tracks and proposed the LFTransMamba network. Readers can refer to [63] for more details of their proposed method.

Track 1: Inspired by the L^2 FIMamba method [64], their LFTransMamba further incorporates a Transformer-based spatial enhancement module to improve SAI spatial feature modeling capabilities, while retaining the lightweight and spatial-angular collaborative design, as shown in Fig. 2.

To further model the global context relationships among SAIs, they introduce a spatial Transformer block in the LF-VSSM module, named LF-TransVSSM. The spatial Transformer block is built upon a multi-head self-attention mechanism that performs attention operations across all SAIs to enhance spatial modeling across regions. This module enables the network to capture long-range dependencies and global structural information, effectively addressing the spatial awareness limitations of the original network. Each spatial Transformer block consists of multiple stacked self-attention layers, with the number of layers denoted by the parameter T .

Masked Light Field Image Modeling: To further improve spatial-angular context modeling, a lightweight training strategy named Masked Light Field Image Modeling (MLFIM) is proposed. Inspired by SimMIM, this strategy requires no additional supervision or pretraining, and can be directly integrated into existing LFSR frameworks for end-to-end optimization. After the initial SpaConv, random masking is independently applied to the feature maps of each SAI. A fixed mask ratio of spatial positions in each SAI is randomly selected and replaced with a learnable mask token, defaulting to 25%. The masked features are then passed through the subsequent LF-TransVSSM modules for spatial-angular interaction and contextual completion. It should be noted that MLFIM is applied only during training, and no masking is used during inference. The network is still optimized with the standard SR reconstruc-

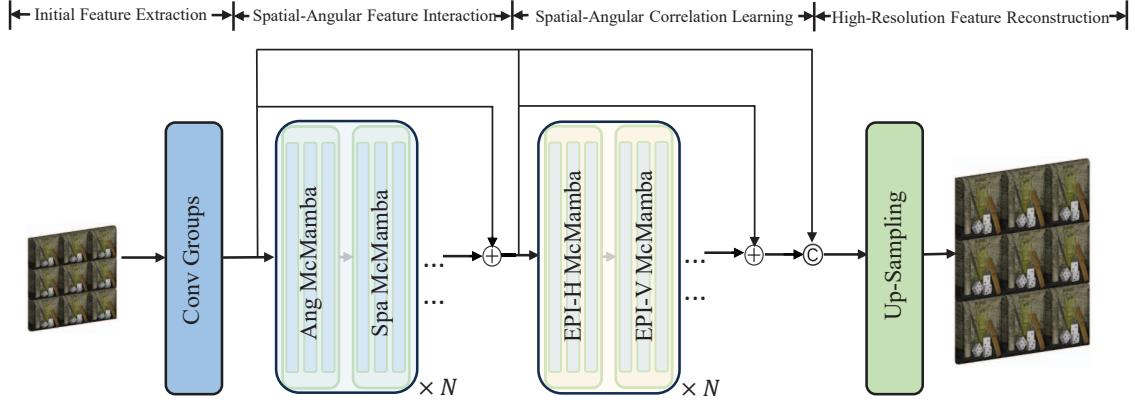


Figure 3. An overview of the MCMamba network.

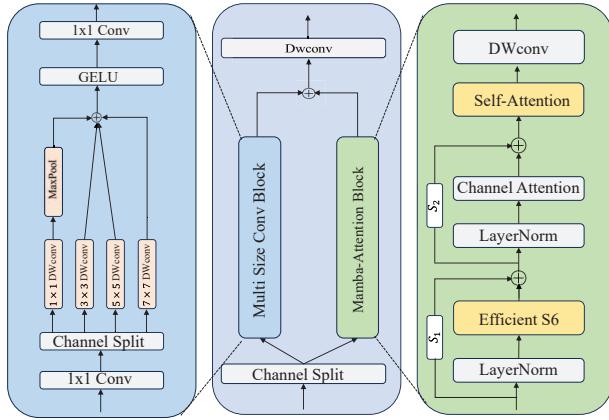


Figure 4. Illustration of the McMamba block.

tion loss on the final output I_{hr} , without any additional loss branches.

Enhanced Position-Sensitive Windowing: Although prior work on PSW [60] demonstrated that eschewing any padding operations during the division stage can yield commendable results, it neglected the treatment of boundary pixels during the integration stage. However, they observe that the network’s inference quality over each patch is not uniform: pixels located near the center of the network’s input patch tend to have higher reconstruction fidelity, while those towards the periphery exhibit lower reliability. To address this issue, they propose EPSW which replaces the conventional uniform weighting in the integration process with a Gaussian weighting scheme. The Gaussian weight function is defined as

$$G(x, y) = \exp\left(-\frac{(x - x_0)^2 + (y - y_0)^2}{2\sigma^2}\right), \quad (1)$$

which assigns higher weights to pixels closer to the patch center (x_0, y_0) and lower weights towards the borders, reflecting the spatial variance in reconstruction accuracy.

Accordingly, the final integrated image $I_{hr}(p)$ at pixel p is reconstructed via a weighted aggregation of overlapping patch estimates $f_i(p)$:

$$I_{hr}(p) = \frac{\sum_{i=1}^N G_i(p) \cdot f_i(p)}{\sum_{i=1}^N G_i(p)}, \quad (2)$$

where $G_i(p)$ denotes the Gaussian weight at pixel p for the i th patch, and N is the total number of patches determined by the stride S and patch size P .

Inference: During inference, they ensemble LF-DET and LFTrans-Mamba models with multiple resolution inference and use the TTA method to boost results.

Track 2: Similar to their approach in Track 1, they also employed the LFTransMamba network in Track 2. It is worth noting that, to reduce computational overhead, the parameter T was set to 0. To ensure a clean and efficient inference pipeline, no Test Time Augmentation (TTA) or ensemble methods were employed.

Track 3: Similar to their approach in Tracks 1 and 2, they also employed the LFTransMamba network in Track 3. For Track 3 participants, two synthetic LF datasets (DLFD and SLFD) [65] were employed. These datasets were generated using Blender and include ground-truth disparity annotations. In line with previous protocols [59], 21 scenes from DLFD and 22 from SLFD were used. During inference, they used the multiple resolution inference and TTA method to boost performance.

4.2. BITSMBU: MCMamba (Tracks 1★, 2★, 3★)

Tracks 1, 3: The BITSMBU team competed in three tracks with their approach, Multi-Scale Context Aggregation.

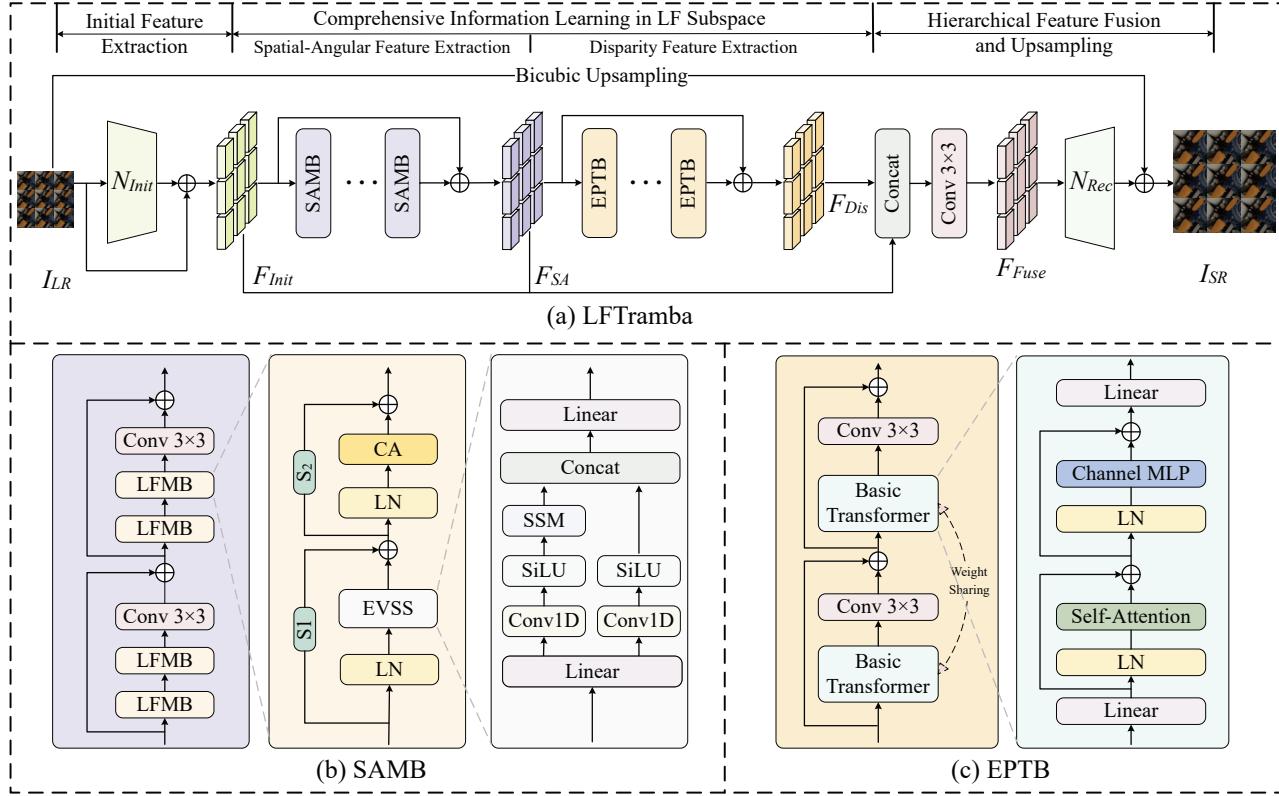


Figure 5. (a) The network architecture of the proposed LFTramba; (b) Illustration of the proposed Spatial-Angular Mamba Block; (c) Illustration of the proposed Epipolar Plane Transformer Block.

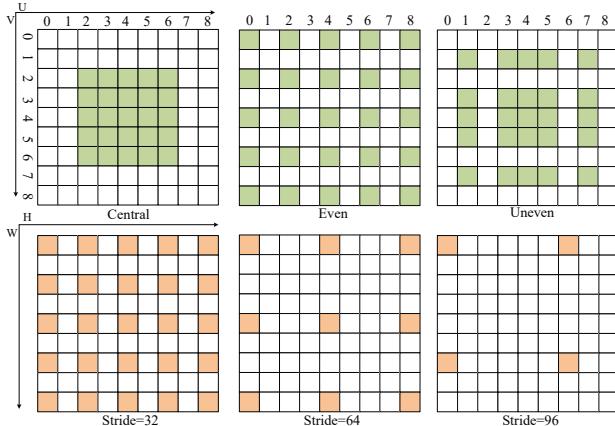


Figure 6. Illustration of the stride-optimized data resampling, including central, even, and uneven sampling with strides of 32, 64, and 96, respectively.

tion Mamba (MCMamba), as shown in Fig. 3. It follows the conventional framework for LF image SR and consists of four main components: initial feature extraction, spatial-angular feature interaction, spatial-angular correlation learning and high-resolution feature reconstruction.

The detailed structure of MCMamba block is shown in Fig. 4. MCMamba block consists of two components: the multi-size conv block and mamba-attention block. In the multi-size conv block, following [66], they adopt a multi-branch structure to enhance feature diversity and strengthen the representation of multi-scale local patterns within the input features. They first apply a 1×1 convolution layer to expand the dimensionality of F_{ms} , producing $F'_{ms} \in R^{B \times 2C_1 \times H \times W}$. These four sub-features are fed into separate Depth Wise (DW) convolution branches to extract features at different spatial scales. After feature extraction is completed in the four branches, the results from the these branches are concatenated, followed by a GELU activation and a 1×1 convolution to produce the final output. In Mamba-Attention Block, the feature F_{ma} is processed through the Efficient S6 [67] module and Channel Attention to obtain the final output. Next, the results from the Mamba-Attention Block and the Multi-Size Conv Block are concatenated, and the feature extraction phase is completed by applying a DW convolution to the combined output.

During inference, they performed PSW++: Position-Sensitive Windowing Strategy proposed by Fidelity-LF-DET [14] to preserve the parallax structure of the border region when cropping the full LF image into patches. They

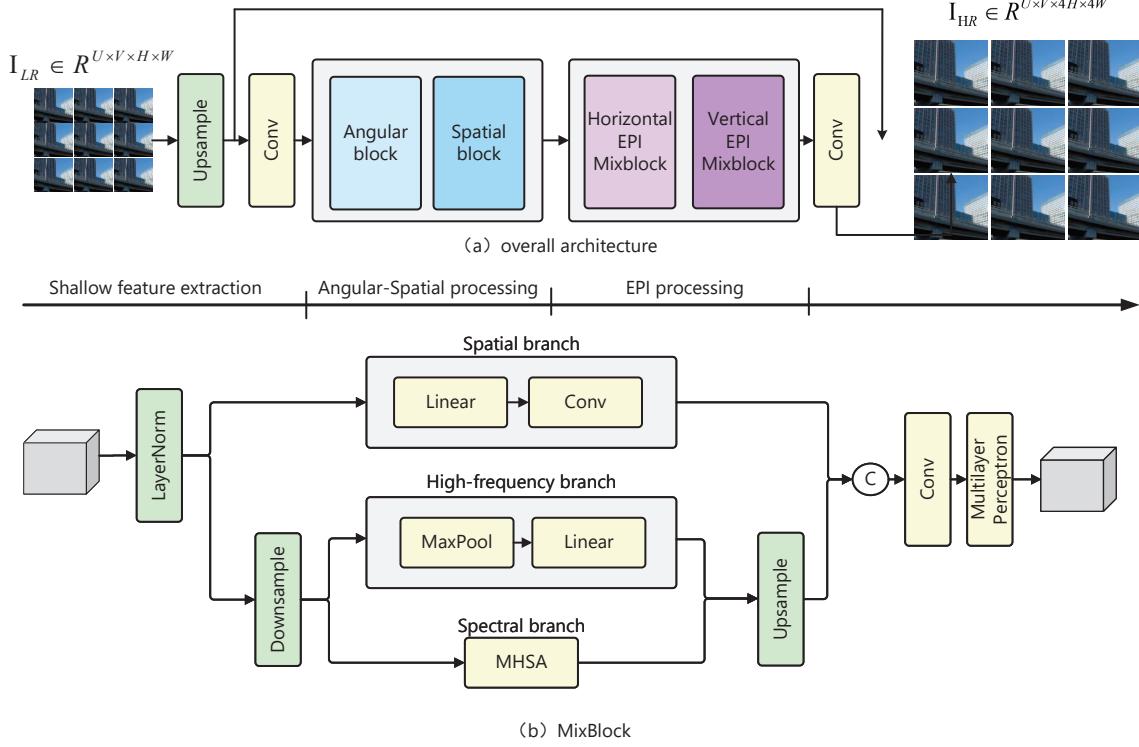


Figure 7. An overview of their LFMix network.

also adopted TTA to further improve the reconstruction quality.

Track 2: In this track, they use the same architecture as in Track 1. The only differences are 1) they use fewer blocks and lower hidden dimensions, 2) they remove the Mamba layer in MCMamba block and only perform self-attention.

4.3. SmartVIPLab: LFTramba (Tracks 1★, 2)

This team participated in two tracks with the proposed LFTramba and its lightweight version, LFTramba-tiny, as shown in Fig. 5(a). Readers can refer to [63] for more details of their proposed method. LFTramba consists of three main components: initial feature extraction, comprehensive information learning in LF subspace, and hierarchical feature fusion and upsampling. The first part follows the approach of prior work EPIT [58]. In the final part, hierarchical features are concatenated along the channel dimension and processed by a convolution to generate the fused feature.

The comprehensive information learning in the LF subspace is achieved through the Spatial-Angular Mamba Block (SAMB) and the Epipolar Plane Transformer Block (EPTB). (1) SAMB, as shown in Fig. 5(b), employs a spatial-angular separable modeling approach to effectively capture both spatial and angular information. The Light

Field Mamba Block (LFMB) is based on the Efficient Visual State Space (EVSS) module with channel attention for better inter-channel feature interactions. Unlike the standard Mamba architecture [68], EVSS replaces causal convolutions with standard convolutions and introduces a symmetric branch to reduce information loss in sequential modeling. By fusing outputs from both branches, EVSS improves feature representation. (2) EPTB, shown in Fig. 5(c), is inspired by the Non-Local Cascading Block [58]. It employs a single-layer spatial convolution to enhance local features while maintaining the parallax structure.

Imbalanced disparity distributions in LF datasets limit model generalization. Prior methods [60, 62] mainly sampled large-disparity regions, which increased training time. To address this, a stride-optimized resampling strategy (Fig. 6) is introduced, using strides of 32, 64, and 96 for central, even, and uneven resampling, respectively. This approach improves disparity coverage, scene texture diversity, and reduces training time and memory usage.

Regularization: LFTramba was optimized using the L1 loss function and the Adam optimizer [69] ($\beta_1 = 0.9$, $\beta_2 = 0.999$) with a learning rate of 2×10^{-4} , which was halved every 15 epochs. For Track 1, the model architecture comprised 2 SAMBs, 10 EPTBs, and 128 channels, trained over 75 epochs with a batch size of 16. For Track 2, the network configuration included 4 SAMBs, 3 EPTBs, and 32

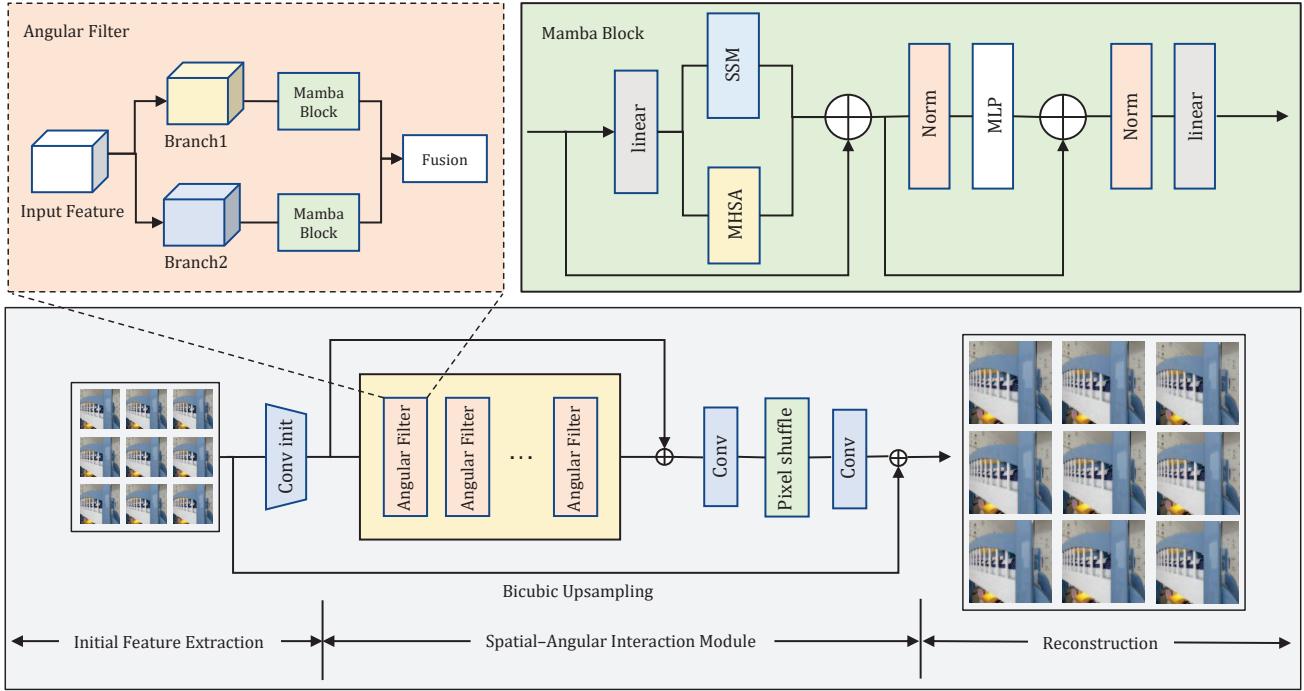


Figure 8. Team BuptMM: The network architecture of the proposed MambaLFSR (Tracks 1, 3).

channels, trained for 90 epochs with a batch size of 2.

Inference: During the inference phase, the Position-Sensitive Windowing (PSW) strategy from DistgEPIT [60] was employed to preserve the parallax structure when partitioning LF images into patches. For Track 1, TTA and a multi-model ensemble strategy [62] were incorporated to enhance reconstruction quality. In contrast, for Track 2, only PSW was utilized to maintain computational efficiency while ensuring structural consistency.

4.4. LFSR-DASE: LFMix (Track 2★)

The LFSR-DASE team proposed a method called LFMix, as shown in Fig. 7. Readers can refer to [70] for more details of their proposed method. Their approach addresses LF image SR through a novel hybrid architecture that jointly processes SAI, MacPI, and EPI representations. The core innovation lies in the MixBlock, which integrates three specialized branches: 1) a spatial branch employing convolutions to preserve full-resolution local details and angular correlations; 2) a spectral branch utilizing self-attention mechanisms on strategically downsampled features to capture global structural patterns with reduced computational complexity; and 3) a high-frequency branch that extracts sharp edge information through max-pooling followed by linear projections. This tri-branch design enables comprehensive modeling of spatial textures, angular consistency, and geometric constraints while maintaining compu-

tational efficiency. By applying controlled downsampling to both the spectral and high-frequency branches prior to feature processing, they significantly reduce FLOPs without sacrificing reconstruction fidelity. The architecture further employs adaptive fusion to combine multi-scale features from different frequency domains, ensuring synergistic utilization of low-frequency structural information and high-frequency details.

4.5. BuptMM: MambaLFSR (Tracks 1, 3★), PDistgF2 (Track 2)

The MambaLFSR method, proposed by the BuptMM team, is depicted in Fig. 8. It consists of three main modules: an initial feature extraction module, a Mamba-based spatial-angular interaction module, and an upsampling reconstruction module. To effectively model long-range dependencies in the spatial-angular domain, Mamba [71] is integrated into the spatial-angular interaction module. Since the unidirectional nature limits its ability to model bidirectional spatial correlations, a dual-branch structure is employed: one branch uses Mamba for global sequence modeling, while the other applies adaptive window multi-head self-attention to capture local interactions. The outputs from these branches are fused and further enhanced through a feed-forward network and residual connections. This hybrid design enables the learning of both long-range and fine-grained spatial-angular representations, thereby signifi-

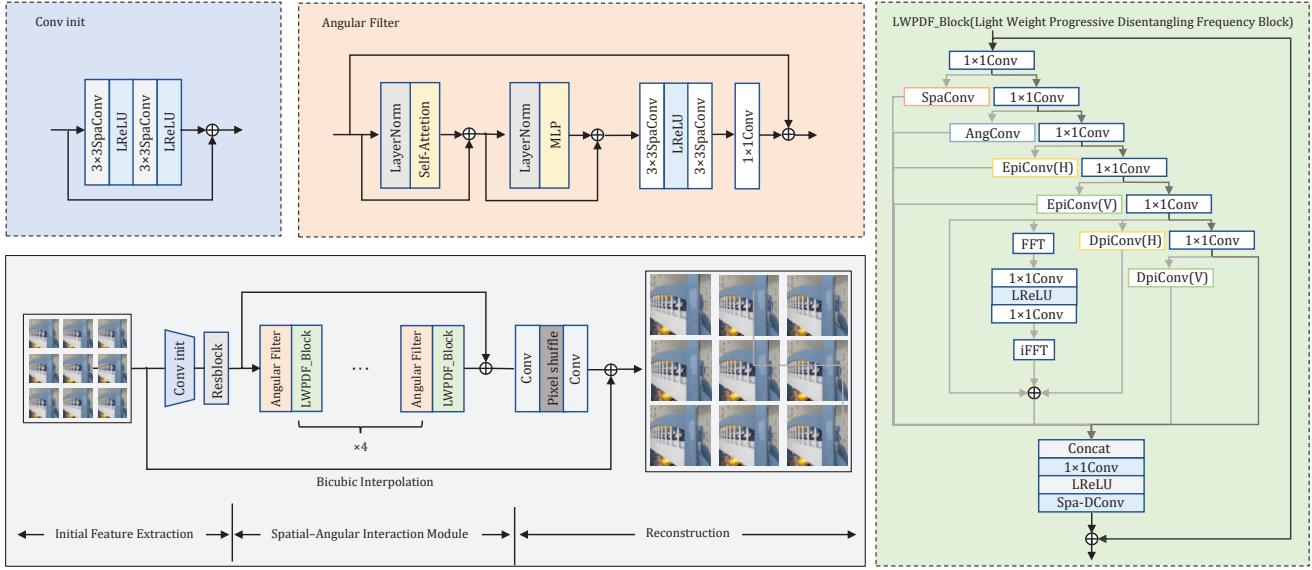


Figure 9. Team BuptMM: The network architecture of the proposed PDistgF2 (Track 2).

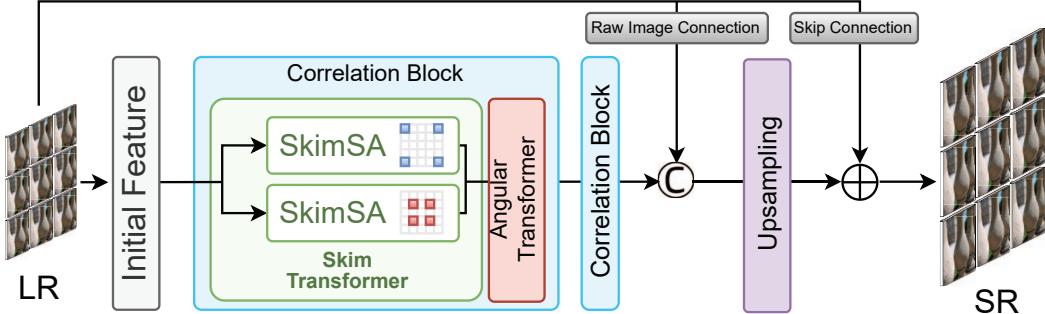


Figure 10. Team Only My Railgun: The network architecture of the proposed SkimLFSR (Tracks 1, 2).

cantly improving performance in light field super-resolution tasks.

For Track 2, the BuptMM team presents PDistgF2 (depicted in Fig. 9), an advanced network building on the success of PDistgNet [72]. PDistgF2 is tailored to achieve high restoration fidelity within strict efficiency constraints. The network has three main stages: initial feature extraction, spatial-angular correlation learning, and reconstruction. A lightweight convolution module with reduced channels and residual connections extracts intra-view features, minimizing computational cost while maintaining effective feature extraction. The core spatial-angular correlation learning stage employs four cascaded blocks, each comprising an angular Transformer (AngTrans) and a progressive disentangling block (LWC42 Conv). The AngTrans captures angular correlations across views using multi-head self-attention (MHSAs) and position encoding, while the LWC42 Conv disentangles features into multiple subspaces. A novel FFT-

based module enhances low-frequency feature extraction in the virtual-slit domain, improving overall SR quality without increasing computational cost. Fine-tuned activation functions further stabilize gradient propagation. PDistgF2 thus balances high performance with efficiency, meeting Track 2’s constraints.

4.6. Only My Railgun: SkimLFSR (Tracks 1, 2)

The Many-to-Many Transformer (M2MT-Net) [73] presents a novel approach for modeling correlations in LF images and achieves state-of-the-art performance in LF image SR with low memory and inference costs. However, similar to DistgSSR [54], LFT [57], and EPIT [58], M2MT-Net processes all information uniformly in a single pass, which leads to inefficiencies, especially considering the high data volume of LF images. This issue is exacerbated in M2MT-Net due to its reliance on heavy linear layers for correlation encoding, which projects a tensor from $U \times V \times C$

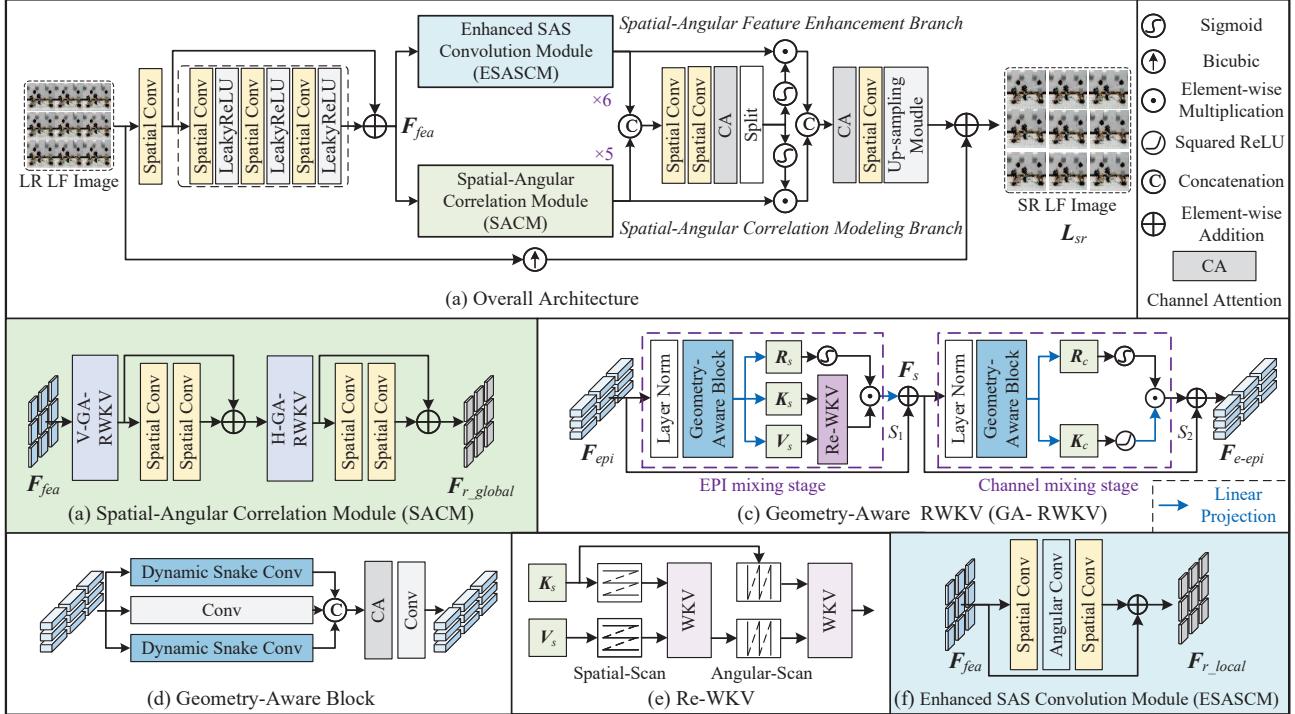


Figure 11. Team NBULFLab: The network architecture of the proposed EPI-RWKV (Track 1).

to a predefined C_{Cor} .

To address these limitations, SkimLFSR was proposed, as illustrated in Fig. 10. Like LFT, EPIT, and M2MT-Net, SkimLFSR employs a series of correlation blocks, each comprising a Skim Transformer and an angular Transformer. The Skim Transformer, a key contribution of this work, is designed to improve efficiency by selectively modeling correlations, while the angular Transformer is adopted from LFT and M2MT-Net to capture angular dependencies. The network concludes with convolutional layers and a pixel shuffler to upsample the spatial resolution and produce the final SR image. To enhance information flow, a raw image connection is introduced to concatenate the input image with the extracted features before upsampling [74], along with a skip connection after upsampling to facilitate residual learning.

This team utilizes the aforementioned model architecture in both Track 1 and 2+. The distinction lies in that Track 2 model is a streamlined version of the Track 1 model. Specifically, in Track 2, this team reduces the number of channels and correlation blocks to decrease model size and computational cost, while incorporating a channel attention block to maintain performance.

4.7. NBULFLab: EPI-RWKV (Track 1)

This team proposes a dual-branch EPI-RWKV (Receptance Weighted Key Value) network for LF image SR

to improve fidelity. The input LR LF image is first mapped to the feature space through a spatial convolution and a spatial residual block consisting of three spatial convolutions, resulting in the feature F_{fea} (as shown in Fig. 11(a)). Subsequently, F_{fea} is processed by two independent branches: the spatial-angular feature enhancement branch and the spatial-angular correlation modeling branch. The spatial-angular feature enhancement branch, composed of multiple enhanced spatial-angular separable (SAS) convolution modules, focuses on exploring spatial and angular information in the LF features (see Fig. 11(f)). The spatial-angular correlation modeling branch utilizes multiple spatial-angular correlation modules to capture global spatial-angular correlations (as shown in Fig. 11(b)). After feature extraction, the features from both branches are concatenated along the channel dimension and adaptively fused using channel attention. Finally, a pixel shuffling-based upsampling module reconstructs the super-resolution LF image (see Fig. 11). To more effectively capture features in both spatial and angular dimensions, an enhanced spatial-angular separable convolution module (ESASM) is designed. Additionally, to deeply model long-range spatial-angular dependencies, a geometric-aware RWKV is constructed within the module (see Fig. 11(c)). Within this geometric-aware RWKV, a geometric-aware block, consisting of dynamic snake convolution (see Fig. 11(d)), is used to explore the LF geometry.

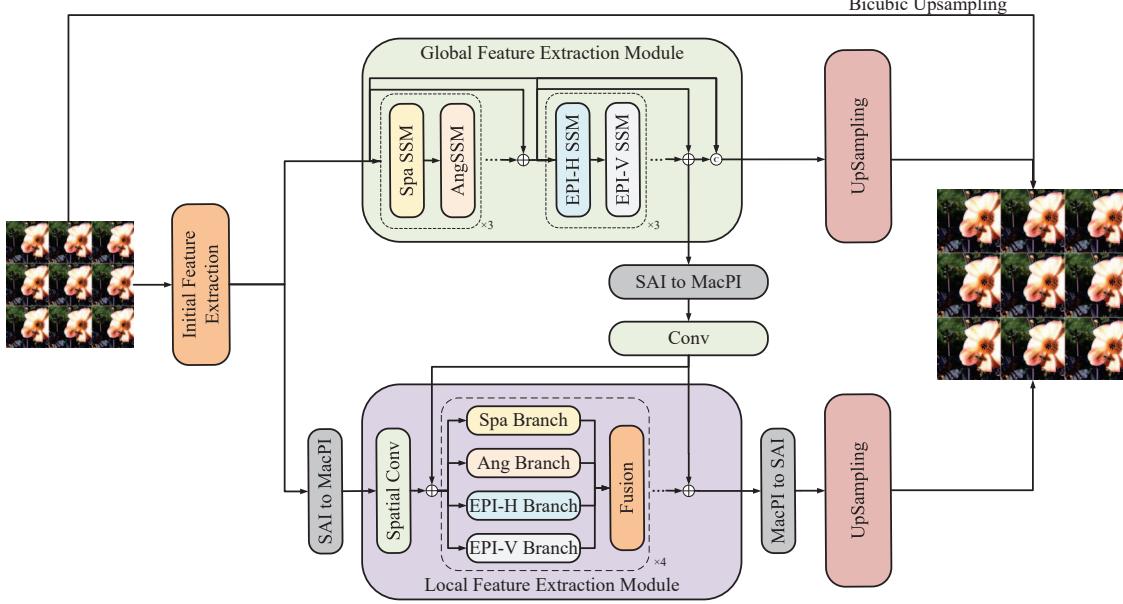


Figure 12. Team Icais-AI-team: The network architecture of the proposed DistgMamba (Track 1).

4.8. Icais-AI-team: DistgMamba (Track 1)

This team proposed DistgMamba, which combines global and local feature extraction, to improve SR performance. The core of this method is a hybrid framework that integrates a Mamba-based global feature extraction module (GFEM) and a CNN-based local feature extraction module (LFEM). The GFEM employs LFMamba [67] to capture global dependencies in the LF image, while the LFEM utilizes DistgSSR [54] to extract local details. Additionally, the LFEM leverages the global information extracted by the GFEM to enhance its perception of global structures, thereby improving overall performance.

The network architecture, as shown in Fig. 12, mainly consists of an initial feature extraction module, GFEM, LFEM, and two independent upsampling modules. The GFEM and LFEM are responsible for extracting global and local features, respectively. The GFEM effectively captures long-range dependencies with the linear complexity of the Mamba model, while the LFEM focuses on modeling local details. Finally, the two upsampling modules upsample the features extracted by the GFEM and LFEM separately, and the results are averaged to generate the final super-resolution LF image. During training, DistgMamba uses the L1 loss and Adam optimizer, and employs data augmentation strategies including random horizontal flipping, vertical flipping, and 90-degree rotation to expand the training data. In the testing phase, TTA [60] is applied to further enhance model performance through transformations such as horizontal flipping, vertical flipping, and rotation. Additionally, a position-sensitive windowing (PSW) oper-

ation [60] is used to improve the structural consistency of disparity.

4.9. SpaceVision: Deep Ensemble of multiscale LF-DET and BigEPIT (Tracks 1, 3)

The SpaceVision team proposed a method named ‘‘Deep Ensemble of multiscale LF-DET and BigEPIT’’. This method integrates two popular network architectures, LF-DET [59] and BigEPIT [62], to address the limitations of each when used individually. LF-DET tends to underperform on real datasets, while BigEPIT faces similar issues on synthetic datasets. To overcome these challenges, this team developed a deep ensemble strategy that combines a multiscale LF-DET with the conventional BigEPIT. The multiscale LF-DET model incorporates two parallel Conv2D branches with kernel sizes of 3×3 and 5×5 in the local feature extraction module of the original LF-DET to handle spatial features at different scales (see Fig. 13). The outputs of these models are then combined using a mean-based approach, where pixel-wise weights are calculated based on the absolute error to produce the final SR image. Both models were implemented in PyTorch, trained with an L1 loss and Adam optimizer, with LF-DET being trained for 100 epochs and BigEPIT for 200 epochs.

4.10. HawkeyeGroup: Big LF-SAET (Track 1)

The HawkeyeGroup team proposed Big LF-SAET, an enhanced version of LF-SAET, which scales its depth and width to improve performance. To effectively leverage spatial, angular [57], and EPI [58] Transformers for global feature extraction, they introduce a specially designed SAET

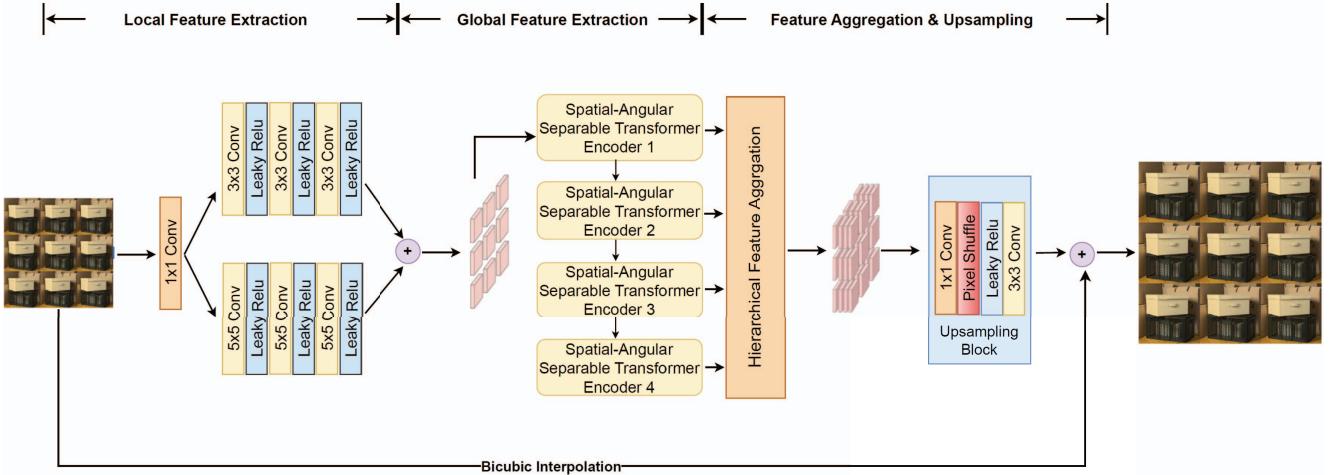


Figure 13. Team SpaceVision: The network architecture of the proposed Deep Ensemble of multiscale LF-DET & BigEPIT (Tracks 1, 3).

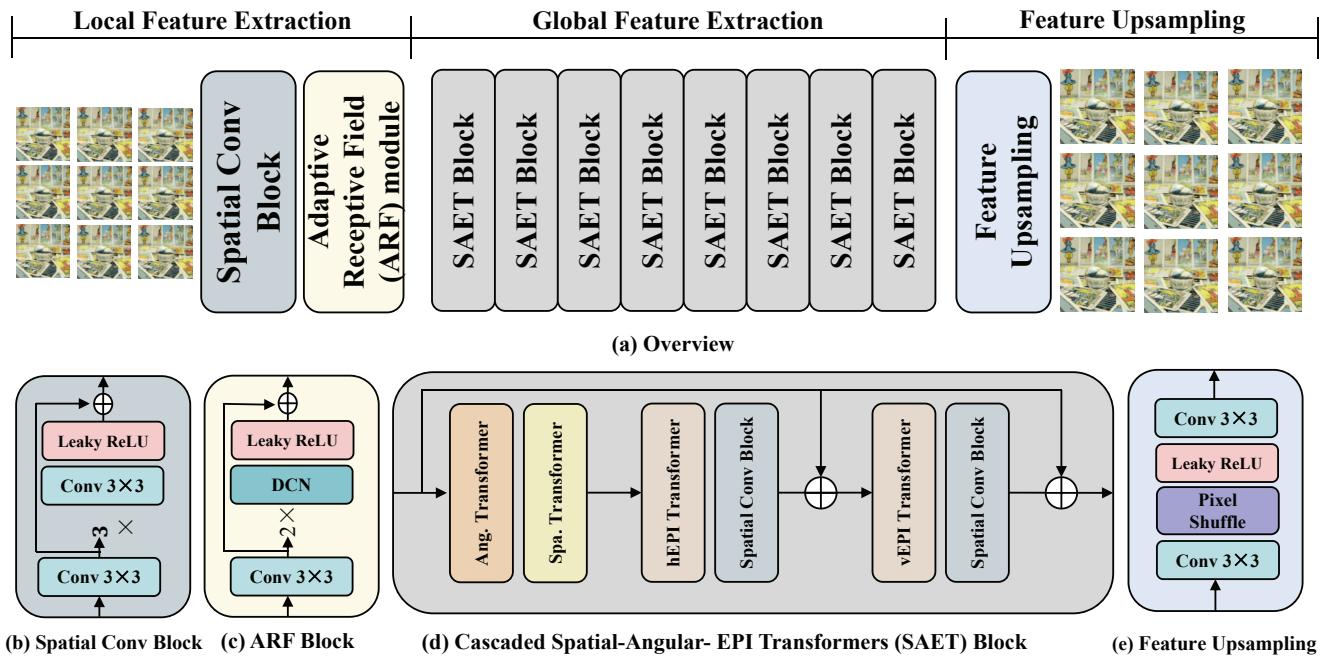


Figure 14. Team HawkeyeGroup: The network architecture of the proposed Big LF-SAET (Track 1).

block that extracts comprehensive information with low computational costs, surpassing traditional convolutional filters confined to local regions. To enhance SR performance, they incorporate parameter-sharing *SpatialConvBlock* in each *SAET* block to integrate spatial information. The architecture comprises 8 stacked *SAET* blocks for sequential global feature extraction, where each block's output serves as the next block's input, facilitating gradual feature refinement, as shown in Fig. 14.

The model is trained using the L1 loss function and the batch size is set to 2 for 4 \times SR. Data augmentation in-

cludes random horizontal flips, vertical flips, and 90-degree rotations. During testing, super-resolved sub-images are merged to reconstruct HR LF images. The model is implemented in PyTorch and trained on a single NVIDIA Tesla V100 GPU.

4.11. SZU-VS: IIATNet (Tracks 1, 3)

The SZU-VS team proposed IIATNet to participate in Tracks 1 and 3.

While Transformers have demonstrated remarkable performance in LF image SR by leveraging self-attention for

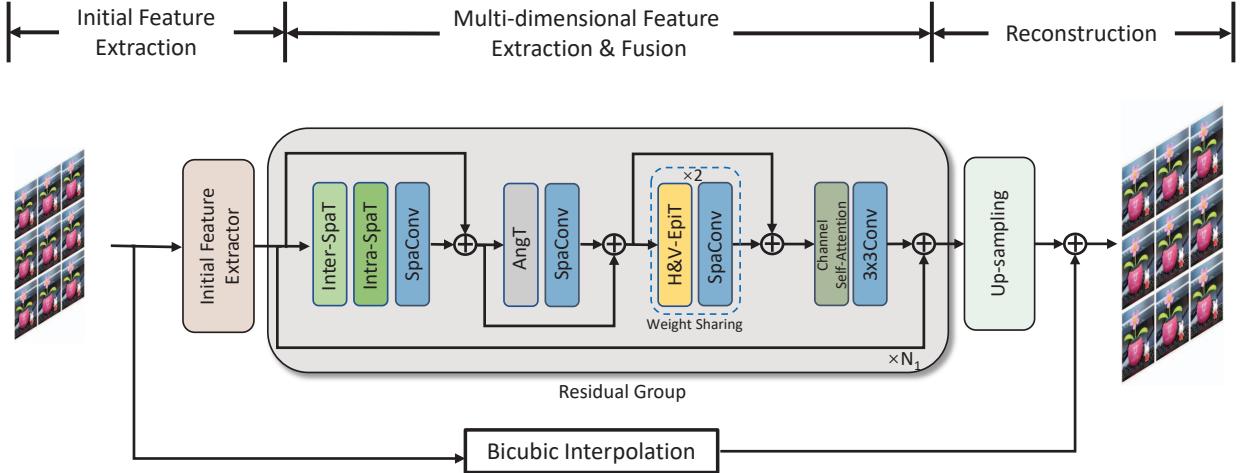


Figure 15. Team SZU-VS: The network architecture of the proposed IIATNet (Tracks 1, 3).

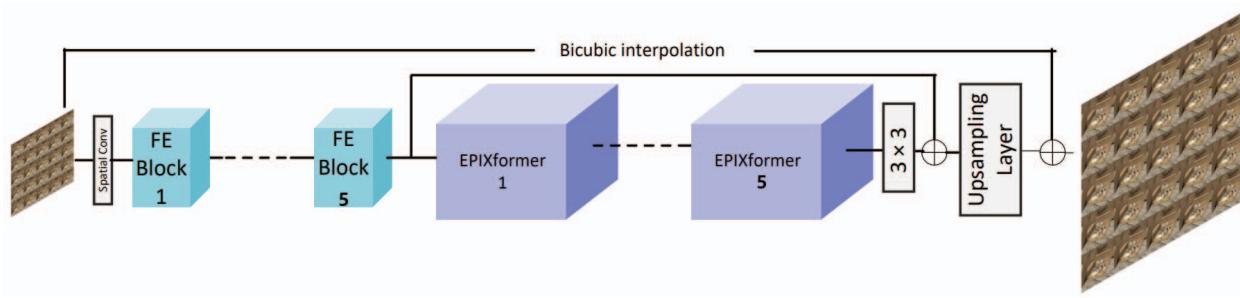


Figure 16. Team CBNM-MIP&VC-Labs: The network architecture of the proposed Light-PILFSSR (Track 2).

global modeling within a single view, they often struggle to capture the complex dependencies across views. To address this limitation, this team introduced an *inter-frame attention mechanism* that enhances the feature representation of the primary view using information from auxiliary views. Specifically, while inputting different views sequentially for computation, they input the central view of the LF image as an auxiliary view into the inter-frame attention simultaneously. By calculating the cross-covariance between these views, an attention weight matrix is generated, representing the inter-view dependencies. This matrix is then used to perform weighted feature aggregation, allowing the model to effectively utilize information from auxiliary views to refine the primary view's features.

As shown in Fig. 15, IIATNet extracts spatial and angular information through a combination of inter-frame and intra-frame Transformers. Spatial convolutions are applied to enhance local detail extraction. The LF image is then reshaped into a macro-pixel image (MacPI) format for further angular feature extraction using Transformers. Additionally, inspired by EPIT [58], the network learns epipolar features to improve robustness to disparity variations.

For training, all LF images were cropped into patches of size 32×32 using bicubic downsampling, with a stride of 32. Data augmentation techniques, including random horizontal flipping, vertical flipping, and 90-degree rotations, were applied while ensuring spatial and angular dimensions were consistently adjusted to maintain LF structure. The network was trained using the L1 loss and a batch size of 1 on a single NVIDIA RTX 3090 Ti GPU.

4.12. CBNM-MIP&VC-Labs: Light-PILFSSR (Track 2)

The CBNM-MIP&VC-Labs team proposed Light-PILFSSR to participate in Track 2.

Following the recent PILFSSR method [75], the team adopts the LF subspace known as virtual-slit images (VSI), enhancing sub-aperture images with sub-pixel information. As shown in Fig. 16, the method leverages the abundant correlation in four-dimensional data through an ensemble representation of LF subspaces for effective feature extraction. The geometry-aware decoder, EPIXformer, utilizes LF physical priors to super-resolve image structures from undersampled LF data. To reduce model complexity, the team

experimented with various configurations and determined to *increase the number of layers while reducing the number of channels improved performance*. Specifically, a 5-layer model with 32 channels outperformed a single-layer model with 64 channels by an average of 0.14 dB PSNR, demonstrating a favorable trade-off between performance and efficiency.

Following the PILFSSR methodology, the central 5×5 views of each LF image were cropped into 128×128 patches for $4 \times$ SR. After converting the images from RGB to YCbCr color space, only the Y channel was used for training and evaluation. Bicubic interpolation was applied to generate LR images. Data augmentation techniques such as random horizontal and vertical flipping and 90-degree rotation were applied. The network was optimized using the L1 loss and a batch size of 4. The training was conducted for 100 epochs on an NVIDIA RTX 4090 GPU, with the initial learning rate being set to 2×10^{-4} and halved every 15 epochs.

4.13. IMAG: LF-HAN (Track 2)

The IMAG team proposed Light Field Hybrid Attention Network (**LF-HAN**) to participate in Track 2.

Their approach adapts the HAT [76] to handle LF imaging with two key modifications: 1) Angular attention is integrated with window and shifted window attention to capture angular dependencies in LF data. 2) Domain-specific channel attention mechanisms are used to disentangle features across spatial, angular, and epipolar dimensions, improving representation learning. Additionally, the network replaces traditional convolutional layers with MBCConv [77] blocks, for enhancing efficiency.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (No. 62401590) and the Humboldt Foundation. We thank the NTIRE 2025 sponsors: ByteDance, Meituan, Kuaishou, and University of Wurzburg (Computer Vision Lab).

5. Teams and Affiliations

Challenge Organizers

Members:

Yingqian Wang¹ (wangyingqian16@nudt.edu.cn),
 Zhengyu Liang¹ (zyliang@nudt.edu.cn),
 Fengyuan Zhang¹ (zhangfengyuan24a@nudt.edu.cn),
 Lvli Tian¹ (tll8023@nudt.edu.cn),
 Longguang Wang² (wanglongguang15@nudt.edu.cn),
 Juncheng Li³ (junchengli@shu.edu.cn),
 Jungang Yang¹ (yangjungang@nudt.edu.cn),
 Radu Timofte⁴ (radu.timofte@uni-wuerzburg.de),
 Yulan Guo⁵ (guoyulan@sysu.edu.cn).

Affiliations:

- ¹National University of Defense Technology
- ²Aviation University of Air Force
- ³Shanghai University
- ⁴Computer Vision Lab, University of Würzburg
- ⁵Sun Yat-sen University

(1) Team OpenMeow - Tracks 1★, 2★, 3★

Members:

Kai Jin¹ (jinkai@bigo.sg), Zeqiang Wei^{2,3}, Angulia Yang¹, Di Wu², Mingzhi Gao¹, Xiuzhuang Zhou^{2,4}

Affiliations:

- ¹Bigo Technology Pte. Ltd.
- ²Beijing University of Posts and Telecommunications
- ³Global Explorer Ltd., Suzhou China
- ⁴Beijing Ketai Industrial Intelligence (BKII), China

(2) Team BITSMBU - Tracks 1★, 2★, 3★

Members:

Yue Yan^{1,2} (3220231312@bit.edu.cn), Yuhao Wang^{1,2}, Shuang Chen^{1,2}, Zeping Tian², Yizhi Hu², Yao Lu^{1,2}

Affiliations:

- ¹Beijing Institute of Technology
- ²Shenzhen SMU-BIT University

(3) Team SmartVIPLab - Tracks 1★, 2

Members:

Haosong Liu¹ (hsliu@stu.hqu.edu.cn), Xiancheng Zhu^{1,2}, Huanqiang Zeng¹, Jianqing Zhu¹, Yifan Shi¹, Junhui Hou³

Affiliations:

- ¹Huaqiao University
- ²Xiamen University of Technology
- ³City University of Hong Kong

(4) Team LFSR-DASE - Track 2★

Members:

Mingyang Yu¹ (yumingyang@stu.ecnu.edu.cn), Zhijian Wu¹, Dingjiang Huang¹

Affiliations:

- ¹East China Normal University

(5) Team BuptMM - Tracks 1, 2, 3★

Members:

Wenli Zheng¹ (joyzheng@bupt.edu.cn), Zekai Xu¹, Huiyuan Fu¹, Heng Zhang², Zhijuan Huang², Hongyuan Yu²

Affiliations:

- ¹Beijing University of Posts and Telecommunications
- ²Multimedia Department, Xiaomi Inc

(6) Team Only My Railgun - Tracks 1, 2

Members:

Zeke Zexi Hu¹ (zexi.hu@sydney.edu.au), Haodong Chen¹, Vera Yuk Ying Chung¹, Xiaoming Chen²

Affiliations:

¹The University of Sydney

²Beijing Technology and Business University

(7) Team NBULFLab- Tracks 1

Members:

Zean Chen¹ (chenzean2024@126.com), Yeyao Chen¹, Gangyi Jiang¹, Haiyong Xu¹, Ting Luo¹, Guanglong Liao¹

Affiliations:

¹Ningbo University

(8) Team icais-AI-team - Track 1

Members:

Danhuo Zhang¹ (zdh0136@gmail.com), Siyu Zhang¹, Wending Mao², Zhongfeng Wang^{1,2}

Affiliations:

¹Nanjing University

²Sun Yat-sen University

(9) Team SpaceVision - Track 1,3

Members:

Sunita Arya¹ (sunita33@sac.isro.gov.in), Abhishek Kumar Sinha¹, S Manthira Moorthi¹

Affiliations:

¹Space Applications Centre, Indian Space Research Organisation, Ahmedabad, India

(10) Team HawkeyeGroup - Track 1

Members:

Hao Zhang¹ (zhang_hao@buaa.edu.cn), Hao Sheng^{1,2,3}, Da Yang¹, Zhenglong Cui¹, Shuai Wang¹

Affiliations:

¹Data Science and Intelligent Computing Laboratory, Hangzhou International Innovation Institute, Beihang University;

²State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University

³Faculty of Applied Sciences, Macao Polytechnic University

(11) Team SZU-VS - Tracks 1, 3

Members:

Haotian Zhang¹ ([\(2310295037@email.szu.edu.cn\)](mailto:(2310295037@email.szu.edu.cn)),

Xingzheng Wang¹, Yuanbo Huang¹, Jiahao Lin¹, Yuhang Lin¹

Affiliations:

¹Shenzhen University

(12) Team CBNU-MIP&VC-Labs - Track 2

Members:

Ahmed Salem¹ (ahmeddiefy@chungbuk.ac.kr), Ebrahim Elkady¹, Hatem Ibrahem², Jae-Won Suh¹, Hyun-Soo Kang¹

Affiliations:

¹Chungbuk National University

²Toronto Metropolitan University

(13) Team IMAG - Track 2

Members:

Changguang Wu¹ (changguangwu@njust.edu.cn), Hao Hou¹, Pengpeng Li¹, Peng Huang¹, Jiangxin Dong¹, Jin-hui Tang¹

Affiliations:

¹Nanjing University of Science and Technology

References

- [1] Vaibhav Vaish, Bennett Wilburn, Neel Joshi, and Marc Levoy. Using plane+ parallax for calibrating dense camera arrays. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2004. [1](#)
- [2] 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](#)
- [3] Chun Zhao and Byeungwoo Jeon. Compact representation of light field data for refocusing and focal stack reconstruction using depth adaptive multi-cnn. *IEEE Transactions on Computational Imaging*, 10:170–180, 2024. [1](#)
- [4] 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](#)
- [5] 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)*, pages 5206–5215, 2022. [1](#)
- [6] Wentao Chao, Xuechun Wang, Yingqian Wang, Guanghui Wang, and Fuqing Duan. Learning sub-pixel disparity distribution for light field depth estimation. *IEEE Transactions on Computational Imaging*, pages 1126–1138, 2023. [1](#)
- [7] Ryan S Overbeck, Daniel Erickson, Daniel Evangelatos, 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](#)

- [8] Jingyi Yu. A light-field journey to virtual reality. *IEEE MultiMedia*, 24(2):104–112, 2017. 1
- [9] Wu G, Liu Y, Fang L, and Chai T. Revisiting light field rendering with deep anti-aliasing neural network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 5430–5444, 2022. 1
- [10] Vincent Sitzmann, Semon Rezhikov, 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
- [11] 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)*, pages 626–636, 2022. 1
- [12] 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)*, pages 19819–19829, 2022. 1
- [13] Yingqian Wang, Longguang Wang, Zhengyu Liang, Jungang Yang, Radu Timofte, Yulan Guo, Kai Jin, Zeqiang Wei, Angulia Yang, Sha Guo, et al. Ntire 2023 challenge on light field image super-resolution: Dataset, methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1320–1335, 2023. 1, 2, 3
- [14] Yingqian Wang, Zhengyu Liang, Qianyu Chen, Longguang Wang, Jungang Yang, Radu Timofte, Yulan Guo, Wentao Chao, Yiming Kan, Xuechun Wang, et al. Ntire 2024 challenge on light field image super-resolution: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 6218–6234, 2024. 1, 3, 7
- [15] Martin Rerabek and Touradj Ebrahimi. New light field image dataset. In *International Conference on Quality of Multimedia Experience (QoMEX)*, 2016. 2, 3
- [16] 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. 2, 3
- [17] 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. 2, 3
- [18] Mikael Le Pendu, Xiaoran Jiang, and Christine Guillenmot. Light field inpainting propagation via low rank matrix completion. *IEEE Transactions on Image Processing*, 27(4):1981–1993, 2018. 2, 3
- [19] Vaibhav Vaish and Andrew Adams. The (new) stanford light field archive. *Computer Graphics Laboratory, Stanford University*, 6(7), 2008. 2, 3
- [20] Florin-Alexandru Vasluianu, Tim Seizinger, Zhuyun Zhou, Zongwei Wu, Radu Timofte, et al. NTIRE 2025 ambient lighting normalization challenge. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [21] Kangning Yang, Jie Cai, Ling Ouyang, Florin-Alexandru Vasluianu, Radu Timofte, Jiaming Ding, Huiming Sun, Lan Fu, Jinlong Li, Chiu Man Ho, Zibo Meng, et al. NTIRE 2025 challenge on single image reflection removal in the wild: Datasets, methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [22] Florin-Alexandru Vasluianu, Tim Seizinger, Zhuyun Zhou, Cailian Chen, Zongwei Wu, Radu Timofte, et al. NTIRE 2025 image shadow removal challenge report. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [23] Lei Sun, Andrea Alfarano, Peiqi Duan, Shaolin Su, Kaiwei Wang, Boxin Shi, Radu Timofte, Danda Pani Paudel, Luc Van Gool, et al. NTIRE 2025 challenge on event-based image deblurring: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [24] Lei Sun, Hang Guo, Bin Ren, Luc Van Gool, Radu Timofte, Yawei Li, et al. The tenth ntire 2025 image denoising challenge report. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [25] Xiaohong Liu, Xiongkuo Min, Qiang Hu, Xiaoyun Zhang, Jie Guo, et al. NTIRE 2025 XGC quality assessment challenge: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [26] Nickolay Safonov, Alexey Bryntsev, Andrey Moskalenko, Dmitry Kulikov, Dmitriy Vatolin, Radu Timofte, et al. NTIRE 2025 challenge on UGC video enhancement: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [27] Egor Ershov, Sergey Korchagin, Alexei Khalin, Artyom Panashin, Arseniy Terekhin, Ekaterina Zaychenkova, Georgiy Lobarev, Vsevolod Plokhotnyuk, Denis Abramov, Elisey Zhdanov, Sofia Dorogova, Yasin Mamedov, Nikola Banic, Georgii Perevozchikov, Radu Timofte, et al. NTIRE 2025 challenge on night photography rendering. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [28] Zheng Chen, Kai Liu, Jue Gong, Jingkai Wang, Lei Sun, Zongwei Wu, Radu Timofte, Yulun Zhang, et al. NTIRE 2025 challenge on image super-resolution (x4): Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [29] Zheng Chen, Jingkai Wang, Kai Liu, Jue Gong, Lei Sun, Zongwei Wu, Radu Timofte, Yulun Zhang, et al. NTIRE 2025 challenge on real-world face restoration: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2

- [30] Bin Ren, Hang Guo, Lei Sun, Zongwei Wu, Radu Timofte, Yawei Li, et al. The tenth NTIRE 2025 efficient super-resolution challenge report. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [31] Pierluigi Zama Ramirez, Fabio Tosi, Luigi Di Stefano, Radu Timofte, Alex Costanzino, Matteo Poggi, Samuele Salti, Stefano Mattoccia, et al. NTIRE 2025 challenge on hr depth from images of specular and transparent surfaces. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [32] Sangmin Lee, Eunpil Park, Angel Canelo, Hyunhee Park, Youngjo Kim, Hyungju Chun, Xin Jin, Chongyi Li, Chun-Le Guo, Radu Timofte, et al. NTIRE 2025 challenge on efficient burst hdr and restoration: Datasets, methods, and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [33] Yuqian Fu, Xingyu Qiu, Bin Ren Yanwei Fu, Radu Timofte, Nicu Sebe, Ming-Hsuan Yang, Luc Van Gool, et al. NTIRE 2025 challenge on cross-domain few-shot object detection: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [34] Xin Li, Kun Yuan, Bingchen Li, Fengbin Guan, Yizhen Shao, Zihao Yu, Xijun Wang, Yiting Lu, Wei Luo, Suhang Yao, Ming Sun, Chao Zhou, Zhibo Chen, Radu Timofte, et al. NTIRE 2025 challenge on short-form ugc video quality assessment and enhancement: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [35] Xin Li, Xijun Wang, Bingchen Li, Kun Yuan, Yizhen Shao, Suhang Yao, Ming Sun, Chao Zhou, Radu Timofte, and Zhibo Chen. NTIRE 2025 challenge on short-form ugc video quality assessment and enhancement: Kwaisr dataset and study. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [36] Shuhao Han, Haotian Fan, Fangyuan Kong, Wenjie Liao, Chunle Guo, Chongyi Li, Radu Timofte, et al. NTIRE 2025 challenge on text to image generation model quality assessment. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [37] Xin Li, Yeying Jin, Xin Jin, Zongwei Wu, Bingchen Li, Yufei Wang, Wenhan Yang, Yu Li, Zhibo Chen, Bihan Wen, Robby Tan, Radu Timofte, et al. NTIRE 2025 challenge on day and night raindrop removal for dual-focused images: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [38] Varun Jain, Zongwei Wu, Quan Zou, Louis Florentin, Henrik Turbell, Sandeep Siddhartha, Radu Timofte, et al. NTIRE 2025 challenge on video quality enhancement for video conferencing: Datasets, methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [39] Xiaoning Liu, Zongwei Wu, Florin-Alexandru Vasluiianu, Hailong Yan, Bin Ren, Yulun Zhang, Shuhang Gu, Le Zhang, Ce Zhu, Radu Timofte, et al. NTIRE 2025 challenge on low light image enhancement: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [40] Yingqian Wang, Zhengyu Liang, Fengyuan Zhang, Lvli Tian, Longguang Wang, Juncheng Li, Jungang Yang, Radu Timofte, Yulan Guo, et al. NTIRE 2025 challenge on light field image super-resolution: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [41] Jie Liang, Radu Timofte, Qiaosi Yi, Zhengqiang Zhang, Shuaizheng Liu, Lingchen Sun, Rongyuan Wu, Xindong Zhang, Hui Zeng, Lei Zhang, et al. NTIRE 2025 the 2nd restore any image model (RAIM) in the wild challenge. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [42] Marcos Conde, Radu Timofte, et al. NTIRE 2025 challenge on raw image restoration and super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [43] Marcos Conde, Radu Timofte, et al. Raw image reconstruction from RGB on smartphones. NTIRE 2025 challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 2
- [44] 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
- [45] Reuben A Farrugia, Christian Galea, and Christine Guillenmot. 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
- [46] 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
- [47] Mattia Rossi and Pascal Frossard. Graph-based light field super-resolution. In *International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6, 2017. 2
- [48] Youngjin Yoon, Hae-Gon Jeon, Donggeun Yoo, Joon-Young Lee, and In So Kweon. Light-field image super-resolution using convolutional neural network. *IEEE Signal Processing Letters*, 24(6):848–852, 2017. 2
- [49] 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
- [50] 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

- [51] Nan Meng, Hayden K.-H. So, Xing Sun, and Edmund Y. Lam and. High-dimensional dense residual convolutional neural network for light field reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 873–886, 2021. 2
- [52] 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
- [53] 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)*, pages 290–308, 2020. 2
- [54] 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*, pages 425–443, 2023. 2, 3, 4, 10, 12
- [55] Gaosheng Liu, Huanjing Yue, Jiamin Wu, and Jingyu Yang. Intra-inter view interaction network for light field image super-resolution. *IEEE Transactions on Multimedia*, pages 256–266, 2023. 2
- [56] Shunzhou Wang, Tianfei Zhou, Yao Lu, and Huijun Di. Detail-preserving transformer for light field image super-resolution. In *AAAI Conference on Artificial Intelligence (AAAI)*, pages 10500–10507, 2022. 2
- [57] Zhengyu Liang, Yingqian Wang, Longguang Wang, Jungang Yang, and Shilin Zhou. Light field image super-resolution with transformers. *IEEE Signal Processing Letters*, pages 563–567, 2022. 2, 10, 12
- [58] Zhengyu Liang, Yingqian Wang, Longguang Wang, Jungang Yang, Shilin Zhou, and Yulan Guo. Learning non-local spatial-angular correlation for light field image super-resolution. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12376–12386, 2023. 3, 4, 8, 10, 12, 14
- [59] Ruixuan Cong, Hao Sheng, Da Yang, Zhenglong Cui, and Rongshan Chen. Exploiting spatial and angular correlations with deep efficient transformers for light field image super-resolution. *IEEE Transactions on Multimedia*, pages 1421–1435, 2023. 3, 6, 12
- [60] 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)*, pages 1373–1383, 2023. 3, 4, 6, 8, 9, 12
- [61] Yao Lu, Shunzhou Wang, Ziqi Wang, Peiqi Xia, Tianfei Zhou, et al. Lfmamba: light field image super-resolution with state space model. *arXiv preprint arXiv:2406.12463*, 2024. 3
- [62] Wentao Chao, Yiming Kan, Xuechun Wang, Fuqing Duan, and Guanghui Wang. Bigepit: Scaling epit for light field image super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 6187–6197, 2024. 4, 5, 8, 9, 12
- [63] Kai Jin, Zeqiang Wei, Angulia Yang, Di Wu, , Mingzhi Gao, and Xiuzhuang Zhou. Lftransmamba: A hybrid mamba-transformer model for light field image super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 5, 8
- [64] Zeqiang Wei, Kai Jin, Zeyi Hou, Kuan Song, and Xiuzhuang Zhou. L^2 fmamba: Lightweight light field image super-resolution with state space model. *arXiv preprint arXiv:2503.19253*, 2025. 5
- [65] Jinglei Shi, Xiaoran Jiang, and Christine Guillemot. A framework for learning depth from a flexible subset of dense and sparse light field views. *IEEE Transactions on Image Processing*, 28(12):5867–5880, 2019. 6
- [66] Ziqi Wang, Yao Lu, Shunzhou Wang, Wang Xia, Peiqi Xia, and Wenjing Wang. Trident transformer for light field image super-resolution. In *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2024. 7
- [67] Yao Lu, Shunzhou Wang, Ziqi Wang, Peiqi Xia, Tianfei Zhou, et al. Lfmamba: light field image super-resolution with state space model. *arXiv preprint arXiv:2406.12463*, 2024. 7, 12
- [68] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv:2312.00752*, 2023. 8
- [69] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning and Representation (ICLR)*, 2015. 8
- [70] Dingjiang Huang Mingyang Yu, Zhijian Wu. Lfmix: A lightweight hybrid architecture model for light field super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2025. 9
- [71] Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, and Yunfan Liu. Vmamba: Visual state space model. *arXiv preprint arXiv:2401.10166*, 2024. 9
- [72] Gaosheng Liu, Huanjing Yue, and Jingyu Yang. Efficient light field image super-resolution via progressive disentangling. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6277–6286, 2024. 10
- [73] Zeke Zexi Hu, Xiaoming Chen, Vera Yuk Ying Chung, and Yiran Shen. Beyond subspace isolation: Many-to-many transformer for light field image super-resolution. *IEEE Transactions on Multimedia*, 27:1334–1348, 2025. 10
- [74] Zexi Hu, Xiaoming Chen, Henry Wing Fung Yeung, Yuk Ying Chung, and Zhibo Chen. Texture-Enhanced Light Field Super-Resolution With Spatio-Angular Decomposition Kernels. *IEEE Transactions on Instrumentation and Measurement*, 71:1–16, 2022. 11
- [75] Manchang Jin, Gaosheng Liu, Kunshu Hu, Xin Luo, Kun Li, and Jingyu Yang. Physics-informed ensemble representation for light-field image super-resolution. *arXiv preprint arXiv:2305.20006*, 2023. 14
- [76] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. *IEEE/CVF Conference on Computer*

Vision and Pattern Recognition (CVPR), pages 5769–5778,
2023. 15

- [77] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. pages 1780–1789, 2017. 15