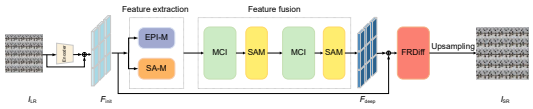
OEE | 光电工程 Opto-Electronic Engineering

Article

2026 年 4 月, 第 53 卷, 第 4 期

融合 Mamba 子空间扫描与扩散模型的光场图像超分辨率方法



王　飞*, 邹　希, 高建邦, 高国旺

西安石油大学电子工程学院, 陕西 西安 710065

摘要：针对光场超分辨中高频细节易丢失与跨视角一致性难保持的问题，本文提出一种融合 Mamba 子空间扫描与扩散生成的重建框架。采用 EPI-Mamba 与 SA-Mamba 两个子空间的双向扫描，分别高效捕获光场的 EPI 结构与空间-角度相关性；随后将两路扫描得到的序列化特征输入多尺度交互融合模块 (MCI) 实现深层信息互补与耦合，经过空间–角度调制模块 (SAM) 对融合结果进行空间–角度双重标定，在此基础上引入频域增强机制，利用 FFT 特征对融合结果进行高频补偿，以强化特征关系并抑制无关或噪声信息。将增强后的特征输入到扩散模型去噪网络中，上采样之后得到超分辨率结果。实验结果表明，本方法在多个定量指标和视觉评估中均表现优越，在 2×任务中与其他方法相比获得了最高的分数 39.43/0.987，在更复杂的 4×任务中也获得了最高的分数 33.70/0.945，PSNR 比现有方法提升了 1.44 dB。在定性视觉效果中，细节保持、图像清晰度等取得了显著效果，尤其在高频纹理与结构保持方面展现出了明显优势。

关键词：光场；超分辨；Mamba；频域增强；扩散模型

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250263 | CSTR: 32245.14.oee.2026.250263 | 中图分类号：O436 | 文献标识码：A

王飞，邹希，高建邦，等. 融合 Mamba 子空间扫描与扩散模型的光场图像超分辨率方法 [J]. 光电工程，2026，**53**(4): 250263

Wang F, Zou X, Gao J B, et al. Light field super-resolution with Mamba subspace scanning and diffusion modeling[J]. *Opto-Electron Eng*, 2026, **53**(4): 250263

1　引　言

目前，光场图像超分辨重建算法主要分为两类：传统重建方法和基于深度学习的重建方法。传统方法侧重于通过优化重建过程中的图像处理技术来提高图像质量，而基于深度学习的方法则通过学习图像特征，自动提取并重建更高质量的图像，显著提升超分辨率重建的效果和适应性。Lim 等[1] 提出的算法通过利用亚像素偏移信息，并通过投影到凸集来提高光场图像的空间分辨率。Nava 等人[2] 将傅里叶切片变换与多视角深度估计相结合，进行超分辨率图像重建。Zhou 等[3] 则通过提取角度图像中的模糊和亚像素偏移信息，建立了基于这些信息的重建模型。Wang 等[4] 提出了一种新的视差与剪切位移映射函数，有效提升了光场图像的空间和角度分辨率。基于先验知识的方法则引入外部信息来约束重建过程，增强了超分辨率效果。Bishop 等[5] 在变分贝叶斯框架中加入了朗伯特表面和纹理保持先验，通过深度信息提高了光场图像的质量。Mitra 等[6] 利用高斯混合模型对光场

收稿日期：2025-09-03；修回日期：2025-12-05；录用日期：2025-12-10；网络出版日期：2026-04-25

基金项目：陕西省自然科学基础研究计划 一般项目 (青年) (2025JC-YBQN-935)；陕西省重点研发计划项目 (2024GX-YBYM-507)；西安石油大学研究生创新与实践能力培养计划立项项目 (YCS23214260)

*通信作者：王飞，200102@xsyu.edu.cn

图像块进行建模，并通过亚空间投影技术快速估计视差值，提升了重建质量。Rossi 等[7] 结合图正则化方法和不同视角的信息，增强了光场图像的几何结构，进一步改善了超分辨率效果。Farrugia 等[8] 则通过视差估计缩小匹配区域，提高了重建精度。尽管传统方法在一定程度上取得了成功，但由于对视差或深度信息的高精度要求，它们往往较为复杂且计算量较大，且泛化能力较弱，难以适应不同的光场场景。这些方法的性能通常不及深度学习方法，且计算速度较慢。

近年来，卷积神经网络的快速发展以及光场数据集的广泛应用，基于深度学习的算法通过自动提取特征和学习图像间的复杂关系，通过端到端的训练方式，从大量数据中获取有效的信息，实现了更高效、更精确的光场重建。Jin 等人[9] 提出了 LF-ATO 算法，通过将光场中每个视图与目标视图拼接，并从所有视图的嵌入表示中独立提取相应信息。该方法还引入了结构感知损失，对光场的 EPI 图像进行视差约束，以更好地保留光场结构，从而提升重建质量。Liu 等人[10] 则进一步提出，光场重建可划分为视图内信息的提取与融合，以及视图间信息的提取与融合两个过程。他们设计了 LF-IINet 算法，采用两个并行分支分别提取光场的全局特征和单视图的独立表示，并通过特征交互机制动态更新信息，这种分支交互结构能够有效保护光场的视差特性。LF-IINet 引入了 3D 空洞卷积池化金字塔结构，以增强对光场图像长距离依赖关系的建模能力，在处理大视差光场图像的重建任务时展现出了优异性能。HLFSSR[11] 则通过引入三种类型的二维特征提取模块，提升了对光场四维数据中空间和角度信息的处理能力，该方法的设计使得每个模块能够有效地捕捉不同维度的信息。Wang 等人[12] 提出的 Distg 算法通过分离空间和角度的光场信息，将其解耦为多个子空间，利用卷积层处理每个子空间的特征，具体结构如图 1 所示。这种方法不仅解决了空间和角度信息之间的耦合问题，还在考虑 EPI 平面的基础上，使用 EFE-H 与 EFE-V 卷积核增强光场几何信息的约束，进而优化了超分辨率重建效果。

现有的光场超分辨率重建方法在提升重建质量方面取得了一定进展，但仍面临一些局限性。许多方法过于侧重于利用不同视角的互补信息，忽视了单视图内部细节的潜力，未能充分提取和利用 4D 光场数据的深层次特征，对于纹理边缘与高频细节的重建效果也不理想。与此同时，复杂的模型设计和较高的计算开销增加了训练和推理的难度，尤其是基于多分支和 Transformer 网络的方法，往往导致较大的模型规模和较慢的推理速度。本文的主要贡献有：

1) 提出基于 Mamba 的光场子空间双向扫描建模框架。将 Mamba 结构引入光场超分辨率任务，设计了 EPI-Mamba 与 SA-Mamba 两种子空间扫描机制。通过双向扫描有效捕捉空间-角度全局依赖关系，减少了长序列建模带来的梯度消失和效率问题。

2) 设计多尺度交互融合模块 (Mamba-based cross interaction, MCI) 与空间–角度调制 (spatial–angular modulator, SAM) 模块。MCI 模块实现了 EPI 与 SA 特征的高效交互与融合，避免模型偏向单一维度特征。SAM 模块通过空间和角度双重注意力机制，进一步强化关键特征、抑制噪声与冗余信息，提升特征判别性。

3) 提出频域增强残差扩散生成模块 (frequency-domain enhanced residual diffusion, FRDiff)。不同于现有方法仅在图像域约束生成，本文在残差域和频域同时进行扩散建模。扩散过程由 Mamba 提取的全局–局部一致性特征引导，使得生成过程不仅依赖噪声预测，还保持跨视角一致性和结构稳定性。此设计使模型在重建中更加聚焦于纹理边缘与高频细节，有效缓解模糊与伪影问题，呈现更佳的视觉效果。结合 Mamba 特征对齐作为条件输入，引导扩散过程保持跨视角一致性和全局结构稳定。

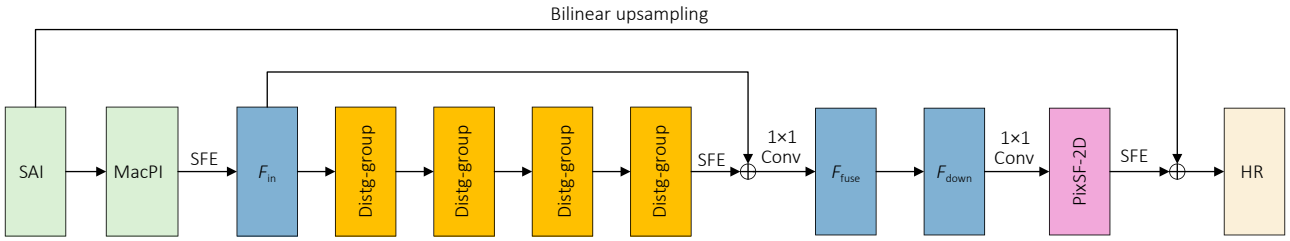


图 1 典型光场图像超分辨率重建结构图

Fig. 1 Schematic diagram of super-resolution reconstruction for typical light-field images

2 方 法

2.1 模型概述与整体框架

本文提出了一种结合 Mamba 子空间扫描建模[13] 与频域增强残差扩散生成的光场超分辨率新框架。传统方法往往依赖卷积或注意力机制在图像域进行直接建模，难以同时兼顾全局-局部依赖与高频细节恢复。而本文通过在子空间建模与扩散生成两个层面进行创新突破，既有效降低了光场 4D 全局建模的计算复杂度，又显著增强了纹理细节与跨视角一致性。

本文由特征提取与融合模块、频域增强残差扩散生成模块以及上采样模块构成。首先利用 EPI-Mamba 与 SA-Mamba 两个子空间双向扫描机制，分别高效地提取光场的 EPI 结构特征与空间-角度相关性。随后，将两路扫描得到的序列化特征输入到多尺度交互融合模块 (MCI)，以实现深层次的信息互补与耦合；并通过空间-角度调制模块 (SAM) 对融合后的特征进行空间与角度的双重标定。在此基础上，引入频域增强机制，利用 FFT 特征对融合结果进行高频补偿，从而强化特征关联并有效抑制无关或噪声信息。最终，将增强后的特征输入扩散模型的去噪网络，配合 Mamba 提供的全局–局部一致性特征，扩散模型在生成过程中能够更好地保持几何一致性和结构稳定性，最终实现高保真、高鲁棒性的光场重建。网络主要框架如图 2 所示，各个模块的网络结构与功能将在下文介绍。

2.2 特征提取与融合模块

2.2.1 Mamba 子空间双向扫描

光场数据可表示为 $I \in \mathbb{R}^{U\times V\times H\times W\times C}$，其中 (U,V) 表示角度维度，(H,W) 表示空间维度，C 为通道数。如果直接在完整的 4D 光场数据上进行全局建模，计算和存储代价极高。将光场特征分别展开到 EPI 子空间和 SA 子空间。以 EPI-Mamba 扫描为例，先固定一个角度维，例如固定 U，得到水平或垂直 EPI 切片 $\mathrm{EPI\_H}(u,v) = I(u,:,h:)$，并将其展平成一维序列 $S_{\mathrm{E}}\mathrm{PI} = \mathrm{Flatten}(\mathrm{EPI}) \in \mathbb{R}^{L\times C}$，$L$ 为序列长度，这样可以捕捉空间-角度交互信息。SA-Mamba 也是一样，沿空间维或角度维扫描，从而捕捉图像纹理细节。可以获得更强的全局空间-角度依赖，当在 EPI-H (或 EPI-W) 子空间上执行双向扫描时，可以同时整合空间与角度维度的垂直 (或水平) 信息，原本需要进行的四方向空间 (或角度) 扫描，能够被重新分解为两类双向扫描：其一是作用于空间 (或角度) 维度的扫描，其二则是沿 EPI 维度的扫描。使用 EPI-Mamba 和 SA-Mamba 两种扫描方式分别进行序列建模，然后再用交互机制进行融合。这种方式显著缩短了扫描路径，缓解了长序列建模中长期依赖带来的困难，同时仍能保持 4D 全局信息的完整性。

将这些序列输入到 MCI，把 EPI-Mamba 和 SA-Mamba 的输出进行交互融合；相当于双分支处理后，在全局层面融合不同维度的信息，避免只偏向某个维度再通过 SAM，对融合后的特征再做一次空间-角度的调制与增强，强调关键的信息、补充残差，进一步提升特征表达

2.2.2 多尺度交互融合模块

为了说明细节，本文以 EPI-Mamba 为例展开介绍，因为它与 SA-Mamba 在结构上基本一致，仅输入形式存在差异。EPI-Mamba 的结构如图 3 所示。

假设输入特征为第 $i-1$ 个 SAM 的输出 f_{l}^{i-1}，首先将其展开为 EPI-H 的 token 序列 $T_{\mathrm{h}}^{i} \in \mathbb{R}^{BVW\times UH\times C}$，其中 B 表示 batch 大小，随后，在 EPI-H 与 EPI-W 子空间上依次引入两个双向子空间扫描 (BiSS) 模块与卷积层，以实现跨 token 的全局依赖建模。具体计算如式 (1) 和式 (2) 所示。其中，EPI-W 序列 $T_{\mathrm{w}}^{i} \in \mathbb{R}^{BVH\times VW\times C}$ 由 $\hat{T}_{\mathrm{h}}^{i}$ 重新整形得到，$\hat{T}_{\mathrm{h}}^{i}$ 与 $\hat{T}_{\mathrm{w}}^{i}$ 均为增强后的特征表示。接着，将 $\hat{T}_{\mathrm{w}}^{i}$ 重塑为 f_{g}^{i}，

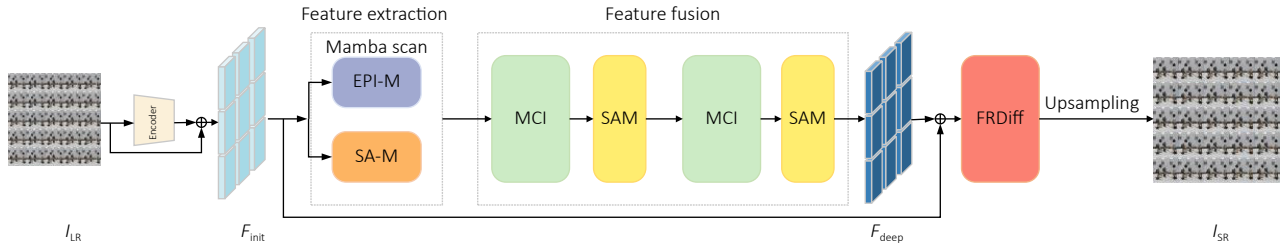


图 2 网络整体框架

Fig. 2 The overall framework of the network

作为第 i 个 MCI 模块的输出。

$$\hat{T}_{\mathrm{h}}^{i}=Conv(Biss(T_{\mathrm{h}}^{i}))+T_{\mathrm{h}}^{i}, \tag{1}$$

$$\hat{T}_{\mathrm{w}}^{i}=Conv(Biss(T_{\mathrm{w}}^{i}))+T_{\mathrm{w}}^{i}. \tag{2}$$

BiSS 模块的内部结构如图 4 所示，输入序列首先经过 LayerNorm[14]，再进入双向扫描模块 Bi-scan[15]，之后再加一层 LayerNorm 与通道注意力 (CA) 用于通道交互[16]。

在 SA-Mamba 中，只需将 EPI token 序列 $T_{\mathrm{h}}^{i},T_{\mathrm{w}}^{i}$ 分别替换为空间序列 $T_{\mathrm{s}}^{i}\in\mathbb{R}^{BUV\times HW\times C}$ 和角度序列 $T_{\mathrm{a}}^{i}\in\mathbf{R}^{BHW\times UV\times C}$。此外，在 EPI-Mamba 内部，让两次 BiSS 与卷积层共享权重，以便沿光场的 EPI 直线传递隐含信息[17]，同时也提升了网络效率。最终，EPI-Mamba 与 SA-Mamba 的结合，使得来自空间–角度维度和 EPI 维度的互补信息能够充分融合，从而实现高效的 4D 全局关系建模。

2.2.3 空间-角度调制模块

在完成 EPI-Mamba 与 SA-Mamba 特征的跨维度交互融合后，得到的全局表示中仍可能存在冗余或特征分布不均衡的问题。为进一步提升特征的判别性与有效性，本文引入空间-角度调制模块 (SAM)。该模块的核心思想是对融合后的特征进行一次空间维度与角度维度的双重标定，从而强化关键依赖关系并抑制无关或噪声信息。具体结构如图 5 所示。

在经过 MCI 模块后得到的特征 $f_{\mathrm{g}}^{i}\in\mathbb{R}^{B\times U\times V\times H\times W\times C}$ 首先被重排至空间子空间，并通过一个卷积层后接 Sigmoid 激活函数 (σ)，生成空间注意力图 $Attn_{\mathrm{s}}$。⊙代表逐元素相乘，即每个对应元素相乘。利用逐元素乘法对 f_{g}^{i} 进行调制，得到空间增强特征：

$$Attn_{\mathrm{s}}=\sigma(Conv(f_{\mathrm{g}}^{i})), \tag{3}$$

$$\tilde{f}_{\mathrm{g}}^{i}=f_{\mathrm{g}}^{i}\odot Attn_{\mathrm{s}}+f_{\mathrm{g}}^{i}, \tag{4}$$

式中：$\tilde{f}_{\mathrm{g}}^{i}$ 表示经过空间维度自适应加权后的结果。在角度子空间上采用相同的流程，得到角度注意力图 $Attn_{\mathrm{a}}$。最终输出 f_{l}^{i} 可表示为：

$$f_{\mathrm{l}}^{i}=\tilde{f}_{\mathrm{g}}^{i}\odot Attn_{\mathrm{a}}+\tilde{f}_{\mathrm{g}}^{i}. \tag{5}$$

这样一来，SAM 模块能够在空间与角度两个层面分别施加注意力调制，从而进一步强化关键区域的特征表达，提升整体的表示能力。

2.3 频域增强残差扩散生成模块

2.2.1 模型输入与条件化

扩散去噪网络以当前时刻的噪声光场作为主输入，并将增强特征 f_{l}^{i} 作为条件特征，通过特征拼接或跨注意

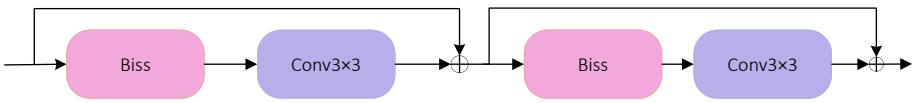


图 3 EPI-Mamba 具体结构

Fig. 3 The specific structure of EPI-Mamba

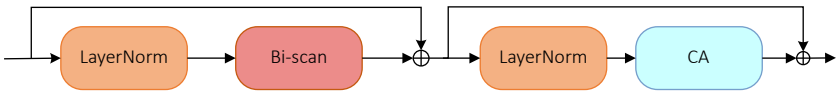


图 4 双向扫描模块

Fig. 4 Bidirectional scanning module

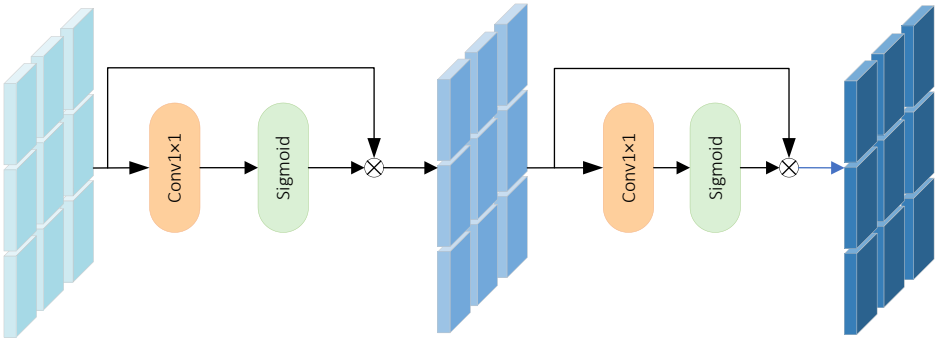


图 5 SAM 模块结构

Fig. 5 Structure of SAM modules

力机制注入到去噪网络的各层中，从而在去噪过程中引入频域与空间–角度双重条件。通过时间编码融入网络，用于指导不同时步的噪声抑制与细节恢复策略。该过程可表示如式 (6) 所示，其中：$\hat{\varepsilon}_\theta$表示网络预测的噪声；r_t表示在扩散过程第 t 步的加噪图像，由真实残差x_0逐步加入噪声后得到；t 为扩散步数，$UNet_\theta$为带参数的 UNet 去噪网络，其本质是一个学习从加噪残差图像r_t到真实噪声ε的映射函数，即真实残差逐步加噪后的结果；网络的参数集合记作θ，在训练过程中，通过最小化预测噪声与真实噪声的差异进行更新；$\gamma(t)$为对时间步 t 的嵌入，将离散的扩散步数映射为连续的向量表示，通过位置编码方式实现，使网络能够学习与扩散阶段相关的去噪模式。

$$\hat{\varepsilon}_\theta(r_t,t,f_1^i)=UNet_\theta(r_t,\gamma(t),f_1^i)\,. \tag{6}$$

2.3.2 FRDiff 模块

传统的扩散模型在超分任务中往往直接在图像域进行噪声预测，难以充分恢复高频细节。而光场图像的精细纹理与角度一致性高度依赖于频域特征，因此，本文提出频域增强残差扩散生成模块 (FRDiff)。该模块在残差域进行建模，并同时引入频域约束，从而在保持光场几何一致性的同时，增强对纹理细节的恢复能力。利用 Mamba 融合特征对齐作为条件输入，引导扩散过程保持跨视角一致性和全局结构稳定。

前向扩散过程：与标准扩散框架一致，将目标残差图像$r_0=I^{\mathrm{HR}}-\hat{I}^{\mathrm{LR}}$，也就是高分辨率与上采样低分辨率之差逐步加噪，得到扩散序列$q(r_t|r_0)=\mathcal{N}(\sqrt{\bar{\alpha}_t}r_0,(1-\bar{\alpha}_t)I)$，其中$t\in\{1,\cdots,T\}$，$\alpha_t$由余弦调度函数生成，$\bar{\alpha}=\prod_{i=1}^{t}\alpha_i$。

逆向生成过程：在反向生成中，模型需要预测噪声ε_θ，逐步恢复无噪残差：

$$r_{t-1}=\frac{1}{\sqrt{\alpha_t}}\left(r_t-\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\varepsilon_\theta(r_t,t,f_1^i)\right)+\sigma_t z\,. \tag{7}$$

融合模块输出的条件特征f_1^i包含 Mamba 子空间扫描与多尺度融合得到的全局–局部一致性信息。这样，扩散模型在生成过程中不仅依赖噪声预测，还被强制对齐到 Mamba 的特征空间，保证角度一致性。不同于现有方法仅在图像域约束生成，本文在残差域和频域同时进行监督。模型预测的残差$\hat{r}$不仅需逼近真实残差r_{gt}，还需在频率域保持一致性，如式 (8) 所示：

$$\mathcal{L}_{\mathrm{total}}=\lambda_{\mathrm{spatial}}\left\|r_{\mathrm{gt}}-\hat{r}\right\|_1+\lambda_{\mathrm{freq}}\left\|\mathcal{F}(r_{\mathrm{gt}})-\mathcal{F}(\hat{r})\right\|_1\,, \tag{8}$$

式中：$\mathcal{F}(\cdot)$为快速傅里叶变换 (FFT)。该设计使得模型在重建过程中更加关注纹理边缘与高频细节。通过仅对残差建模，降低生成难度，加速训练收敛。同时在空间与频域约束，提升高频细节还原能力。利用 Mamba 提取的全局–局部一致性特征作为扩散条件，保证跨视角一致性与全局结构稳定。在前述特征提取与多尺度频域融合模块获得增强特征后，本文在扩散模型框架下进行逐步去噪重建，以实现从高噪声初始输入逐步生成高质量的高分辨率光场。

2.4 上采样模型及损失函数设计

低分辨率光场图像输入到网络中后，通过 Pixel Shuffle 操作，将图像的空间分辨率提升到目标高分辨率，通道数从 C 个转换成r^2C个，经过 Pixel Shuffle 的输出图像进入卷积层，进一步提取特征并恢复图像细节，提升图像质量。最终得到的图像即为上采样后的高分辨率图像，它包含了更多的细节和高频信息。通过空间重排的方式提升分辨率，并通过卷积层恢复了图像的细节和纹理。不仅保证了计算效率，还在保持图像细节的同时，增强了光场图像的空间–角度一致性。

本文的损失函数采用 L1 损失，如式 (9) 所示，充分利用其在超分辨率任务中的鲁棒性和优越表现[18]。

$$L_{\mathrm{loss}}=\sum_{U,V}\sum_{H,W}\left(\left\|I^{\mathrm{SR}}(H,W,U,V)-I^{\mathrm{HR}}(H,W,U,V)\right\|_1\right). \tag{9}$$

3 实验结果

本文实验采用五个当前光场超分辨率研究中广泛使用的公开数据集，其中，HCInew[19] 和 HCIold[20] 为合成数据集，由计算机渲染生成，具有丰富的纹理和精确的几何结构，常用于评估模型在理想条件下的性能。EPFL[21]、INRIA[22] 和 STFgantry[23] 为真实数据集，通过光场相机或相机阵列采集，包含复杂多变的真实场景，有助于验证模型在实际应用中的鲁棒性。这些数据集的角度分辨率均为 9×9，即包含 81 个视角。为保证训练数据的质量并降低计算成本，仅选取每个光场图像的中心 5×5 视角 (共 25 张子孔径图像) 用于训练和测试，使所有数据的角度分辨率保持一致。在训练数据集的生成过程中，首先对光场图像进行裁剪。对于每个子孔径图像，采用滑动窗口裁剪方法，将图像裁剪为固定大小的小块。在放大系数为 2×和 4×时，裁剪大小分别为 64 pixel×64 pixel 和 128 pixel×128 pixel，步长设置为 32 或 64。裁剪完成后，训练集中的光场图像从 RGB 颜色空间转换为

YCbCr 空间。实验中仅提取 Y 通道用于超分辨率重建。由于人眼对亮度更为敏感，仅对 Y 通道进行超分辨率处理可以在减少计算量的同时保持图像的感知效果。

为了生成低分辨率图像，每个裁剪后的 Y 通道图像将进行双三次下采样操作。针对不同的超分辨率因子(如 2 倍或 4 倍)，分别进行 0.5×或 0.25×的下采样，以生成对应的 32×32 低分辨率小块图像。再使用本文方法会低分辨率光场进行超分辨率重建。对于测试数据集，首先对光场图像的子孔径图像 Y 通道进行双三次下采样，生成对应的低分辨率图像。使用空间超分辨率方法重建 Y 通道，以恢复其高分辨率细节。Cb 和 Cr 通道采用相同倍率的下采样处理，并通过上采样恢复至原始尺寸。最后将重建后的 Y 通道与处理后的 CbCr 通道结合，并转换回 RGB 颜色空间，以便对超分辨率重建结果进行直观对比，验证模型在不同放大倍率下的复原效果。

为了全面评估所提方法的效果，本文对比了当前最前沿的超分辨率重建技术，并进行了系统性分析。对比的算法包括单图像超分辨率方法 (如 VDSR[24]、EDSR[25]、RCAN[16])、传统光场超分辨率方法 (如 GB[26])，以及基于深度学习的光场超分辨率方法 (如 LFSSR[27]、resLF[28]、LFSSR-ATO[9]、MEG-Net[29]、LF-IINet[10]、MFSRNet[30])。

进行模型预训练时，优化器采用 Adam，将 batchsize 设为 4，学习率设为 2×10^{-4}，每 100 k 次迭代减少一半。本文对裁剪得到的光场图像块进行数据增强，采用水平翻转 (×2)、垂直翻转 (×2) 以及 90°旋转 (×3) 的方式，使模型能够学习到更加丰富的纹理和结构特征，提升在不同场景下的适应能力，并使用 NVIDIA RTX 4090 GPU 进行了 20 万次迭代训练。本文采用峰值信噪比 (peak signal-to-noise ratio, PSNR) 和结构相似性指数 (structural similarity index, SSIM) 作为定量评价指标，并重点关注 Y 通道的质量表现。

3.1 定量结果分析

本文在多个光场数据集上进行了实验，比较了不同超分辨率重建方法在 2×和 4×放大倍数下的定量结果。定量结果如表 1 所示，采用峰值信噪比 (PSNR) 和结构相似性指数 (SSIM) 作为评价指标，表格中以红色标注最优结果，蓝色标注次优结果，以便直观呈现各方法在不同数据集上的表现差异。实验覆盖多个光场数据集，能够全面反映各方法在不同场景和视差条件下的性能，并验证所提方法的稳定性与泛化能力。结果表明，本文方法在多个数据集上均取得了最优或次优表现，尤其在具有较大视差和复杂结构的场景中展现出更强的鲁棒性。相比现有主流方法，所提模型在更复杂的 4×放大倍数下始终保持较高的 PSNR 和 SSIM 得分，表现出良好的综合性能，体现了其在光场超分辨率中的应用潜力。

3.2 定性结果分析

通过在不同数据集和放大倍数下对重建结果进行图像展示与对比，直观展示了所提方法在细节恢复、纹理重建以及复杂场景处理方面的优势。在更具挑战性的 4×超分辨率任务中，本文提出的方法同样在复杂纹理和细节恢复方面表现出色。如图 6 所示，本文对比了不同方法在 INRIA 数据集的 Hublais Decoded 场景和 HCInew 数据集的 Bicycle 场景下的重建结果。

在 INRIA 数据集的 Hublais Decoded 场景中，本文提出的方法在复杂纹理和细节的恢复方面表现突出，能够准确重建字符等精细结构，有效避免边缘模糊与细节丢失。该结果充分体现了所提模型在高频信息重建方面的优势，使得超分辨率后的图像在视觉效果上更接近原始高分辨率图像。相比之下，VDSR[24]、EDSR[25]、RCAN[16] 和 resLF[28] 等方法在重建结果中普遍存在模糊与伪影，特别是在字符边缘和复杂纹理区域，细节恢复不完整，导致图像清晰度下降，真实感不足。在 HCInew 数据集的 Bicycle 场景中，所提方法同样展现了良好的细节还原能力，能够稳定保留物体的纹理结构。尤其是在望远镜、书籍等具有丰富边缘和纹理特征的区域，所提方法能准确重建其清晰的轮廓与表面细节。而其他方法在该场景下表现较弱，重建图像常伴随边缘模糊、纹理失真或细节缺失等问题，整体视觉质量下降，难以还原场景的真实结构。

定性结果表明，本文方法在多个场景中能够准确还原高频细节和复杂结构，显著提升图像的视觉质量。相较于传统的单图像超分辨率方法及现有光场超分方法，所提模型在细节保持、图像清晰度方面更具优势，同时有效抑制了模糊与伪影现象。为了更清晰地对比各方法在细节恢复上的差异，通过局部放大图突出显示关键区域的纹理和边缘信息，以便于直观评估不同方法在复杂结构和细微纹理重建上的表现。

3.3 计算效率分析

在 2×与 4×超分辨率任务中，本文对比了多种方法

表 1 不同方法在 2×和 4×超分辨率任务上的定量结果 (PSNR [dB]/SSIM)

Table 1 Quantitative results of different methods on 2× and 4× super-resolution tasks

Method	2×					
	EPFL	HCInew	HCIold	INRIA	STFgantry	Average
Bicubic	29.66/0.938	31.82/0.936	37.61/0.978	31.25/0.958	30.98/0.950	32.26/0.952
VDSR	32.50/0.960	34.37/0.956	40.61/0.987	34.43/0.974	35.54/0.979	35.49/0.971
EDSR	33.09/0.963	34.83/0.960	41.01/0.988	34.97/0.977	36.29/0.982	36.04/0.974
RCAN	33.16/0.964	34.98/0.960	41.05/0.988	35.01/0.977	36.33/0.983	36.11/0.974
GB	31.31/0.958	34.75/0.964	40.10/0.987	32.94/0.972	34.95/0.980	34.81/0.972
LFSSR	33.68/0.975	36.83/0.975	43.79/0.994	35.28/0.983	38.00/0.990	37.52/0.983
resLF	33.46/0.971	36.59/0.974	43.34/0.993	35.20/0.981	37.98/0.981	37.31/0.982
LFSSR-ATO	34.24/0.975	37.22/0.976	44.05/0.994	36.12/0.984	39.54/0.993	38.23/0.984
MEG-Net	34.31/0.977	37.37/0.978	44.06/0.994	36.09/0.985	38.75/0.991	38.12/0.985
LF-IINet	34.69/0.977	37.75/0.979	44.84/0.992	36.57/0.985	39.87/0.994	38.74/0.985
MFSRNet	34.78/0.979	37.85/0.978	44.71/0.995	36.59/0.986	40.51/0.994	38.89/0.986
OURS	35.54/0.978	38.39/0.980	44.96/0.995	37.13/0.985	41.12/0.995	39.43/0.987
Method	4×					
	EPFL	HCInew	HCIold	INRIA	STFgantry	Average
Bicubic	25.17/0.832	27.63/0.852	32.45/0.934	26.85/0.886	25.96/0.845	27.61/0.870
VDSR	27.25/0.878	29.31/0.883	34.81/0.952	29.19/0.921	28.51/0.901	29.81/0.907
EDSR	27.84/0.886	29.60/0.887	35.18/0.954	29.66/0.926	28.70/0.908	30.20/0.912
RCAN	27.88/0.886	29.64/0.888	35.23/0.954	29.77/0.927	28.92/0.912	30.29/0.913
GB	26.31/0.866	28.99/0.883	33.98/0.951	28.08/0.912	28.13/0.899	29.10/0.902
LFSSR	28.34/0.910	30.76/0.913	36.74/0.969	30.36/0.946	30.23/0.940	31.29/0.936
resLF	28.11/0.901	30.57/0.909	36.59/0.968	30.20/0.940	29.95/0.934	31.08/0.930
LFSSR-ATO	28.52/0.912	30.88/0.914	37.00/0.970	30.71/0.949	30.61/0.943	31.54/0.938
MEG-Net	28.73/0.916	31.08/0.917	37.25/0.972	30.65/0.949	30.74/0.945	31.69/0.940
LF-IINet	29.09/0.919	31.35/0.921	37.54/0.973	31.01/0.951	31.22/0.950	32.04/0.943
MFSRNet	29.32/0.920	31.42/0.922	37.64/0.973	31.51/0.952	31.42/0.942	32.26/0.942
OURS	30.70/0.913	33.49/0.926	37.63/0.974	32.23/0.955	34.47/0.959	33.70/0.945

在计算开销与重建质量方面的表现，指标包括参数规模 (params)、浮点运算量 (FLOPs) 以及 PSNR/SSIM。所有实验均基于 PyTorch 框架，在 AMD EPYC 9754 (128 核，主频 2.25 GHz)、512 GB 内存以及 NVIDIA GeForce RTX 4090 D GPU 的硬件环境下完成。表 2 汇总了具体的对比数据，其中计算量是以输入光场图像块大小 5×5×32×32 pixels 为基准，PSNR 和 SSIM 为五个测试数据集的平均值。结果展示中，最佳的 PSNR [dB]/SSIM 以红色突出，次优结果则以蓝色标注。

在 2×超分辨率实验中，本文方法的参数量仅为 2.08 M，计算量为 503.8 GFLOPs，同时在各对比模型中实现了最高的平均 PSNR/SSIM (39.43 dB/0.987)。在 4×任务下，参数规模上升至 2.49 M，复杂度为 548.5 GFLOPs，依旧获得了最佳的平均 PSNR/SSIM (33.70 dB/0.945)。

3.4 消融实验

3.4.1 MCI 模块的有效性

为验证多尺度交互融合模块 (MCI) 的有效性，在空

图 6 4×超分辨率结果的视觉效果对比

Fig. 6 Visual comparison of 4× super-resolution results

表 2 不同方法在 2×和 4×超分辨率任务中的参数量、计算量及平均 PSNR/SSIM 值

Table 2 Parameters, computational cost, and average PSNR/SSIM of different methods on 2× and 4× super-resolution tasks

Methods	2×			4×		
	Params/M	FLOPs/G	Ave.PSNR/SSIM	Params/M	FLOPs/G	Ave.PSNR/SSIM
EDSR	38.31	987.6	36.04/0.974	38.49	1021	30.20/0.912
RCAN	15.31	391.0	36.11/0.974	15.42	407.6	30.29/0.913
LFSSR	0.81	25.7	37.52/0.983	1.62	128.4	31.29/0.936
resLF	6.35	37.1	37.31/0.982	6.79	39.7	31.08/0.930
LFSSR-ATO	1.51	597.7	38.23/0.984	1.66	687.0	31.54/0.938
MEG-Net	1.69	48.4	38.12/0.985	1.77	102.2	31.69/0.940
LF-IINet	4.84	56.2	38.74/0.985	4.89	57.42	32.04/0.943
MFSRNet	1.22	45.6	38.89/0.986	1.25	107.6	32.26/0.942
OURS	2.08	503.8	39.43/0.987	2.49	548.5	33.70/0.945

间-角度调制模块、频域增强残差扩散生成等模块保持不变的情况下移除该模块，直接将 EPI-Mamba 与 SA-Mamba 提取的特征进行简单拼接后输入后续网络。实验结果表明，在缺少 MCI 的情况下，模型在 PSNR 和 SSIM 指标上均出现明显下降，如表 3 所示。这说明 MCI 在促进空间–角度双维度特征的深层交互与补充方面发挥了关键作用，能够有效避免模型偏向单一维度特征，从而增强特征的判别性和鲁棒性。主观视觉结果进一步表明，移除 MCI 后的重建结果存在纹理细节缺失和边缘模糊现象，如图 7 所示，验证了该模块在高频信息恢复与跨视角一致性保持中的重要性。

3.4.2 SAM 模块的有效性

为了评估空间–角度调制模块 (SAM) 的有效性，在完整模型的基础上移除该模块，仅使用 MCI 融合后的特征直接输入后续网络进行重建。实验结果如表 4 所示，缺少 SAM 的模型在 PSNR 和 SSIM 指标均有所下降，

表 3 移除 MCI 模块的网络变体在 2×上采样任务中的定量结果

Table 3 Quantitative results of the network variant without the MCI module on the 2× upsampling task

Methods	Params/M	EPFL	HCInew	HCIold	INRIA	STFgantry	Average
w/o MCI	2.06	35.26/0.977	37.88/0.978	44.91/0.995	36.96/0.984	40.83/0.993	39.16/0.985
MCI	2.08	35.54/0.978	38.39/0.980	44.96/0.995	37.13/0.985	41.12/0.995	39.43/0.987

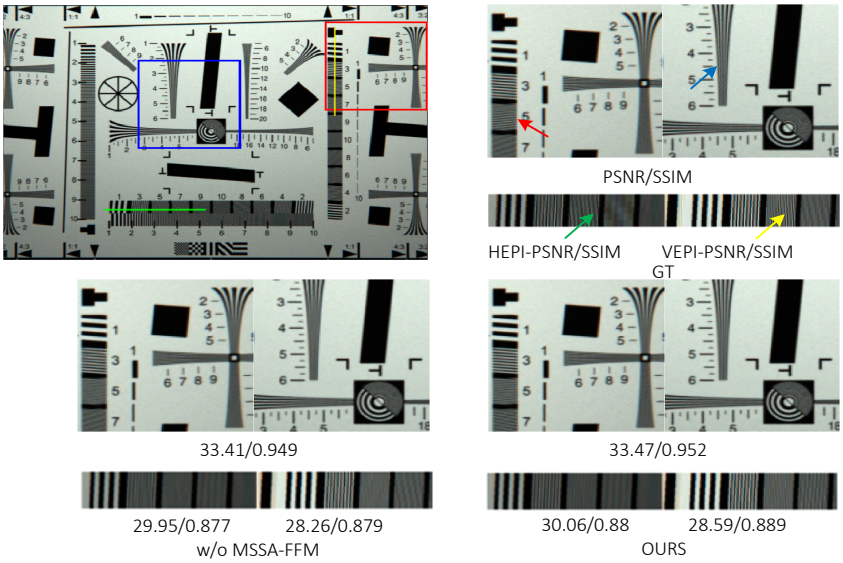


图 7 在 ISO Chart 1 场景的 2×上采样任务中，移除 MCI 的视觉影响

Fig. 7 Visual effects of MCI module removal on the 2× upsampling task in the ISO Chart 1 scene

尤其是在复杂场景下表现出明显的性能劣化，移除 SAM 会导致跨视角一致性减弱，重建图像的几何结构出现偏差。主观视觉结果也表明，在没有 SAM 的情况下，重建结果容易出现噪声增强和纹理混叠问题，如图 8 所示。这表明 SAM 模块在通过空间与角度双重注意力机制强化关键特征、抑制冗余信息方面发挥了关键作用，从而提升了模型在细节恢复和跨视角一致性保持上的性能。

3.4.3 FRDiff 模块的有效性

在移除 FRDiff 模块的设置条件下，对 SAM 模块输出的特征进行 Pixel Shuffle 操作开始上采样，然后通过常规残差学习对上采样结果进行逐像素回归，以弥补部分缺失的高频信息。这种方式虽然能够恢复基本的结构轮廓，但缺乏扩散建模所提供的多样化生成能力，也无法通过频域约束显式补偿高频成分，具体指标如表 5 所示。因此在细节还原、纹理清晰度和跨视角一致性方面存在明显不足，如图 9 所示。

4 结 论

本文提出了一种基于扩散模型的光场图像超分辨率重建网络，提高了光场图像的分辨率，通过 Mamba 双

表 4 移除 SAM 模块的网络变体在 2×上采样任务中的定量结果

Table 4 Quantitative results of the network variant without the SAM module on the 2× upsampling task

Methods	Params/M	EPFL	HCInew	HCIold	INRIA	STFgantry	Average
w/o SAM	2.07	35.48/0.977	38.23/0.979	44.93/0.995	37.05/0.984	41.08/0.993	39.13/0.985
SAM	2.08	35.54/0.978	38.39/0.980	44.96/0.995	37.13/0.985	41.12/0.995	39.43/0.987

表 5 移除 FRDiff 模块的网络变体在 2×上采样任务中的定量结果

Table 5 Quantitative results of the network variant without the FRDiff module on the 2× upsampling task

Methods	Params/M	EPFL	HCInew	HCIold	INRIA	STFgantry	Average
w/o FRDiff	2.04	35.12/0.976	37.85/0.978	44.87/0.994	36.76/0.984	40.68/0.993	39.10/0.985
FRDiff	2.08	35.54/0.978	38.39/0.980	44.96/0.995	37.13/0.985	41.12/0.995	39.43/0.987

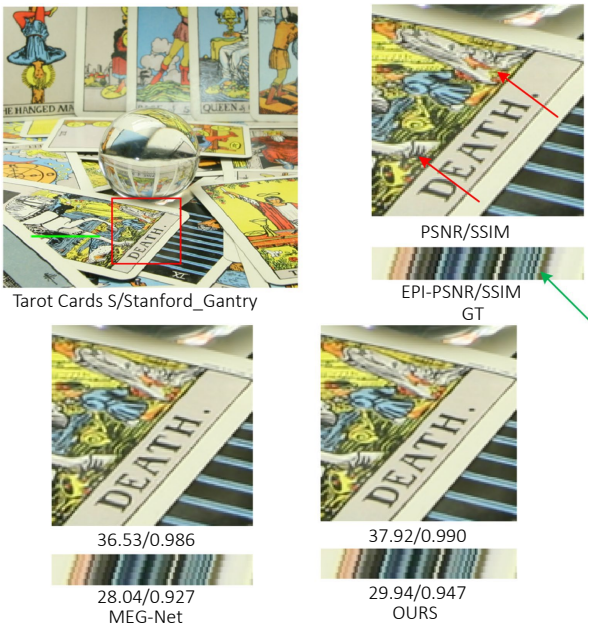


图 8 在 Tarot Cards S 场景的 2×上采样任务中，移除 SAM 的视觉影响

Fig. 8 Visual effects of SAM module removal on the 2× upsampling task in the Tarot Cards S scene

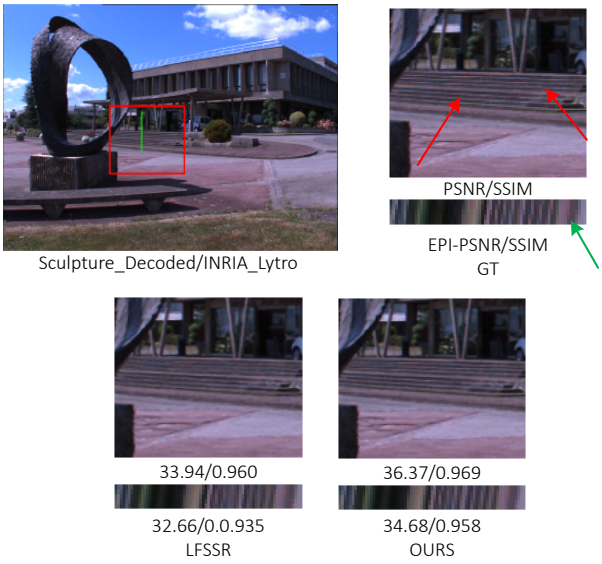


图 9 在 Sculpture_Decoded 场景的 2×上采样任务中，移除 FRDiff 的视觉影响

Fig. 9 Visual effects of FRDiff module removal on the 2× upsampling task in the Sculpture_Decoded scene

向扫描来提取特征，再用 MCI 和 SAM 模块交叠而成的特征融合模块进行特征融合，将得到的特征送入扩散模型，在扩散模型框架下进行逐步去噪重建，同时在空间与频域约束，提升了高频细节还原能力。实验结果表明，该网络在公开的真实光场数据集中表现出了较好的性能，在 4×超分辨重建情况下，PSNR 在 EPFL、HCInew、INRIA、STFgantry 数据集上取得了最高分数，SSIM 指标在 HCInew、STFgantry 取得了最高的分数，与其他方法相比，两项指标的平均值也取得了最高的分数。在定性分析中，视觉上也取得了较好的效果，在多个场景中能够准确还原高频细节和复杂结构，显著提升图像的视觉质量。相较于传统的单图像超分辨率方法及现有光场超分方法，所提模型在细节保持、图像清晰度方面更具优势，同时有效抑制了模糊与伪影现象。与其他算法相比，在最高计算复杂度为 528 G FLOPs 的状况下，依旧取得了最高平均分数 33.70/0.945。消融实验验证了各主要模块的有效性，在定量比较和视觉效果中均表现出较高水平。

作者贡献

王飞：数据分析和研究资料的收集；邹希：研究设计与论文初稿的撰写；高建邦：数据采集和实验操作；高国旺：对研究方法提出了重要的建议。

利益冲突

所有作者声明无利益冲突

参考文献

1. Lim J, Ok H, Park B, et al. Improving the spatail resolution based on 4D light field data[C]//*Proceedings of 2009 16th IEEE International Conference on Image Processing (ICIP)*, Cairo, 2009: 1173–1176. https://doi.org/10.1109/ICIP.2009.5413719.
2. Nava F P, Luke J P. Simultaneous estimation of super-resolved depth and all-in-focus images from a plenoptic camera[C]//*Proceedings of 2009 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video*, Potsdam, 2009: 1–4. https://doi.org/10.1109/3DTV.2009.5069675.
3. Zhou S B, Yuan Y, Su L J, et al. Multiframe super resolution reconstruction method based on light field angular images[J]. *Opt Commun*, 2017, **404**: 189−195.
4. Wang Y L, Hou G Q, Sun Z N, et al. A simple and robust super resolution method for light field images[C]//*Proceedings of 2016 IEEE International Conference on Image Processing (ICIP)*, Phoenix, 2016: 1459–1463. https://doi.org/10.1109/ICIP.2016.7532600.
5. Bishop T E, Favaro P. The light field camera: extended depth of field, aliasing, and superresolution[J]. *IEEE Trans Pattern Anal Mach Intell*, 2012, **34**(5): 972−986.
6. Mitra K, Veeraraghavan A. Light field denoising, light field superresolution and stereo camera based refocussing using a GMM light field patch prior[C]//*Proceedings of 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Providence, 2012: 22–28. https://doi.org/10.1109/CVPRW.2012.6239346.
7. Rossi M, Frossard P. Graph-based light field super-resolution[C]//*Proceedings of 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, Luton, 2017: 1–6. https://doi.org/10.1109/MMSP.2017.8122224.
8. Farrugia R A, Galea C, Guillemot C. Super resolution of light field images using linear subspace projection of patch-volumes[J]. *IEEE J Sel Top Signal Process*, 2017, **11**(7): 1058−1071.
9. Jin J, Hou J H, Chen J, et al. Light field spatial super-resolution via deep combinatorial geometry embedding and structural consistency regularization[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 2020: 2257–2266. https://doi.org/10.1109/CVPR42600.2020.00233.
10. Liu G S, Yue H J, Wu J M, et al. Intra-inter view interaction network for light field image super-resolution[J]. *IEEE Trans Multimedia*, 2023, **25**: 256−266.
11. Van Duong V, Huu T N, Yim J, et al. Light field image super-resolution network via joint spatial-angular and epipolar information[J]. *IEEE Trans Comput Imaging*, 2023, **9**: 350−366.
12. Wang Y Q, Wang L G, Wu G C, et al. Disentangling light fields for super-resolution and disparity estimation[J]. *IEEE Trans Pattern Anal Mach Intell*, 2023, **45**(1): 425−443.
13. Gao R S, Xiao Z Y, Xiong Z W. Mamba-based light field super-resolution with efficient subspace scanning[C]//*Proceedings of the 17th Asian Conference on Computer Vision*, Hanoi, 2024: 421–437. https://doi.org/10.1007/978-981-96-0917-8_24.
14. Ba J L, Kiros J R, Hinton G E. Layer normalization[Z]. arXiv: 1607.06450, 2016. https://arxiv.org/abs/1607.06450.
15. Zhu L H, Liao B C, Zhang Q, et al. Vision mamba: Efficient visual representation learning with bidirectional state space model[C]//*Proceedings of the 41st International Conference on Machine Learning*, Vienna, 2024: 1–14.
16. Zhang Y L, Li K P, Li K, et al. Image super-resolution using very deep residual channel attention networks[C]//*Proceedings of the 15th European Conference on Computer Vision*, Munich, 2018: 294–310. https://doi.org/10.1007/978-3-030-01234-2_18.
17. Liang Z Y, Wang Y Q, Wang L G, et al. Learning non-local spatial-angular correlation for light field image super-resolution[C]//*Proceedings of 2023 IEEE/CVF International Conference on Computer Vision*, Paris, 2023: 12376–12386. https://doi.org/10.1109/ICCV51070.2023.01137.
18. Anagun Y, Isik S, Seke E. SRLibrary: comparing different loss functions for super-resolution over various convolutional architectures[J]. *J Vis Commun Image Represent*, 2019, **61**: 178−187.
19. Rerabek M, Ebrahimi T. New light field image dataset[C]// 8th International Conference on Quality of Multimedia Experience (QoMEX). 2016
20. Honauer K, Johannsen O, Kondermann D, et al. A dataset and evaluation methodology for depth estimation on 4D light fields[C]//*Proceedings of the 13th Asian Conference on Computer Vision*, Taipei, China, 2016: 19–34. https://doi.org/10.1007/978-3-319-54187-7_2.
21. Wanner S, Meister S, Goldluecke B. Datasets and benchmarks for

densely sampled 4D light fields[C]//*Proceedings of the Vision, Modeling, and Visualization (2013)*, Lugano, 2013: 225–226.
22. Le Pendu M, Jiang X R, Guillemot C. Light field inpainting propagation via low rank matrix completion[J]. *IEEE Trans Image Process*, 2018, **27**(4): 1981−1993.
23. Vaish V, Adams A. The (new) Stanford light field archive[R]. Stanford: Stanford University Computer Graphics Laboratory, 2008.
24. Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 2016: 1646–1654. https://doi.org/10.1109/CVPR.2016.182.
25. Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, 2017: 1132–1140. https://doi.org/10.1109/CVPRW.2017.151.
26. Rossi M, Frossard P. Geometry-consistent light field super-resolution via graph-based regularization[J]. *IEEE Trans Image Process*, 2018, **27**(9): 4207−4218.
27. Yeung H W F, Hou J H, Chen X M, et al. Light field spatial super-resolution using deep efficient spatial-angular separable convolution[J]. *IEEE Trans Image Process*, 2019, **28**(5): 2319−2330.
28. Zhang S, Lin Y F, Sheng H. Residual networks for light field image super-resolution[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 2019: 11038–11047. https://doi.org/10.1109/CVPR.2019.01130.
29. Zhang S, Chang S, Lin Y F. End-to-end light field spatial super-resolution network using multiple epipolar geometry[J]. *IEEE Trans Image Process*, 2021, **30**: 5956−5968.
30. Wang S Z, Sheng H, Yang D, et al. MFSRNet: spatial-angularcorrelation retaining for light field super-resolution[J]. *Appl Intell*, 2023, **53**(17): 20327−20345.

作者简介

【通信作者】王飞 (1985-)，男，博士，副教授，从事图像处理、视频分析、信号处理和各种嵌入式设备相关算法的软件开发。目前主要研究方向为图像处理、信号分析与算法优化。

E-mail：200102@xsyu.edu.cn

邹希 (1999-)，女，硕士研究生，主要从事图像处理、算法优化。

E-mail：zx1110070@163.com

高建邦 (1990-)，男，博士，助理教授，主要研究方向为图像处理、信号处理、故障诊断。

E-mail：gjbang2008@126.com

高国旺 (1977-)，男，博士，教授，主要研究方向为多模态信息处理、微弱信号检测与处理。

E-mail：wwgao@xsyu.edu.cn

Light field super-resolution with Mamba subspace scanning and diffusion modeling

Wang Fei*, Zou Xi, Gao Jianbang, Gao Guowang

School of Electronic Engineering, Xi'an Shiyou University, Xi'an, Shannxi 710065, China

Abstract

Objective Light field super-resolution (LFSR) aims to reconstruct high-resolution light field images from low-resolution observations while preserving both fine spatial details and angular consistency among multiple views. Because spatial and angular information are tightly coupled in light field data, reconstruction is challenged by both feature complexity and view dependency. Existing approaches often suffer from two major limitations. First, high-frequency textures, edge details, and subtle structural patterns are easily degraded during feature extraction and upsampling, resulting in blurred outputs. Second, insufficient modeling of spatial-angular correlations may introduce inconsistencies across viewpoints, impairing geometric fidelity and visual coherence. To address these issues, this paper proposes a light field super-resolution framework that integrates Mamba-based subspace scanning with diffusion-based generative reconstruction. The framework is designed to enhance high-frequency detail recovery, strengthen long-range spatial-angular dependency modeling, and improve reconstruction accuracy and cross-view consistency under different upscaling settings.

Methods The proposed framework adopts a dual-branch subspace scanning strategy based on the Mamba architecture. Considering that light field images exhibit complementary characteristics in different subspaces, two specialized branches are constructed for efficient modeling. The first branch, termed EPI-Mamba, focuses on Epipolar Plane Image (EPI) structures, which explicitly characterize geometric relationships across viewpoints. This branch is therefore used to capture directional continuity and structural variation in epipolar dimensions. The second branch, termed Spatial-Angular Mamba (SA-Mamba), is designed to model correlations between spatial content and angular variation, enabling the network to learn dependencies that are difficult to represent using conventional convolution alone. Both branches perform bidirectional scanning in their respective subspaces, allowing efficient long-range dependency modeling while maintaining relatively low computational complexity.

The serialized features extracted from the two branches are then fed into a Multi-scale Cross Interaction (MCI) module. This module promotes deep information exchange between EPI-aware features and spatial-angular features at multiple scales, thereby enhancing complementary fusion of geometric and texture information. To further refine the fused representations, a Spatial-Angular Modulation (SAM) module is introduced. This module jointly calibrates features from the spatial and angular perspectives, adaptively emphasizing informative responses and suppressing inconsistent activations. As a result, cross-view feature alignment is improved and the coherence of reconstructed light field content is strengthened.

To mitigate the loss of high-frequency details, a frequency-domain enhancement mechanism is incorporated into the framework. Specifically, Fast Fourier Transform (FFT) is applied to the fused features to obtain frequency-domain representations, from which informative high-frequency components are selectively enhanced. This process compensates for detail attenuation, reinforces discriminative structural responses, and suppresses irrelevant or noisy signals. The enhanced features are then input into a diffusion-based denoising network. Benefiting from the strong generative capability of diffusion models, the network progressively restores fine details through iterative denoising and reconstructs high-resolution light field images after upsampling. The cooperation between subspace-aware feature extraction and diffusion-based refinement enables the framework to balance reconstruction fidelity, structural accuracy, and perceptual quality.

Results and Discussions Extensive experiments are conducted on multiple benchmark datasets to evaluate the proposed method. Quantitative comparisons show that the proposed framework consistently outperforms representative state-of-the-art methods under different magnification settings. In the 2× light field super-resolution task, the proposed method achieves a peak signal-to-noise ratio (PSNR) of 39.43 dB and a structural similarity index (SSIM) of 0.987, representing the best performance among the compared methods. In the more challenging 4× task, the proposed approach still attains the highest results, reaching 33.70 dB in PSNR and 0.945 in SSIM. In particular, the PSNR is improved by up to 1.44 dB over existing methods, demonstrating a clear quantitative advantage.

Foundation item: Project supported by Shaanxi Provincial Natural Science Basic Research Program-General Project (Youth) (2025JC-YBQN-935) , Shaanxi Provincial Key Research and Development Program (2024GX-YBYM-507) and Graduate Innovation and Practice Ability Training Program of Xi'an Shiyou University (YCS23214260)

*Correspondence: Wang F, E-mail: 200102@xsyu.edu.cn

Qualitative results further confirm the superiority of the proposed framework. Compared with competing approaches, the reconstructed images exhibit sharper boundaries, clearer local textures, and more faithful structural recovery, especially in regions containing dense lines, repetitive patterns, or complex high-frequency details. In addition, the restored views show stronger cross-view consistency, with fewer artifacts such as blurring or misalignment. These observations indicate that the method not only improves distortion-based metrics but also enhances perceptual quality and geometric coherence, both of which are essential for light field imaging applications.

The performance gains can be explained from several aspects. First, the dual-branch Mamba subspace scanning strategy makes full use of the intrinsic properties of light field data by separately modeling EPI structures and spatial-angular dependencies. Second, the MCI and SAM modules strengthen feature interaction, adaptive fusion, and cross-view calibration, thereby improving both discriminative ability and reconstruction stability. Third, the frequency-domain enhancement mechanism directly compensates for high-frequency information loss, which is especially beneficial for recovering textures and edge details. Finally, the diffusion-based denoising network further refines the reconstructed results by exploiting generative priors, leading to more realistic and visually pleasing outputs. Together, these modules form a unified framework in which each component contributes to performance from a complementary perspective.

Conclusions This paper presents a light field super-resolution framework that combines Mamba-based subspace scanning with diffusion-based generative reconstruction. By jointly capturing EPI structures and spatial-angular correlations, and by integrating multi-scale cross interaction, spatial-angular modulation, and frequency-domain enhancement, the proposed method effectively addresses two core challenges in light field super-resolution: high-frequency detail loss and cross-view inconsistency. Experimental results demonstrate that the method achieves state-of-the-art performance in both objective metrics and subjective visual quality, with notable advantages in texture recovery, edge preservation, and structural consistency. These findings indicate that the proposed framework provides an effective solution for high-quality light field super-resolution. Future work may explore more lightweight architectures and more efficient diffusion strategies to reduce computational cost and extend the framework to related applications such as depth estimation, view synthesis, and light field restoration.

Keywords: light field; super-resolution; mamba; frequency enhancement; diffusion model

Wang F, Zou X, Gao J B, et al. Light field super-resolution with Mamba subspace scanning and diffusion modeling[J]. *Opto-Electron Eng*, 2026, **53**(4): 250263; DOI: 10.12086/oee.2026.250263