Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Arbitrary-Scale Light Field Super-Resolution via Spatio-Angular Consistent Neural Epipolar Geometry Learning

Lili Lin
Zhejiang Gongshang University
Wenyu Li
Zhejiang Gongshang University
Hao Zheng
Zhejiang Gongshang University
Qiming Fang
18200968@qq.com

Hangzhou Dianzi University
Jiahan Meng
Hangzhou Dianzi University
Wenhui Zhou
Hangzhou Dianzi University

Research Article

Keywords: Light Field, Super-resolution, Arbitrary Scale, Neural Epipolar Geometry, Spatio-Angular Consistency

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11144103/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Arbitrary-Scale Light Field Super-Resolution via Spatio-Angular Consistent Neural Epipolar Geometry Learning

Lili Lin[1], Wenyu Li[1], Hao Zheng[1], Qiming Fang[2], Jiahan Meng[2], Wenhui Zhou[2]

[1]School of Information and Electronic Engineering, Zhejiang Gongshang University, Hangzhou, 310018, Zhejiang, China.
[2]School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, 310018, Zhejiang, China.

Abstract

Light field imaging, a promising technology in computer vision, captures both spatial and angular information of light rays. However, the inherent trade-off between spatial and angular resolution limits its applications. This paper introduces a novel spatio-angular consistent local neural epipolar geometry learning framework for arbitrary-scale light field super-resolution. By modelling light field epipolar plane images (EPIs) as continuous local implicit geometry functions, our approach enables simultaneous super-resolution across both angular and spatial domains. Experimental results on synthetic and real-world datasets demonstrate the effectiveness of our model, achieving high geometric structure fidelity under arbitrary upscaling factors.

Keywords: Light Field, Super-resolution, Arbitrary Scale, Neural Epipolar Geometry, Spatio-Angular Consistency

1 Introduction

In contrast to conventional cameras that integrate light rays over the angular domain, light field cameras (also known as plenoptic cameras) [1–3] preserve the angular information of light rays by inserting a microlens array (MLA) between the image sensor and the main lens. This mechanism captures rich visual information of 3D scenes,

enabling multi-view disparity, perspective shifts, and digital refocusing. However, to record the 4D light field distribution, lenslet-array-based light field imaging systems trade spatial resolution for angular resolution [4] by multiplexing pixels on a 2D image sensor. This inherent spatio-angular trade-off hinders light field imaging from meeting the high-resolution requirements of downstream visual applications, such as near-eye displays, 3D rendering, and scene analysis.

To address these physical trade-off limitations, light field super-resolution (SR) has attracted significant attention in recent years [5–8], with Transformer-based methods such as ViewFormer [9] showing impressive reconstruction performance. However, existing approaches are typically designed and trained with predefined, fixed upsampling scales in either the spatial or angular domain, categorized as *light field spatial super-resolution* or *light field angular super-resolution*, respectively. Moreover, the spatial SR models can only super-resolve to fixed spatial resolutions (e.g., $\times 2$ or $\times 4$), while angular SR models synthesize specific densely-sampled sets of sub-aperture images (SAIs) from a sparse set (e.g., $2 \times 2 \rightarrow 5 \times 5$ or 7×7).

More recently, implicit neural representations (INRs) and neural radiance fields (NeRFs) have provided a promising avenue for light field SR, offering a novel perspective on continuous signal sampling and reconstruction. However, INRs and NeRFs are generally applied to large-baseline light fields. With existing INR/NeRF methodologies, it remains challenging to achieve satisfactory performance on narrow-baseline 4D light fields [10, 11], particularly those captured by lenslet-array-based light field cameras. A primary bottleneck is that the volumetric rendering quality of INRs/NeRFs heavily relies on ground-truth view poses (or highly accurate estimates, e.g., via COLMAP), which are generally unavailable in lenslet-array-based imaging due to the ultra-small baselines of microlens arrays [12, 13]. Furthermore, both volumetric representations and implicit image functions struggle to effectively model the complex spatio-angular correlations inherent to light fields [7, 11], including the geometric consistency and constraints across sub-aperture images (SAIs).

Accordingly, this paper aims to establish an effective continuous neural representation with spatio-angular consistency for lenslet-array-based light field imaging, thereby achieving flexible and arbitrary-scale joint angular and spatial super-resolution. Its core idea leverages the well-known property that projections of corresponding pixels in light field SAIs onto epipolar planes exhibit distinctive linear structures related to scene geometry. That is, geometric cues and constraints of the underlying scene are intrinsically embedded in the sharp, structured, and continuous linear patterns of light field epipolar plane images (EPIs). Thus, a intuitive idea is to learn continuous neural implicit functions from light field EPIs to capture spatio-angular correspondence at arbitrary spatial and angular resolutions. However, as a 2D spatio-angular slice of 4D light field obtained by fixing one spatial and one angular coordinate, each EPI contains only local spatio-angular information. Consequently, the primary challenge lies in preserving global spatio-angular consistency during local neural learning, as this consistency is not only implicit within individual EPIs but also manifested through strong correlations across different EPIs.

To achieve spatio-angular consistent local neural epipolar geometry learning, this paper first designs a spatio-angular aware geometry encoder to explicitly extract

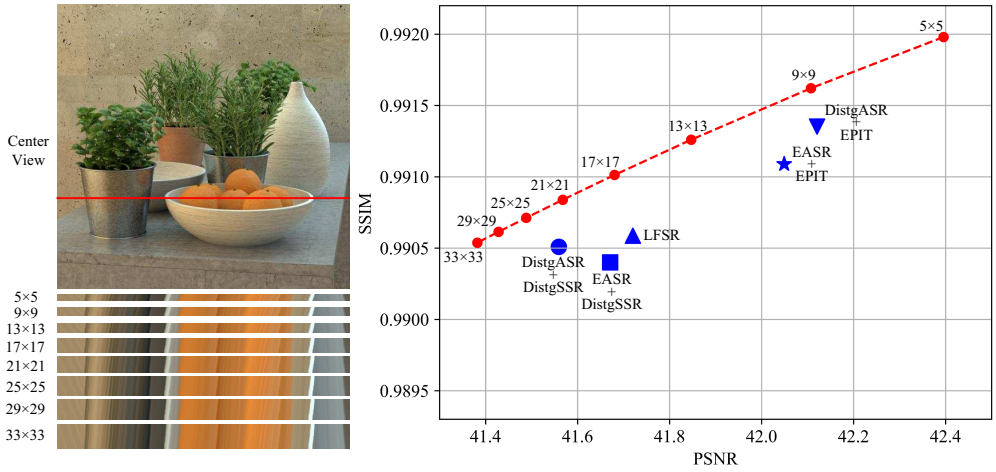


Fig. 1 Left: Visual results of our model at different angular resolutions alongside $\times 2$ spatial super-resolution. The EPIs corresponding to the red lines in the Center View are upsampled to an appropriate scale for better visualization. **Right:** Performance comparisons (SSIM/PSNR) of different methods over "HCIOld" dataset for angular super-resolution from 2×2 to 5×5 with $\times 2$ spatial super-resolution. Note that our model supports arbitrary-scale angular super-resolution, and its SSIM and PSNR scores across various angular resolution (ranging from 5×5 to 33×33) are marked by solid red dots. Furthermore, our performance remains stable without significantly drops as spatial and angular resolutions increase, demonstrating that our model maintains satisfactory geometric structure fidelity for arbitrary-scale joint angular and spatial super-resolution.

latent codes of light field epipolar geometry. Next, neural implicit learning is employed to generate continuous epipolar geometry representations from these latent codes. Finally, an expansion rendering module reconstructs the light field image with both high spatial resolution and dense angular sampling. Experimental results demonstrate that our model achieves arbitrary-scale light field super-resolution while maintaining satisfactory geometric structure fidelity, as illustrated in Fig. 1.

Our main contributions are the following:

(1) A spatio-angular consistent local neural epipolar geometry learning framework is proposed for arbitrary-scale light field super-resolution. It learns continuous implicit neural representations of light field epipolar geometry by mapping EPIs into an intermediate, interpolable latent space.

(2) A spatio-angular aware geometry encoder is designed to extract latent geometry codes from a light field. It explicitly integrates both local and global dependencies of light field epipolar geometry, embedding spatio-angular correlations into the latent geometry codes via a spatio-angular aware Transformer.

(3) A divide-and-conquer local neural epipolar geometry learning strategy is presented. Directly learning a full 4D implicit function is challenging and computationally inefficient, leading to a large memory footprints. Instead, we decompose the 4D implicit representation learning into the joint optimization of two weight-shared 2D implicit functions for horizontal EPIs (LIGF_h) and vertical EPIs (LIGF_v), respectively.

2 Related Work

2.1 Light field super-resolution

Light field super-resolution can be generally categorized into light field angular super-resolution (LFASR) and spatial super-resolution (LFSSR).

LFASR methods typically includes SAI-based and EPI-based approaches. SAI-based methods take a sparse set of SAIs as input, utilizing deep neural network to extract spatio-angular features to synthesize novel SAIs at desired angular positions [5, 14–21]. Although these methods have achieved state-of-the-art performance, they are constrained to reconstructing light fields with fixed angular resolution, where model complexity is tightly coupled with the target resolution. On the other hand, EPI-based LFASR methods treat view synthesis as the continuous reconstruction or interpolation of EPI lines [5, 7, 22–27], with the main idea being to extract and learn the linear structural patterns inherent to EPIs. The primary challenge for EPI-based approaches stems from the difficulty of modeling complex spatio-angular correlation, as spatial and angular information are highly intertwined within 4D light fields.

Meanwhile, for LFSSR, convolutional neural networks (CNNs) have shown strong performance in LFSSR tasks [5, 14, 16, 17, 19, 28, 28, 29]. However, due to the limited receptive field of CNNs, these methods generally struggle to capture long-range correspondences across SAIs, focusing instead on local feature representations. To address this, Liang et al. [21] designed the Angular and Spatial Transformer to aggregate information from all angular views while capturing long-range spatial dependencies within each SAI. Furthermore, in its extended work, EPIT [7] models the dependencies between all pixel pairs on an EPI, effectively learning joint spatio-angular correlations. Nevertheless, much like angular super-resolution task, previous spatial super-resolution methods remain restricted to pre-defined, fixed upsampling scales.

Most recently, several studies have focused on joint light field angular and spatial super-resolution, aiming to simultaneously enhance both spatial and angular resolution. Meng et al. [30] introduced high-order convolution to achieve spatial and angular super-resolution across multiple scales. Chen et al. [31] utilized a cascaded CNN architecture within a two-stage learning framework to progressively recover high-resolution light field images. Wang et al. [32] employed separate Transformer cascades for spatial and angular feature extraction, reshaped the features into epipolar plane images (EPIs), and introduced an epipolar Swin Transformer to capture spatio-angular correlations. Additionally, Liu et al. [33] proposed an adaptive aggregation strategy to select and combine advantageous pixels from input light field images across different sub-apertures.

2.2 Implicit neural representation

Recent studies have shown that INR offer a promising alternative for view synthesis and scene rendering. The core concept behind INR is to parameterize a continuous, differentiable 2D/3D signal (such as an image or a 3D scene) using multi-layer perceptrons (MLPs) [34]. By leveraging the continuous representation capabilities of MLPs,

INR-based methods exhibit impressive neural rendering and reconstruction performance across various tasks. These include continuous image super-resolution [35–37], signed distance function (SDF) for 3D structure modeling [38–40], and neural radiance fields (NeRF) for novel view synthesis [41–43]. 0 To reduce the computational complexity and memory footprint of neural rendering, neural 4D light field representations have been proposed [11, 44, 45]. These methods parameterize each light ray in 4D coordinates and map them to its corresponding color values. However, most of existing methods are tailored for wide-baseline light fields captured by sparse camera arrays, rather than lenslet-array-based light fields. Although Attal et al.[11] proposed a ray-space embedding network to enable small-baseline view synthesis without explicit geometric constraints, it relies on a dense input of 5×5 views for high-quality rendering; sparser input sequences (e.g., 2×2) fail to interpolate effectively. Furthermore, this method is strictly applicable to camera-array-based light fields, such as the Stanford Light Field dataset [46]. Note that to handle the lack of poses, they have to hand-design different scene coordinates (including camera coordinates and object plane coordinates) and scene depth ranges as initial two-plane ray parameterizations for each scene, as described in the Appendix of [11].

Most recently, several promising attempts have been made toward the implicit representation of lenslet-array-based light fields [47–49]. However, maintaining spatio-angular consistency during neural implicit learning remains a significant challenge.

3 Methodology

The overall architecture of our network is depicted in Fig. 2. It takes a sparsely-sampled low-resolution light field $\mathcal{L}(u, v, x, y) \in \mathbb{R}^{U \times V \times W \times H \times c}$ as input, and outputs a densely-sampled high-resolution light field $\mathcal{L}' \in \mathbb{R}^{U' \times V' \times W' \times H' \times c}$, where (W, H) and (W', H') represent the spatial resolutions, while (U, V) and (U', V') represent the angular resolutions. Our network consists of three primary modules: a spatio-angular aware geometry encoder, a local neural epipolar geometry learning module, and an expansion rendering module.

3.1 Spatio-angular Aware Geometry Encoder

The primary objective of the spatio-angular aware geometry encoder is to map light field EPIs into an intermediate, interpolable latent space. Since spatial and angular information in a light field are highly intertwined, it is essential to integrate both the local and global geometric dependencies as well as the spatio-angular correlations. In our approach, the former is achieved by designing a hybrid CNN-Transformer network that explicitly extracts geometric and spatio-angular features. The latter is accomplished by leveraging an attention mechanism to model interactions between EPI feature tokens and spatio-angular feature tokens, thereby effectively capturing their complementary relationships.

Our geometry encoder consists of two main components: a *spatial-angular-EPI (SAE) feature extraction* module, and a *spatio-angular aware Transformer* module. It decomposes the 4D light field space into a set of 2D subspaces. Specifically, the SAE module extracts spatio-angular features on the (W, H) and (U, V) subplanes, as well

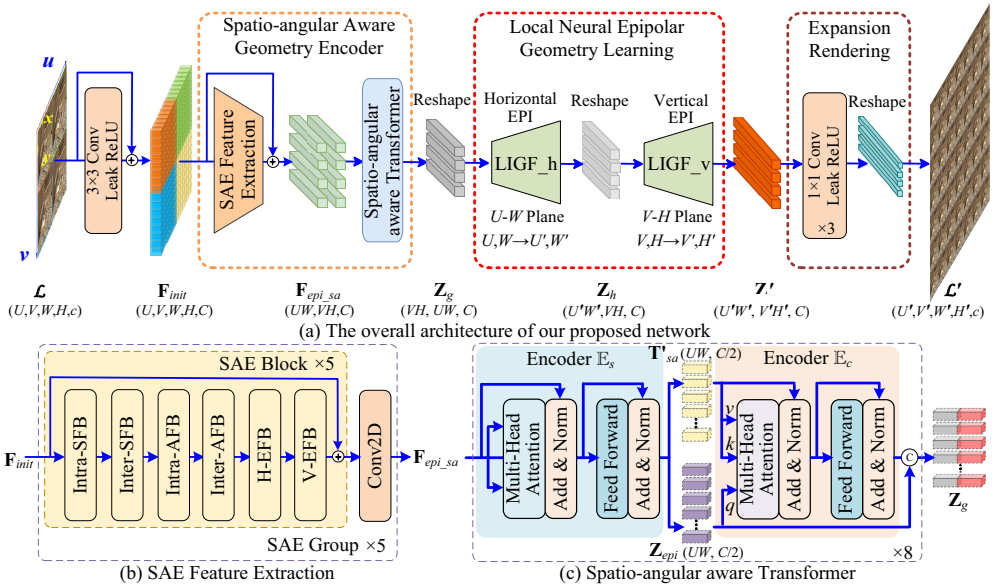


Fig. 2 The overall architecture of our proposed network.

as EPI geometric features on the (U, X) and (V, Y) subplanes. First, initial features $\mathbf{F}_{init}$ are extracted from input $\mathcal{L}$ using a convolutional layer with 3×3 kernels and C output channels. Next, $\mathbf{F}_{init}$ is fed into the SAE module to generate the feature $\mathbf{F}_{epi_sa}$, which serves as an aggregation of EPI and spatio-angular features. Finally, the spatio-angular aware Transformer module takes $\mathbf{F}_{epi_sa}$ as input to perform global dependency modeling and attention-based feature fusion.

3.1.1 The SAE Module

As depicted in Fig. 2(b), $\mathbf{F}_{init}$ is fed into 5 cascaded SAE groups, each consisting of 5 SAE blocks. As the fundamental structure, each SAE block cascades 6 distinct convolutional modules, which are categorized into three types: inter-intra spatial (intra-SFB and inter-SFB), inter-intra angular (intra-AFB and inter-AFB), and horizontal-vertical EPI convolutional modules (H-EFB and V-EFB). Adopted from [50], these six convolutional modules are sequentially stacked within each SAE block in the order of spatial → angular → EPI modules. Furthermore, a local residual connection adds the input of each module directly to its output before passing it to the subsequent module.

The cascaded inter-intra spatial and angular convolutional modules are designed to extract and fuse spatial and angular light field features. We first perform intra-spatial convolution (intra-SFB) on pixels within the same view using a 3×3 kernel with a stride of 1 and a dilation of A (denoting the angular resolution). Zero padding is applied to maintain equal input and output dimensions. To capture spatial correlations across neighboring views, we employ an inter-spatial convolution (inter-SFB) module

featuring 3×3 convolutional layers with a stride of 1 and a dilation of 1. The subsequent angular feature extraction component comprises an intra-angular (intra-AFB) and an inter-angular (inter-AFB) modules. The intra-angular module utilizes a kernel of size $A \times A$ with a stride of A and zero padding, operating across pixels sharing identical spatial coordinates. The inter-angular module employs a $2A \times 2A$ convolution layer with a stride of $2A$ to extract angular correlations across 2×2 neighboring MacPIs. Both angular modules use a 2D pixel-shuffling layer to ensure dimension alignment between input and output.

The EPI convolutional module consists of a cascade of vertical and horizontal components, designed to explicitly extract epipolar geometry features along the horizontal and vertical epipolar lines. The kernel sizes for the horizontal and vertical modules are set to $1 \times (2A+1)$ and $(2A+1) \times 1$, respectively. To facilitate training, extensive residual structures are integrated throughout the architecture. Specifically, residual connections are added not only between individual convolutional module within each SAE block, but also across multiple blocks and groups. This residual structure preserves direct information flow and enables effective gradient propagation across deep networks, thereby mitigating the vanishing gradient problem.

3.1.2 The Spatio-angular Aware Transformer Module

Constrained by the local receptive fields of convolutional kernels, the SAE feature extraction module lacks sensitivity to global long-range dependencies across light field spatio-angular and geometric features that exhibit strong overall correlation. Inspired by Transformer-based light field spatial super-resolution methods [7, 21], we address this limitation by proposing a spatio-angular aware Transformer module, as illustrated in Fig. 2(c). Our approach is distinguished from previous approaches in that each Transformer block integrates two dedicated encoders $\mathbb{E}_s$ and $\mathbb{E}_c$. The encoder $\mathbb{E}_s$ is a standard Transformer encoder with a self-attention mechanism, which captures the global dependencies of input tokens. The encoder $\mathbb{E}_c$ is a cross-attention encoder which aims to retain the geometrically relevant spatio-angular characteristics while marginalizing out the irrelevant and trivial detail features. To further enhance the global perception ability, our spatio-angular aware Transformer module cascades 8 such Transformer blocks.

Specifically, we first treat the VH dimension of $\mathbf{F}_{epi_sa}$ as the batch dimension. This dimensional permutation folds VH into the batch axis to form UW composite tokens $\mathbf{T}_{epi_sa}$ of length C, which serve as the input to $\mathbb{E}_s$. By doing so, the self-attention mechanism captures long-range dependencies along the UW composite axis without constructing a computationally expensive 4D attention map, thereby significantly reducing computational overhead while maintaining global context awareness. The output of $\mathbb{E}_s$ is subsequently split along the channel dimension back into the latent EPI codes (tokens) $\mathbf{Z}_{epi}$ and the enhanced spatio-angular tokens $\mathbf{T}'_{sa}$. In the encoder $\mathbb{E}_c$, to extract geometrically relevant features from $\mathbf{T}'_{sa}$. $\mathbf{Z}_{epi}$ serves as the query q, while $\mathbf{T}'_{sa}$ serves as the key k and value v. In this design, $\mathbf{Z}_{epi}$ actively retrieves geometrically relevant spatio-angular cues from $\mathbf{T}'_{sa}$, ensuring that the feature representations are guided by epipolar geometry constraints while suppressing interference from irrelevant texture details. The encoder $\mathbb{E}_c$ outputs the geometrically meaningful

latent spatio-angular codes, which are then concatenated with $\mathbf{Z}_{epi}$ along the channel dimension to form the final latent geometry codes $\mathbf{Z}_g$.

3.2 Local Neural Epipolar Geometry Learning

Simultaneously super-resolving angular and spatial resolutions from $U \times V$, $W \times H$ to $U' \times V'$, $W' \times H'$, respectively, could potentially be achieved by learning a 4D implicit function for a continuous light field representation. However, such an approach would be impractical because it hampers learning effectiveness, requires significant computational resources, and converges slowly.

Instead, we propose a divide-and-conquer learning strategy which converts the 4D implicit function learning into the cascade learning of two 2D local implicit geometry functions (LIGFs) that are obtained by separately fixing (u, x) and (v, y). Specifically, LIGF_h is responsible for joint continuous up-sampling of the angular dimension u and the horizontal spatial dimension x on the horizontal epipolar plane, while LIGF_v subsequently performs joint up-sampling of the angular dimension v and the vertical spatial dimension y. This decomposition strategy reduces the intractable 4D up-sampling problem into two manageable 2D continuous up-sampling steps.

The 2D local implicit geometry function g_θ is defined on the latent space representation $\mathbf{Z}_g$, and shared by all light field EPIs. We use an MLP with trainable parameters θ. The continuous representation of a light field EPI is formulated as follows:

$$\mathbf{I}_{epi}\left(u_q\right) = g_\theta\left(z^*, u_q - u^*\right), \tag{1}$$

where z^* is the nearest $\mathbf{Z}_g$ to the query coordinate u_q, and u^* is the coordinate of the nearest latent code.

Note that we use feature unfolding and local ensemble technologies [35] to enrich local information by concatenating the 3×3 neighboring latent codes; the latter is used to ensure smooth transitions between the nearest latent codes.

The structure of our local neural epipolar geometry learning module consists a cascade structure comprising a LIGF_h module and a LIGF_v module. Both modules are 2-layer MLPs with ReLU activation and hidden dimensions of 256. In our experiments, the LIGF_h and LIGF_v modules share their weights.

The LIGF_h module takes the high-resolution horizontal coordinates U', high-resolution image width W' and the latent geometry codes $\mathbf{Z}_g$ as input, with the VH dimension of $\mathbf{Z}_g$ serving as the batch dimension. The LIGF_h module outputs a simultaneous horizontal and vertical enhanced latent geometry codes $\mathbf{Z}_h \in \mathbb{R}^{VH \times U'W' \times C}$. Similarly, the LIGF_v module takes the high-resolution vertical coordinates V', high-resolution image height and $\mathbf{Z}_h$ as input, with the $U'W'$ dimension of $\mathbf{Z}_h$ serving as the batch dimension. The module outputs high-resolution latent geometry codes $\mathbf{Z}' \in \mathbb{R}^{U'W' \times V'H' \times C}$.

Note that, the order of LIGF_h and LIGF_v has no impact on performance because our data augmentation includes $x-y$ flipping.

3.3 The Expansion Rendering Module

The expansion rendering module uses a stack of three convolutional layers with kernel size of 1×1. It progressively squeezes the channel number of $\mathbf{Z}'$ to c to generate a macro-pixel image with high resolution $\mathbf{I} \in \mathbb{R}^{U'W' \times V'H' \times c}$, which can be easily reshaped into the SAI array representation $\mathcal{L}' \in \mathbb{R}^{U' \times V' \times W' \times H' \times c}$.

4 Experimental Results

4.1 Datasets

Following [5], our comparative experiments are carried out on four real-world light field datasets (30Scenes [51], Occlusion [52], Reflective [52] and EPFL [53]) as well as two synthetic light field datasets (HCINew [54, 55] and HCIOld [56]). The split of training and testing sets is identical with that in [5].

In our experiments, we compare our approach with two LFASSR methods (SASR-Net [33] and DSNet [57]), and several two-stage combined LFASR and LFSSR methods. For the first stage, we employ two recent state-of-the-art LFASR methods, LF-EASR [20] and DistgASR [5], and feed their reconstructed outputs into eleven LFSSR methods, including Bicubic interpolation, LF-SAA[58], LF-DFNet[59], LF-InterNet[29], LF-IINet[60], LFT[21], DPT[61], DistgSSR [5], EPIT [7], LFTrans-Mamba [62], OHT-B [63], respectively. For a fair comparison, all aforementioned methods are retrained and evaluated on the same benchmark dataset.

Note that all of the existing methods can only achieve fixed-scale light field angular and spatial super-resolution.

4.2 Implementation Details

Our experiments were conducted on a server with an Intel Xeon E5-2680 v4 CPU, 64G DDR4 memory, and two NVIDIA RTX 3090 GPUs. Our network was trained with an $L1$ loss, and optimized by the Adaptive Moment (Adam) optimizer ($\beta_1 = 0.9$ and $\beta_2 = 0.999$) with a mini-batch size of 4 and an initial learning rate of 2×10^{-4}. The learning rate decreased by half every 15 epochs. Our model generally converges after training for around 80 epochs.

To enable the model to handle arbitrary resolutions, we introduce a variable during training to randomly scale the input size while keeping the spatial and angular resolution of the output image fixed. Specifically, the spatial resolution of each output patch is fixed at 64×64, while the input resolution is randomly selected from the range $[16, 64)$, corresponding to a maximum upscaling factor of 4.

The peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) are used as quantitative evaluation metrics on the Y channel, as is customary in existing approaches. Our method can also be trained on color space and outputs color results with similar performance.

To verify the superiority of the proposed network, we evaluate the performance of our model in the fixed-scale and arbitrary-scale LFASSR tasks, respectively. Considering the combination of spatial and angular super-resolution, we trained the first

Table 1 Quantitative comparisons (PSNR/SSIM) of our model with the state-of-the-art ones under the $SR^{A(2\times2\rightarrow5\times5),S(\times2)}$ task. The best and second-best results across all methods are marked in bold or underlined, respectively.

ASR Method	SSR Method	30Scenes	Occlusions	Reflective	HCINew	HCIOld	EPFL
LF-EASR [20]	Bicubic	25.35/0.8205	20.80/0.6598	23.16/0.8065	32.01/0.9385	38.11/0.9812	29.96/0.9454
	LF-SAA[58]	25.37/0.8200	20.88/0.6738	23.15/0.8071	31.87/0.9353	38.18/0.9812	30.33/0.9447
	LF-DFNet[59]	25.67/0.8180	21.01/0.6736	23.34/0.8065	34.46/0.9590	41.56/0.9901	33.20/0.9687
	LF-InterNet[29]	38.16/0.9824	34.79/0.9703	35.53/0.9677	30.05/0.9454	38.18/0.9813	32.09/0.9394
	LF-IINet[60]	40.72/0.9759	36.85/0.9659	37.15/0.9648	33.49/0.9160	40.24/0.9773	32.66/0.9411
	LFT[21]	40.79/0.9763	36.89/0.9663	37.16/0.9650	33.58/0.9174	40.33/0.9779	32.71/0.9418
	DPT[61]	40.26/0.9882	36.71/0.9797	37.66/0.9729	32.95/0.9437	39.88/0.9854	31.56/0.9581
	DistgSSR [5]	41.93/**0.9920**	38.36/0.9854	38.94/0.9777	34.74/0.9617	41.56/0.9905	32.90/0.9695
	EPIT [7]	41.85/0.9918	38.27/0.9851	38.89/0.9768	35.21/**0.9653**	42.12/0.9914	33.21/0.9703
	LFTransMamba [62]	41.93/0.9915	38.34/0.9850	39.04/0.9767	34.98/0.9628	42.34/0.9918	33.24/0.9702
	OHT-B [63]	41.80/0.9913	38.42/0.9851	39.02/0.9770	35.01/0.9631	42.25/0.9914	33.36/0.9705
DistgASR [5]	Bicubic	25.25/0.8198	20.71/0.6274	23.14/0.8067	31.95/0.9376	38.03/0.9812	29.77/0.9455
	LF-SAA[58]	25.47/0.8195	20.77/0.6257	23.23/0.8052	33.82/0.9546	41.11/0.9893	32.84/0.9665
	LF-DFNet[59]	41.08/0.9905	37.45/0.9829	38.24/0.9748	33.70/0.9520	40.93/0.9888	32.23/0.9665
	LF-InterNet[29]	41.89/0.9919	38.32/0.9855	38.92/0.9772	34.81/0.9623	41.84/0.9910	33.10/0.9701
	LF-IINet[60]	41.74/0.9917	38.12/0.9849	38.81/0.9771	34.15/0.9576	41.39/0.9906	33.05/0.9693
	LFT[21]	40.24/0.9888	36.79/0.9808	37.58/0.9723	33.61/0.9523	40.68/0.9886	32.55/0.9663
	DPT[61]	41.78/0.9918	38.21/0.9852	38.89/0.9773	33.53/0.9499	40.86/0.9884	32.39/0.9657
	DistgSSR [5]	41.86/0.9919	38.21/0.9851	38.94/0.9776	34.66/0.9612	41.67/0.9904	33.17/0.9697
	EPIT [7]	41.80/0.9917	38.15/0.9849	38.93/0.9774	35.06/0.9644	42.05/0.9911	33.32/**0.9718**
	LFTransMamba [62]	41.95/**0.9920**	38.41/0.9857	39.00/0.9766	35.06/0.9634	42.26/0.9916	33.39/0.9711
	OHT-B [63]	41.85/0.9917	**38.51**/**0.9859**	39.08/0.9778	35.08/0.9636	42.09/0.9908	33.21/0.9699
SASR-Net [33]		41.89/0.9919	38.30/0.9853	38.86/0.9772	34.46/0.9599	41.72/0.9907	32.98/0.9692
DSNet [57]		41.92/0.9815	38.23/0.9726	39.09/0.9675	**35.69**/0.9388	39.25/0.9665	32.91/0.9425
Ours		**41.96**/**0.9920**	38.45/0.9857	**39.12**/**0.9786**	35.12/0.9639	**42.40**/**0.9920**	**33.51**/0.9711

three real-world LF datasets and the left three LF datasets separately and obtained two models.

4.3 Fixed-scale Light Field Super-resolution

In the fixed-scale super-resolution ta, we evaluate the synthesis quality of 5×5 light field images reconstructed from 2×2 sparsely-sampled ones, while the spatial scale is enlarged by a factor of 2, $\beta = 2$. The task is named $SR^{A(2\times2\rightarrow5\times5),S(\times2)}$ task. Table 1 lists the comparison results with each method. Meanwhile, Fig. 3 provides some visual comparisons of our approach with other methods. Clearly, our model shows better performance in the representation of pinstripes than other models. Evidently, "Ours" model achieves the highest PSNR and SSIM values on "30Scenes", "Occlusion", "Reflective", "HCIOld" and "EPFL". These results demonstrate that our method possesses good geometric structure fidelity which helps to preserve the linear structure of EPI clearly.

4.4 Arbitrary-scale Light Field Super-resolution

The most significant capability of our method is its ability to construct arbitrary-scale LFASSR. However, it is difficult to evaluate the performance of arbitrary-scale reconstruction due to the lack of ground truth with high angular resolution in the light field datasets.

So we use the available 5×5 ground truth to assess the performance of the reconstructed light field with higher angular resolution. For example, "Ours (9×9)" can

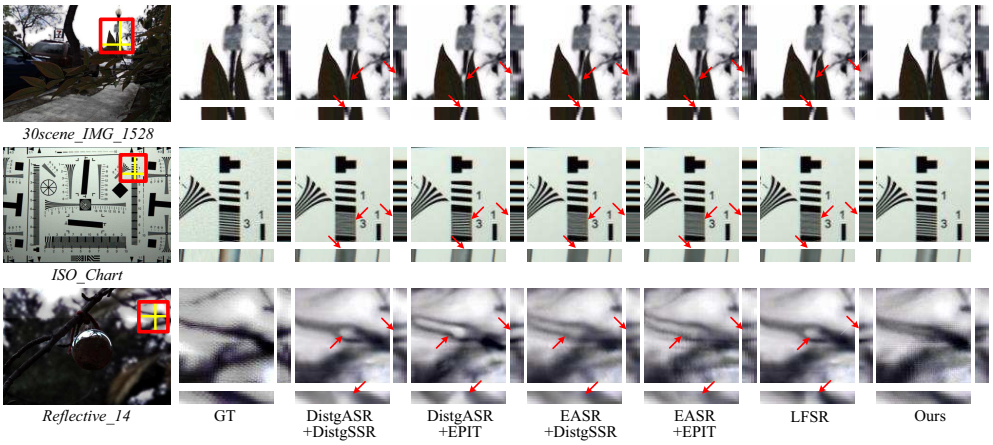


Fig. 3 Visual comparisons of one real-world scene and one synthetic scene for $SR^{A(2\times2\rightarrow5\times5),S(\times2)}$ task. The results show the ground-truth center views (GT), close-up image patches (in the red boxes) and close-up EPI patches (located along the yellow lines within the close-up image patches). The EPI patches are upsampled to an appropriate scale in the angular dimension for better viewing.

obtain the reconstructed light fields with 9×9 angular resolution. It can be regarded as a 9×9 SAI array, from which we get a 5×5 SAI array by a fixed interval extraction (interval=1) that is position-aligned with the 5×5 ground truth in the angular domain. Therefore, we evaluate the performance of the extracted 5×5 SAI array as the performance of "Ours (9×9)". Similarly, from "Ours (13×13)", we can obtain a 5×5 SAI array by a fixed interval extraction (interval=2), and use it for the performance evaluation of "Ours (13×13)", and so on. For arbitrary-scale spatial super-resolution, we downsample the images to the same scale as the ground truth for comparison. Specifically, for the result of the spatial scale is enlarged by a factor of 4, we compute the mean value of each 2×2 sub-pixel as the downsampled pixel value, and finally compare it with the ground truth.

Table 2 Quantitative comparisons (PSNR/SSIM) of our model with different angular resolutions and $\times 2$ spatial super-resolution.

Scenes	9×9	13×13	17×17	21×21	25×25	29×29	33×33
origami	38.70/0.988	38.27/0.988	38.01/0.987	37.85/0.987	37.73/0.987	37.64/0.986	37.58/0.986
monasRoom	42.78/0.993	42.56/0.992	42.41/0.992	42.31/0.992	42.24/0.992	42.19/0.992	42.15/0.992
bedroom	34.03/0.955	33.80/0.953	33.66/0.952	33.57/0.952	33.50/0.951	33.45/0.951	33.42/0.951
buddha	41.43/0.990	41.14/0.990	40.95/0.990	40.82/0.990	40.73/0.989	40.67/0.989	40.62/0.989

Table 2 evaluates the performance of our model with arbitrary-scale angular super-resolution and $\times 2$ spatial super-resolution. The performance of our model does not degrade significantly with the increase in output angular resolution. Notably, our SSIM scores are hardly affected, indicating that our model maintains satisfactory geometric structure fidelity for arbitrary-scale angular super-resolution. Fig. 4 provides visual comparisons of our model with different angular resolutions. The results show

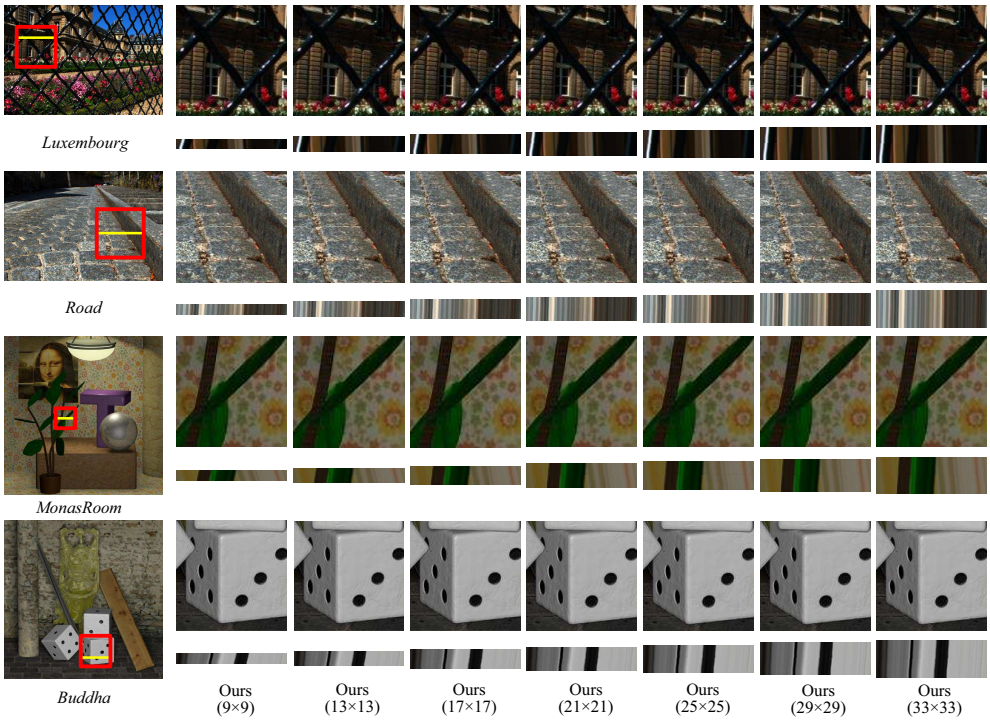


Fig. 4 Visual comparisons of our model with different angular resolutions. "$(N \times N)$" indicates $2 \times 2 \rightarrow N \times N$ angular super-resolution and $\times 2$ spatial super-resolution. The results show the close-up image patches (in the red boxes) and close-up EPI patches (located along the yellow lines within the close-up image patches). The EPI patches are upsampled to an appropriate scale in the angular dimension for better viewing.

Table 3 Quantitative comparisons (PSNR/SSIM) of our model with different spatial resolutions and $2 \times 2 \rightarrow 5 \times 5$ angular super-resolution.

Scenes	$\times 2$	$\times 4$	$\times 8$	$\times 16$
origami	39.28/0.989	39.06/0.989	38.96/0.988	38.93/0.988
monasRoom	43.02/0.993	42.80/0.993	42.72/0.993	42.69/0.993
bedroom	34.27/0.956	34.17/0.955	34.12/0.955	34.11/0.955
buddha	41.77/0.991	41.55/0.990	41.50/0.990	41.48/0.990

almost no degradation in visual quality as the angular resolution increases, and the linear structures of the EPIs with different angular resolutions remain nearly identical, further verifying the effectiveness of our method and its high fidelity in geometric structure.

Table 3 lists the performance comparisons of our model with arbitrary-scale spatial super-resolution and $2 \times 2 \rightarrow 5 \times 5$ angular super-resolution. The performance of our model also does not degrade significantly with the increasing in output spatial resolution, and our SSIM scores are also hardly affected. Our model provides satisfactory geometric structure fidelity for arbitrary-scale super-resolution. Fig. 5 provides visual

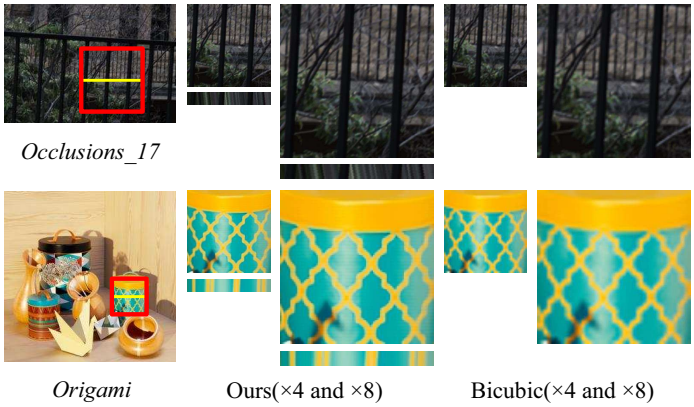


Fig. 5 Visual comparisons of the bicubic method and our model with different spatial resolutions. "$(\times N)$" indicates $\times N$ spatial resolution and angular resolution from 2×2 to 9×9. The results show the close-up image patches (in the red boxes) and close-up EPI patches (located along the yellow lines within the close-up image patches). The EPI patches are upsampled to an appropriate scale in the angular dimension for better viewing.

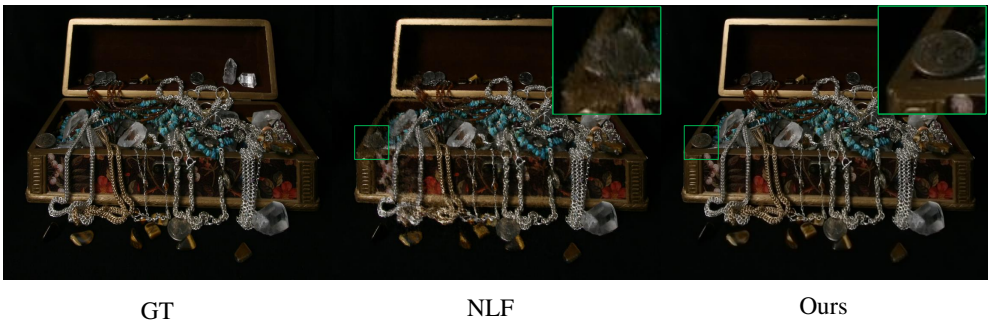


Fig. 6 Visual comparisons of one real-world scene for angular super-resolution task.

comparisons of the bicubic method and our model with different spatial resolutions. "$(\times N)$" indicates $\times N$ spatial resolution and angular resolution from 2×2 to 9×9. Our results maintain high fidelity at both $\times 4$ and $\times 8$ scales. Moreover, we observe that the results of our method contain sharper edges.

4.5 Comparison with Neural Light Field

In order to facilitate a comparative analysis with the NeRF method, we select the NLF [11] method, which is capable of angular super-resolution of the light field dataset, and conduct experimental comparisons on the Stanford dataset [46]. Ten scenes from the Stanford dataset together with the EPFL [53] are used as to train the $SR^{A\times\frac{5}{2},S\times 2}$ model. Our model is then used to test the results of scaling up from 2×2 to 5×5 only in angular resolution (with no change in spatial resolution). Similarly, for the NLF model, the angular resolution of the input image is 2×2, and the output resolution is

Table 4 Quantitative comparisons (PSNR/SSIM) of ablation experiments with different configurations.

	SAConv	$\mathbb{E}_c$	$\mathbb{E}_s$	HCINew	HCIOld	EPFL
Ours (baseline)			√	30.79 / 0.9328	38.83 / 0.9852	31.63 / 0.9615
Ours (w/o $\mathbb{E}_c$)	√		√	34.41 / 0.9596	41.70 / 0.9912	33.26 / 0.9699
Ours (w/o $\mathbb{E}_s$)	√	√		34.55 / 0.9604	41.88 / 0.9912	33.22 / 0.9702
Ours	√	√	√	**35.12 / 0.9639**	**42.40 / 0.9920**	**33.61 / 0.9711**

5×5. Due to the lack of generalization ability of NLF, retraining is required for each different scene. Comparison results are shown in Fig. 6. Compared with our result, the output image of NLF can be seen to be blurred.

5 Ablation Studies

To verify the effectiveness of the proposed structures, Table 4 lists the ablation results with different configurations.

"Ours (baseline)" is a baseline model that only uses the horizontal-vertical EPI convolutional module and a standard Transformer encoder $\mathbb{E}_s$ with a self-attention mechanism. Because the pure EPI-based method cannot model the spatio-angular correlation of light fields well, its performance is not satisfactory.

"Ours (w/o $\mathbb{E}_c$)" adds the inter-intra spatial and angular convolution(SAConv) modules to extract and fuse the light field spatio-angular correlation, which can steadily improve the reconstruction performance. However, the reconstruction performance is affected by irrelevant and trivial spatio-angular features.

"Ours (w/o $\mathbb{E}_s$)" only uses the encoder $\mathbb{E}_c$ without the self-attention encoder $\mathbb{E}_s$, which may result in performance fluctuations due to the lack of global dependencies learning.

"Ours" model combines all of our proposed ideas including the SAE feature extraction module and the spatio-angular aware Transformer encoder, and obtains the highest classification accuracy.

6 Conclusions

This paper proposes a local neural epipolar geometry learning strategy for arbitrary-scale, joint light field angular and spatial super-resolution. It learns the continuous epipolar geometry representations by mapping the light field EPIs to an intermediate, interpolable latent space. To this end, a spatio-angular aware geometry encoder is proposed, which not only explicitly integrates the local and global dependencies of light field epipolar geometry, but also embeds the spatio-angular correlation into the latent geometry codes by a spatio-angular aware Transformer. Finally, a divide-and-conquer local neural epipolar geometry learning strategy is proposed to convert the 4D implicit function learning into a joint learning of two 2D implicit functions with shared weights. To validate the superiority of our network, extensive experiments have been conducted on 6 publicly available light field datasets. Comparison results not only indicate that our method outperforms the state-of-the-art methods, but also demonstrate that our model has high fidelity in geometric structure.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grant 62471437, the "Pioneer" and "Leading Goose" Research and Development Program of Zhejiang under Grant 2023C01030.

References

[1] Ng, Y.-R., Levoy, M., Bredif, M., Duval, G., Horowitz, M., Hanrahan, P.: Light field photography with a handheld plenoptic camera. Computer Science Technical Report **2**(11), 1–11 (2005)

[2] Lumsdaine, A., Georgiev, T.: The focused plenoptic camera. In: IEEE International Conference on Computational Photography, San Francisco, USA, pp. 1–8 (2009)

[3] Wu, G., Masia, B., Jarabo, A., Zhang, Y., Wang, L., Dai, Q., Chai, T., Liu, Y.: Light field image processing: An overview. IEEE Journal of Selected Topics in Signal Processing **11**(7), 926–954 (2017)

[4] Zhu, H., Guo, M., Li, H., Wang, Q., Robles-Kelly, A.: Revisiting spatio-angular trade-off in light field cameras and extended applications in super-resolution. IEEE Transactions on Visualization and Computer Graphics **27**(6), 3019–3033 (2021)

[5] Wang, Y., Wang, L., Wu, G., Yang, J., An, W., Yu, J., Guo, Y.: Disentangling light fields for super-resolution and disparity estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(1), 425–443 (2022)

[6] Jin, J., Hou, J., Yuan, H., Kwong, S.: Learning light field angular super-resolution via a geometry-aware network. In: AAAI, vol. 34, pp. 11141–11148 (2020)

[7] Liang, Z., Wang, Y., Wang, L., Yang, J., Zhou, S., Guo, Y.: Learning non-local spatial-angular correlation for light field image super-resolution. In: ICCV, pp. 12376–12386 (2023)

[8] Li, M., Ma, B., Wang, S.: Hierarchical spatial-angular integration for lightweight light field image super-resolution. Knowledge-Based Systems, 113240 (2025)

[9] Wang, S., Lu, Y., Xia, W., Xia, P., Wang, Z., Gao, W.: Light field angular super-resolution by view-specific queries. The Visual Computer **41**(5), 3565–3580 (2025)

[10] Tancik, M., Srinivasan, P.P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J.T., Ng, R.: Fourier features let networks learn high frequency functions in low dimensional domains. In: NeurIPS (2020)

[11] Attal, B., Huang, J.-B., Zollhöfer, M., Kopf, J., Kim, C.: Learning neural light fields with ray-space embedding. In: CVPR, pp. 19819–19829 (2022)

[12] Zhou, W., Zhou, E., Liu, G., Lin, L., Lumsdaine, A.: Unsupervised monocular depth estimation from light field image. IEEE Transactions on Image Processing **29**, 1606–1617 (2020)

[13] Zhang, Q., Wang, Q., , Li, H., Yu, J.: Ray-space epipolar geometry for light field cameras. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(7), 3705–3718 (2022)

[14] Yoon, Y., Jeon, H.-G., Yoo, D., Lee, J.-Y., Kweon, I.S.: Learning a deep convolutional network for light-field image super-resolution. In: ICCV Workshops, pp. 24–32 (2015)

[15] Yoon, Y., Jeon, H.-G., Yoo, D., Lee, J.-Y., Kweon, I.S.: Light-field image super-resolution using convolutional neural network. IEEE Signal Processing Letters **24**(6), 848–852 (2017)

[16] Wang, Y., Liu, F., Zhang, K., Hou, G., Sun, Z., Tan, T.: Lfnet: A novel bidirectional recurrent convolutional neural network for light-field image super-resolution. IEEE Transactions on Image Processing **27**(9), 4274–4286 (2018)

[17] Zhang, S., Lin, Y., Sheng, H.: Residual networks for light field image super-resolution. In: CVPR, pp. 11038–11047 (2019)

[18] Shi, J., Jiang, X., Guillemot, C.: Learning fused pixel and feature-based view reconstructions for light fields. In: CVPR, pp. 2552–2561 (2020)

[19] Zhang, S., Chang, S., Lin, Y.: End-to-end light field spatial super-resolution network using multiple epipolar geometry. IEEE Transactions on Image Processing **30**, 5956–5968 (2021)

[20] Liu, G., Yue, H., Wu, J., Yang, J.: Efficient light field angular super-resolution with sub-aperture feature learning and macro-pixel upsampling. IEEE Transactions on Multimedia **25**, 6588–6600 (2022)

[21] Liang, Z., Wang, Y., Wang, L., Yang, J., Zhou, S.: Light field image super-resolution with transformers. IEEE Signal Processing Letters **29**, 563–567 (2022)

[22] Vagharshakyan, S., Bregovic, R., Gotchev, A.: Light field reconstruction using shearlet transform. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(1), 133–147 (2018)

[23] Wang, Y., Liu, F., Wang, Z., Hou, G., Sun, Z., Tan, T.: End-to-end view synthesis for light field imaging with pseudo 4DCNN. In: ECCV, pp. 333–348 (2018)

[24] Wu, G., Liu, Y., Fang, L., Dai, Q., Chai, T.: Light field reconstruction using

convolutional network on EPI and extended applications. IEEE transactions on pattern analysis and machine intelligence **41**(7), 1681–1694 (2019)

[25] Wu, G., Liu, Y., Dai, Q., Chai, T.: Learning sheared epi structure for light field reconstruction. IEEE Transactions on Image Processing **28**(7), 3261–3273 (2019)

[26] Yu, Z., Chen, L., Zeng, Z., Yang, K., Luo, S., Chen, S., Zhong, C.: Lgfn: Lightweight light field image super-resolution using local convolution modulation and global attention feature extraction. In: CVPR, pp. 6712–6721 (2024)

[27] Li, Y.-C., Wang, G.-Y., Liu, P., Zhang, S.: Light field images super-resolution method based on hybrid low-dimensional spatial-angular interaction feature and linear complementation epipolar feature. Expert Systems with Applications, 125586 (2024)

[28] Yuan, Y., Cao, Z., Su, L.: Light-field image superresolution using a combined deep cnn based on epi. IEEE Signal Processing Letters **25**(9), 1359–1363 (2018)

[29] Wang, Y., Wang, L., Yang, J., An, W., Yu, J., Guo, Y.: Spatial-angular interaction for light field image super-resolution. In: ECCV, pp. 290–308 (2020). Springer

[30] Meng, N., So, H.K.-H., Sun, X., Lam, E.Y.: High-dimensional dense residual convolutional neural network for light field reconstruction. IEEE transactions on pattern analysis and machine intelligence **43**(3), 873–886 (2019)

[31] Chen, Y., Jiang, G., Jiang, Z., Yu, M., Ho, Y.-S.: Deep light field super-resolution using frequency domain analysis and semantic prior. IEEE Transactions on Multimedia **24**, 3722–3737 (2021)

[32] Wang, S., Sheng, H., Chen, R., Yang, D., Cui, Z., Cong, R., Xiong, Z.: Spatial–angular–epipolar transformer for light field spatial and angular super-resolution. Displays **85**, 102816 (2024)

[33] Liu, G., Yue, H., Li, K., Yang, J.: Adaptive pixel aggregation for joint spatial and angular super-resolution of light field images. Information Fusion **104**, 102183 (2024)

[34] Sitzmann, V., Martel, J.N.P., Bergman, A., Lindell, D.B., Wetzstein, G.: Implicit neural representations with periodic activation functions. In: NeurIPS (2020)

[35] Chen, Y., Liu, S., Wang, X.: Learning continuous image representation with local implicit image function. In: CVPR, pp. 8628–8638 (2021)

[36] Kim, J., Kim, T.-K.: Arbitrary-scale image generation and upsampling using latent diffusion model and implicit neural decoder. In: CVPR, pp. 9202–9211 (2024)

[37] Gao, S., Liu, X., Zeng, B., Xu, S., Li, Y., Luo, X., Liu, J., Zhen, X., Zhang, B.:

Implicit diffusion models for continuous super-resolution. In: Cvpr, pp. 10021–10030 (2023)

[38] Chen, Z., Zhang, H.: Learning implicit fields for generative shape modeling. In: CVPR, pp. 5939–5948 (2019)

[39] Lu, Y., Wan, L., Ding, N., Wang, Y., Shen, S., Cai, S., Gao, L.: Unsigned orthogonal distance fields: An accurate neural implicit representation for diverse 3d shapes. In: CVPR, pp. 20551–20560 (2024)

[40] Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: Deepsdf: Learning continuous signed distance functions for shape representation. In: CVPR, pp. 165–174 (2019)

[41] Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: Representing scenes as neural radiance fields for view synthesis. In: ECCV (2020)

[42] Dihlmann, J.-N., Engelhardt, A., Lensch, H.: Signerf: Scene integrated generation for neural radiance fields. In: CVPR, pp. 6679–6688 (2024)

[43] Liu, T., Ye, X., Shi, M., Huang, Z., Pan, Z., Peng, Z., Cao, Z.: Geometry-aware reconstruction and fusion-refined rendering for generalizable neural radiance fields. In: CVPR, pp. 7654–7663 (2024)

[44] Wizadwongsa, S., Phongthawee, P., Yenphraphai, J., Suwajanakorn, S.: NeX: Real-time view synthesis with neural basis expansion. In: CVPR, pp. 8534–8543 (2021)

[45] Gupta, A., Cao, J., Wang, C., Hu, J., Tulyakov, S., Ren, J., Jeni, L.: Lightspeed: light and fast neural light fields on mobile devices. NeurIPS **36** (2024)

[46] Wilburn, B., Joshi, N., Vaish, V., Talvala, E.-V., Antúnez, E.R., Barth, A., Adams, A., Horowitz, M., Levoy, M.: High performance imaging using large camera arrays. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(3), 765–776 (2005)

[47] You, Z., Geiger, A., Chen, A.: Nelf-pro: Neural light field probes for multi-scale novel view synthesis. In: CVPR, pp. 19833–19843 (2024)

[48] Wu, G., Zhou, Y., Liu, Y., Fang, L., Chai, T.: Geo-NI: Geometry-aware neural interpolation for light field rendering. arXiv:2206.09736 (2022)

[49] Zhong, L., Zong, B., Wang, Q., Yu, J., Zhou, W.: Implicit epipolar geometric function based light field continuous angular representation. In: CVPRW, pp. 10121–10129 (2023)

[50] Zhang, H., Zhou, W., Lin, L., Lumsdaine, A.: Cascade residual learning based

adaptive feature aggregation for light field super-resolution. Pattern Recognition **165**, 111616 (2025)

[51] Kalantari, N.K., Wang, T.-C., Ramamoorthi, R.: Learning-based view synthesis for light field cameras. ACM Transactions on Graphics (SIGGRAPH Asia) **35**(6), 193–119310 (2016)

[52] Raj, A.S., Lowney, M., Shah, R., Wetzstein, G.: Stanford lytro light field archive. LF2016. html (2016)

[53] Rerabek, M., Ebrahimi, T.: New light field image dataset. In: 8th International Conference on Quality of Multimedia Experience (QoMEX) (2016)

[54] Honauer, K., Johannsen, O., Kondermann, D., Goldluecke, B.: A dataset and evaluation methodology for depth estimation on 4D light fields. In: Asian Conference on Computer Vision, pp. 19–34 (2016)

[55] Johannsen, O., Honauer, K., Goldluecke, B., Alperovich, A., Battisti, F., Bok, Y., Brizzi, M., Carli, M., Choe, G., Diebold, M., Gutsche, M., Jeon, H.-G., Kweon, I.S., Neri, A., Park, J., Park, J., Schilling, H., Sheng, H., Si, L., Strecke, M., Sulc, A., Tai, Y.-W., Wang, Q., Wang, T.-C., Wanner, S., Xiong, Z., Yu, J., Zhang, S., Zhu, H.: A taxonomy and evaluation of dense light field depth estimation algorithms. In: IEEE Conference on Computer Vision and Pattern Recognition - LF4CV Workshop, pp. 1795–1812 (2017)

[56] Wanner, S., Meister, S., Goldluecke, B.: Datasets and benchmarks for densely sampled 4d light fields. In: VMV, vol. 13, pp. 225–226 (2013)

[57] Mao, Y., Huang, X., Chen, Y., Liu, D., An, P., Lee, S.: Learning a domain-specialized network for light field spatial-angular super-resolution. IEEE Transactions on Visualization and Computer Graphics **32**(2), 2127–2140 (2026) https://doi.org/10.1109/TVCG.2025.3644930

[58] Yeung, H.W.F., Hou, J., Chen, X., Chen, J., Chen, Z., Chung, Y.Y.: Light field spatial super-resolution using deep efficient spatial-angular separable convolution. IEEE Transactions on Image Processing **28**(5), 2319–2330 (2018)

[59] Wang, Y., Yang, J., Wang, L., Ying, X., Wu, T., An, W., Guo, Y.: Light field image super-resolution using deformable convolution. IEEE Transactions on Image Processing **30**, 1057–1071 (2020)

[60] Liu, G., Yue, H., Wu, J., Yang, J.: Intra-inter view interaction network for light field image super-resolution. IEEE Transactions on Multimedia **25**, 256–266 (2021)

[61] Wang, S., Zhou, T., Lu, Y., Di, H.: Detail-preserving transformer for light field image super-resolution. In: Proceedings of the AAAI Conference on Artificial

Intelligence, vol. 36, pp. 2522–2530 (2022)

[62] Jin, K., Wei, Z., Yang, A., Wu, D., Gao, M., Zhou, X.: Lftransmamba: A hybrid mamba-transformer model for light field image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 1195–1204 (2025)

[63] Zeyu, X., Li, Z., Jia, W.: Occlusion-embedded hybrid transformer for light field super-resolution. Proceedings of the AAAI Conference on Artificial Intelligence **39**(8), 8700–8708 (2025)