PLENOPTIC SUPER-RESOLUTION VIA HYBRID LIGHT-FIELD AND 3D GAUSSIAN PRIORS

Anonymous authors
Paper under double-blind review

ABSTRACT

Plenoptic super-resolution aims to reconstruct a scene at higher spatial-angular resolution from sparse, low-resolution light-field (LF) observations. Prior methods rely on spatio-angular priors like epipolar-plane image (EPI) continuity, but these can break under wide baselines, large disparities, and occlusions, and most models are tied to fixed upsampling factors, limiting resolution control at inference. To overcome these limitations, we complement LF structural priors with an explicit 3D Gaussian representation, which naturally supports continuous angular rendering and occlusion-aware view synthesis. Building on these complementary LF and 3DGS priors, we propose PlenoGS, a two-stage framework for plenoptic super-resolution. The first stage reconstructs a stable 3D Gaussian representation regularized by LF priors. The second stage enriches this representation with high-resolution spatial details through geometry-consistent refinement and scale-aware local implicit decoding, thereby preserving consistency across viewpoints while supporting flexible spatial magnification. On synthetic and real wide-baseline LF images, PlenoGS reduces occlusion ghosting and grid artifacts and improves spatial detail, angular consistency, and novel-view quality.

1 INTRODUCTION

The plenoptic function provides a complete representation of real-world light propagation by describing radiance across spatial position, angular direction, wavelength, and time (Adelson & Bergen, 1991). In practice, however, imaging devices can capture only limited samples of this high-dimensional function. Light fields (LFs), which jointly record spatial and angular radiance information, provide one of its closest measurable and structured approximations (Levoy & Hanrahan, 1996; Gortler et al., 1996). Since practical imaging systems trade off spatial and angular sampling density, acquiring densely sampled, high-resolution LFs remains prohibitively expensive. To overcome these acquisition limitations, *plenoptic super-resolution* seeks to infer a scene representation at higher spatial-angular resolution from sparse, low-resolution LF observations, enabling applications such as immersive 3D perception and novel view synthesis (Mildenhall et al., 2019; Wu et al., 2019; Valenzise et al., 2023). In the light-field literature, this problem is commonly referred to as light-field spatio-angular super-resolution (LF-SASR) (Wu et al., 2017; Sun et al., 2021).

Conventional LF-SASR methods learn mappings in the spatial, angular, or EPI domains, using *i)* high-dimensional CNNs for spatial–angular correlations (Meng et al., 2021; Sun et al., 2021), *ii)* EPI-based networks that leverage cross-view geometry (Wu et al., 2019), and *iii)* Vision Transformers for non-local cross-view interactions (Liang et al., 2022; Wang et al., 2024). However, they have three limitations: *i)* they often assume small disparities and locally continuous EPIs, so wide baselines with large disparities or occlusions can cause blur, ghosting, and structural artifacts; *ii)* they are typically trained for fixed upsampling factors and cannot vary output resolution at inference; and *iii)* dense 4D operations or attention are computationally expensive.

An alternative is to explicitly model 3D scene geometry and render from target viewpoints. 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) enables strong novel-view synthesis via explicit representation, real-time rendering, and continuous viewpoint control. Its geometry-aware rendering naturally handles visibility and occlusions, making it appealing for angular super-resolution. However, 3DGS alone falls short for plenoptic super-resolution: with sparse, wide-baseline training views, it may produce unstable geometry, floating artifacts, and inconsistent details without cross-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

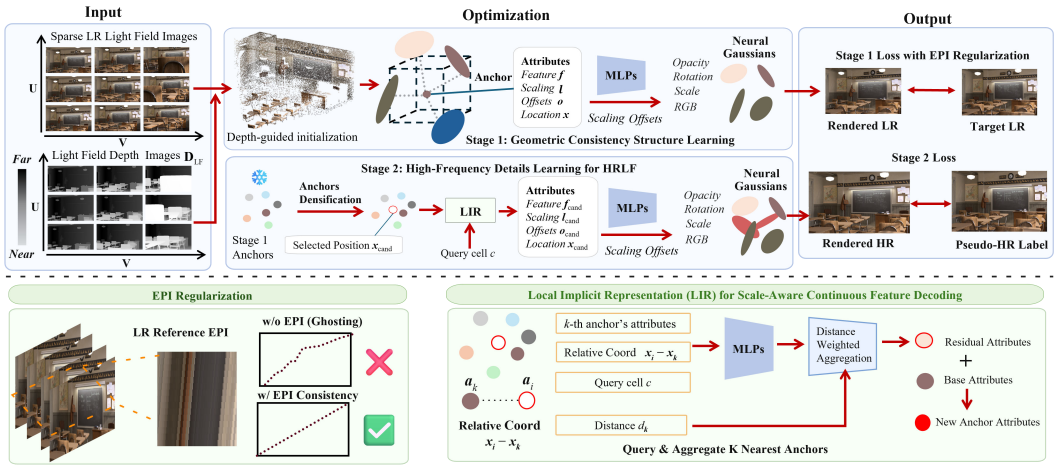

Figure 1: **Overview of PlenoGS.** The pipeline follows a two-stage coarse-to-fine process. **Stage 1** learns a stable 3D Gaussian representation regularized by EPI consistency. **Stage 2** adds high-resolution details via geometry-consistent refinement and scale-aware local implicit decoding, enabling resolution-flexible LF-SASR.

view structural constraints. Existing 3DGS super-resolution methods, such as SuperGS (Xie et al., 2025), often rely on discrete feature structures such as hash or voxel grids (Müller et al., 2022) to recover high frequencies. Interpolating these grids can introduce aliasing near complex textures and occlusion boundaries, especially at large or unseen magnification factors, limiting resolution flexibility.

These complementary strengths and limitations motivate the central research question of this work: *can light-field structural priors and explicit 3D Gaussian representations be effectively combined to achieve robust, resolution-flexible plenoptic super-resolution while retaining efficient rendering?* We answer this question with **PlenoGS**, a hybrid-prior plenoptic super-resolution scheme that exploits the complementary properties of the two representations: light-field priors regularize the reconstruction from sparse observations, while 3DGS provides explicit geometry, occlusion-aware rendering, and continuous angular synthesis. By integrating these properties, PlenoGS reconstructs spatially detailed and angularly consistent light fields under challenging wide-baseline conditions, while supporting flexible spatial magnification and real-time rendering.

As shown in Fig. 1, PlenoGS uses a two-stage coarse-to-fine pipeline to instantiate this idea. In the coarse stage, structurally consistent 3DGS primitives are learned via depth-guided point cloud initialization and EPI structural constraints, enabling dense novel view synthesis, with depth estimated by Video Depth Anything (VDA) (Chen et al., 2025b). Stage 2 freezes this geometry and adopts the error-guided multi-view voting mechanism of SuperGS (Xie et al., 2025) for controlled anchor densification. The attributes of newly added anchors are then refined using a local implicit representation (LIR) over their surrounding K anchors. By modeling attribute extraction as a locally continuous function of relative coordinates, the local implicit representation mitigates grid aliasing and enables resolution-flexible joint LF-SASR. Experiments on synthetic and real-world wide-baseline LF images show that PlenoGS substantially reduces occlusion ghosting and grid artifacts while improving spatial detail, angular consistency, and the quality of synthesized novel views.

The main contributions of this work are as follows:

- We introduce PlenoGS, a hybrid-prior approach to sparse, wide-baseline plenoptic super-resolution that combines explicit 3D Gaussian geometry with light-field structural priors, enabling occlusion-aware angular synthesis and resolution-flexible spatial super-resolution.
- We develop an LF-aware geometry stabilization strategy that combines depth-guided initialization with EPI-based regularization to improve sparse wide-baseline reconstruction.
- We design a scale-aware anchor attribute decoder that uses continuous relative-coordinate conditioning to enable arbitrary-resolution reconstruction while preserving high-frequency details.

2 Related Work

Light Field Super-Resolution. The spatial–angular trade-off in LF cameras (e.g., Lytro, Raytrix) limits plenoptic applications (Ng et al., 2005). Early LF super-resolution targeted either angular density or spatial detail. Angular SR (novel view synthesis) progressed from depth-image-based rendering (Wanner & Goldluecke, 2014), patch-based synthesis (Mitra & Veeraraghavan, 2012), and phase-based interpolation (Zhang et al., 2015) to CNNs that model disparity or EPIs (Kalantari et al., 2016; Wu et al., 2019; Yoon et al., 2015). Spatial SR evolved from 2D SISR to separable spatial–angular convolutions (Yeung et al., 2018), macro-pixel upsampling (Liu et al., 2023), cross-view interaction modeling (Wang et al., 2020; 2018), and disentangled representations (Wang et al., 2023). To reduce redundancy and error accumulation in two-stage pipelines, recent work uses joint SASR with end-to-end 4D CNNs (Meng et al., 2021; 2020; Mo et al., 2022), epipolar Vision Transformers (Liang et al., 2023; 2022; Wang et al., 2024), and LF-specific state-space models (Mao et al., 2026), with enhanced cross-view interactions for angular coherence (Chen et al., 2025a). However, most methods rely on EPI priors, which break down under large disparities and occlusions, causing ghosting/tearing, and typically support only fixed upsampling ratios, requiring retraining per scale.

3DGS-based Novel View Synthesis. 3DGS (Kerbl et al., 2023) is a leading approach (Fei et al., 2025) for real-time, high-fidelity novel view synthesis using anisotropic Gaussians. Extensions improve anti-aliasing (Yu et al., 2024), reduce memory (Lee et al., 2024), and support sparse views (Zhu et al., 2024). For light-field imaging, LF-GS (Huang et al., 2025) incorporates sub-aperture depth/normal priors. Meng et al. (2026) perform LF-SASR by splatting 2D Gaussians on projected EPI features, but it relies on micro-baselines and degrades with wide-baseline occlusions. For wide-baseline LF, global geometric priors (e.g., monocular depth) and multi-view constraints (Chung et al., 2024) provide stronger 3D structure. For high-resolution reconstruction, SRGS (Feng et al., 2026) upsamples with 3DGS; SuperGS (Xie et al., 2025) reduces multi-view inconsistency via error-guided voting; and SplatSuRe (Asthana et al., 2026) uses scene geometry to refine undersampled regions with multi-view-consistent detail. However, many 3DGS-based SR methods use discrete features (e.g., hash grids), which can introduce artifacts beyond the training resolution or at non-integer scales.

Continuous Implicit Neural Representations. Implicit Neural Representations (INRs) model discrete signals as continuous functions with coordinate-based MLPs (Sitzmann et al., 2020), decoupling scenes from fixed spatial resolutions. Meta-SR (Hu et al., 2019) and ArbSR (Wang et al., 2021) enabled arbitrary-scale magnification via spatial-aware interpolation, later improved by LIIF (Chen et al., 2021), LTE (Lee & Jin, 2022), ITSRN (Yang et al., 2021), and LFSR methods such as SEIIF (Cong et al., 2025) and SAEIIF (Cong et al., 2026). These works show that continuous-coordinate decoding enables infinite-resolution SR via arbitrary queries. However, extending 2D INRs to multi-view images breaks 3D consistency, while pure 3D INRs such as NeRF (Mildenhall et al., 2020) are too slow at inference. Efficient 3D representations (Müller et al., 2022; Chen et al., 2022) discretize structure, reintroducing resolution limits and causing high-frequency interpolation artifacts.

3 Method

3.1 Preliminaries

Problem Formulation. Given sparsely sampled low-resolution (LR) light field views $\{I_{\text{LR}}^{(v)}\}_{v=1}^{V}$ with calibrated intrinsics $\mathbf{K}_v$ and extrinsics $(\mathbf{R}_v, \boldsymbol{t}_v)$, we reconstruct a dense high-resolution light field across spatial and angular dimensions. Unlike LF-SASR networks that learn a fixed discrete mapping from sparse LR to dense HR views, PlenoGS uses a geometry-aware, resolution-flexible representation to synthesize novel views and support arbitrary magnification without retraining.

3D Gaussian Representation. Vanilla 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) models a scene with anisotropic Gaussian primitives, each defined by a 3D mean $\boldsymbol{\mu}$, covariance $\boldsymbol{\Sigma}$, opacity α, and view-dependent color coefficients. Pixel color C is rendered via front-to-back α-blending:

$$C = \sum_{n=1}^{N} \boldsymbol{c}_n \alpha_n \prod_{m=1}^{n-1} (1 - \alpha_m). \tag{1}$$

To obtain a more structured and compact representation, we adopt the anchor-based formulation of Scaffold-GS (Lu et al., 2024). A set of 3D anchors $\mathcal{A} = \{\boldsymbol{a}_i\}$ stores learnable features $\boldsymbol{f}_i$, from which local Gaussian attributes are decoded by lightweight MLPs. The basic optimization objective consists of photometric reconstruction, structural similarity, and volume regularization:

$$\mathcal{L}_{\text{base}} = \mathcal{L}_1 + \lambda_{\text{ssim}}(1 - \text{SSIM}) + \lambda_{\text{vol}}\mathcal{L}_{\text{vol}}. \tag{2}$$

Although this representation naturally supports continuous angular rendering through camera-space rasterization, its learned anchor features are still optimized from finite-resolution observations.

Error-Guided Multi-View Densification. For spatial super-resolution, we use 2D pseudo-HR labels as high-frequency supervision. Because independently generated pseudo-labels can be inconsistent across views, we follow SuperGS (Xie et al., 2025) and use error-guided multi-view support to filter candidate anchors before insertion. Our focus is on continuously refining the attributes of these densified anchors for resolution-flexible reconstruction.

The uncertainty-weighted reconstruction loss is

$$\mathcal{L}_{\text{rec}} = \sum_{(u,v)} \exp(-U(u,v)) \left\| \hat{I}(u,v) - I_{\text{pseudo}}(u,v) \right\|_1, \tag{3}$$

where $\hat{I}$ is the rendered HR image and $U(u,v)$ is the uncertainty map. Each anchor learns a log-variance for latent-feature uncertainty, optimized in Stage 2 and alpha-composited into $U(u,v)$ (Section 3.2.2). As in SuperGS, higher-uncertainty pixels are down-weighted. Unlike grid-based SR, PlenoGS decodes new anchor features with a local continuous implicit function, reducing dependence on fixed voxel or hash-grid resolution.

3.2 PlenoGS Model Development

PlenoGS first builds a stable LF-aware 3D anchor field from sparse wide-baseline views, then refines it at flexible resolutions by adding geometrically verified anchors with features from a scale-aware continuous local implicit function. Stage 1 enforces geometric consistency; Stage 2 restores high-frequency details without fixed-grid interpolation artifacts.

3.2.1 Stage 1: Geometric Consistency Structure Learning

Stage 1 establishes a robust 3D geometric foundation using anchor-based 3DGS, guided by LF-specific depth geometry initialization and regularized by EPI physical priors.

Depth-Guided Geometry Initialization. Dense SfM often fails on wide-baseline LFs due to occlusions and large disparities. We use depth foundation models such as Video Depth Anything (VDA) (Chen et al., 2025b) to estimate dense relative depth maps $\mathbf{D}_{\text{LF}}$, and use sparse SfM geometry to calibrate their scale and shift; LF consistency is enforced via EPI regularization. With calibrated intrinsics $\mathbf{K}$ and extrinsics $(\mathbf{R}, \boldsymbol{t})$, we back-project the calibrated dense depth maps to 3D to form an initial dense point cloud. To remove duplicates and floaters from depth errors, we voxel-fuse points and keep the median depth per voxel, producing the structured anchor set $\mathcal{A}_{\text{base}} = \{\boldsymbol{a}_i\}$.

EPI-guided Multi-View Regularization. Optimizing views independently leads to angular inconsistencies. In LF imaging, EPIs exhibit piecewise-planar structures (Bolles et al., 1987). To enforce multi-view geometric consistency, we introduce an EPI-based reconstruction regularizer. In Stage 1, we randomly sample spatial scanlines $\mathcal{S}$. For each $l \in \mathcal{S}$, we construct a rendered EPI patch $\hat{\boldsymbol{E}}_l$. The reference EPI $\boldsymbol{E}_{\text{ref},l}$ is built only from available LR input views. The EPI loss enforces pixel- and gradient-level consistency:

$$\mathcal{L}_{\text{EPI}} = \frac{1}{|\mathcal{S}|} \sum_{l \in \mathcal{S}} \Big(\|\hat{\boldsymbol{E}}_l - \boldsymbol{E}_{\text{ref},l}\|_1 + \lambda_{\text{grad}} \|\nabla \hat{\boldsymbol{E}}_l - \nabla \boldsymbol{E}_{\text{ref},l}\|_1 \Big). \tag{4}$$

Here, $|\mathcal{S}|$ is the total number of sampled scanlines per training iteration, and ∇ is the 2D finite-difference gradient over the spatial-angular axes of the EPI patch. This term enforces consistent local EPI edge orientations and slope structures, preventing 3DGS optimization from collapsing into noisy, unstructured splats. The Stage 1 objective $\mathcal{L}_{\text{stage1}}$ is:

$$\mathcal{L}_{\text{stage1}} = \mathcal{L}_1 + \lambda_{\text{ssim}}(1 - \text{SSIM}) + \lambda_{\text{vol}}\mathcal{L}_{\text{vol}} + \lambda_{\text{epi}}\mathcal{L}_{\text{EPI}}. \tag{5}$$

3.2.2 STAGE 2: HIGH-FREQUENCY DETAILS LEARNING FOR HRLF

Stage 1 reconstructs reliable low-frequency geometry. Stage 2 uses HR pseudo-labels I_{pseudo} from a 2D SISR network, e.g., SwinIR (Liang et al., 2021) or LIIF (Chen et al., 2021), only for sparse training views. Following SuperGS (Xie et al., 2025), we adopt its error-guided multi-view voting rule for anchor densification, together with uncertainty-aware optimization.

Anchor Densification and Freezing Strategy. High-error pixels are back-projected to 3D, voxelized, and converted into new anchors only if their voxel is supported by at least τ_{vote} views, filtering single-view candidates. To avoid geometric drift, Stage 2 freezes all Stage-1 geometry (anchor positions, offsets, and geometric MLPs), allowing only gated anchor insertion and small jitter.

New anchor features are sampled from a Gaussian:

$$\boldsymbol{f}_i \sim \mathcal{N}(\boldsymbol{f}_k, \mathrm{diag}(\boldsymbol{\sigma}_i^2)),$$

with mean set to the nearest-neighbor feature $\boldsymbol{f}_k$ and local offsets reset to zero. The log-variance $\log \boldsymbol{\sigma}_i^2$ is learned with Stage-2 anchor features and yields per-anchor uncertainty u_i. We alpha-composite u_i to form an uncertainty map $U(u, v)$ and reweight the reconstruction loss by $W(u, v) = \exp(-U(u, v))$ (Eq. 3). Anchors with high uncertainty ($u_i > \tau_u$) are jittered within $s_{\text{vox}}/4$ to add support. To preserve structure under high upsampling, we also apply LPIPS (Zhang et al., 2018) on downsampled outputs. The Stage-2 objective is:

$$\mathcal{L}_{\text{stage2}} = \mathcal{L}_{\text{rec}} + \lambda_{\text{ssim}}(1 - \text{SSIM}) + \lambda_{\text{vol}}\mathcal{L}_{\text{vol}} + \lambda_{\text{lpips}}\mathcal{L}_{\text{lpips}}. \tag{6}$$

Scale-Aware Continuous Feature Decoding. While vanilla 3DGS can be rasterized at arbitrary resolutions by increasing the output grid density, a model trained only on low-resolution views cannot reliably recover high-frequency details at test time. The learned Gaussian attributes and anchor features are biased to the training sampling frequency, so rendering them on a denser grid often produces over-smoothed textures instead of true spatial super-resolution. To address this, we rescale the anchor features to the target sampling scale before rasterization.

We propose a scale-aware continuous feature decoding module in 3D anchor space, inspired by LIIF (Chen et al., 2021). For each newly spawned anchor $\boldsymbol{a}_i$, we retrieve its K nearest Stage-1 context anchors $\mathcal{N}_K(\boldsymbol{a}_i) = \{\boldsymbol{a}_k\}$. Decoding is conditioned on a target-scale query cell descriptor $\boldsymbol{c}_i \in \mathbb{R}^3$, which encodes the local 3D sampling footprint of one target-resolution pixel around the anchor. Instead of a discrete scale label, $\boldsymbol{c}_i$ maps the requested magnification to a depth-aware 3D sampling interval. For magnification s, anchor depth z_i, and focal lengths (f_x, f_y), we set:

$$\boldsymbol{c}_i(s) = \left[\frac{z_i}{s f_x}, \frac{z_i}{s f_y}, \frac{z_i}{s \bar{f}}\right], \quad \bar{f} = \frac{f_x + f_y}{2}. \tag{7}$$

$\boldsymbol{c}_i$ is a compact footprint descriptor (not an exact voxel size): larger values imply coarser sampling, smaller values imply denser sampling. This follows the pinhole model, where a pixel at depth z_i spans a 3D interval proportional to z_i/f, yielding a geometry-based resolution-to-cell conversion rather than a heuristic rule.

We form the local query vector $\boldsymbol{q}_k$ by concatenating the frozen base attributes (feature $\boldsymbol{f}_k$, scaling $\boldsymbol{l}_k$, and offsets $\boldsymbol{o}_k$), the relative 3D offset $\boldsymbol{x}_{\text{cand}} - \boldsymbol{x}_k$ (where $\boldsymbol{x}_k$ is the spatial position of the k-th anchor), the query cell descriptor $\boldsymbol{c}$, and the scalar distance $d_k = \|\boldsymbol{x}_{\text{cand}} - \boldsymbol{x}_k\|_2$:

$$\boldsymbol{q}_k = [\boldsymbol{f}_k,\ \boldsymbol{l}_k,\ \boldsymbol{o}_k,\ \boldsymbol{x}_{\text{cand}} - \boldsymbol{x}_k,\ \boldsymbol{c},\ d_k], \quad \boldsymbol{h}_k = \text{MLP}_{\text{local}}(\boldsymbol{q}_k). \tag{8}$$

The local hidden states are fused using distance-aware weights:

$$w_k = \frac{\exp(-d_k/T)}{\sum_{j=1}^{K} \exp(-d_j/T)}, \tag{9}$$

where T is a temperature hyperparameter that controls the smoothness of the weight distribution among the K nearest anchors. A smaller T enforces a sharper aggregation dominated by the closest anchor, whereas a larger T yields a smoother, more uniform blending of the local neighborhood. The fused local representation is then decoded into scale-conditioned residual attributes:

$$\boldsymbol{g} = \text{MLP}_{\text{fuse}}\left(\sum_{k=1}^{K} w_k \boldsymbol{h}_k\right), \quad [\Delta\boldsymbol{f},\ \Delta\boldsymbol{l},\ \Delta\boldsymbol{o}] = \text{MLP}_{\text{head}}([\boldsymbol{g}, \boldsymbol{c}]). \tag{10}$$

Final attributes for the densified anchor at $\boldsymbol{x}_{\text{cand}}$ are obtained via residual refinement:

$$\boldsymbol{f}_{\text{cand}} = \boldsymbol{f}_{\text{base}} + \Delta\boldsymbol{f}, \quad \boldsymbol{l}_{\text{cand}} = \boldsymbol{l}_{\text{base}} + \Delta\boldsymbol{l}, \quad \boldsymbol{o}_{\text{cand}} = \boldsymbol{o}_{\text{base}} + \Delta\boldsymbol{o}, \tag{11}$$

where $\boldsymbol{f}_{\text{base}}$, $\boldsymbol{l}_{\text{base}}$, and $\boldsymbol{o}_{\text{base}}$ are the feature, scale, and offsets copied from the nearest Stage-1 neighbor. The tuple $(\boldsymbol{x}_{\text{cand}}, \boldsymbol{f}_{\text{cand}}, \boldsymbol{l}_{\text{cand}}, \boldsymbol{o}_{\text{cand}})$ is then fed to MLPs to decode the Neural Gaussians.

Unlike naive 2D post-processing, this approach injects target-scale information into the 3D representation before Gaussian attribute decoding, instead of filtering after rendering. Conditioning the anchor feature residual on the physical footprint $\boldsymbol{c}_i$ lets the network adapt Gaussian attributes to different sampling densities: smaller cells at higher resolutions preserve sharp, high-frequency details, while larger footprints at coarser resolutions enforce smoothness to reduce aliasing and oversharpening. As a result, one trained model supports arbitrary, non-integer magnifications without changing the optimized 3DGS rasterization pipeline.

Target-Scale Feature Precomputation. At a target scale, the continuous decoder runs once per anchor and stores scale-adapted features as anchor attributes before rasterization, avoiding per-pixel MLP queries. Because these attributes stay with shared 3D anchors, rendering remains multi-view consistent while supporting resolution-flexible spatial decoding.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate our framework on two wide-baseline Light Field (LF) datasets:
1) *FIT Synthetic Dataset* (Chlubna et al., 2024): A high-quality Blender-rendered dataset with accurate camera parameters and ground truth. We use three challenging scenes (*Classroom*, *Buildings*, *Macro*) with severe occlusions and complex textures.
2) *Stanford Real-World Dataset* (Vaish & Adams, 2008): To assess robustness on uncalibrated real captures, we use three classical scenes (*Lego Knights*, *Jewelry*, *Tarot*). Since camera parameters in the form required by our 3DGS pipeline are not directly available for the selected Stanford captures, we estimate camera poses using hloc (Sarlin et al., 2019) and then apply the same depth-guided initialization pipeline.

Baselines. We compare PlenoGS with SOTA methods in two paradigms: 1) *Traditional image-domain LF-SASR:* the robust cascaded baseline **LF-EASR + EPIT** (Liu et al., 2023; Liang et al., 2023) and **SAEIIF** (Cong et al., 2026). 2) *3D neural rendering:* vanilla **3DGS** (Kerbl et al., 2023), anchor-based **Scaffold-GS** (Lu et al., 2024), LF-specific **LF-GS** (Huang et al., 2025), **SuperGS** (Xie et al., 2025), and **SplatSuRe** (Asthana et al., 2026), together with our **PlenoGS**.

Evaluation Protocols. To enable rigorous, fair comparison across diverse baselines, we use:
1) *Metrics and Color Spaces:* Traditional LF-SASR methods process only the luminance (Y) channel. Following standard practice, we report PSNR/SSIM on Y only and exclude these methods from RGB-based LPIPS to avoid bias. For visualization, we convert outputs to RGB by bicubically upsampling the low-resolution Cb/Cr channels.
2) *Reference and Primary Baselines:* Traditional image-domain LF-SASR methods depend on fixed angular settings and discrete scales, so we evaluate official pre-trained models in their native setups. They serve as task-transfer reference baselines, not matched training comparisons. For arbitrary-scale evaluation, 3DGS-based methods are the main directly comparable baselines.
3) *Efficiency Assessment:* Since scene complexity varies, Params and FPS are reported at the dataset level over all scenes. Params are averaged across the inference models of individual scenes, while FPS is reported as the arithmetic mean of the scene-wise throughput measured over all test views. All view-dependent computations required to render a frame are included, while model loading, disk I/O, and metric computation are excluded.

Training Details. PlenoGS is trained in two stages. **Stage 1:** optimize the anchor-based Gaussian backbone for $30{,}000$ iterations with frozen anchor positions (LR 0). Offset LR starts at 1×10^{-2} and decays; feature/opacity/scaling/rotation LRs are 7.5×10^{-3}, 2×10^{-2}, 7×10^{-3}, and 2×10^{-3}. With EPI consistency, apply it every 10 iterations with $\lambda_{\text{epi}} = 0.2$, using 3 angular views and 4 scanlines per step; randomly choose the EPI axis (spatial stride 1).

Stage 2: starting from Stage 1, optimize the continuous feature field for $15{,}000$ iterations while freezing Stage-1 Gaussian geometry and attribute MLPs. Train only the LIR residual field, Stage-2

Table 1: Quantitative results for the $9 \times 9 \rightarrow 15 \times 15$ angular and $\times 2$ spatial SR task. Best and second-best scores are marked in RED and YELLOW. Traditional image-domain methods are evaluated only on the Y-channel and are thus excluded from LPIPS.

Method	Metric	FIT Dataset (Chlubna et al., 2024) (Synthetic)						Stanford Dataset (Vaish & Adams, 2008) (Real-World)					
		Classrm	Build.	Macro	**Avg**	Params↓	FPS↑	Lego	Jewel.	Tarot	**Avg**	Params↓	FPS↑
LF-EASR (Liu et al., 2023) + EPIT (Liang et al., 2023)	PSNR↑	26.90	26.12	24.62	25.88	7.86	1.72	27.21	27.82	26.37	27.13	7.86	1.39
	SSIM↑	0.811	0.820	0.632	0.754			0.820	0.890	0.846	0.852		
SAEIIF (Cong et al., 2026)	PSNR↑	27.20	28.57	27.42	27.73	2.60	0.12	26.95	28.75	27.08	27.59	2.60	0.10
	SSIM↑	0.822	0.848	0.731	0.800			0.808	0.907	0.859	0.858		
3DGS (Kerbl et al., 2023)	PSNR↑	23.94	24.88	27.82	25.55	11.09	386.77	23.99	26.91	21.60	24.16	5.52	439.05
	SSIM↑	0.790	0.866	0.919	0.858			0.854	0.912	0.835	0.867		
	LPIPS↓	0.218	0.267	0.132	0.206			0.173	0.107	0.234	0.172		
Scaffold-GS (Lu et al., 2024)	PSNR↑	25.03	27.56	30.74	27.78	2.02	323.73	22.84	25.07	19.42	22.44	5.19	300.08
	SSIM↑	0.858	0.906	0.936	0.900			0.841	0.899	0.807	0.849		
	LPIPS↓	0.135	0.164	0.096	0.131			0.183	0.100	0.239	0.174		
LF-GS (Huang et al., 2025)	PSNR↑	21.83	27.31	32.57	27.24	17.51	340.45	21.88	26.51	19.89	22.76	16.20	387.20
	SSIM↑	0.757	0.907	0.953	0.872			0.772	0.897	0.829	0.833		
	LPIPS↓	0.220	0.193	0.131	0.181			0.281	0.132	0.203	0.205		
SuperGS (Xie et al., 2025)	PSNR↑	33.33	36.30	36.36	35.33	12.79	249.17	36.12	35.96	31.88	34.65	16.46	195.36
	SSIM↑	0.949	0.958	0.968	0.959			0.964	0.974	0.946	0.961		
	LPIPS↓	0.073	0.118	0.061	0.084			0.065	0.046	0.105	0.072		
SplatSuRe (Asthana et al., 2026)	PSNR↑	30.75	30.68	37.47	32.97	11.04	278.07	38.20	36.11	33.30	35.87	5.39	318.29
	SSIM↑	0.887	0.924	0.975	0.929			0.956	0.962	0.939	0.952		
	LPIPS↓	0.158	0.223	0.068	0.150			0.101	0.084	0.172	0.119		
PlenoGS (Ours)	PSNR↑	35.11	34.16	37.20	35.49	27.56	100.24	38.56	37.03	33.36	36.32	42.16	79.57
	SSIM↑	0.959	0.941	0.974	0.958			0.974	0.979	0.958	0.970		
	LPIPS↓	0.060	0.114	0.050	0.075			0.049	0.036	0.091	0.059		

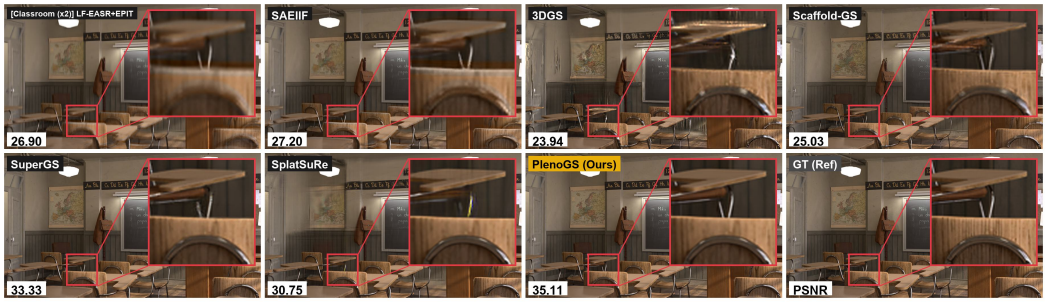

Figure 2: Comparison on the *Classroom* for $9 \times 9 \rightarrow 15 \times 15$ angular and $\times 2$ spatial SR. Zoomed views show PlenoGS preserves sharp tabletop edges and thin chair details, while other methods exhibit over-smoothing, distortion, or artifacts.

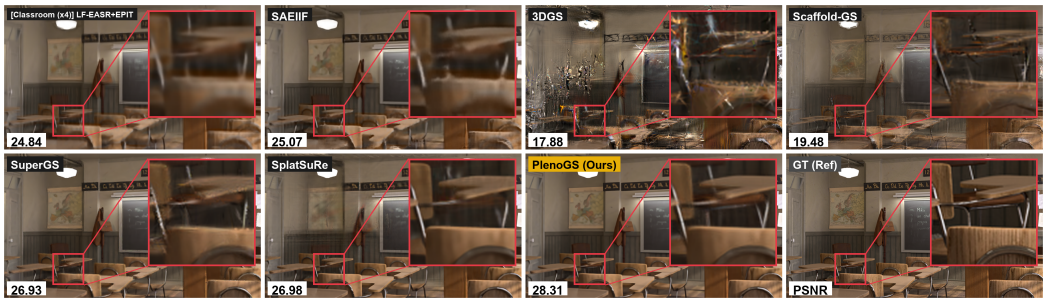

Figure 3: Comparison on the *Classroom* for $6 \times 6 \rightarrow 15 \times 15$ angular and $\times 4$ spatial SR. Zoomed-in views show PlenoGS better preserves thin structures and sharp boundaries while reducing artifacts and geometric distortions.

anchor feature mean/log-variance, and a lightweight residual color decoder (LR 1×10^{-3} for LIR + Stage-2 features; 3×10^{-5} for the decoder). Add new anchors every 100 iterations until iteration 2,000, then run 1,000 iterations of uncertainty-based refinement. Hyperparameters: voxel size 0.03, per-view candidate limit 1500, $\tau_{\text{err}} = 0.1$, $\tau_{\text{vote}} = 3$, max 1000 new anchors. Refinement uses

Table 2: Quantitative evaluation on the highly sparse $6\times6 \rightarrow 15\times15$ angular and extreme $\times4$ spatial SR task. Best and second-best results are highlighted in RED and YELLOW, respectively.

Method	Metric	FIT Dataset (Chlubna et al., 2024) (Synthetic)						Stanford Dataset (Vaish & Adams, 2008) (Real-World)					
		Classrm	Build.	Macro	**Avg**	Params↓	FPS↑	Lego	Jewel.	Tarot	**Avg**	Params↓	FPS↑
LF-EASR (Liu et al., 2023) + EPIT (Liang et al., 2023)	PSNR↑	24.84	24.99	24.22	24.68	9.76	7.91	25.63	25.75	24.67	25.35	7.44	7.91
	SSIM↑	0.720	0.790	0.617	0.709			0.746	0.846	0.768	0.787		
SAEIIF (Cong et al., 2026)	PSNR↑	25.07	26.48	25.32	25.62	2.60	0.72	26.01	26.97	25.12	26.04	2.60	0.46
	SSIM↑	0.741	0.810	0.672	0.741			0.757	0.878	0.783	0.806		
3DGS (Kerbl et al., 2023)	PSNR↑	17.88	19.29	19.52	18.90	8.79	399.13	13.57	21.82	14.40	16.60	4.57	456.73
	SSIM↑	0.490	0.678	0.625	0.598			0.468	0.816	0.550	0.611		
	LPIPS↓	0.480	0.427	0.402	0.436			0.494	0.212	0.473	0.393		
Scaffold-GS (Lu et al., 2024)	PSNR↑	19.48	20.95	23.45	21.29	1.38	336.03	14.22	20.16	13.00	15.80	3.24	318.15
	SSIM↑	0.604	0.741	0.728	0.691			0.478	0.771	0.518	0.589		
	LPIPS↓	0.367	0.344	0.324	0.345			0.509	0.219	0.489	0.406		
LF-GS (Huang et al., 2025)	PSNR↑	14.78	20.65	24.61	20.01	15.02	364.73	12.78	20.94	13.53	15.75	19.43	387.76
	SSIM↑	0.411	0.761	0.817	0.663			0.449	0.799	0.568	0.605		
	LPIPS↓	0.508	0.338	0.290	0.379			0.497	0.225	0.424	0.382		
SuperGS (Xie et al., 2025)	PSNR↑	26.93	30.56	29.90	29.13	11.81	262.23	26.45	27.85	24.48	26.26	12.65	286.35
	SSIM↑	0.819	0.874	0.850	0.848			0.834	0.911	0.813	0.853		
	LPIPS↓	0.243	0.266	0.208	0.239			0.297	0.151	0.310	0.253		
SplatSuRe (Asthana et al., 2026)	PSNR↑	26.98	27.28	30.90	28.39	9.69	282.57	32.17	32.24	27.71	30.70	5.16	328.33
	SSIM↑	0.785	0.855	0.873	0.837			0.911	0.933	0.877	0.907		
	LPIPS↓	0.283	0.312	0.196	0.264			0.186	0.124	0.258	0.189		
PlenoGS (Ours)	PSNR↑	28.31	31.72	30.61	30.21	7.45	185.01	31.24	31.63	26.85	29.91	8.29	184.06
	SSIM↑	0.852	0.882	0.866	0.867			0.910	0.943	0.871	0.908		
	LPIPS↓	0.208	0.265	0.188	0.220			0.183	0.106	0.240	0.176		

uncertainty threshold 0.02 and adds up to 1024 child anchors per step. Supervision uses pseudo-HR images (e.g., LIIF). The loss is

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + 0.2\mathcal{L}_{\text{ssim}} + 0.01\mathcal{L}_{\text{vol}} + 0.05\mathcal{L}_{\text{lpips}}, \tag{12}$$

where $\mathcal{L}_{\text{rec}}$ is uncertainty-weighted. No additional image-gradient loss is used beyond the Stage-1 EPI constraint. Log-variance is bounded to avoid variance explosion, so no explicit variance penalty is needed. Experiments use NVIDIA RTX 4090 GPUs.

4.2 Main Results: Joint Spatial-Angular SR

We design two main experimental settings: a standard SR task ($9\times9 \rightarrow 15\times15$ angular, $\times2$ spatial) and a more challenging sparse-input, large-magnification task ($6\times6 \rightarrow 15\times15$ angular, $\times4$ spatial).

1) $9 \times 9 \rightarrow 15 \times 15$ *angular and* $\times2$ *spatial SR.* We evaluate the standard $9 \times 9 \rightarrow 15 \times 15$ angular and $\times2$ spatial SR setting. Table 1 reports quantitative results, and Fig. 2 shows a qualitative comparison on *Classroom*. Table 1 shows that among 3D methods, PlenoGS achieves the best average PSNR and lowest LPIPS on FIT, while its SSIM (0.958) is slightly below SuperGS (0.959). On Stanford, PlenoGS achieves the best average PSNR, SSIM, and LPIPS. In Fig. 2, zoomed-in views show clear reconstruction differences. LF-EASR+EPIT and SAEIIF over-smooth, blurring thin chair legs. 3DGS distorts the tabletop and chair; Scaffold-GS still produces blurred, inaccurate boundaries. SuperGS is sharper but slightly deforms thin structures. SplatSuRe is cleaner but adds color and edge artifacts around the chair leg. PlenoGS best preserves the straight tabletop edge and slender chair structures.

2) $6 \times 6 \rightarrow 15 \times 15$ *angular and* $\times4$ *spatial SR.* We evaluate the challenging $6 \times 6 \rightarrow 15 \times 15$ angular and $\times4$ spatial SR setting. Quantitative results are in Table 2, with visual comparisons in Fig. 3. Under $\times4$, PlenoGS achieves the best average PSNR, SSIM, and LPIPS on FIT. On Stanford, SplatSuRe attains the highest average PSNR (30.70 dB), while PlenoGS yields the best SSIM (0.908) and LPIPS (0.176). Fig. 3 shows PlenoGS produces more faithful novel-view SR than prior light-field and Gaussian-based methods: LF-EASR+EPIT and SAEIIF over-smooth high-frequency details, and Gaussian-based methods either introduce structural artifacts (3DGS) or blur/distort local geometry (Scaffold-GS, SuperGS). SplatSuRe is cleaner but still misses fine structures. PlenoGS reconstructs sharper boundaries and thin structures while better preserving scene geometry.

Overall, combining explicit stable 3D geometry with scale-aware continuous refinement improves fidelity and perceptual quality on both synthetic and real wide-baseline scenes.

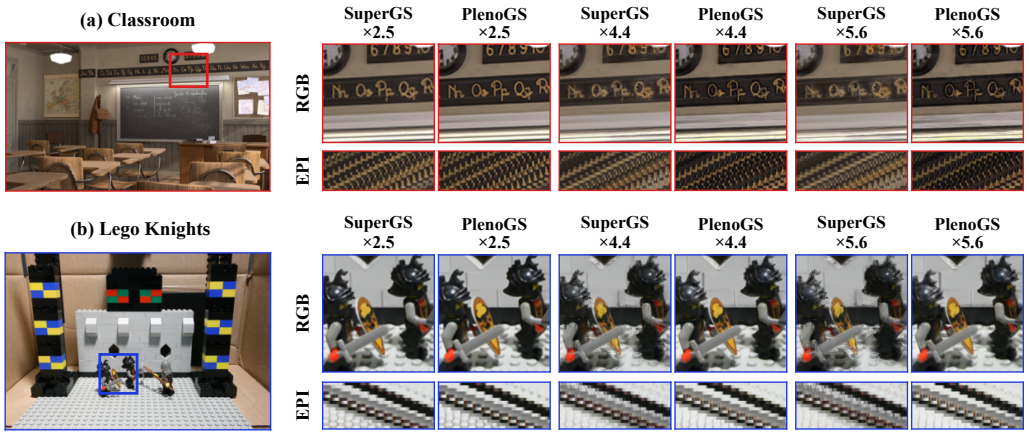

Figure 4: Qualitative comparison of arbitrary-scale super-resolution at $\times 2.5$, $\times 4.4$, and $\times 5.6$. Compared with SuperGS, PlenoGS maintains sharper RGB details and clearer boundaries at non-integer scales. Its EPI slices show more continuous, better-aligned structures.

Table 3: Ablation study of PlenoGS components on the *Classroom* scene for $\times 2$ spatial SR.

Metrics	SuperGS-S1 + LIR	w/o EPI Loss	w/o LIR	PlenoGS
PSNR↑	34.72	34.56	33.94	**35.11**
SSIM↑	0.956	0.957	0.955	**0.959**
LPIPS↓	0.066	0.062	0.064	**0.060**

4.3 Discussions

1) *Qualitative Demonstration of Resolution-Flexible Rendering.* To evaluate resolution flexibility, we test SuperGS and PlenoGS at unseen non-integer magnifications and report visual results (no ground truth at arbitrary scales). In Fig. 4, we query the feature field at $\times 2.5$, $\times 4.4$, and $\times 5.6$, showing rendered crops and horizontal EPIs. PlenoGS produces more continuous outputs with smoother edges and EPI trajectories, while the discrete SuperGS can introduce artifacts. Without retraining, both methods render plausible details, but PlenoGS better preserves sharp structures and smoother EPIs, indicating stronger cross-view consistency at arbitrary scales.

2) *Ablation Studies.* We validate the key design modules with an ablation on *Classroom* for $9 \times 9 \rightarrow 15 \times 15$ angular and $\times 2$ spatial SR, as shown in Table 3. **1. SuperGS-S1 + LIR** To test whether LIR alone suffices, we insert the SuperGS Stage-1 model into an explicit 3DGS and apply LIR-style residual refinement to baked Gaussians, which yields only minor local gains. **2. w/o EPI Loss** Removing the Stage-1 EPI loss ($\mathcal{L}_{\text{EPI}}$) weakens cross-view geometric and ray consistency, leading to more severe floating artifacts in novel views. **3. w/o LIR** Removing LIR worsens reconstruction, showing that local implicit refinement is key to recovering high-frequency details.

5 Conclusion

We revisit plenoptic super-resolution: reconstructing a high-resolution spatial-angular scene representation from sparse, low-resolution views. We propose PlenoGS, a hybrid approach that combines light-field structural priors for cross-view coherence with an explicit 3D Gaussian representation for geometry- and occlusion-aware rendering and continuous view synthesis. PlenoGS uses a two-stage pipeline: in stage 1, depth-guided initialization with EPI-based regularization to stabilize the 3D representation, and in stage 2, geometry-preserving refinement with scale-aware continuous anchor decoding to recover fine details. This enables flexible upscaling while maintaining multi-view consistency and reducing occlusion and grid artifacts common in light-field and 3DGS methods. Experiments on synthetic and real wide-baseline light fields show better spatial detail, angular consistency, and novel-view quality, demonstrating the value of coupling light-field structure with explicit 3D geometry for high-fidelity plenoptic reconstruction.

AI USE STATEMENT

Generative AI tools assisted with manuscript preparation tasks, including literature search and organization, analysis of experimental results, and revising and polishing parts of the text. The authors reviewed and, where applicable, verified all AI-assisted content against experiments, original sources, or other materials. AI tools were not authors and did not determine the study's conclusions. The authors take full responsibility for the methods, experiments, results, and claims in the paper.

ETHICS STATEMENT

This work focuses on light-field super-resolution and 3D scene representation and does not involve human-subject studies, private personal information, or sensitive data. All experiments are conducted on publicly available research datasets and follow their intended research use. We are not aware of significant ethical risks introduced by the proposed method.

REPRODUCIBILITY STATEMENT

We provide the main architectural design, training objectives, dataset settings, evaluation protocols, and implementation details in the main paper and supplementary material. Additional implementation details and experimental configurations are included where relevant to facilitate understanding and reproduction of the reported experiments.

REFERENCES

Edward H. Adelson and James R. Bergen. The plenoptic function and the elements of early vision. In Michael S. Landy and J. Anthony Movshon (eds.), *Computational Models of Visual Processing*, pp. 3–20. MIT Press, Cambridge, MA, 1991.

Pranav Asthana, Alex Hanson, Allen Tu, Tom Goldstein, Matthias Zwicker, and Amitabh Varshney. SplatSuRe: Selective super-resolution for multi-view consistent 3D gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11840–11849, June 2026.

Robert C Bolles, H Harlyn Baker, and David H Marimont. Epipolar-plane image analysis: An approach to determining structure from motion. *International Journal of Computer Vision*, 1(1): 7–55, 1987.

Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. TensoRF: Tensorial radiance fields. In *European Conference on Computer Vision*, pp. 333–350. Springer, 2022.

Rongshan Chen, Hao Sheng, Da Yang, Ruixuan Cong, Zhenglong Cui, and Sizhe Wang. Multiplane-based cross-view interaction mechanism for robust light field angular super-resolution. *IEEE Transactions on Visualization and Computer Graphics*, 31(10):8159–8173, 2025a. doi: 10.1109/TVCG.2025.3564643.

Sili Chen, Hengkai Guo, Shengnan Zhu, Feihu Zhang, Zilong Huang, Jiashi Feng, and Bingyi Kang. Video depth anything: Consistent depth estimation for super-long videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22831–22840, June 2025b.

Yinbo Chen, Sifei Liu, and Xiaolong Wang. Learning continuous image representation with local implicit image function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8628–8638, 2021.

Tomáš Chlubna, Tomáš Milet, and Pavel Zemčík. Lightweight all-focused light field rendering. *Computer Vision and Image Understanding*, 244:104031, 2024.

Jaeyoung Chung, Jeongtaek Oh, and Kyoung Mu Lee. Depth-regularized optimization for 3D Gaussian splatting in few-shot images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 811–820, 2024. doi: 10.1109/CVPRW63382.2024.00086.

Ruixuan Cong, Yu Wang, Mingyuan Zhao, Da Yang, Rongshan Chen, and Hao Sheng. Rethinking the upsampling process in light field super-resolution with spatial-epipolar implicit image function. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7559–7569, October 2025.

Ruixuan Cong, Hao Sheng, Yu Wang, Da Yang, Zhenglong Cui, Weifeng Lyv, Yi Zhang, and Wei Ke. Learning three-domain implicit image function for arbitrary-scale light field super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(8):9397–9414, 2026. doi: 10.1109/TPAMI.2026.3679405.

Ben Fei, Jingyi Xu, Rui Zhang, Qingyuan Zhou, Weidong Yang, and Ying He. 3D Gaussian splatting as a new era: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 31(8):4429–4449, 2025.

Xiang Feng, Yongbo He, Linxi Chen, Yan Yang, Chengkai Wang, Yifei Chen, Yixuan Zhong, Zhenzhong Kuang, Jiajun Ding, Xufei Yin, and Yanming Zhu. SRGS: Super-resolution 3D gaussian splatting, 2026. URL https://arxiv.org/abs/2404.10318v3.

Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen. The lumigraph. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, pp. 43–54, New York, NY, USA, 1996. ACM. doi: 10.1145/237170.237200.

Xuecai Hu, Haoyuan Mu, Xiangyu Zhang, Zilei Wang, Tieniu Tan, and Jian Sun. Meta-SR: A magnification-arbitrary network for super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1575–1584, 2019.

Yixu Huang, Rui Zhong, Ségolène Rogge, and Adrian Munteanu. LF-GS: 3D Gaussian splatting for view synthesis of multi-view light field images. *IEEE Signal Processing Letters*, 32:3555–3559, 2025.

Nima Khademi Kalantari, Ting-Chun Wang, and Ravi Ramamoorthi. Learning-based view synthesis for light field cameras. *ACM Transactions on Graphics*, 35(6):1–10, 2016.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):1–14, 2023.

Jaewon Lee and Kyong Hwan Jin. Local texture estimator for implicit representation function. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1929–1938, 2022.

Joo Chan Lee, Daniel Rho, Xiangyu Sun, Jong Hwan Ko, and Eunbyung Park. Compact 3D Gaussian representation for radiance field. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21719–21728, 2024.

Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, pp. 31–42, New York, NY, USA, 1996. ACM. doi: 10.1145/237170.237199.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using swin transformer. In *2021 IEEE/CVF international conference on computer vision workshops (ICCVW)*, pp. 1833–1844. IEEE, 2021.

Zhengyu Liang, Yingqian Wang, Longguang Wang, Jungang Yang, and Shilin Zhou. Light field image super-resolution with Transformers. *IEEE Signal Processing Letters*, 29:563–567, 2022.

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 *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12376–12386, October 2023.

Gaosheng Liu, Huanjing Yue, Jiamin Wu, and Jingyu Yang. Efficient light field angular super-resolution with sub-aperture feature learning and macro-pixel upsampling. *IEEE Transactions on Multimedia*, 25:6588–6600, 2023.

Tao Lu, Mulin Yu, Linning Xu, Yuanbo Xiangli, Limin Wang, Dahua Lin, and Bo Dai. Scaffold-GS: Structured 3D Gaussians for view-adaptive rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20654–20664, 2024.

Yifan Mao, Xinpeng Huang, Yilei Chen, Deyang Liu, Ping An, and Sanghoon Lee. Learning a domain-specialized network for light field spatial-angular super-resolution. *IEEE Transactions on Visualization and Computer Graphics*, 32(2):2127–2140, 2026. doi: 10.1109/TVCG.2025.3644930.

Jiahan Meng, Chi Yang, Chuanjun Li, Lili Lin, and Wenhui Zhou. Light field gaussian splatting representation and joint spatial-angular super-resolution for low-altitude scenarios. *Journal of Beijing University of Aeronautics and Astronautics*, 52(9):3189–3201, 2026. doi: 10.13700/j.bh.1001-5965.2025.0797.

Nan Meng, Xiaofei Wu, Jianzhuang Liu, and Edmund Lam. High-order residual network for light field super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 11757–11764, 2020.

Nan Meng, Hayden K.-H. So, Xing Sun, and Edmund Y. Lam. High-dimensional dense residual convolutional neural network for light field reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(3):873–886, 2021. doi: 10.1109/TPAMI.2019.2945027.

Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics*, 38(4):1–14, 2019.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision*, pp. 405–421, 2020.

Kaushik Mitra and Ashok Veeraraghavan. Light field denoising, light field superresolution and stereo camera based refocussing using a GMM light field patch prior. In *CVPR Workshops*, pp. 22–28, 2012.

Yu Mo, Yingqian Wang, Chao Xiao, Jungang Yang, and Wei An. Dense dual-attention network for light field image super-resolution. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(7):4431–4443, 2022. doi: 10.1109/TCSVT.2021.3121679.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 41(4):1–15, 2022.

Ren Ng, Marc Levoy, Mathieu Brédif, Gene Duval, Mark Horowitz, and Pat Hanrahan. Light field photography with a hand-held plenoptic camera. Technical Report CSTR 2005-02, Stanford University Computer Science Department, April 2005.

Paul-Edouard Sarlin, Cesar Cadena, Roland Siegwart, and Marcin Dymczyk. From coarse to fine: Robust hierarchical localization at large scale. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12716–12725, 2019.

Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems*, volume 33, pp. 7462–7473, 2020.

Yuqi Sun, Ri Cheng, Bo Yan, and Shili Zhou. Space-angle super-resolution for multi-view images. In *Proceedings of the ACM International Conference on Multimedia*, pp. 750–759, 2021.

Vaibhav Vaish and Andrew Adams. The (new) stanford light field archive. Stanford Graphics Laboratory, Stanford University, 2008. URL `https://lightfield.stanford.edu/data/LF/`. Accessed: 2026-09-23.

Giuseppe Valenzise, Martin Alain, Emin Zerman, and Cagri Ozcinar (eds.). *Immersive Video Technologies*. Academic Press, 2023.

Longguang Wang, Yingqian Wang, Zaiping Lin, Jungang Yang, Wei An, and Yulan Guo. Learning a single network for scale-arbitrary super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4801–4810, October 2021.

Sizhe Wang, Hao Sheng, Rongshan Chen, Da Yang, Zhenglong Cui, Ruixuan Cong, and Zhang Xiong. Spatial–angular–epipolar transformer for light field spatial and angular super-resolution. *Displays*, 85:102816, 2024.

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*, pp. 290–308. Springer, 2020.

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*, 45(1):425–443, 2023.

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. doi: 10.1109/TIP.2018.2834819.

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, 2014. doi: 10.1109/TPAMI.2013.147.

Gaochang Wu, Belen Masia, Adrian Jarabo, Yuchen Zhang, Liangyong Wang, Qionghai Dai, Tianyou Chai, and Yebin Liu. Light field image processing: An overview. *IEEE Journal of Selected Topics in Signal Processing*, 11(7):926–954, 2017.

Gaochang Wu, Yebin Liu, Lu Fang, Qionghai Dai, and Tianyou Chai. Light field reconstruction using convolutional network on EPI and extended applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(7):1681–1694, 2019.

Shiyun Xie, Zhiru Wang, Yinghao Zhu, Xu Wang, Chengwei Pan, and Xiwang Dong. SuperGS: Consistent and detailed 3D super-resolution scene reconstruction via Gaussian splatting. *arXiv preprint arXiv:2505.18649*, 2025.

Jingyu Yang, Sheng Shen, Huanjing Yue, and Kun Li. Implicit transformer network for screen content image continuous super-resolution. *Advances in Neural Information Processing Systems*, 34:13304–13315, 2021.

Henry Wing Fung Yeung, Junhui Hou, Jie Chen, Yuk Ying Chung, and Xiaoming Chen. Fast light field reconstruction with deep coarse-to-fine modeling of spatial-angular clues. In *European Conference on Computer Vision*, pp. 138–154. Springer, 2018.

Youngjin Yoon, Hae-Gon Jeon, Donggeun Yoo, Joon-Young Lee, and In So Kweon. Learning a deep convolutional network for light-field image super-resolution. In *Proceedings of the IEEE international conference on computer vision workshops*, pp. 24–32, 2015.

Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, and Andreas Geiger. Mip-Splatting: Alias-free 3D Gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19447–19456, 2024.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018.

Zhoutong Zhang, Yebin Liu, and Qionghai Dai. Light field from micro-baseline image pair. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3800–3809, 2015.

Zehao Zhu, Zhiwen Fan, Yifan Jiang, and Zhangyang Wang. FSGS: Real-time few-shot view synthesis using Gaussian splatting. In *European Conference on Computer Vision*. Springer, 2024.