中图法分类号:TP391;TP18　　**文献标识码**: A　　**文章编号**: 1006-8961(2026)11-4037-16

论文引用格式:Zhao G B, Yu K and Wang F. 2026. Local-aware state-space model for remote-sensing image super-resolution. Journal of Image and Graphics, 31(11):4037-4052(赵公博, 于珂, 王峰. 2026. 面向遥感图像超分辨率的局部感知状态空间模型. 中国图象图形学报, 31(11):4037-4052)[DOI:10.11834/jig.250580]

面向遥感图像超分辨率的局部感知状态空间模型

赵公博,于珂,王峰*

复旦大学未来信息创新学院,上海　200433

摘　要:**目的**　遥感图像超分辨率(remote sensing image super-resolution,RSISR)旨在从低分辨率观测中重建高分辨率影像,实现细节纹理恢复与全局结构保持。然而,现有方法难以兼顾全局依赖建模与局部特征表达。尽管状态空间模型(state space model,SSM)在长程依赖建模方面具有线性复杂度优势,但其一维展开机制易导致局部语义破坏与像素次序扰动,并在性能与计算效率之间陷入权衡。**方法**　为应对上述局限,本文提出基于局部感知的状态空间重建框架 LAMA-SR(local-aware Mamba for super-resolution)。首先,对输入遥感影像显式施加二维旋转位置编码(two-dimensional rotary position embedding,RoPE-2D),以建立像素之间的相对空间约束与跨区域关联性,从而缓解一维扫描导致的空间邻接关系丢失问题。随后,编码后的特征通过二维选择性扫描进行全局依赖建模,以在线性计算复杂度下刻画长程上下文。同时,引入轻量卷积式局部信息聚合器(local information aggregator,LIA),以增强局部上下文聚合与细粒度纹理恢复,实现对细节区域的补偿性重建。该设计在全局建模能力与局部细节保真度之间形成协同。**结果**　在多个遥感图像基准数据集上,LAMA-SR 相较于主流卷积神经网络(convolutional neural network,CNN)和 Transformer 模型,在峰值信噪比(peak signal-to-noise ratio,PSNR)与结构相似性指数(structural similarity index,SSIM)等指标上均取得显著提升。同时,该模型保持了较低的参数规模与计算开销,展现出优越的效率—效果平衡性。**结论**　实验结果验证了所提出 LAMA-SR 框架在遥感图像超分辨率任务中的有效性与普适性。其在保留 Mamba 长程依赖建模优势的同时,显著改善了局部语义一致性与细节重建能力。

关键词:深度学习;超分辨率重建;Mamba;位置编码;注意力机制

Local-aware state-space model for remote-sensing image super-resolution

Zhao Gongbo, Yu Ke, Wang Feng*

College of Future Information Technology, Fudan University, Shanghai　200433, China

Abstract: **Objective** Remote-sensing image super-resolution (RSISR) seeks to reconstruct high-resolution (HR) images from low-resolution (LR) observations, thereby enhancing spatial detail while preserving structural integrity across large-scale scenes. Although convolutional neural networks (CNNs) and Transformer-based architectures have achieved impressive progress, existing approaches still face an intrinsic trade-off between reconstruction fidelity and computational efficiency. CNNs capture fine-grained local texture effectively but are limited by small receptive fields, whereas Transformers model long-range dependencies well but incur quadratic complexity, making them expensive for high-resolution remote-sensing imagery. The recently proposed selective state-space model, Mamba, offers linear-time sequence modeling with strong global dependency representation. However, when applied to vision tasks, Mamba still has several shortcomings,

收稿日期:2025-11-19;**修回日期**:2026-02-03;**预印本日期**:2026-02-10

*通信作者:王峰　fengwang@fudan.edu.cn

基金项目:上海市协同创新项目(XTCX-KJE001-2025-08)

Supported by:Shanghai Collaborative Innovation Program(XTCX-KJE001-2025-08)

including unidirectional information flow, disrupted pixel ordering caused by one-dimensional unfolding, and weak preservation of local semantic structure. These limitations motivate the design of an architecture that can jointly strengthen local-detail representation and global spatial consistency under constrained computation. **Method** To this end, we propose local-aware Mamba for super-resolution (LAMA-SR), a novel state-space-based reconstruction framework for RSISR. Rather than encoding positional information inside the scanning module itself, LAMA-SR first applies an explicit two-dimensional rotary position embedding (RoPE-2D) to input feature maps to encode relative spatial relationships and inter-pixel geometric dependencies along both axes. This design preserves structural continuity between neighboring regions and reduces the spatial distortion introduced by one-dimensional scanning. The position-enhanced features are then processed by a two-dimensional selective scan (SS2D) mechanism, which models long-range dependencies efficiently with linear computational complexity. To strengthen fine-grained spatial representation, a multi-scale mixture-of-experts (MoE) local information aggregator is introduced before the Mamba module. This component divides input features into small-, medium-, and large-receptive-field branches, allowing adaptive extraction of local patterns at different spatial scales. Unlike conventional MoE designs that rely on explicit routing, a lightweight channel-attention-based soft-routing mechanism dynamically adjusts the contribution of each expert according to the input content. This allows the network to emphasize small-scale experts for texture-rich regions and larger-scale experts for structural continuity, thereby enriching local representation and providing stronger input for subsequent state-space propagation. Collectively, these components create a cooperative mechanism that combines efficient global modeling with high local fidelity within a compact Mamba-based backbone. **Result** Extensive experiments were conducted on four widely used benchmark datasets—UCMerced, RSSCN7, AID, and WHU-RS19—under upscaling factors of ×2, ×3, and ×4. Quantitative evaluation using peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) shows that LAMA-SR consistently outperforms representative CNN-based, Transformer-based, and Mamba-based methods. For example, on the UCMerced dataset at a ×4 scale, LAMA-SR achieves 29.26 dB PSNR with only 8.10 M parameters and 20.23 G FLOPs, surpassing MambaIR, which uses 20.42 M parameters and 79.70 G FLOPs, while also delivering better reconstruction fidelity. Ablation studies further demonstrate that RoPE-2D and the local information aggregator (LIA) provide complementary benefits: RoPE-2D improves spatial coherence and relative position modeling, while LIA enhances texture recovery and channel selectivity. Visual results and class-wise analysis confirm that LAMA-SR produces sharper object boundaries and more natural textures, particularly in complex scenes such as airports, industrial zones, and vegetation-rich areas, validating its robustness and generalization across diverse remote-sensing scenarios. **Conclusion** LAMA-SR provides a state-space-based framework for remote-sensing image super-resolution that improves both reconstruction fidelity and computational efficiency. By promoting more consistent spatial information propagation and targeted refinement of local structures, the model better preserves neighborhood continuity and scene layout during reconstruction than conventional Mamba-style baselines. Experiments on multiple benchmark datasets show that LAMA-SR delivers sharper textures, clearer object boundaries, and a more favorable accuracy-efficiency trade-off across different upscaling factors. These results indicate that state-space models can be adapted effectively to high-resolution remote-sensing imagery while maintaining fine-grained spatial semantics, offering a stable and scalable direction for future research on lightweight super-resolution models.

Key words: deep learning; super-resolution reconstruction; Mamba; position embedding; attention mechanism

0 引　言

遥感图像超分辨率重建(remote sensing image super-resolution,RSISR)旨在从低分辨率遥感影像中恢复出具有更高空间分辨率的图像,从而提升下游任务如地物识别(Helber等,2019)、目标检测、环境监测(Zhao等,2020)、灾害管理(Ghaffarian等,2018)和目标检测(Dong等,2020)。与自然图像不同,遥感图像往往覆盖大范围场景,包含尺度差异极大的地物目标(道路网、建筑群、农田纹理与水系边界等),并受空气散射、成像平台抖动、光照和传感器物理极限等因素影响而存在细节缺失与噪声污染(Li等,2024)。这使得RSISR不仅需要恢复高频纹理,还必须保持大尺度空间结构和几何一致性,即既不能把边缘“磨平”,也不能让目标结构出现形变或错

位(Zhang等,2018)。因此,如何在有限计算预算下同时实现局部细节还原和全局结构保持,成为当前国内外该领域的核心科学问题(Liang等,2021)。

传统方法主要依赖卷积神经网络(convolutional neural network,CNN)(He等,2016)进行单幅图像的超分辨率重建(Dong等,2016;Kim等,2016;李云飞等,2017;Lei和Shi,2022;Li等,2022;Lim等,2017;Wang等,2022b;Zhang等,2018;宋霄罡 等,2025)。这些CNN类方法的优势在于:在局部纹理恢复方面具有很强的精细建模能力,尤其适合恢复房屋边缘、道路骨架等高频结构(Lim等,2017)。但它们也有共性局限:卷积的感受野本质上是局部的,即便通过堆叠加深网络或引入空洞卷积,也很难在长距离范围内保持全局一致性,这在遥感场景中表现为"大范围结构容易断裂或不连续",例如道路网络变形、农田块边界被打散、城市街区轮廓不连贯(Zhang等,2018)。因此,单纯依靠卷积堆叠很难同时建模遥感图像中的长程依赖关系和跨区域一致性(Liang等,2021)。

Transformer结构引入的自注意力机制为长距离依赖建模提供了一种显式的全局建模能力(Vaswani等,2017)。在超分辨率场景中,基于自注意力的模型能够同时观察到广域上下文,避免仅根据局部邻域进行"盲目锐化",从而在恢复大尺度地物形态和语义连贯性方面优于纯CNN方法(Cai等,2023;Lei等,2022;Liang等,2021;Lu等,2022;Xiao等,2024)。尤其是窗口化或分层式视觉Transformer结构通过分区注意力和跨窗口交流,实现了对大范围纹理模式的一致建模,同时一定程度上控制了注意力计算的规模。不过,自注意力的计算与显存开销依然随空间分辨率近似平方增长。在遥感场景中,原图往往是超大幅面切块推理,高分辨率补偿本身就非常昂贵,这使得Transformer类RSISR方法在边缘设备、机载/星载实时判读以及地面快速应急监测等场景下难以直接部署(Wu等,2025)。

为同时获得长程建模能力与较低复杂度,近年来兴起的状态空间模型(state space model,SSM),尤其是Mamba结构,被引入到遥感图像超分辨率中并迅速成为新的研究焦点(Gu等,2022)。Mamba通过选择性状态传递实现跨长距离的依赖建模,其推理复杂度在序列长度上接近线性规模,相比自注意力的全局两两交互更加高效(Gu和Dao,2023)。这为RSISR带来了一个新的可能路径:不再依赖纯卷积的局部堆叠,也不完全依赖全注意力的二次复杂度,而是用线性时间的全局建模去维持大尺度地物结构的连贯性。但由于Mamba最初是从序列建模场景发展而来,它天然依赖对输入进行"扫描式"的状态更新,这在视觉任务中通常意味着将二维影像按行、列或块展平成一维序列,再按序推进状态。这种展平会破坏原始空间邻接关系,导致局部几何细节(例如屋顶边缘的转角、道路的交叉口)在重建时出现模糊、粘连或错位(Jiang等,2025)。

针对Mamba在遥感超分中的局限,不同研究工作从结构融合、注意力增强到频域建模等方向展开了改进探索,形成了一系列具有代表性的方法。SR-Mambaformer提出将卷积分支、Transformer注意力分支与Mamba分支并行融合,分别负责局部纹理恢复、显著区域加权与全局一致性保持(Zhi等,2024)。该结构能较好兼顾细节与语义,但多分支耦合导致模型推理路径冗长,计算与内存开销增大,不利于轻量化部署。Efficient Mamba-Attention Network更注重效率,通过显著性引导与多尺度轻量特征提取模块,将计算聚焦于关键地物区域,并在Mamba状态传播中嵌入通道—空间注意力以强化长程依赖(Wu等,2025)。其优势在于在保持近线性复杂度的同时提升判读精度,但依赖注意力补偿使模型仍存在人工权衡;当场景纹理规则或重复时,显著性机制易压制真实边界,造成局部平滑。MFEM(multiscale frequency-enhanced Mamba)在Mamba主干内部引入多尺度卷积与频域增强模块,直接提升模型的局部与高频感知能力(Chen等,2025)。该方法能有效缓解"全局一致而局部模糊"的问题,使Mamba既具长程建模又能保持纹理锐利;但多尺度卷积与频域分支增加了结构复杂性与推理负担。Rep-Mamba则从空间结构保持性出发,提出跨尺度状态传播与重参数化卷积机制,在推理阶段折叠多分支卷积以降低运行成本(Jiang等,2025)。该方法强化了全局—局部协同,对遥感图像中复杂几何边界表现优异;但其补偿仍发生于Mamba外部,未从根本上解决二维影像在一维扫描中空间关系受损的问题。

综上所述,现有基于Mamba的遥感超分辨率方法虽然在不同层面提升了模型的适应性,但普遍依赖外部卷积、注意力或频域分支进行修正,结构复杂、推理开销大,且缺乏一种能够在Mamba内部同时实现二维空间建模与局部几何保持的统一机制。

为此，本文提出一种基于二维旋转位置编码（rotary position embedding，RoPE-2D）（Su 等，2024）与轻量局部信息聚合器（local information aggregator，LIA）的 Mamba 遥感超分辨率模型。该方法在状态传播过程中显式注入空间相对位置信息，使 Mamba 具备局部几何感知与边缘连续性；同时，LIA 模块在进入 Mamba 前以混合专家（mixture of experts，MoE）（Shazeer 等，2017）结构自适应聚合不同尺度的局部邻域信息。具体而言，输入特征按通道划分为多组并分别经小、中、大感受野的专家卷积分支提取特征，再由通道注意力机制（Hu 等，2018）实现轻量“软路由”加权，从而在低计算开销下自适应选择最合适的局部感受野。该机制显式建模多尺度局部依赖与边缘连续性，为后续的 Mamba 状态传播提供结构保持性更强的输入表征。整体上，RoPE-2D 提升了 Mamba 的二维空间建模能力，而 LIA 则弥补其局部几何感知的不足，使模型能够在保持线性复杂度的同时兼顾细节纹理与全局结构一致性。如图 1 所示，本文方法在 UCMerced ×4 数据集上对多个典型超分辨率模型进行了峰值信噪比（peak signal-to-noise ratio，PSNR）、参数量与每秒浮点运算次数（floating point operations per second，FLOPs）对比。实验结果表明，本文方法在重建质量和计算效率之间实现了良好平衡：在显著降低参数数量与计算开销的同时，取得了当前最优的重建性能，充分展示了其在性能与效率上的综合优势。

本文主要贡献如下：1）提出一种融合 RoPE-2D 与状态空间建模的遥感图像超分方法，从结构层面提升 Mamba 的二维空间感知能力；2）基于混合专家模型设计轻量局部聚合模块（LIA），在保持线性复杂度的同时，增强了模型对地物边缘与纹理的恢复效果；3）设计了完整的实验方案，在多个遥感数据集上验证了所提模型在性能—效率平衡方面的优越性，为轻量化高保真 RSISR 提供了一种可推广的解决方案。

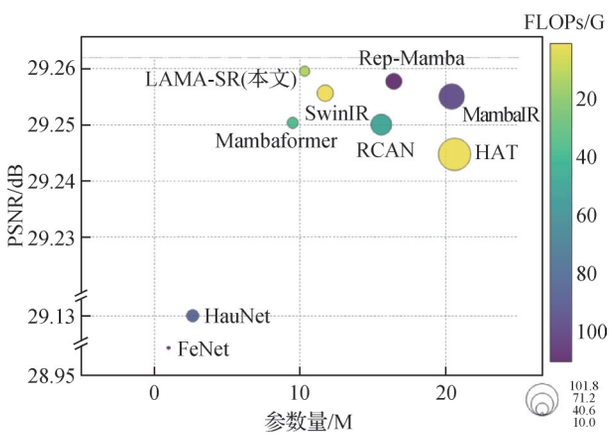


图 1　模型复杂度与重建性能对比图

Fig. 1　Comparison of model complexity and reconstruction performance

1　本文方法

本节介绍本文提出的超分辨率（super-resolution，SR）框架，其目标是从低分辨率输入中重建高分辨率图像。该网络主要由 3 个模块组成：浅层特征提取器、基于视觉 Mamba 模块（vision Mamba block，VMB）构建的深层特征提取主干网络，以及图像重建模块。为了进一步增强特征表示能力，框架中集成了 3 个关键组件：局部信息聚合器（LIA）、旋转位置编码（RoPE）以及视觉状态空间块（visual state space block，VSSB）。

1.1　整体框架

如图 2 所示，本文网络专为遥感图像超分辨率任务设计，主要由 3 个阶段组成：浅层特征提取、深层特征提取和高分辨率图像重建。网络的输入为一幅低分辨率图像 $\boldsymbol{I}_{\mathrm{LR}} \in \mathbf{R}^{H \times W \times 3}$，其中，$H$ 和 W 分别表示图像的高度和宽度，3 表示 RGB 通道数。首先，该图像经过一个深度可分离卷积（depthwise separable convolution，DWConv）层处理，用于提取低级或浅层特征，记为 $\boldsymbol{F}_{\mathrm{S}} = DWConv(\boldsymbol{I}_{\mathrm{LR}})$，输出表示为 $\boldsymbol{F}_{\mathrm{S}} \in \mathbf{R}^{H \times W \times C}$，$C$ 表示特征通道的数量。

这些浅层特征随后被送入一系列 VMB 中进行深层特征提取。每个 VMB 包含多个 VML（vision Mamba layers），而每个 VML 由 3 个关键组件组成：RoPE、LIA 和 VSSB。首先，将位置编码（$\boldsymbol{PE}$）添加到输入特征中，以向特征注入空间位置信息；然后，特征图通过 LIA 以捕捉短程依赖关系，从而增强局部上下文信息；接着，VSSB 模块对整幅图像建模长程空间依赖，实现高效的全局特征表示。第 i 个 VMB 中第 1 个 VML 的完整计算过程可表示为

$$\boldsymbol{F}^{(i,1)} = VSSB(LIA(\boldsymbol{F}^{(i,1-1)} + \boldsymbol{PE})) \tag{1}$$

在第 i 个 VMB 中经过所有 VML 处理后，会在 VMB 末尾应用一个 DWConv 层，以进一步优化局部

特征信息。最终，为了保持训练的稳定性，将输入与VMB的输出通过残差连接进行相加，具体为

$$\boldsymbol{F}^{(i)} = DWConv(\boldsymbol{F}^{(i,L)}) + \boldsymbol{F}^{(i-1)} \tag{2}$$

式中，L表示每个VMB中VML的数量。上述过程重复执行N次，得到最终的深层特征图$\boldsymbol{F}_{\mathrm{D}}$。整个深层特征提取阶段可总结为

$$\boldsymbol{F}_{\mathrm{D}} = VMB_N(\cdots(VMB_1(\boldsymbol{F}_{\mathrm{S}}))) \tag{3}$$

然后，将浅层特征与深层特征通过逐元素相加进行融合，具体为

$$\boldsymbol{F}_{\mathrm{fused}} = \boldsymbol{F}_{\mathrm{D}} + \boldsymbol{F}_{\mathrm{S}} \tag{4}$$

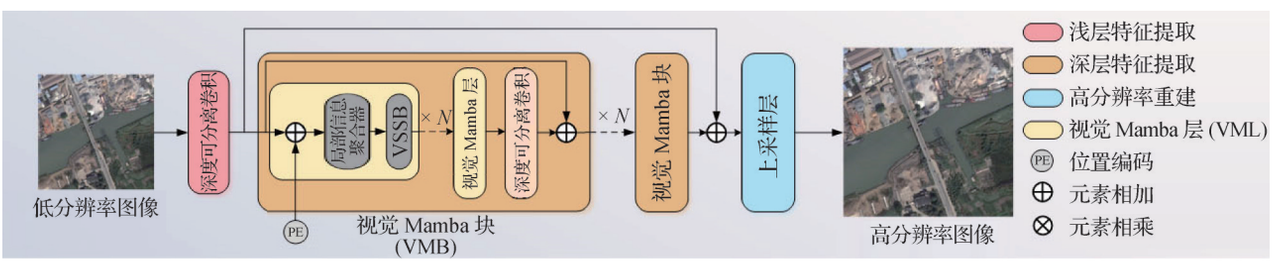
最后，通过上采样层对融合后的特征图进行重建，生成高分辨率图像，具体为

$$\boldsymbol{I}_{\mathrm{SR}} = Upsample(\boldsymbol{F}_{\mathrm{fused}}) \tag{5}$$

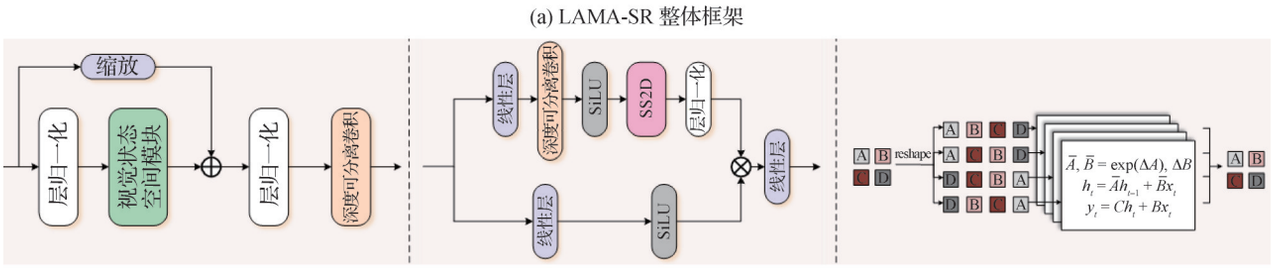
该网络通过最小化预测高分辨率图像$\boldsymbol{I}_{\mathrm{SR}}$与真实图像$\boldsymbol{I}_{\mathrm{HR}}$之间的$L1$损失函数进行训练，从而提升像素级精度与结构保真度，损失函数定义如下

$$L(\theta) = \frac{1}{K}\sum_{i=1}^{K}\left\|\boldsymbol{I}_{\mathrm{HR}}^{(i)} - \boldsymbol{I}_{\mathrm{SR}}^{(i)}\right\|_1 \tag{6}$$

式中，θ表示模型的参数，K为训练图像的总数量。



(a) LAMA-SR 整体框架



(b) 视觉状态空间块 (VSSB)　(c) 视觉状态空间模块 (VSSM)　(d) 二维选择性扫描 (SS2D)

图2　LAMA-SR整体框架

Fig. 2　Framework of the LAMA-SR((a)framework of the LAMA-SR;(b)VSSB;(c)VSSM;(d)SS2D)

1.2　二维位置编码

RoPE最初是为一维序列数据设计的，其核心思想是通过复指数函数表示位置编码(Su等，2024)。对于位置m处的嵌入向量$\boldsymbol{x}_m$，其位置编码变换定义如下

$$f(\boldsymbol{x}_m, m) = \boldsymbol{R}_{\theta,m}\boldsymbol{x}_m \tag{7}$$

式中，$\boldsymbol{R}_{\theta,m}$是旋转矩阵，通过复指数函数对位置信息进行编码。该矩阵定义如下

$$\boldsymbol{R}_{\theta,m} = \begin{Bmatrix} \cos m\theta_1 & -\sin m\theta_1 & 0 & 0 & \cdots & 0 & 0 \\ \sin m\theta_1 & \cos m\theta_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cos m\theta_2 & -\sin m\theta_2 & \cdots & 0 & 0 \\ 0 & 0 & \sin m\theta_2 & \cos m\theta_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \cos m\theta_{d/2} & -\sin m\theta_{d/2} \\ 0 & 0 & 0 & 0 & \cdots & \sin m\theta_{d/2} & \cos m\theta_{d/2} \end{Bmatrix} \tag{8}$$

该变换在嵌入空间的二维子空间中施加一个逆时针旋转，旋转角度为θ_i，定义如下

$$\theta_i = 10\,000^{-2(i-1)/d_{\mathrm{model}}},\ i \in [1,\cdots,d_{\mathrm{model}}/2\,] \tag{9}$$

式中，i表示嵌入维度的索引，d是嵌入的总维度。式(9)确保不同的嵌入维度以不同的速率进行旋转：较低索引的维度编码更细粒度的位置差异，而较高索引的维度则用于捕捉更宽泛、粗粒度的位置关系。

旋转矩阵$\boldsymbol{R}_{\theta,m}$在复向量空间中起作用，其中嵌入的每个二维子空间都会根据其位置索引执行一个受控旋转。该变换可重写为复指数形式如下

$$\boldsymbol{R}_{\theta,m} = \mathrm{e}^{\mathrm{i}m\theta} \tag{10}$$

该形式将位置变换表示为复平面中的纯相位偏移，这意味着在改变角度关系的同时保持向量模不变。因此，依赖于嵌入向量之间内积的自注意力机制，能通过这种旋转相位偏移自然地捕捉相对位置信息。

一种处理图像中位置编码的自然方法是将二维特征图展平为一维序列，并应用标准的一维位置编码。然而，这种方法会引入相对空间关系的不一致性。为了解决这一问题，RoPE被扩展到了二维空间（Heo等，2024）。在一维RoPE中，位置信息通过乘

$$\boldsymbol{R}_{\theta,(p^x,p^y)}=\begin{Bmatrix}\cos m\theta_1 & -\sin m\theta_1 & 0 & 0 & \cdots & 0 & 0\\ \sin m\theta_1 & \cos m\theta_1 & 0 & 0 & \cdots & 0 & 0\\ 0 & 0 & \cos n\theta_1 & -\sin n\theta_1 & \cdots & 0 & 0\\ 0 & 0 & \sin n\theta_1 & \cos n\theta_1 & \cdots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & 0 & 0 & \cdots & \cos n\theta_{d/4} & -\sin n\theta_{d/4}\\ 0 & 0 & 0 & 0 & \cdots & \sin n\theta_{d/4} & \cos n\theta_{d/4}\end{Bmatrix} \tag{12}$$

由于RoPE由一维序列扩展到二维坐标后，单轴位置索引的尺度通常从长度 L 变为约 $\sqrt{L}$ 量级，为保持与原始一维RoPE相近的相位变化范围，需要对式（9）中的常数系数取平方根，从而得到更新后的定义，具体为

$$\theta_i=100^{-4(i-1)/d_{\text{model}}},\ i\in[1,\cdots,d_{\text{model}}/4] \tag{13}$$

二维RoPE保证了不同嵌入维度在适当比例下进行旋转，使得模型能够同时捕捉局部与全局的空间关系。由于旋转变换分别作用于 x 和 y 坐标，模型在两个维度上都能保持等距的空间关系，避免了展平成一维序列所引入的空间扭曲（Heo等，2024）。

1.3 局部信息聚合器

如图3所示，本文提出了一种混合专家（MoE）驱动的局部信息聚合器，用于在特征输入Mamba模块之前充分整合其邻域像素信息，从而弥补状态空间模型在局部几何结构建模方面的不足。

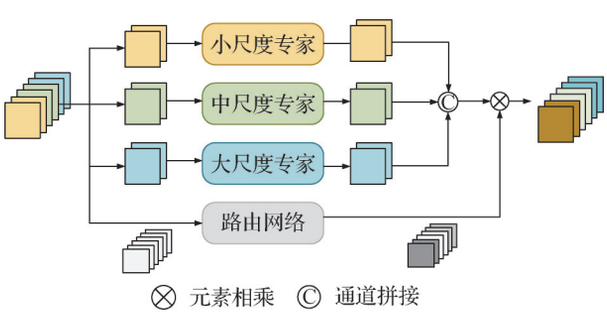


图3 通道注意力驱动的多尺度局部信息聚合器

Fig. 3 Multi-scale local information aggregator driven by channel attention

法旋转变换进行编码，从而使相对位置信息自然融入嵌入向量中。将这一概念扩展到二维时，需要在两个空间维度上分别施加独立的旋转变换，同时保留相对位置结构。

对于具有网格表示的图像，特征图中每个嵌入向量 $\boldsymbol{x}_{(m,n)}$ 对应于位置 (m,n)。由此，二维RoPE的变换形式为

$$f(\boldsymbol{x}_{(m,n)},m,n)=\boldsymbol{R}_{\theta,(m,n)}\boldsymbol{x}_{(m,n)} \tag{11}$$

式中，$\boldsymbol{R}_{\theta,(m,n)}$ 为二维旋转矩阵，其定义为

给定输入特征 $\boldsymbol{X}\in\mathbf{R}^{C\times H\times W}$，首先按照通道维度将其划分为3组 $\{\boldsymbol{X}^{\text{S}},\boldsymbol{X}^{\text{M}},\boldsymbol{X}^{\text{L}}\}$，分别对应小、中、大3类局部感受野。随后，这3组特征分别送入卷积核大小为3、5、7的多尺度专家，以捕获不同空间尺度下的局部结构与纹理细节。具体为

$$\begin{cases}\boldsymbol{Z}_3=f_{3\times3}(\boldsymbol{X}^{\text{S}})\\ \boldsymbol{Z}_5=f_{5\times5}(\boldsymbol{X}^{\text{M}})\\ \boldsymbol{Z}_7=f_{7\times7}(\boldsymbol{X}^{\text{L}})\end{cases} \tag{14}$$

与传统MoE结构依赖显式路由（hard routing）不同，传统方法通常通过Top-k等离散门控对每个位置、特征进行“硬选择”，仅将输入分发给少数专家参与计算，并需要额外的负载均衡约束来缓解专家塌缩等训练不稳定问题；并且，在超分重建中，同一空间位置往往需要同时借助小感受野分支恢复细纹理、并借助大感受野分支维持结构轮廓，而离散硬路由的“单一尺度选择”容易丢失多尺度互补信息。相比之下，本文在LIA中以通道注意力作为门控，生成对各尺度（3/5/7卷积）分支的连续权重并进行加权融合，形成无需离散分发、梯度可同时回传至所有分支的“软路由”机制，使多尺度信息能够按内容自适应地平滑混合，等效于动态选择有效感受野并保留互补细节，从而在保持轻量化的同时获得更平滑的专家选择。

具体而言，首先，对输入特征进行全局通道汇聚，得到压缩后的通道描述子，具体为

$$\boldsymbol{r}=GAP(\boldsymbol{X})\in\mathbf{R}^C \tag{15}$$

式中，*GAP* 表示全局平均池化操作（global average pooling）。

然后，通过两层瓶颈式线性变换生成对应于3个尺度专家的通道级注意力权重，具体为

$$[a^{\mathrm{S}}, a^{\mathrm{M}}, a^{\mathrm{L}}] = sigmoid(\boldsymbol{W}_2\varphi(\boldsymbol{W}_1\boldsymbol{r})) \tag{16}$$

式中，$a^{(\cdot)} \in \mathbf{R}^{C(\cdot)}$ 分别对应小、中、大尺度专家的通道权重，$\varphi(\cdot)$ 为 ReLU（rectified linear unit）激活函数。得到的小、中、大尺度专家权重在空间维度上进行广播，并分别与各尺度专家的输出进行逐通道加权，最后沿通道维拼接形成最终的聚合特征。具体为

$$Y = Concat(a^{(\mathrm{S})}\boldsymbol{Z}_3, a^{(\mathrm{M})}\boldsymbol{Z}_5, a^{(\mathrm{L})}\boldsymbol{Z}_7) \tag{17}$$

通过这种通道注意力驱动的多尺度特征汇聚机制，局部信息聚合器具备以下显著优势：1）自适应选择局部感受野：在几乎不增加计算开销的情况下，模型能够根据输入内容自动选择合适的尺度分支。对于纹理细密区域，更多权重将分配给小尺度专家；在结构跨度较大的区域，中、大尺度专家将获得更高权重，这使得模型能够更灵活地处理不同类型的局部结构。2）增强多尺度局部依赖建模能力：通过显式融合多尺度感受野，模型能够更好地保持边缘连续性和局部几何一致性，提升对复杂空间的表达能力。3）为状态空间传播提供更优特征表示：经过多尺度聚合后得到的特征在局部纹理、边缘信息及空间结构上更为丰富与稳定，为后续 Mamba 模块的长距离状态传播提供了高质量的输入。

2 实验与分析

2.1 数据集与实验设置

2.1.1 数据集

为评估所提出的超分辨率方法的性能，本文选取4个广泛使用的遥感图像数据集。这些数据集包含多种地物类型和场景变化的高分辨率航拍图像，为遥感图像超分辨率模型提供了全面的测试平台：

1）UCMerced（UC Merced land use dataset）（Yang和Newsam，2010）。该数据集图像来源于美国地质调查局的国家地图。该数据集涵盖21类地物使用类型，例如农业用地、飞机、海滩、建筑、森林、高速公路和停车场等，每类包含100幅图像。数据集分为训练集和测试集，各包含1 050幅图像，其中训练集的20%被进一步用做验证集。

2）RSSCN7（remote sensing scene classification dataset with 7 classes）（Zou等，2015）。包含2 800幅遥感图像，涵盖7类场景类型，分别为：草地、森林、农田、工业区、居民区、河流与湖泊和停车场。每类场景包含400幅图像，图像采集于不同季节和大气条件下，以增强数据的多样性。所有图像的空间分辨率为400×400像素。该数据集被平均划分为训练集和测试集，各包含1 400幅图像。

3）AID（aerial image dataset）（Xia等，2017）。包含10 000幅图像，图像采集自Google Earth，空间分辨率为600×600像素。该数据集涵盖30类地物类型，包括机场、港口、公园、教堂、高架桥和广场等。其中20%的图像被用做测试集，剩余80%用于训练，并从每个类别的训练集中随机选取5幅图像作为验证集。

4）WHU-RS19（Wuhan University remote sensing 19-class dataset）（Dai和Yang，2011）。由1 005幅分辨率为600×600像素的航拍图像组成，涵盖19类地物类型，包括机场、港口、沙漠、草地、河流和农田等。该数据集具有高分辨率和多样化的地貌场景，为评估超分辨率模型在遥感图像中重建细节能力提供了良好的测试基础。该数据集用做跨数据集评估。

2.1.2 实验设置

在上述数据集中，超分辨率（SR）任务采用3种不同的放大倍率：×2、×3和×4。低分辨率（low resolution，LR）图像通过对原始高分辨率（high resolution，HR）图像进行双三次插值降采样生成。训练后的模型将这些低分辨率图像重建为高分辨率图像，其性能通过PSNR（Gao等，2009）与结构相似性指数（structural similarity index，SSIM）（Wang等，2004），进行评估，从而能够在多个数据集和放大倍率下对不同的超分辨率方法进行全面比较。模型采用Adam优化器（Kingma和Ba，2015）并使用$L1$损失函数进行优化，超参数设置为$\beta_1 = 0.9$以及$\beta_2 = 0.99$。在训练过程中，应用了水平翻转和随机旋转等数据增强技术，以提升模型的泛化能力。所有实验基于PyTorch框架实现，并在4张NVIDIA RTX 4090 GPU上运行。展示的实验结果均为更换随机种子训练3次后取平均值所得。

2.2 对比实验

为评估所提方法在性能与效率方面相较于主流超分辨率模型的表现，实验选取了多种具有代表性

的基线模型进行对比，包括FeNet（feature enhancement network）（Wang 等，2022b）、HauNet（hybrid attention based U-shaped network）（Wang 等，2023）、MambaIR（Mamba based image restoration）（Guo 等，2024）、RCAN（very deep residual channel attention networks）（Zhang 等，2018）、HAT（hybrid attention Transformer）（Chen 等，2023），SwinIR（image restoration using Swin Transformer）（Liang 等，2021）、SR-Mambaformer（a lightweight model for remote-sensing image super-resolution）（Zhi 等，2024）和 Rep-Mamba（re-parameterization in vision Mamba for lightweight remote sensing image super-resolution）（Jiang 等，2025）。

表1给出了UCMerced数据集×4放大下各方法的参数量与FLOPs以及对应的PSNR/SSIM。本文方法在PSNR上取得最高值29.269 5 dB，同时FLOPs仅为20.23 G、参数量8.10 M，在达到SOTA（state-of-the-art）级别精度的同时显著降低计算开销。对比来看，SwinIR（51.33 G）与MambaIR（79.70 G）虽然在PSNR上接近（分别为29.255 6 dB与29.255 0 dB），但计算量显著更高；而FeNet与HauNet等轻量模型的PSNR分别为28.997 5 dB与29.129 8 dB，均未达到29.15 dB。在SSIM指标上，本文方法取得0.798 0的最高值。总体而言，本文方法在保持低计算复杂度的前提下，展示了更优的精度—效率权衡。

此外，本文在4个遥感数据集—UCMerced、RSSCN7、AID和WHU-RS19上进行了全面评估，涵盖多个放大倍率（×2、×3和×4），以测试所提方法的性能与泛化能力。与体量较小的UCMerced和RSSCN7相比，AID和WHU-RS19包含更大规模的场景与更多类别的多样性。特别地，WHU-RS19用于跨数据集评估，即所有模型在AID数据集上训练后，直接在WHU-RS19上进行测试。相关结果如表2、表3和表4、表5所示。

图4—图7提供了在4个数据集上进行不同网络定性对比的结果，以补充定量实验结果。在这些基准测试中，本文方法展现出稳健且一致的性能表现。在UCMerced和RSSCN7数据集上，模型在所有放大倍率下始终获得最高或接近最高的PSNR和SSIM分数。在更大规模的AID数据集上，尽管本文模型相比于HAT和SwinIR等基于Transformer的方法具有明显更少的参数量，但仍取得了具有竞争力的性能。特别地，与参数量远超本文模型的HAT模型相比，本文方法在PSNR上仍能取得优势。值得注意的是，尽管LAMA-SR与MambaIR、Rep-Mamba、SR-Mambaformer都基于Mamba架构，但在PSNR指标上，LAMA-SR始终优于其他基于Mamba架构的模型，进一步验证了所提出设计的有效性。此外，在用于跨数据集评估的WHU-RS19数据集上，LAMA-SR在×3和×4放大倍率下取得了最高的PSNR和SSIM，表明其在训练数据分布之外的出色泛化能力。

表1　基于UCMerced×4的不同方法的参数量、FLOPs、内存占用以及PSNR对比

Table 1　Comparison of parameters, FLOPs, memory usage, and PSNR on UCMerced ×4

方法	参数量/M	FLOPs/G	PSNR/dB	SSIM
FeNet	0.37	1.44	28.997 5	0.783 2
HAT	20.62	101.72	29.258 7	0.795 7
HauNet	2.66	38.95	29.129 8	0.789 5
MambaIR	20.42	79.70	29.255 0	0.795 6
RCAN	15.59	65.25	29.250 0	0.794 2
SwinIR	11.75	51.33	29.255 6	0.794 9
SR-Mambaformer	9.52	34.82	29.254 2	0.793 3
Rep-Mamba	16.45	50.04	29.261 7	0.796 4
LAMA-SR（本文）	8.10	20.23	**29.269 5**	**0.798 0**

注：加粗字体表示各列最优结果。

2.3　消融实验

2.3.1　Mamba block数量消融

在网络深度（Mamba block数量 N）的设定上，本文遵循“任务表征能力—计算复杂度—跨场景泛化”的三方权衡。

首先，遥感超分常面对大幅面高分辨率影像，训练与推理通常采用滑窗或切块方式，网络深度增加会带来近线性的参数与FLOPs增长，并显著提升显存占用与推理时延；因此本文对主干深度设置上限，以满足实际应用中对效率与可扩展性的约束。

其次，从任务适配性出发，RSISR不仅需要长程依赖以保证道路、建筑群等大尺度结构的连续性，也需要多尺度局部建模以恢复纹理与边缘细节。本文在单个基本单元中通过RoPE-2D显式注入二维相对位置约束，通过LIA完成多尺度局部细节聚合，并结合Mamba进行全局依赖建模，使得每个Block本身

表 2　不同方法在 UCMerced 和 RSSCN7 数据集上的 PSNR 性能比较

Table 2　PSNR comparisons of different methods on the UCMerced and RSSCN7 datasets

/dB

方法	UCMerced			RSSCN7		
	×2	×3	×4	×2	×3	×4
Bicubic	30.658 5	27.378 7	25.585 9	29.326 4	27.411 0	26.298 3
FeNet	35.879 6	31.374 6	28.997 5	31.694 6	29.560 0	28.359 6
RCAN	36.122 4	31.728 2	29.250 0	31.850 3	29.691 3	28.291 1
HauNet	36.048 6	31.528 5	29.129 8	31.742 8	29.628 3	28.454 6
SwinIR	36.164 1	31.763 2	29.255 6	31.821 6	29.672 4	28.448 5
HAT	36.172 1	31.770 4	29.258 7	31.862 7	29.700 8	28.488 6
MambaIR	36.137 6	31.698 4	29.255 0	31.854 5	29.694 6	28.493 0
SR-Mambaformer	36.143 2	31.732 1	29.254 2	31.848 7	29.694 1	28.483 2
Rep-Mamba	36.165 9	31.778 7	29.261 7	31.874 3	**29.707 0**	28.499 3
LAMA-SR（本文）	**36.175 0**	**31.791 6**	**29.269 5**	**31.907 6**	29.706 8	**28.509 5**

注：加粗字体表示各列最优结果。

表 3　不同方法在 UCMerced 和 RSSCN7 数据集上的 SSIM 性能比较

Table 3　SSIM comparisons of different methods on the UCMerced and RSSCN7 datasets

方法	UCMerced			RSSCN7		
	×2	×3	×4	×2	×3	×4
Bicubic	0.889 3	0.781 1	0. 689 2	0.797 5	0.695 2	0.619 3
FeNet	0.941 9	0.859 2	0. 783 2	0.828 2	0.736 2	0.674 1
RCAN	0.944 0	0.865 6	0.794 2	0.833 9	0.741 9	0.671 1
HauNet	0.943 1	0. 864 9	0.789 5	0.830 5	0.740 1	0.680 8
SwinIR	0.944 9	0.867 1	0.794 9	0.833 0	0.742 2	0.680 3
HAT	**0.945 1**	0. 867 1	0.795 7	0.834 0	0.741 6	**0.682 5**
MambaIR	0.943 7	0. 867 0	0.795 6	0.833 6	0.743 1	0.682 0
SR-Mambaformer	0.944 3	0.866 0	0.793 3	0.832 5	0.741 9	0.681 1
Rep-Mamba	0.943 8	0.866 9	0.796 4	0.833 3	0.742 0	0.681 3
LAMA-SR（本文）	0.945 0	**0.868 4**	**0.798 0**	**0.835 1**	**0.743 6**	0.681 6

注：加粗字体表示各列最优结果。

已具备“局部增强 + 全局一致”的协同表达能力，从而降低了单纯依赖堆叠更深网络来扩展感受野的必要性。

基于上述原则，本文在合理复杂度范围内对 N 进行验证：如图 8 所示，当 N 从较小值增加至 6 时，模型表征能力增强、重建质量显著提升；但继续加深会带来收益递减并增加过拟合风险——在规模较小、场景相对集中的 UCMerced 上性能明显下降，而在规模更大、场景更复杂的 AID 上增益趋于饱和并出现轻微回落。综合精度提升、复杂度约束与跨数据集泛化表现，本文最终采用 $N = 6$ 作为默认配置，以在重建质量与计算开销之间取得更优折中。

2.3.2　不同位置编码方法对比

为验证不同位置编码策略对模型性能的影响，本文在相同网络结构与训练配置下，在 UCMerced 数据集上进行×2 实验以对多种位置编码方式进行系

表4　不同方法在AID和WHU-RS19数据集上的PSNR性能比较

Table 4　PSNR comparison of different methods on the AID and WHU-RS19 datasets

/dB

方法	AID			WHU-RS19		
	×2	×3	×4	×2	×3	×4
Bicubic	32.376 2	29.068 1	27.276 9	30.086 7	27.565 2	26.089 6
FeNet	36.885 5	32.791 6	30.535 0	31.946 0	29.521 2	28.051 9
RCAN	37.033 5	33.019 4	30.592 1	31.935 9	29.510 7	28.071 3
HauNet	37.005 5	32.976 0	30.781 2	31.941 4	**29.575 0**	28.074 9
SwinIR	37.080 7	33.051 5	30.854 4	31.951 2	29.530 5	28.057 7
HAT	37.090 4	**33.072 1**	30.822 9	31.953 4	29.543 7	28.056 3
MambaIR	37.060 5	33.039 7	30.805 9	31.953 0	29.531 0	28.075 2
SR-Mambaformer	37.073 6	33.058 7	30.814 8	31.950 2	29.536 4	28.069 3
Rep-Mamba	37.091 0	33.067 9	30.861 6	31.953 1	29.543 2	28.077 8
LAMA-SR(本文)	**37.100 2**	33.071 1	**30.865 2**	**31.961 0**	29.558 3	**28.086 9**

注:加粗字体表示各列最优结果。

表5　不同方法在AID和WHU-RS19数据集上的SSIM性能比较

Table 5　SSIM comparison of different methods on the AID and WHU-RS19 datasets

方法	AID			WHU-RS19		
	×2	×3	×4	×2	×3	×4
Bicubic	0.900 9	0.803 3	0.718 3	0.840 1	0.736 4	0.654 3
FeNet	0.944 0	0.871 1	0.803 3	0.846 3	0.761 1	0.696 8
RCAN	0.945 1	0.876 1	0.804 8	0.847 0	0.761 2	0.697 8
HauNet	0.945 2	0.875 1	0.811 0	0.847 5	0.761 7	0.697 9
SwinIR	0.946 3	0.877 4	0.812 8	0.848 0	0.761 0	0.697 2
HAT	0.946 4	**0.878 0**	0.813 0	**0.848 1**	0.761 7	0.697 7
MambaIR	0.945 6	0.876 0	0.811 3	0.847 3	0.761 5	0.698 0
SR-Mambaformer	0.945 3	0.875 5	0.810 7	0.847 2	0.761 3	0.697 5
Rep-Mamba	0.945 9	0.876 5	0.809 6	0.847 8	0.761 5	0.697 7
LAMA-SR(本文)	**0.946 6**	0.877 2	**0.814 3**	**0.848 1**	**0.762 1**	**0.698 3**

注:加粗字体表示各列最优结果。

统对比。模型中的位置编码分别替换为sinusoidal-1D、sinusoidal-2D和RoPE-2D,实验结果如表6所示。结果显示,引入sinusoidal-1D编码后,相较于未使用位置编码的模型,性能得到一定提升;当采用sinusoidal-2D编码时,模型的PSNR进一步提高,说明在需要空间一致性的任务中,相较一维编码,捕捉二维空间位置信息(横向与纵向)更加有效。值得注意的是,RoPE-2D编码在所有配置中表现最为优异,取得了最高的PSNR。该性能提升归因于RoPE-2D能够建模更复杂的空间关系与旋转不变特征,相较于传统正弦位置编码,在保持空间一致性和捕捉高阶位置依赖方面具有更显著优势。

相较于基线模型(36.053 9 dB),仅引入RoPE-2D后PSNR提升至36.123 3 dB,且参数量保持不变(7.85 M),FLOPs仅增加约0.01 G,表明二维相对位置建模能够以极低开销增强空间结构一致性与几何

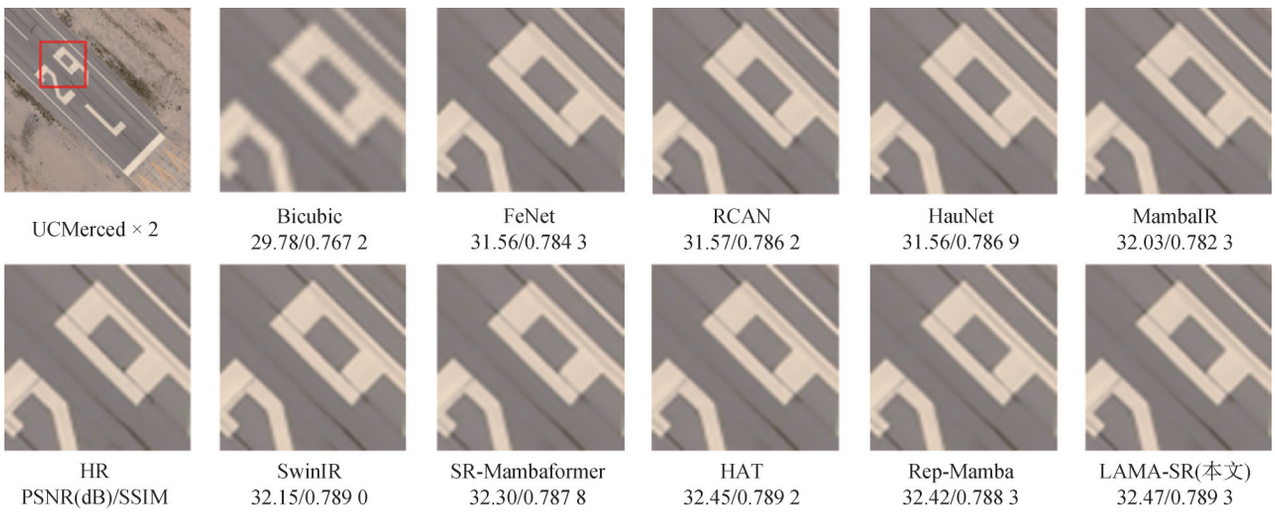


图 4　不同方法在 UCMerced 数据集上的视觉对比

Fig. 4　Visual comparisons of different methods on UCMerced dataset

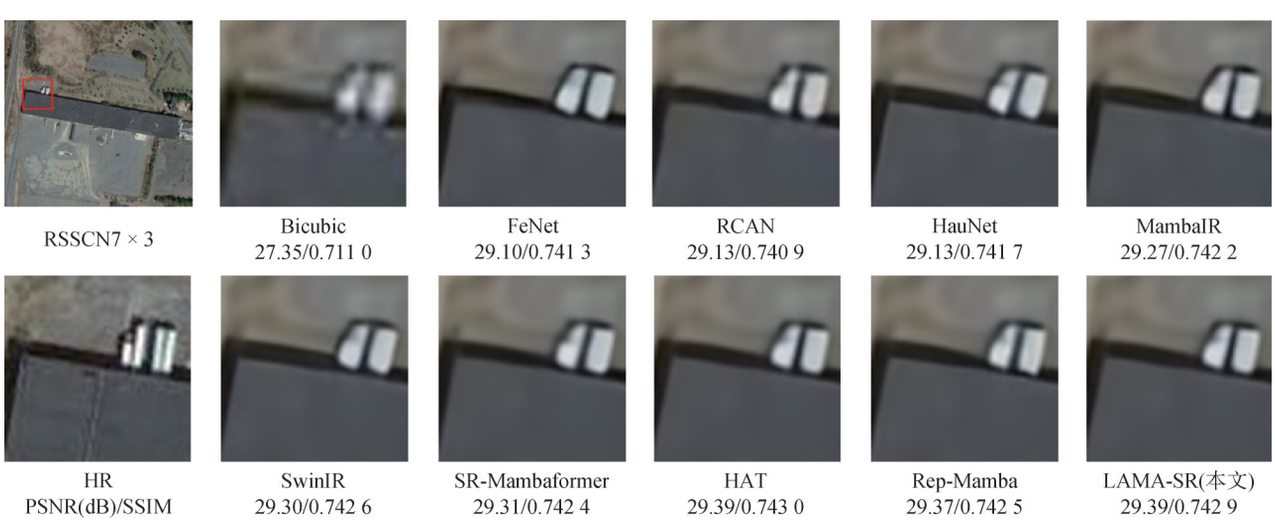


图 5　不同方法在 RSSCN7 数据集上的视觉对比

Fig. 5　Visual comparisons of different methods on RSSCN7 dataset

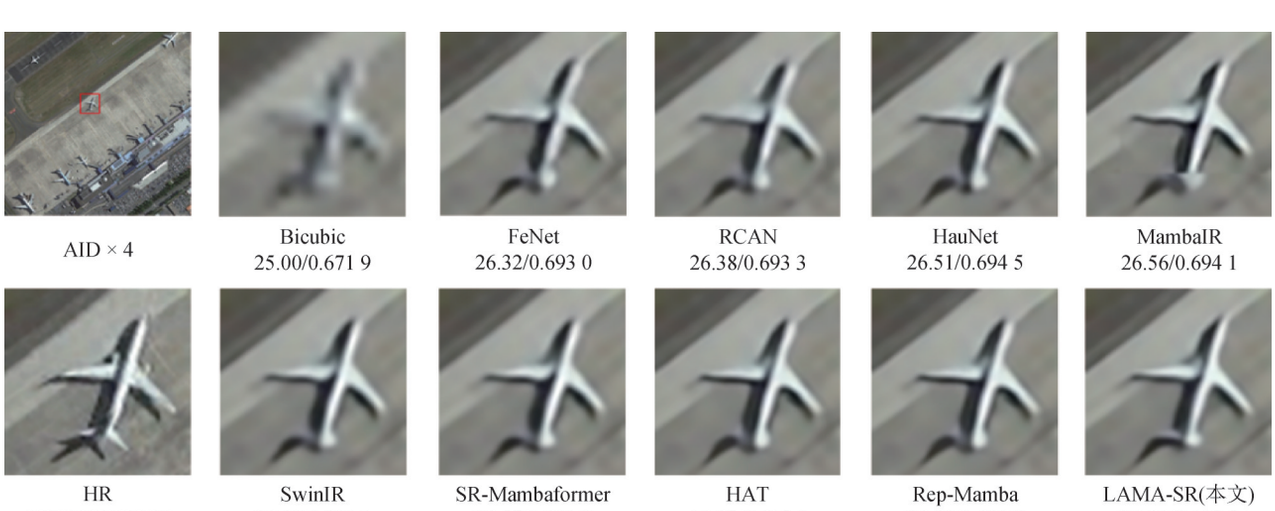


图 6　不同方法在 AID 数据集上的视觉对比

Fig. 6　Visual comparisons of different methods on AID dataset



图7　不同方法在WHU-RS19数据集上的视觉对比

Fig. 7　Visual comparisons of different methods on WHU-RS19 dataset

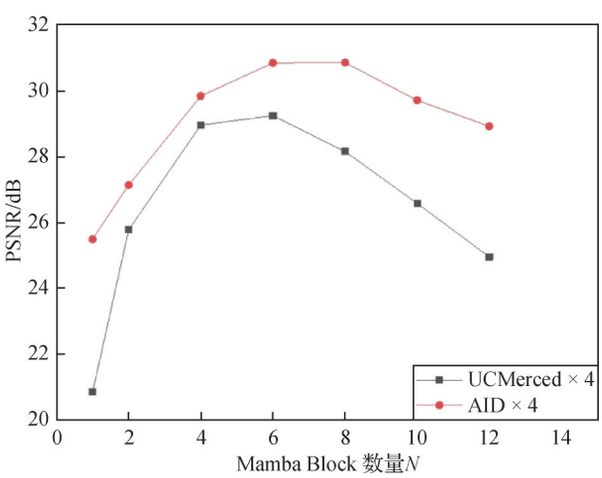


图8　Mamba block数量对性能的影响

Fig. 8　Effect of Mamba block number on model performance

约束。仅引入LIA时PSNR提升至36.167 4 dB,同时参数量增加约0.25 M、FLOPs增加约2.83 G,说明多尺度局部信息聚合对纹理与边缘细节恢复贡献更显著,但计算开销主要由该模块带来。进一步地,RoPE-2D与LIA联合使用时取得最高PSNR(36.175 0 dB),而额外FLOPs仅增加0.01 G,显示RoPE-2D可作为几乎无参数的轻量增强,为LIA的局部聚合提供更稳定的二维位置参照,从而在不显著增加复杂度的前提下进一步提升重建质量与空间一致性。

总体而言,实验结果强调了在局部聚合机制中引入位置编码的重要性。尽管RoPE-2D本身能够增强空间感知能力,但与LIA的结合进一步提升了模型在高质量图像重建中的表现,这一点由显著提高的PSNR结果所验证。

表6　不同位置编码方法比较

Table 6　Comparison of different positional encoding methods

网络	PSNR/dB
基线 + sinusoidal-1D	36.074 6
基线 + sinusoidal-2D	36.117 3
基线 + RoPE-2D	**36.123 3**

注:加粗字体表示最优结果。

2.3.3　RoPE和LIA的有效性分析

为评估RoPE-2D与所提出的LIA模块的有效性,在基线模型的基础上分别集成RoPE-2D和LIA,进行了系列消融实验。在UCMerced数据集×2超分辨率任务的消融结果如表7所示。

进一步地,为了从可解释性的角度验证RoPE-2D与LIA模块在特征表达与空间建模方面的作用,

表7　不同RoPE-2D与LIA组合的性能比较

Table 7　Performance comparison of different combinations of RoPE-2D and LIA

RoPE-2D	LIA	参数量/M	FLOPs/G	PSNR/dB
×	×	7.85	17.39	36.053 9
√	×	7.85	17.40	36.123 3
×	√	8.10	20.22	36.167 4
√	√	8.10	20.23	**36.175 0**

注:加粗字体表示最优结果。“√”和“×”分别表示包含和不包含该模块。

本文在UCMerced ×4数据集上进行了基于局部归因图(local attribution map，LAM)(Gu等，2021)的可视化分析，如图9所示。

LAM结果直观地揭示了模型在重建特定区域时所关注的输入像素分布范围及强度。可以观察到，基线模型的归因响应范围较窄，主要集中在目标边缘附近，表明其对上下文信息的利用有限；仅引入LIA模块时，归因响应明显扩散，模型能够感知到更宽的局部邻域，但由于缺乏明确的位置关系约束，关注区域呈现一定的漂移；仅引入RoPE-2D时，模型在空间结构上表现出更稳定的响应分布，能够保持合理的边界定位；而在RoPE-2D与LIA联合使用的情况下，显著图既呈现更广的扩散范围，又能保持精确的空间聚焦，说明两者在位置建模与局部信息聚合方面形成了互补。

此外，基于LAM中提出的扩散指数(diffusion index，DI)对不同模型的特征聚合范围进行了量化。结果显示，如图9所示，联合模型的DI值达39.72，显著高于基线模型的21.34，表明在重建过程中，模型能够综合利用更大范围的输入像素信息。结合PSNR与SSIM的同步提升(29.50 dB/0.971 8)，进一步验证了RoPE-2D与LIA在结构保持与细节恢复方面的有效协同。

综上，LAM可视化与DI指标均从可解释性视角印证了模块设计的合理性。RoPE-2D通过引入旋转不变的位置编码增强了空间感知，LIA则通过局部信息自适应聚合扩展了特征感受野，两者协同作用使模型在保持边缘锐度的同时，更充分地利用上下文特征，从而实现了更高质量的重建表现。

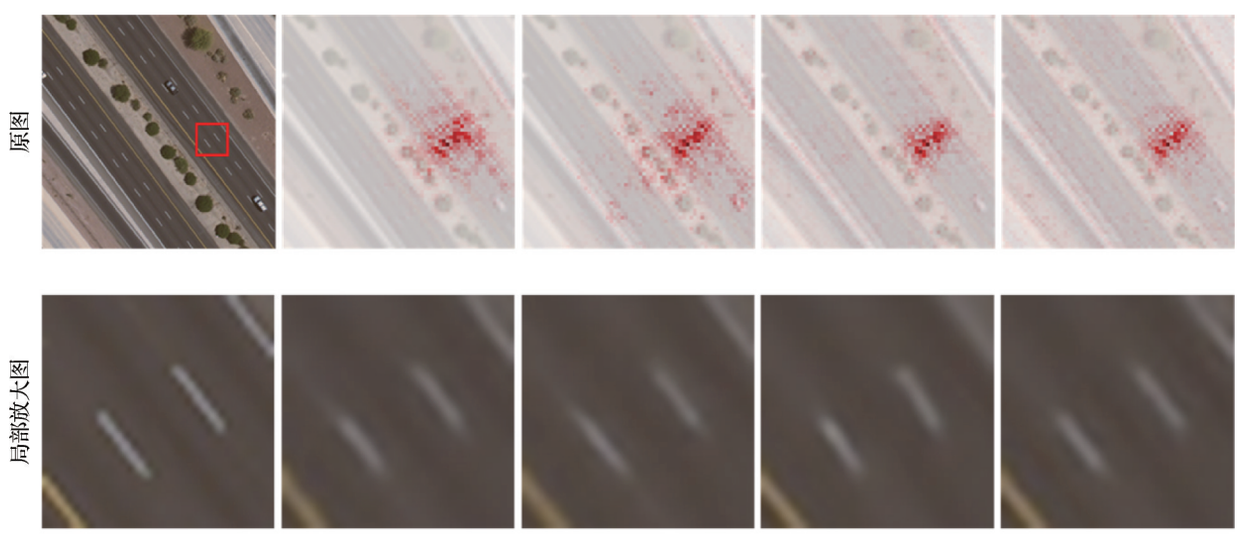


图9　局部归因图可视化

Fig. 9　Visualization of LAMs

3　结　论

本文提出了一种新型的基于Mamba架构的遥感图像超分辨率网络—LAMA-SR。该模型在选择性状态空间主干网络中引入旋转位置编码(RoPE-2D)与局部交互聚合模块(LIA)，能够在捕获全局依赖关系的同时，有效保留细粒度的局部细节信息。基于4个典型遥感数据集的实验结果表明，RoPE-2D能在二维特征空间中引入显式的相对位置信息，使模型在结构保持与纹理细节恢复方面较基线模型均有显著提升；LIA模块通过自适应局部特征聚合机制，使网络在细粒度区域的表达更充分，边缘与纹理重建更加自然。当两者协同作用时，模型在PSNR、SSIM等定量指标上均取得最佳表现，验证了位置建模与局部聚合机制在空间感知与重建一致性方面的互补效应。

基于LAM的可解释性分析进一步揭示了模型

的内部响应规律。相较于基线模型，集成RoPE-2D的模型在归因图上展现出更集中的结构性响应，而引入LIA后，响应区域显著扩展，能够覆盖更广泛的上下文特征。当两模块联合使用时，模型在归因分布上呈现出兼具聚焦性与扩散性的特征，反映出在局部与全局依赖建模之间实现了良好的平衡。结合扩散指数DI指标的量化结果可见，LAMA-SR在特征扩散范围与信息利用效率方面均优于其他配置，表现出更强的空间关联建模能力。

尽管模型在重建性能与可解释性方面均取得显著提升，研究仍存在一定局限。LIA模块在高分辨率任务中引入了额外的计算开销，对推理效率造成一定影响；RoPE-2D的位置编码形式仍基于固定函数映射，对复杂几何变换或非规则地物场景的适应性有限。此外，LAM分析主要聚焦于局部特征层面的响应，尚未充分揭示深层语义特征间的关系。未来的工作将进一步优化位置编码与聚合机制的可学习性，并探索其在多模态融合与实时遥感应用中的潜力，以实现更高效、更具泛化能力的超分辨率重建模型。

参考文献（References）

Cai H, Li J Y, Hu M Y, Gan C and Han S. 2023. EfficientViT: lightweight multi-scale attention for high-resolution dense prediction//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE: 17256-17267 [DOI: 10.1109/ICCV51070.2023.01587]

Chen W Y, Luo L G, Qu S M and Dang C X. 2025. MFEM: multiscale frequency-enhanced Mamba for lightweight remote sensing image super-resolution. IEEE Geoscience and Remote Sensing Letters, 22: #8002805 [DOI: 10.1109/LGRS.2025.3587491]

Chen X Y, Wang X T, Zhou J T, Qiao Y and Dong C. 2023. Activating more pixels in image super-resolution transformer//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 22367-22377 [DOI: 10.1109/CVPR52729.2023.02142]

Dai D X and Yang W. 2011. Satellite image classification via two-layer sparse coding with biased image representation. IEEE Geoscience and Remote Sensing Letters, 8(1): 173-176 [DOI: 10.1109/LGRS.2010.2055033]

Dong C, Loy C C, He K M and Tang X O. 2016. Image super-resolution using deep convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2): 295-307 [DOI: 10.1109/TPAMI.2015.243928]

Dong Z P, Wang M, Wang Y L, Zhu Y and Zhang Z Q. 2020. Object detection in high resolution remote sensing imagery based on convolutional neural networks with suitable object scale features. IEEE Transactions on Geoscience and Remote Sensing, 58(3): 2104-2114 [DOI: 10.1109/TGRS.2019.2953119]

Gao X B, Lu W, Tao D C and Li X L. 2009. Image quality assessment based on multiscale geometric analysis. IEEE Transactions on Image Processing, 18(7): 1409-1423 [DOI: 10.1109/TIP.2009.2018014]

Ghaffarian S, Kerle N and Filatova T. 2018. Remote sensing-based proxies for urban disaster risk management and resilience: a review. Remote Sensing, 10(11): #1760 [DOI: 10.3390/rs10111760]

Gu A and Dao T. 2023. Mamba: linear-time sequence modeling with selective state spaces [EB/OL]. [2025-11-15]. https://arxiv.org/pdf/2312.00752v1.pdf

Gu A, Goel K and Ré C. 2022. Efficiently modeling long sequences with structured state spaces//Proceedings of the 10th International Conference on Learning Representations. [s.l.]: ICLR: 1-27

Gu J J and Dong C. 2021. Interpreting super-resolution networks with local attribution maps//Proceedings of 2021 IEEE/CVF Conference on Computer Vision. Nashville, USA: IEEE: 9195-9204 [DOI: 10.1109/CVPR46437.2021.00908]

Guo H, Li J M, Dai T, Ouyang Z H, Ren X D and Xia S T. 2024. MambaIR: a simple baseline for image restoration with state-space model//Proceedings of the 18th European Conference on Computer Vision. Milan, Italy: Springer: 222-241 [DOI: 10.1007/978-3-031-72649-1_13]

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Helber P, Bischke B, Dengel A and Borth D. 2019. EuroSAT: a novel dataset and deep learning benchmark for land use and land cover classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(7): 2217-2226 [DOI: 10.1109/JSTARS.2019.2918242]

Heo B, Park S, Han D and Yun S. 2024. Rotary position embedding for vision Transformer//Proceedings of the 18th European Conference on Computer Vision. Milan, Italy: Springer: 289-305 [DOI: 10.1007/978-3-031-72684-2_17]

Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]

Jiang K, Yang M R, Xiao Y, Wu J B, Wang G C, Feng X C, et al. 2025. Rep-Mamba: re-parameterization in vision Mamba for lightweight remote sensing image super-resolution. IEEE Transactions on Geoscience and Remote Sensing, 63: #5637012 [DOI: 10.

1109/TGRS.2025.3597745］

Kim J，Lee J K and Lee K M. 2016. Accurate image super-resolution using very deep convolutional networks//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition（CVPR）. Las Vegas，USA：IEEE：1646-1654［DOI：10.1109/CVPR.2016.182］

Kingma D P and Ba J. 2015. Adam：a method for stochastic optimization//Proceedings of the 3rd International Conference on Learning Representations. San Diego，USA：ICLR：1-15

Lei S and Shi Z W. 2022. Hybrid-scale self-similarity exploitation for remote sensing image super-resolution. IEEE Transactions on Geoscience and Remote Sensing，60：#5401410［DOI：10.1109/TGRS.2021.3069889］

Lei S，Shi Z W and Mo W J. 2022. Transformer-based multistage enhancement for remote sensing image super-resolution. IEEE Transactions on Geoscience and Remote Sensing，60：#5615611［DOI：10.1109/TGRS.2021.3136190］

Li H T，Deng W H，Zhu Q Q，Guan Q F and Luo J C. 2024. Local-global context-aware generative dual-region adversarial networks for remote sensing scene image super-resolution. IEEE Transactions on Geoscience and Remote Sensing，62：#5402114［DOI：10.1109/TGRS.2024.3355419］

Li Q，Gong M G，Yuan Y and Wang Q. 2022. Symmetrical feature propagation network for hyperspectral image super-resolution. IEEE Transactions on Geoscience and Remote Sensing，60：#5536912［DOI：10.1109/TGRS.2022.3203749］

Li Y F，Fu R D，Jin W and Ji N. 2017. Image super-resolution using multi-channel convolution. Journal of Image and Graphics，22（12）：1690-1700（李云飞，符冉迪，金炜，纪念. 多通道卷积的图像超分辨率方法. 2017. 中国图象图形学报，22（12）：1690-1700［DOI：10.11834/jig.170325］

Liang J Y，Cao J Z，Sun G L，Zhang K，Van Gool L and Timofte R. 2021. SwinIR：image restoration using swin transformer//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal，Canada：IEEE：1833-1844［DOI：10.1109/iccvw54120.2021.00210］

Lim B，Son S，Kim H，Nah S and Lee K M. 2017. Enhanced deep residual networks for single image super-resolution//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops（CVPRW）. Honolulu，USA：IEEE：1132-1140［DOI：10.1109/CVPRW.2017.151］

Lu Z S，Li J C，Liu H，Huang C Y，Zhang L L and Zeng T Y. 2022. Transformer for single image super-resolution//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops（CVPRW）. New Orleans，USA：IEEE：456-465［DOI：10.1109/CVPRW56347.2022.00061］

Shazeer N，Mirhoseini A，Maziarz K，Davis A，Le Q，Hinton G，et al. 2017. Outrageously large neural networks：the sparsely-gated mixture-of-experts layer//Proceedings of the 5th International Conference on Learning Representations. Toulon，France：ICLR：1-19

Song X G，Zhang P F，Liu W B，Lu X F and Hei X H. 2025. Image superresolution reconstruction based on multiscale large-kernel attention feature fusion network. Journal of Image and Graphics，30（4）：1084-1099（宋霄罡，张鹏飞，刘万波，鲁晓锋，黑新宏. 2025. 多尺度大核注意力特征融合网络的图像超分辨率重建. 中国图象图形学报，30（4）：1084-1099）［DOI：10.11834/jig.240042］

Su J L，Ahmed M，Lu Y，Pan S F，Bo W and Liu Y F. 2024. RoFormer：enhanced transformer with rotary position embedding. Neurocomputing，568：#127063［DOI：10.1016/j.neucom.2023.127063］

Vaswani A，Shazeer N，Parmar N，Uszkoreit J，Jones L，Gomez A N，et al. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach，USA：Curran Associates Inc.：6000-6010

Wang J R，Wang B L，Wang X X，Zhao Y Q and Long T. 2023. Hybrid attention-based U-shaped network for remote sensing image super-resolution. IEEE Transactions on Geoscience and Remote Sensing，61：#5612515［DOI：10.1109/TGRS.2023.3283769］

Wang S Z，Zhou T F，Lu Y and Di H J. 2022a. Contextual transformation network for lightweight remote-sensing image super-resolution. IEEE Transactions on Geoscience and Remote Sensing，60：#5615313［DOI：10.1109/TGRS.2021.3132093］

Wang Z Y，Li L L，Xue Y，Jiang C C，Wang J，Sun K P，et al. 2022b. FeNet：feature enhancement network for lightweight remote-sensing image super-resolution. IEEE Transactions on Geoscience and Remote Sensing，60：#5622112［DOI：10.1109/TGRS.2022.3168787］

Wang Z，Bovik A C，Sheikh H R and Simoncelli E P. 2004. Image quality assessment：from error visibility to structural similarity. IEEE Transactions on Image Processing，13（4）：600-612［DOI：10.1109/TIP.2003.819861］

Wu T R，Zhao R D，Lyu M，Jia Z H，Li L L，Liu M Q，et al . 2025. Efficient Mamba-attention network for remote sensing image super-resolution. IEEE Transactions on Geoscience and Remote Sensing，63：#5627814［DOI：10.1109/TGRS.2025.3578879］

Xia G S，Hu J W，Hu F，Shi B G，Bai X，Zhong Y F，et al. 2017. AID：a benchmark data set for performance evaluation of aerial scene classification. IEEE Transactions on Geoscience and Remote Sensing，55（7）：3965-3981［DOI：10.1109/TGRS.2017.2685945］

Xiao Y，Yuan Q Q，Jiang K，He J，Lin C W and Zhang L P. 2024. TTST：a top-k token selective transformer for remote sensing image super-resolution. IEEE Transactions on Image Processing，33：738-752［DOI：10.1109/TIP.2023.3349004］

Yang Y and Newsam S. 2010. Bag-of-visual-words and spatial extensions for land-use classification//Proceedings of the 18th SIGSPATIAL

International Conference on Advances in Geographic Information Systems. San Jose, USA: ACM: 270-279 [DOI: 10.1145/1869790.1869829]

Zhang Y L, Li K P, Li K, Wang L C, Zhong B N and Fu Y. 2018. Image super-resolution using very deep residual channel attention networks//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 294-310 [DOI: 10.1007/978-3-030-01234-2_18]

Zhao M J, Si F Q, Wang Y, Zhou H J, Wang S M, Jiang Y, et al. 2020. First year on-orbit calibration of the Chinese environmental trace gas monitoring instrument onboard GaoFen-5. IEEE Transactions on Geoscience and Remote Sensing, 58(12): 8531-8540 [DOI: 10.1109/TGRS.2020.2988573]

Zhi R C, Fan X P and Shi J Y. 2024. MambaFormerSR: a lightweight model for remote-sensing image super-resolution. IEEE Geoscience and Remote Sensing Letters, 21: #6015705 [DOI: 10.1109/LGRS.2024.3453428]

Zou Q, Ni L H, Zhang T and Wang Q. 2015. Deep learning based feature selection for remote sensing scene classification. IEEE Geoscience and Remote Sensing Letters, 12(11): 2321-2325 [DOI: 10.1109/LGRS.2015.2475299]

作者简介

赵公博,男,工程师,主要研究方向为遥感图像处理、超分辨率重建。E-mail:gbzhao23@m.fudan.edu.cn

王峰,通信作者,男,教授,主要研究方向为SAR/ISAR成像、目标识别与无人机载遥感。E-mail:fengwang@fudan.edu.cn

于珂,女,工程师,主要研究方向为多模态地物分类。E-mail:kyu23@m.fudan.edu.cn