

文章编号 1004-924X(2026)09-1468-28

## 基于 Mamba 架构的遥感图像超分辨率 重建方法综述

李秉豪<sup>1,2</sup>, 姜肖楠<sup>1,2\*</sup>, 傅瑶<sup>1,2</sup>, 王亚楠<sup>1,2</sup>, 万龙腾<sup>1,2</sup>, 吴凡路<sup>1,2\*</sup>

(1. 中国科学院 长春光学精密机械与物理研究所, 吉林长春 130033;

2. 中国科学院大学, 北京 100049)

**摘要:** 基于选择性状态空间模型的视觉架构 Mamba, 通过选择性扫描与状态更新机制, 可在线性复杂度下建模长程依赖, 为遥感图像超分辨率提供兼顾全局表征与计算效率的新路径。本文系统梳理了基于 Mamba 的超分辨率理论基础与方法框架, 并将代表性工作归纳为四类改进方向: 频域建模与频谱增强、结构融合与注意力增强、跨模态与多源建模、轻量化与重参数化。在统一合成退化评测设置下, 结合 AID, DIOR 和 UCAS\_AOD 等数据集, 对代表性方法的重建性能、感知质量及适用特点进行比较分析。结果表明, Mamba 系列方法在结构连续保持、几何一致性恢复和伪纹理抑制等方面表现出较强潜力, 并在多数据集对比中展现出较好的重建性能。最后, 本文讨论真实退化、跨域泛化与可复现基准等挑战与趋势, 为后续研究与应用落地提供的参考。

**关键词:** 遥感图像; 超分辨率重建; 视觉状态空间模型; Mamba; 频域增强; 跨模态融合; 轻量化

**中图分类号:** TP391; TP751 **文献标识码:** A

**doi:** 10. 37188/OPE. 20263409. 1468

**CSTR:** 32169. 14. OPE. 20263409. 1468

## Review of Mamba-based methods for remote sensing image super-resolution

LI Binghao<sup>1,2</sup>, JIANG Xiaonan<sup>1,2\*</sup>, FU Yao<sup>1,2</sup>, WANG Yanan<sup>1,2</sup>, WAN Longteng<sup>1,2</sup>, WU Fanlu<sup>1,2\*</sup>

(1. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences,

Changchun 130033, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China)

\* Corresponding author, E-mail: jiangxn@ciomp. ac. cn; flwu@ciomp. ac. cn

**Abstract:** The visual architecture Mamba, built on selective state-space models, enables long-range dependencies to be modeled with linear complexity through selective scanning and state updating, providing a promising approach for balancing global representation learning and computational efficiency in remote sensing image super-resolution. This paper systematically reviews the theoretical foundations and methodological framework of Mamba-based super-resolution, summarizes representative studies, and categorizes them into four major directions: frequency-domain modeling and spectral enhancement, structural fusion and attention enhancement, cross-modal and multi-source modeling, and lightweight design with re-param-

收稿日期: 2026-01-24; 修订日期: 2026-03-03.

基金项目: 国家自然科学基金面上项目 (No. 62371352); 国家重点研发计划资助项目 (No. 2022YFB3902300)

eterization. Under a unified synthetic degradation evaluation protocol, representative methods are comparatively analyzed in terms of reconstruction performance, perceptual quality, and applicability on benchmark datasets, including AID, DIOR, and UCAS\_AOD. The results demonstrate that Mamba-based methods exhibit substantial potential for preserving structural continuity, restoring geometric consistency, and suppressing pseudo-textures, while achieving competitive reconstruction performance across multiple datasets. Finally, key challenges and future trends, including real-world degradations, cross-domain generalization, and reproducible benchmarks, are discussed, providing valuable references for future research and practical deployment.

**Key words:** remote sensing images; super-resolution reconstruction; visual state space models; Mamba; frequency domain enhancement; cross-modal fusion; lightweight

## 1 引言

遥感卫星与航空成像系统受限于传感器物理尺寸、轨道、航高以及成像链路的调制传递函数等因素,获取的影像往往存在空间分辨率不足的问题,这一限制会直接影响土地利用监测、灾害评估、城市精细识别、海洋要素反演等下游应用的可靠性与精度。因此,在不显著增加硬件成本的前提下,通过软件算法提升遥感图像的空间细节表达能力,成为遥感智能解译链路中的关键环节之一<sup>[1]</sup>。

遥感图像超分辨率重建(Remote Sensing Image Super-Resolution, RSISR)旨在从低分辨率(Low Resolution, LR)观测中恢复高分辨率(High Resolution, HR)细节,其本质可视为成像退化过程的逆问题。不同的HR影像在模糊、下采样与噪声扰动等退化作用下可能映射为相同或相近的LR观测,因此,RSISR属于典型的病态逆问题。与自然图像相比,遥感图像还具有更强的多尺度性与场景异质性,同一幅大视场影像中往往同时存在规则重复纹理(农田条带、建筑阵列)、非结构化边缘(河岸、道路、海岸线)以及小目标密集分布(车辆、船舶、飞机)等复杂地物;同时,多源遥感,包括光学、多光谱、高光谱、合成孔径雷达(Synthetic Aperture Radar, SAR)及红外等数据在光谱响应、噪声形态与退化模型上差异显著,这使得细节重建、光谱一致性、结构保真与计算可行性之间的矛盾更为突出。此外,遥感超分不仅追求更高的峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)、结构相似性(Structural Similarity, SSIM)等客观指标,还需要关注结构几何

一致性、纹理真实感、跨尺度目标可辨识度以及对真实退化的稳定性<sup>[2-11]</sup>。

围绕RSISR的研究进展,现有方法大体可归为两条技术路线:传统方法与深度学习方法。传统方法(如插值放大、基于先验的重建、浅层学习等)通常具有实现简单、计算开销低与工程部署友好等优点,在硬件资源受限或需要快速预处理的场景仍有价值,但其多依赖局部邻域或固定先验,难以在高倍率放大下稳定恢复足够的高频细节,往往产生模糊、振铃或伪影等问题。近年来,深度学习方法依托大规模数据与端到端学习能力,在细节刻画与重建精度方面显著超越传统算法,并逐步形成以卷积神经网络(Convolutional Neural Network, CNN)、生成对抗网络(Generative Adversarial Network, GAN)、扩散模型与Transformer为代表的技术谱系<sup>[12]</sup>。其中,Transformer通过自注意力机制提升了长程依赖建模能力,但其计算、显存开销随图像尺度快速增长,在大幅面遥感图像的高效推理方面仍存在瓶颈<sup>[13-15]</sup>。

值得关注的是,自2023年以来出现的Mamba为遥感超分提供了新的技术思路<sup>[16]</sup>。该类模型以近线性复杂度实现长序列建模,在保持全局建模能力的同时显著改善计算效率,为全局感知-高效推理的矛盾提供了新的解决思路。不过,Mamba在高频纹理恢复、多尺度特征协同和复杂退化适应等方面仍存在一定局限,因此仍需进一步优化其建模机制,以提升其在遥感超分任务中的适应性与实用价值。目前,遥感超分领域已出现一批围绕Mamba的结构改造与任务适配研究,如特征增强、结构交互、辅助信息

引入和模型优化等,已初步显示出在复杂地物纹理恢复、长距离结构一致性与大图高效推理方面的潜力<sup>[17]</sup>。但由于该方向仍处于快速发展阶段,不同研究在扫描机制、频谱补偿、跨尺度融合及退化建模等方面差异较大,相关成果尚缺乏面向遥感任务特性的系统梳理与比较分析。近两年,已有综述主要从遥感图像超分辨率总体进展、数据资源整理或遥感 Mamba 跨任务应用等角度展开,为领域研究提供了重要参

考<sup>[18-21]</sup>。与现有侧重遥感超分总体研究进展或遥感 Mamba 跨任务应用归纳的综述不同,本文主要面向遥感图像超分辨率重建任务展开综述,重点围绕传统方法、主流深度模型与 Mamba 模型演化之间的衔接关系,以及 Mamba 相关方法在频域增强、结构融合、跨模态建模和轻量化设计等关键方向上的改进思路进行归纳与讨论,遥感图像超分辨率重建方法发展历程如图 1 所示。

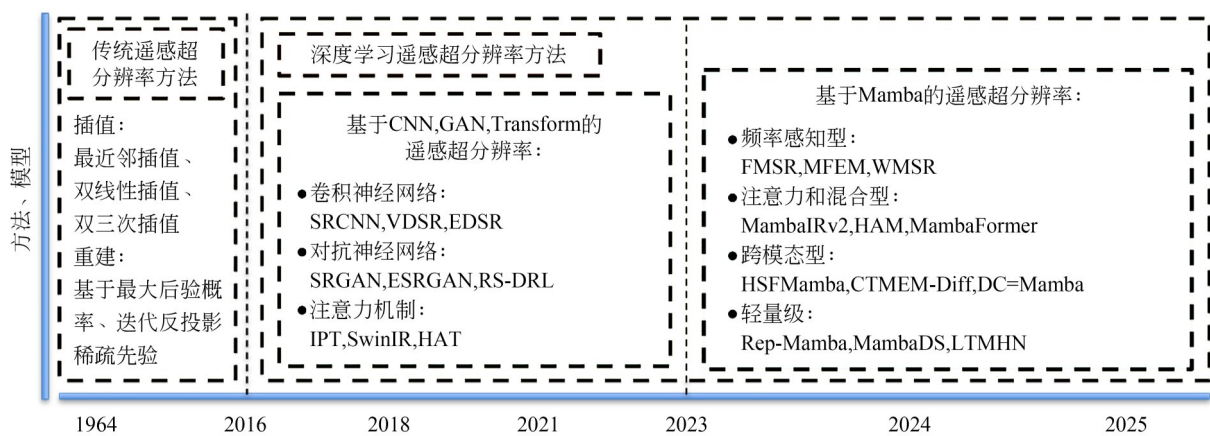


图 1 遥感图像超分辨率方法发展历程

Fig. 1 Development history of super resolution methods for remote sensing images

本文面向遥感图像超分辨率重建任务,系统回顾从传统插值方法、基于先验的重建方法到深度学习模型的发展演变脉络,突出不同技术路线在遥感场景中的适用特点与演进关系;以 Mamba 为主线,对其在遥感超分中的代表性改进方向进行归纳,包括频域增强、结构融合、跨模态与多源信息注入以及轻量化与可部署设计等,并分析不同设计对纹理细节恢复、全局结构一致性和推理效率的影响;结合典型公开数据集与代表性方法的实验结果,对以 Mamba 系列为代表的遥感超分辨率方法在 AID, DIOR 与 UCAS\_AOD 等典型数据集上的重建性能进行比较分析,并结合定量指标与主观视觉对比对重建效果进行讨论,从而为后续相关研究提供参考。

## 2 图像超分辨率重建

本章首先简要回顾插值与先验约束等传统重建方法,随后概述 CNN、GAN、Transformer 等

深度学习框架在遥感超分领域的代表性进展,进而系统梳理 Mamba 及其改进模型在遥感图像重建中的关键设计与演化趋势。需要说明的是,文中涉及的高光谱重建、跨模态超分及连续尺度重建等内容,主要用于补充展示 Mamba 框架在不同遥感任务中的扩展方式。由于其输入数据形态、任务目标与退化假设与单幅光学遥感超分存在差异,本文仅从模型结构与改进思路角度进行归纳讨论。

### 2.1 传统方法

早期的遥感图像超分主要依赖传统方法,包括插值放大和基于图像先验的重建两大类。

插值作为最基本的图像增强手段,可分为线性和非线性两类,整体的插值方法如图 2 所示。线性插值如最近邻、双线性和双三次插值等通过邻近像素的加权求和估计未知像素值;非线性插值则融合频域或边缘信息,如基于小波系数的多分辨率插值和基于边缘检测的方向性插值。传统插值利用已知像素对图像进行上采样,算法实

现简单且计算开销低,在要求实时性的遥感应应用中具有一定价值。然而,由于插值算法仅能利用局部邻域的像素强度分布,无法创造出图像丢失的高频细节,因此往往会引入平滑效应或伪影,导致边缘细节模糊不清。例如,双三次插值在放大过程中对边缘进行卷积平滑,易产生模糊过渡。为此,有研究人员提出边缘定向插值,在插值过程中检测并沿着边缘方向进行像素推断,以减轻锯齿和模糊。Carrato 等<sup>[22]</sup>最早提出了利用非线性插值结合边缘方向约束的方法,在合成和实测数据上均取得了更清晰的重建结果,但计算复杂度也显著增加。随后, Li 等<sup>[23]</sup>在边缘定向插值基础上引入协方差系数自适应调整插值权重,对边缘区域采用较大权重插值以保持轮廓清晰,而非边缘区域仍采用双线性插值以降低计算量。该方法在一定程度上兼顾了边缘清晰度和运算效率,但内部纹理细节仍可能因线性插值而过度平滑。总体而言,大量改进型插值算法聚焦于抑制传统插值的模糊和伪影,例如结合小波变换提升高频分量重建、采用径向基函数实现自适应插值等。插值方法计算快捷,在大规模遥感图像初步放大处理中仍具参考价值,但其高频重构能力有限,难以满足高精度应用需求<sup>[24]</sup>。

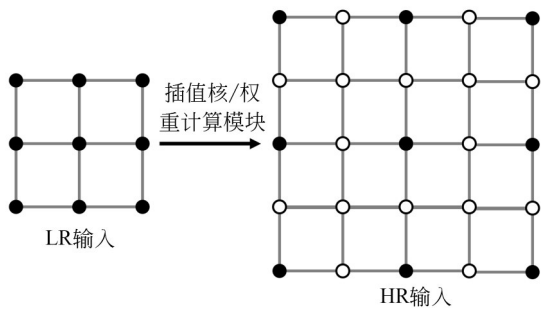


图 2 插值超分辨率示意图

Fig. 2 Schematic diagram of interpolation super-resolution

针对插值方法无法重构细节的问题,研究人员提出利用图像先验知识进行超分辨率重建,包括稀疏表示、字典学习及非局部相似性等代表性手段,整体基于先验的重建方法如图 3 所示。

稀疏表示方法通常假设高分辨率与低分辨率图像在同一冗余字典空间中具有一致的稀疏表示形式,即二者可以共享相同的稀疏系数进行线性组合表示。Yang 等<sup>[25]</sup>开创性地提出了将稀

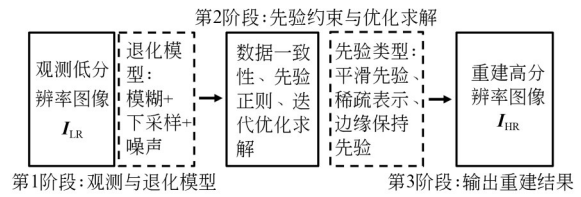


图 3 基于先验约束的遥感图像超分辨率重建示意图

Fig. 3 Schematic diagram of prior-constraint-based super-resolution reconstruction for remote sensing images

疏编码引入单帧图像超分辨率 (Super-Resolution, SR) 的思路,通过在 HR 和 LR 图像对上训练一对冗余字典,并针对每个 LR 小块求解其稀疏表示,再用对应 HR 字典重构得到高分细节。具体而言,该方法通过计算输入低分辨率图像小块在低分辨率字典下的稀疏编码系数,并将同一系数应用到高分辨率字典上,以重构出对应的高分辨率小块,最终拼接形成整幅图像。这种稀疏表示策略有效利用了先验训练库中的信息来弥补观测的欠缺,实现了比纯插值更精细的纹理重建。随后,大量改进的字典学习 SR 方法被提出以适应遥感数据的特点。Zheng 等<sup>[26]</sup>提出将稀疏表示同时用于遥感图像的超分和去噪,通过联合训练的 LR、HR 字典在稀疏重构中保持一致性,采用批量正交匹配追踪算法提高稀疏编码效率。Zhang 等<sup>[27]</sup>将过完备字典拆分为基础字典和残差字典两对,先用基础字典重构 LR 图像的初步 HR 估计,再训练残差字典刻画高频误差部分,从而提升重建精度。此外, Gou 等<sup>[28]</sup>结合非局部相似性提出成对的字典学习模型:在稀疏编码时,不仅利用训练样本,还利用待重建图像中与当前图像块相似的其他区域来共同约束解,提高了稀疏重建的稳定性。这类非局部先验方法利用遥感图像内部重复的纹理模式(如规则农田、建筑阵列等)来辅助超分,是对外部样本先验的有益补充。

另一方面,针对传统稀疏表示需保证 HR、LR 稀疏系数一致而较难满足的缺陷, Shao 等<sup>[29]</sup>引入耦合稀疏自编码器直接学习 LR 到 HR 稀疏编码的非线性映射,以减少稀疏系数不一致带来的重建误差。Deka 等<sup>[30]</sup>则提出一种自适应联合稀疏表示方法,巧妙地在不依赖外部训练样本的情况下,仅利用输入 LR 图像自身进行字典学习和稀疏重构,充分挖掘图像内含信息实现超分辨

率。Wang 等<sup>[31]</sup>提出的稀疏结构流形嵌入方法将几何先验融入邻域选择,通过在稀疏编码中加入局部几何约束来更好地恢复遥感图像的结构和边缘细节。总体来说,基于稀疏表示和先验约束的传统方法相较简单插值,在遥感图像细节重建方面取得了更佳效果。但这些方法也存在明显局限:例如算法往往计算复杂度高、依赖精心训练的字典或先验模型,当面对超大规模遥感数据时实时性不足;此外,对噪声和模型不匹配较敏感,鲁棒性有限。在遥感大数据时代,随着高分辨率遥感图像数量激增,这些传统方法的发展受到了极大限制。

尽管传统方法具有较强可解释性,但其性能高度依赖先验假设与退化模型的准确性,难以适应遥感图像复杂纹理与多样退化。因此,研究逐步转向端到端学习框架,以数据驱动方式联合建模退化与重建映射。

## 2.2 基于深度学习模型的方法

近年来,深度学习引领的技术革命为图像超分辨率领域注入了强劲动力。2014年 Dong 等<sup>[13]</sup>提出的超分辨率卷积神经网络(Super-Resolution Convolutional Neural Network, SRCNN)标志着 CNN 首次应用于超分辨率任务,实现端到端的非线性映射,重建效果显著优于传统方法。此后,引入残差学习的超深度超分辨率网络(Very Deep Super-Resolution, VDSR)、利用递归网络加深网络层数的深度递归残差网络(Deep Recursive Residual Network, DRRN)等<sup>[32]</sup>改进的 CNN 模型相继问世,在自然图像和遥感图像超分上均取得了持续提升,CNN 整体超分辨率网络结构如图 4 所示。

与传统方法相比,CNN 能够自动从大规模

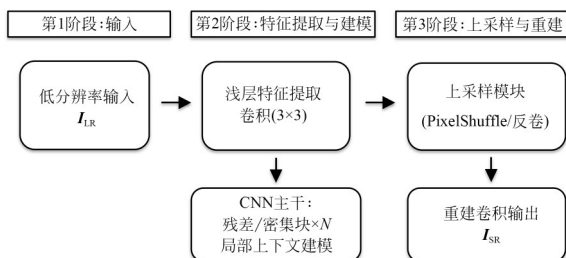


图 4 CNN 超分辨率网络结构示意图

Fig. 4 Schematic diagram of CNN super-resolution network architecture

数据中学习图像的多层次特征表示,有效提取复杂纹理和结构信息,从而提升重建质量。除了 CNN 外,GAN 策略也被引入遥感超分,以追求更逼真的感知效果。GAN 通过生成器和判别器的博弈训练,可以合成具有丰富细节和逼真纹理的图像。例如 Ledig 等<sup>[33]</sup>提出的超分辨率生成对抗网络引入对抗损失使重建图像的视觉效果显著改善,在遥感图像细节纹理保持方面表现出色。但 GAN 易出现伪影,需要精心平衡损失来避免过度假象纹理的产生。总的来说,CNN 注重提升客观指标,GAN 侧重提升主观视觉质量,二者为遥感图像超分提供了不同取向的解决方案<sup>[34]</sup>。

近年来,Transformer 架构在计算机视觉领域异军突起,给遥感图像超分带来了新的契机<sup>[35]</sup>。Transformer 最早在自然语言处理中凭借自注意力机制取得成功,Dosovitskiy 等<sup>[36]</sup>提出的视觉 Transformer(Vision Transformer, ViT)将其引入图像分类领域,证明 Transformer 在充足数据下可达到甚至超过 CNN 的性能。如图 5 所示,CNN 的特征聚合主要发生在局部邻域,需要多层堆叠才能逐步扩大感受野;而 Transformer 的自注意力可在单次计算中实现全局位置间的信息交互,从而更有效建模长程依赖。对于遥感图像这类大视角、高复杂度的数据,自注意力机制有助于捕获图像中跨区域的相关性,提高重建对全局结构和纹理的刻画能力<sup>[37-39]</sup>。

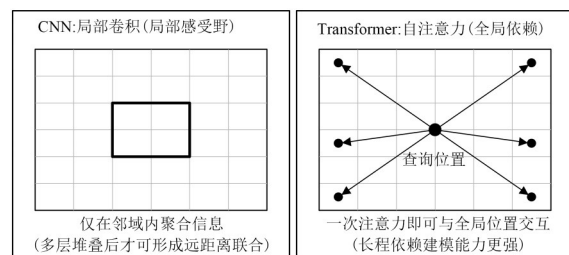


图 5 CNN 局部感受野与 Transformer 全局自注意力建模长程依赖的对比示意图

Fig. 5 Schematic diagram for comparing CNN local receptive field with transformer global self-attention for modeling long-range dependencies

图像超分领域中,Chen 等<sup>[40]</sup>于 2021 年提出了图像处理 Transformer (Image Processing

Transformer, IPT), 利用 ImageNet 大规模预训练 Transformer 用于包括超分在内的多种低级视觉任务。IPT 通过多任务预训练和多头多尾结构, 实现了单模型在多种分辨率提升任务上的性能超越当时最优 CNN 方法。在参考图像辅助的超分任务中, Yang 等<sup>[41]</sup>提出了纹理 Transformer 超分辨率网络(Texture Transformer for Image Super-Resolution, TTSR), 将待超分图像和参考高分图像分别作为 Transformer 中的查询和键, 设计硬注意力迁移高分辨率纹理, 再通过软注意力融合纹理细节, 从而大幅改善生成图像的质感, 如图 6 所示。这些开创性工作证明了 Transformer 在图像超分任务中的可行性和卓越潜力。

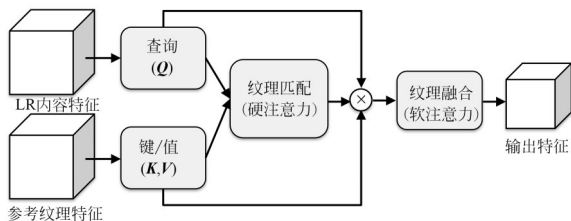


图 6 TTSR 参考纹理迁移注意模块示意图

Fig. 6 TTSR reference texture transfer attention module

针对遥感图像超分, Transformer 架构同样开始展现优势。Cai 等<sup>[42]</sup>率先将自注意力机制应用于高光谱(Hyperspectral Image, HSI)超分, 利用光谱自相似特性提高了重建的光谱保真度。Deng 等<sup>[43]</sup>提出了基于 Swin-Transformer 的遥感超分网络, 通过窗口划分和重新排列策略在多尺度窗口间聚合信息, 提升了对不同尺度细节的重建能力。Chen 等<sup>[44]</sup>将 HSI 沿光谱维拆分并嵌入可变形卷积 Transformer 模块, 以捕获长程光谱相关性, 实现跨波段的精细重建。Sun 等<sup>[45]</sup>设计了光谱-纹理混合注意力金字塔网络, 缓解弱纹理区域细节不足的问题。此外, 一系列面向自然图像但具有普适价值的 Transformer 超分模型也值得关注。Chen 等<sup>[15]</sup>在 CVPR2023 提出混合注意力 Transformer (Hybrid Attention Transformer, HAT), 融合通道注意力和窗口自注意力, 激活更多潜在像素关系以提升超分精度; 随后 Chen 等<sup>[46]</sup>在 ICCV2023 提出双重聚合 Transformer (Dual Aggregation Transformer, DAT), 面向图像超分同时聚合空间维与通道维两类信息; 其在

相邻 Transformer 块中交替采用空间自注意力与通道自注意力实现跨块特征聚合, 并在块内引入自适应交互模块与空间门控前馈网络进一步增强局部非线性表达, 从而在多个超分基准上取得领先性能。

在图像复原领域, Wang 等<sup>[47]</sup>提出了 U 型结构的 Uformer, 利用编码器-解码器 Transformer 高效整合多尺度信息用于降噪和超分。Zamir 等<sup>[48]</sup>在 CVPR2022 推出图像复原 Transformer (Restoration Transformer, Restormer), 通过改进自注意力和前馈模块, 大幅降低计算复杂度, 使 Transformer 能高效处理高分辨率图像, 同时在去雨、去模糊等任务上取得领先。这些 Transformer 架构在遥感图像超分任务中同样具有借鉴意义。一些研究已将上述模型应用于遥感图像并取得优异结果, 例如将 Restormer 用于合成孔径雷达成像的超分去噪, 证明 Transformer 在保持细节同时具有出色的全局建模能力。

Transformer 虽然优势明显, 但也存在局限<sup>[49]</sup>。首先, 自注意力计算开销随图像尺寸平方级增长, 在处理高分辨率遥感图像时内存和算力需求巨大。对尺寸为  $HW$  的特征图计算自注意力需  $(HW)^2$  的复杂度, 严重制约了 Transformer 在大幅面遥感图像上的实时应用。为此, 例如移位窗口 Transformer<sup>[50]</sup>这类改进架构通过将自注意力限制在局部窗口并引入分层结构, 降低了计算量, 但仍难以从根本上避免长序列建模的瓶颈。其次, Transformer 模型参数量通常远超同等规模的 CNN, 需要大量训练数据支撑。遥感领域标注高质量 HR、LR 训练对往往匮乏, 使 Transformer 在小数据集上可能出现欠拟合或不稳定训练。另外, Transformer 缺少 CNN 的某些归纳偏置, 例如平移不变性, 对于遥感图像中的细粒度模式学习并不一定总是优势。因此, 有研究开始探索融合 Transformer 与其他模型的混合架构, 以兼顾各自优点, 如 CNN 搭配 Transformer 的混合网络既利用卷积提取局部细节又利用自注意力捕获全局依赖。综合来看, Transformer 为遥感图像超分带来了新思路, 其强大的全局建模能力有望提升重建效果, 但在计算效率和数据依赖方面的不足也需要通过架构改进和训练策略加以克服。

面对 Transformer 计算效率的挑战,自 2023 年起出现的 Mamba 架构为遥感超分提供了全新的突破口<sup>[16]</sup>。Mamba 源于选择性状态空间模型 (Selective State Space Model, SSM),通过在经典连续状态空间模型中引入选择性门控机制,实现了对长序列依赖的近线性复杂度建模。与仅有局部感知的 CNN 和计算复杂度随图像尺寸平方级增长的 Transformer 相比, Mamba 在保持全局建模能力的同时显著提高了计算效率。这种特性使其能够在处理大幅面遥感图像时,突破 CNN 感受野受限与 Transformer 显存爆炸的双重矛盾。尽管 Mamba 在低层视觉任务上的应用尚处起步阶段,但其在效率与性能平衡上的巨大潜力,促使研究者们开始探索将其引入遥感图像超分领域,以期构建新一代的高效重建范式。

### 2.3 Mamba 基础理论

尽管 Transformer 依托自注意力机制提升了遥感图像超分辨率重建中的全局依赖建模能力,但其计算量和显存开销会随输入尺寸快速增长,在高分辨率、大幅面遥感影像处理中面临明显的效率瓶颈<sup>[51-53]</sup>。针对这一问题,具有长序列建模优势的状态空间模型为兼顾全局表征与计算效率提供了新的理论路径。与自注意力通过显式计算位置间相关性不同,状态空间模型通过隐状态递推实现历史信息压缩与跨位置传播,从而在较低计算开销下建立长距离上下文联系,其连续时间形式通常表示为:

$$\dot{h}(t) = Ah(t) + Bx(t), \quad (1)$$

$$y(t) = Ch(t) + Dx(t), \quad (2)$$

其中:  $x(t)$  表示输入,  $h(t)$  表示隐状态,  $\dot{h}(t)$  表示隐状态关于时间  $t$  的一阶导数,反映状态随时间的变化率,  $y(t)$  表示输出,  $A, B, C, D$  分别为状态转移矩阵、输入映射矩阵、输出映射矩阵和直通项矩阵。该形式以动态系统方式刻画输入、状态与输出之间的关系,为长距离依赖建模提供了统一表达。

为适应离散序列计算,连续状态空间模型通常需进一步离散化,其离散形式可写为:

$$h_k = \bar{A}h_{k-1} + \bar{B}x_k, \quad (3)$$

$$y_k = Ch_k + Dx_k, \quad (4)$$

其中:  $\bar{A}$  和  $\bar{B}$  分别表示离散化后的状态转移项和输入项。离散化后,模型能够沿序列递推更新隐

状态,并将前序信息持续传播至后续位置,从而在较长范围内建立上下文联系。与主要依赖局部感受野的卷积网络不同,这种递推机制更适合长序列信息建模。

在状态空间模型的发展过程中, Gu 等<sup>[54]</sup>提出的结构化状态空间序列模型 S4 通过对状态矩阵进行结构化参数化,提高了状态空间模型的稳定性与长序列表达能力,为后续选择性状态空间模型的发展奠定了基础。Gu 等<sup>[16]</sup>进一步提出的 Mamba 在 S4 的基础上引入选择性状态空间机制,使输入映射、输出映射以及离散步长能够随当前输入动态变化,具体的公式为:

$$B_k = f_B(x_k), C_k = f_C(x_k), \Delta_k = f_\Delta(x_k), \quad (5)$$

其中:  $B_k, C_k$  和  $\Delta_k$  分别由当前输入自适应生成。与固定参数状态空间模型相比,这种输入依赖的参数化方式使模型能够依据内容差异选择性地保留、更新或抑制信息,从而增强对关键纹理、边缘结构和长程依赖关系的建模能力。

与 Transformer 通过自注意力显式建模任意位置间两两相关性不同, Mamba 采用基于状态空间的递推更新机制完成信息传播。对于长度为  $N$  的序列,自注意力通常需要建立规模随  $N^2$  增长的相关性矩阵,而 Mamba 在每个位置仅执行当前输入与隐状态的递推更新,无需显式构造全局交互矩阵,因此其计算复杂度通常随序列长度近似线性增长。该机制在保持长程依赖建模能力的同时显著降低了高分辨率输入下的计算负担,也为后续视觉 Mamba 结构及其在遥感图像超分辨率重建中的扩展应用提供了理论基础。

### 2.4 基于 Mamba 架构的方法

#### 2.4.1 视觉 Mamba 基础架构

随着 Mamba 由序列建模逐步扩展到视觉任务,研究人员围绕二维图像结构建模进一步发展出多种视觉状态空间模型。与自然语言等一维序列不同,视觉数据同时具有局部空间邻域关系和长距离上下文依赖,因而如何在保持近线性复杂度优势的同时有效建模二维结构,已成为视觉 Mamba 架构设计的关键问题。围绕这一目标,现有代表性工作主要从双向状态建模、二维扫描机制与局部窗口建模等角度展开改进,为后续遥感图像超分辨率重建中的扫描策略设计、结构连续性恢复与局部细节增强提供了基础。

Zhu 等<sup>[55]</sup>提出的双向状态空间模型 Vision Mamba,通过构建双向状态空间视觉骨干网络,将图像块序列表示与双向 Mamba 模块结合,以缓解单向扫描对视觉全局表征的限制。该方法说明,基于状态空间模型的双向信息传播同样能够形成有效的视觉表征,从而为 Mamba 在视觉任务中的应用奠定了基础。围绕视觉数据中二维结构难以由简单一维序列展开充分刻画的问题,Liu 等<sup>[56]</sup>进一步提出了视觉状态空间模型 (Visual State Space Model, VMamba)。该方法面向二维视觉数据引入视觉状态空间模块,并设计二维选择性扫描机制 SS2D,通过沿多个方向对二维特征进行展开与扫描,建立跨空间位置的上下文联系。相较于直接将图像线性展开为一维序列的方式,VMamba 在二维结构保持和全局依赖聚合方面更具代表性,也使视觉状态空间模型具备了更强的图像建模能力。尽管 VMamba 增强了二维上下文建模能力,但传统序列化扫描仍可能破坏相邻像素或相邻块之间的局部邻接关系,并导致局部依赖建模路径过长。针对这一问题,Huang 等<sup>[57]</sup>提出了窗口选择性扫描视觉状态空间模型 (Visual State Space Model with Windowed Selective Scan, LocalMamba)。该方法采用窗口化选择性扫描策略,在局部窗口内优先建模邻域结构,再结合分层扫描实现更大范围的信息传播,从而在保持全局建模潜力的同时增强对局部细节与边缘结构的表征能力。

面向低层视觉任务,Guo 等<sup>[58]</sup>提出了基于状态空间模型的图像复原基线网络 (A Simple Baseline for Image Restoration with State-Space Model, MambaIR),率先将具有选择性扫描机制的视觉状态空间模型引入图像重建任务。MambaIR 的核心思想是通过一次序列扫描实现全局依赖建模,从而在全局感知能力与计算效率之间取得更优平衡。如图 7 所示,视觉状态空间模型通常首先将二维空间特征图按一定顺序展开为空间序列,随后在序列域执行状态空间更新,并通过状态的逐步传播建立长距离依赖,最终将序列输出回写为二维特征图,从而在不显式构建全局注意力矩阵的条件下实现跨区域信息聚合。由于状态在扫描过程中持续累积并向前传递,单次扫描即可使任意位置之间建立间接信息通路,

这也为后续遥感图像超分辨率重建中的高效结构建模提供了基础范式。

基于上述视觉 Mamba 基础架构,研究人员进一步将其扩展到遥感图像超分辨率重建任务,并围绕频域信息补偿、网络结构融合、跨模态建模和模型轻量化等方向提出了一系列改进方法<sup>[59-65]</sup>。为便于梳理相关研究脉络,下文将其归纳为频域建模与频谱增强、结构融合与注意力增强、跨模态与多源遥感建模以及轻量化与重参数化结构四类,并分别进行介绍。

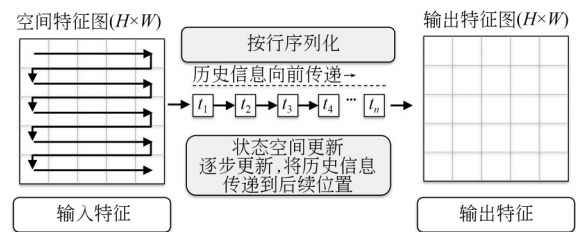


图 7 视觉 SSM 一次扫描建模全局依赖的流程示意图  
Fig. 7 Schematic diagram of process for visual state space model to model global dependencies using a single scan

#### 2.4.2 频域建模与频谱增强类

Mamba 模型虽然具备全局建模能力,但其序列扫描与状态更新过程的局部感知方式导致一定高频细节损失<sup>[66-68]</sup>。为弥补这一不足,Xiao 等<sup>[69]</sup>提出的快速傅里叶 Mamba 超分辨率网络 (Frequency-Assisted Mamba for Remote Sensing Image Super-Resolution, FMSR) 在 Mamba 模块外部引入了频域分支,利用快速傅里叶变换 (Fast Fourier Transform, FFT) 提取频谱信息并融合到重建过程中,在仅增加少量计算的情况下显著缓解了高频细节丢失的问题。基于类似思路,Zhang 等<sup>[70]</sup>提出的 WMSR-med 方法 (Wavelet-based Mini-Grid Network for Medical Image Super-Resolution) 改进了频域融合方式,引入维纳全局滤波的注意力机制来自适应增强频率成分。相比 FMSR 仅在 Mamba 块外简单的频域连接,WMSR-med 通过全局维纳滤波注意力更高效地建模图像的频率分量,不仅提升重建精度且参数量也完成了一定的减少。上述频域增强策略均在一定程度上改善了 Mamba 的频率感知,但由于频域处理模块置于主干之外,仍未从根本

上解决 Mamba 内部的信息损失。

Chen 等<sup>[71]</sup>提出了多尺度频率增强 Mamba 轻量遥感图像超分辨率网络(Multiscale Frequency-Enhanced Mamba for Lightweight Remote Sensing Image Super-Resolution, MFEM)。MFEM 将频域融合直接嵌入 Mamba 主干中,包含两个核心模块:多尺度可分蒸馏块和频率 Mamba 增强块。前者通过设计 3 种规格的多尺度残差块实现不同尺度下的特征提取与融合,利用可分卷积和通道拆分保持模型轻量级的同时,增强对遥感图像复杂结构的表征能力;后者则构建自适应频域处理分支,对特征进行频域变换并通过门控单元选择性增强高频分量,同时引入 2D 选择性扫描机制与时域特征交互,从内部有效缓解了 Mamba 全局建模中局部信息缺失的问题。总体而言,MFEM 在保持模型轻量化的同时,实现了频域信息与空间特征的分组融合,在多尺度复杂地物细节重建上相比同类方法有明显优势。除此之外,Chen 等<sup>[72]</sup>针对海洋遥感数据提出了小波辅助 Mamba 超分辨率框架(Wavelet-assisted Mamba Super-Resolution, WMSR)。为便于理解其核心设计,图 8 给出了 WMSR 的关键创新模块示意图。该方法首先通过离散小波变换将输入特征分解为低频与高频两类子带;低频分支采用 SSM 建模全局背景模式,高频分支通过高频增强模块(High-Frequency Enhancement Module, HFEM)提取纹理细节,最终在融合后完成超分辨率重建。该分解建模与融合范式能够同时兼顾平滑背景与细微高频结构,尤其适用于海表温度等具有强低频主导且局部细节关键的遥感数据超分任务。这种频域分解结合状态空间建模的方式针对海表温度遥感图像的平滑背景和细微涡流同时优化,在 SST 超分实验和 3 个海洋数据集

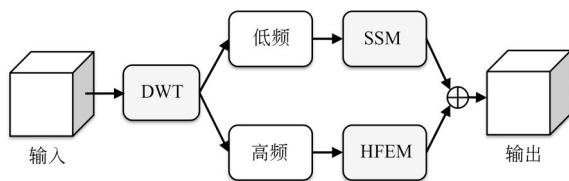


图 8 小波辅助 Mamba 频域分解(LF, HF)与分支融合的创新模块示意图

Fig. 8 Schematic diagram of novel module in wavelet-assisted mamba for frequency domain decomposition (LF, HF) and branch fusion

中表现出色,生成的高分辨率温度图能更清晰地呈现海洋涡旋等细节。

综上,频域建模与频谱增强类 Mamba 方法通过在频率域对全局信息与细节纹理进行补偿建模,能够有效提升遥感图像的细节重建能力。这类方法通常在较小额外开销下改善高频纹理恢复与边缘清晰度,并体现出频域信息引入对多尺度特征重建的积极作用。

从信息表征角度看, Mamba 的状态更新更擅长在长程范围内聚合低频背景与结构语义,而高频纹理在跨层传播过程中更容易被平滑或被门控削弱,表现为细节能量不足与边缘对比度下降。频域增强类方法的有效性主要来源于将纹理恢复从纯空间域匹配转化为对幅相或子带能量的显式调制:一方面,FFT, DWT 等变换能够将细节与背景解耦,使网络在统一尺度下直接对高频分量进行补偿;另一方面,门控或注意机制可对不同频段的贡献进行自适应分配,避免将噪声一并放大。其中,频域分支通常对规则纹理与重复结构更敏感,而在低信噪比或退化非平稳的情形下,高频补偿可能引入振铃或过锐化倾向,因此需要与空间一致性约束协同设计,以提升真实细节恢复的稳定性。

相较于依赖局部空间域表征的 CNN, 频域增强类方法能够更直接地补偿高频细节;相较于更强调全局依赖建模的 Transformer, 其在边缘纹理恢复方面通常更具针对性。因此, 该类方法更适用于纹理丰富、边缘密集且对局部细节恢复要求较高的遥感场景。

#### 2.4.3 结构融合与注意力增强类

尽管 Mamba 模型具备较强的长程依赖建模能力,但在多尺度信息协同、局部细节强化和全局上下文交互方面仍存在一定局限。为此,部分研究通过融合注意力机制与改进网络结构设计,进一步提升模型的特征表达能力和空间结构恢复能力。首先,Guo 等<sup>[73]</sup>提出的 MambaIRv2 针对原始 Mamba 序列建模的因果性限制提出了解决方案。传统 Mamba 在扫描图像时每个像素仅依赖其序列前驱像素,导致未扫描区域的信息无法利用,需要多方向多次扫描来弥补感受野。MambaIRv2 引入了注意状态空间模型,其核心是在状态空间方程的输出项中加入可学习的提示

向量,从而打破严格因果约束。通过在MambaIR的基础上引入提示矩阵 $P$ ,实现了更为灵活的特征提取。具体公式如下:

$$h_i = Ah_{i-1} + Bx_i, \quad (6)$$

$$y_i = (C + P)h_i + Dx_i, \quad (7)$$

其中: $A, B, C, D$ 分别为状态转移、输入、输出和直通项矩阵, $P$ 为融合全局语义的提示矩阵。通过在输出投影中叠加全局提示,MambaIRv2使当前像素的重建既依赖已扫描序列,又能够查询尚未扫描的相关像素语义,从而实现类似ViT的非因果全局建模。

MambaIRv2还提出语义引导邻近化(Semantic Guided Neighboring, SGN)机制,根据像素语义相似度动态重排序列顺序,将原本距离遥远但语义相近的像素临近排列,进一步缓解了长距离依赖随距离衰减的问题。为直观展示提示矩阵与状态空间输出投影的结合方式,以及SGN对序列组织的辅助作用,如图9所示。得益于多方向扫描策略,并结合全局提示机制与语义邻域建模的设计,MambaIRv2在保持线性复杂度的同时获得了接近Transformer的全局感知能力。MambaIRv2的研究证明,通过引入类注意力机制的改造,SSM模型同样能突破因果瓶颈,在图像重建中取得更高的精度。

另一代表性工作是Li等<sup>[74]</sup>提出的层次注意力Mamba网络(Hierarchical Attention Mamba, HAM)。HAM针对遥感图像超分存在的多尺度结构难以重建和局部-全局相关提取不足等问题,在Mamba骨架中融合了分层特征交互和频域增强注意力机制。其创新点在于设计分层聚合注意力模块(Hierarchical Aggregation Attention, HAA)和空间-频域信息交互模块(Spatial-Frequency Information Interaction Module, SFIIM)。HAA以金字塔方式逐级聚合不同层次特征,增强多尺度特征的传播和利用。SFIIM则同时融合频域和空间域信息,一方面通过动态频谱增强技术将特征在傅里叶幅-相空间分解强调判别性频率成分,另一方面采用双分支通道门控的混合空间特征聚合器提取空间细节。通过HAA和SFIIM的协同作用,HAM实现了局部细节与全局语义的同步提取,大幅提升了重建图像的空间一致性和高频细节逼真度。在航空场景分类

数据集(Aerial Image Dataset, AID)、光学遥感目标检测基准(Detection in Optical Remote, DIOR)、高分辨率航空目标检测数据集(UCAS High Resolution Aerial Object Detection Dataset, UCAS\_AOD)等6个遥感图像超分基准上的评估表明,HAM相比现有方法取得了较高的重建指标,重建图像的边缘纹理更清晰且结构失真更少,同时模型推理效率与基础Mamba相当。这说明,将Transformer式的分层注意力融合进Mamba,可以充分发挥两者所长,实现精度与效率的平衡。

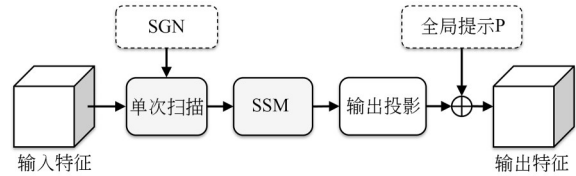


图9 MambaIRv2提示增强注意状态空间模块示意图  
Fig.9 Schematic of prompt-augmented attentive state-space module in MambaIRv2

Zhi等<sup>[75]</sup>设计的轻量级遥感图像超分辨率模型(A Lightweight Model for Remote-Sensing Image Super-Resolution, MambaFormerSR)将Mamba与Transformer显式融合,提出了一种轻量级遥感图像超分模型。Zhi等认为Mamba的全局建模能力可与注意力机制互补,因而设计了状态空间与注意力融合模块(State Space and Attention Fusion Module, SAFM)来结合Mamba的长程依赖捕获与Transformer的全局自注意力,以充分提取图像特征。随后,引入改进的前馈网络模块卷积傅里叶前馈网络(Convolutional Fourier Feedforward Network, CTFFN),通过在前馈层中融合卷积(提取局部信息)和频域注意力(聚焦关键信息频段)来精细强化特征表示。最后增加一个辅助重建分支,用于先重建粗略图像信息,再让主干专注细节重构。与同等复杂度模型相比,MambaFormerSR在保持模型小型化的同时有效融合了SSM和Transformer两种框架的优势,证明了异构融合在遥感超分领域的可行性。所提出的SAFM和CTFFN模块为后续将Mamba与其他机制结合提供了思路。

最后,Transformer和Mamba结合的思路也

应用于高光谱重建等任务中<sup>[76-78]</sup>。如 Lin 等<sup>[79]</sup>提出的 SR-CSCA 网络 (Spectral Reconstruction via Dual Cross-Scanning and Cross-Attention Mechanisms), 将双重交叉扫描与交叉注意力机制用于光谱重建。该方法利用 Mamba 的选择性扫描分别沿空间和光谱两个维度对高光谱数据进行序列展开和特征提取, 得到多分辨率的上下文信息。随后, 通过 Transformer 式的跨模态交叉注意力融合光谱与空间特征, 实现高效的高光谱重建。SR-CSCA 本质上构建了 Mamba 与 Transformer 的融合架构, 在保持接近线性时间复杂度的同时, 捕获了遥感高光谱数据的跨维度关联。实验表明, 与传统 CNN 和纯 Transformer 方法相比, SR-CSCA 在从 RGB 影像重建 HSI 的任务中取得了更高的光谱重建准确度, 尤其在细节纹理和光谱一致性方面有明显优势, 同时计算开销较低。这进一步验证了将注意力机制融入 Mamba 架构能够提升复杂遥感任务性能。此类方法状态空间建模与注意力交互协同的共同范式, 可采用双分支融合框架进行统一描述, 如图 10 所示。

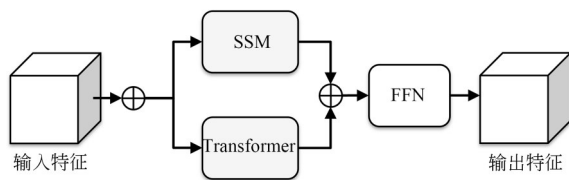


图 10 Mamba 与 Transformer 双分支协同融合模块示意图  
Fig. 10 Schematic of dual-branch fusion module combining Mamba and Transformer

总体而言, 结构融合与注意力增强类方法通过引入非因果建模、分层注意力模块以及 Transformer 融合等机制, 进一步提升了 Mamba 模型对多尺度信息和长程依赖的建模能力, 并在多种遥感超分场景中表现出较强的重建优势。其核心动机在于缓解序列扫描过程中信息可见性受限以及长距离依赖随路径衰减的问题。单向因果扫描使局部重建更依赖前驱上下文, 未被覆盖区域难以参与当前像素判别, 因此在大尺度结构恢复和跨区域一致性保持方面存在一定局限。为此, 引入提示向量、语义引导重组或分层交互结构, 本质上是在状态空间输出映射或序列组织层面增强全局信息的可访问性, 使模型在一次扫描

过程中能够建立更充分的跨位置关联。

此类改造虽然提高长边界连续保持能力与重复结构重建稳定性, 但其性能仍较大程度依赖于语义相似度估计和层级交互策略的可靠性。当场景中存在较强非结构纹理或局部细节极弱时, 过强的全局交互也可能带来纹理迁移偏差, 因此, 需要在全局聚合与局部细化之间保持合理的权重分配。相较于受限局部感受野的 CNN, 结构融合与注意力增强类方法在全局结构保持、跨区域关联建模和多尺度信息协同方面更具优势; 相较于标准 Transformer, 其在保持较强全局建模能力的同时, 通常还能以相对较低的计算代价实现长序列建模。因此, 该类方法更适用于场景结构复杂、尺度变化显著且对整体几何一致性要求较高的通用遥感超分任务。

#### 2.4.4 跨模态与多源遥感建模类

除单幅光学遥感超分外, 近年来也有研究将 Mamba 引入跨模态与多源数据的重建场景。此类问题通常涉及不同传感器成像机理与退化分布差异, 方法设计更强调特征对齐与信息选择性融合。典型场景之一是光学与 SAR 的跨模态超分重建, Zhao 等<sup>[80]</sup>提出的分层选择融合 Mamba 网络 (Hierarchical Selective Fusion Mamba Network, HSF Mamba) 旨在利用高分辨率光学图像指导低分辨率、有噪声的 SAR 图像实现联合超分和去噪重建。HSF Mamba 的核心在于两个模块: 跨模态特征选择模块 (Cross-Modal Feature Selection Module, CFMS) 和频率-空间自适应聚合模块 (Frequency-Spatial Adaptive Aggregation Module, FSAM)。其中, CFMS 从状态空间模型参数入手, 引入输入依赖的门控机制来动态调整 Mamba 的状态矩阵, 使之针对 SAR 图像的特点提取所需的光学引导信息。简单来说, CFMS 可在不同行扫描阶段精细挑选来自光学图像的多尺度特征, 并通过门控单元控制融合强度, 从而在各层逐步注入有用的引导信息。随后, FSAM 在融合阶段同时考虑频域和空间域: 一方面在频域上自适应强调光学引导的高频细节, 另一方面在空间域利用双分支结构减弱 SAR 斑点噪声的影响。光学引导 SAR 的跨模态融合流程如图 11 所示。这种分层逐级融合的策略使 HSF Mamba 能够在抑制斑点噪声的同时尽可能恢复 SAR 图

像的细节结构。在构建的大规模光学-SAR超分去噪数据集 OGSSR 上,HSFMamba 在各种降质模式下均取得了优于现有方法的定量指标;主观视觉上,重建的 SAR 图像边缘更锐利、纹理更加逼近真实光学影像。这表明,通过跨模态信息选择与频空域融合,Mamba 架构可以成功应用于多源遥感数据的联合重建任务。

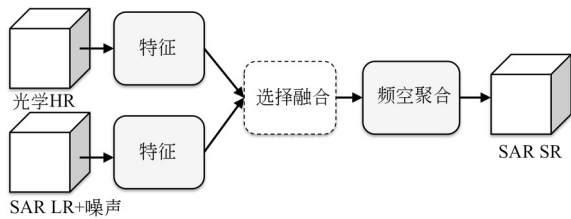


图 11 光学引导 SAR 跨模态选择融合与频空聚合示意图  
Fig. 11 Schematic of optical-guided SAR cross-modal selective fusion and spatial-frequency aggregation

针对 HSI 与多光谱 (Multispectral Image, MSI) 数据的跨模态融合, Xu 等<sup>[81]</sup>提出了跨模态纹理匹配与增强 Mamba 扩散模型 (Cross-Modal Texture Matching and Enhancement Mamba Diffusion Model, CTMEM-Diff), 面向无配对 HSI 图像超分, 将扩散概率模型与 Mamba 有机结合<sup>[82]</sup>。具体来说, CTMEM-Diff 以预训练的扩散生成模型为基础, 在逐步去噪生成高分辨率 HSI 的过程中, 引入多源数据的纹理迁移。每次扩散迭代时, 都将低分辨率 HSI 与高分辨率 MSI 分别输入一个退化匹配块 (Degradation Matching Block, DMB), 通过针对两种模态设计不同退化函数将它们映射到统一特征空间, 从而在统计上匹配可比的区域。然后, 设计纹理一致 Mamba 去噪模块 (Texture-Consistent Mamba Denoising Module, TCMM), 利用 Mamba 的全局上下文建模能力和选择性门控机制, 将匹配到的 MSI 细节纹理高效地转移到 HSI 上。TCMM 在每个扩散步通过门控单元融合来自 MSI 的细节信息与当前 HSI 特征, 使模型在迭代去噪中逐步增强 HSI 的细节。图 12 直观展示了扩散迭代中跨模态纹理迁移的整体流程。这样的框架无需成对训练数据, 就能让 HSI 从 MSI 中汲取高频纹理<sup>[83-85]</sup>。

CTMEM-Diff 构建了 3 个包含多源遥感数据的新数据集进行测试, 结果显示, 在所有常用评

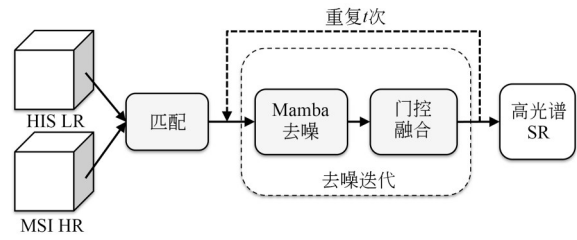


图 12 扩散模型结合 Mamba 实现跨模态纹理迁移的去噪迭代示意图

Fig. 12 Denoising iteration diagram for cross-modal texture transfer using a diffusion model combined with Mamba

价指标上该方法均优于现有无配对高光谱超分辨率方法。尤其是在纹理复杂的遥感场景下, CTMEM-Diff 生成的高光谱图像细节更加清晰、光谱曲线更加精确, 验证了跨模态纹理增强策略的有效性。CTMEM-Diff 充分利用了扩散模型逐步求精的生成过程, 将 Mamba 引入扩散框架实现跨模态特征对齐和纹理迁移, 开创了遥感跨模态超分的新方向<sup>[86-88]</sup>。

Si 等<sup>[89]</sup>提出了降解感知型跨模态 Mamba 框架 (Degradation-aware Cross-modality Mamba, DC-Mamba)。DC-Mamba 的一大特色是显式考虑了成像退化的复杂性, 引入退化感知的跨模态融合策略。具体而言, DC-Mamba 首先设计了一个自监督降质学习模块, 直接从低分辨率红外图像中估计模糊核、噪声水平等退化信息, 并将其编码为指导信号。这些降质先验随后用于指导 Mamba 模块的恢复过程: 该框架嵌入若干降质感知模块, 将上述降质信息融合到状态空间建模中, 使模型对失焦、噪声等失真区域更加敏感, 从而有针对性地加强这些区域的重建。与此同时, DC-Mamba 引入可见光引导, 利用配准后的高分辨率可见光图像, 与热红外特征进行选择性交叉融合, 以补充热图像中缺失的细节纹理。Mamba 的全局建模能力确保了这种跨模态融合可以在像素级对齐不完美的情况下仍保持全局一致性。实验结果表明, 在包含真实复杂退化的热红外遥感图像上, DC-Mamba 相比仅用单模态的方法重建效果取得了大幅提升。在典型场景下, DC-Mamba 重建的红外图像无论在边缘清晰度还是纹理保真度上都明显优于传统 CNN 和 Transformer 基准方法; 同时由于采用线性复杂度的

SSM 单元,模型在高分辨率输入下的推理效率也优于自注意力模型。DC-Mamba 的研究展示了将 Mamba 用于跨模态盲超分的潜力,通过融合降质先验和可见光信息,成功应对了实际应用中多样复杂的退化情况。

总体而言,跨模态与多源建模类方法通过引入辅助模态提供的互补信息,为遥感超分重建提供了超出单一光学观测的先验支撑,因而在复杂退化、信息缺失或目标结构不完整等条件下展现出较大潜力。这类方法的关键并不在于简单叠加多源信息,而在于如何有效处理不同传感器在噪声分布、空间分辨率、辐射响应和几何一致性等方面的差异。

从融合机制看,该类方法通常通过选择性融合策略在特征层面对引导信息进行可信度调制,使网络能够在纹理可迁移区域更充分利用辅助模态,而在配准误差较大或噪声占优区域抑制不可靠的引导成分。Mamba 的优势在于其长程上下文聚合能力可在较大范围内提供结构一致性约束,有助于在弱配准条件下维持整体几何形态稳定。但多源融合增益通常与跨模态对齐质量高度相关,若辅助模态与目标模态的结构对应关系不稳定,模型仍易产生结构错配或细节幻觉,因此,需要结合对齐模块、退化感知建模或不确定性估计来提升融合稳健性。

相较于仅依赖单幅光学图像的 CNN 或 Transformer 方法,跨模态与多源建模类方法能够利用辅助模态提供的互补信息,在复杂退化或信息缺失条件下获得更稳定的重建增益。因此,该类方法更适用于多源数据较完备、模态间具有较好结构对应关系,且对复杂场景重建精度要求较高的应用场景。

#### 2.4.5 轻量化与重参数化结构类

Mamba 模型在提升效果的同时也带来了较高的模型复杂度,因而另一研究重点是如何优化 Mamba 结构以降低参数量和计算量<sup>[90-94]</sup>。Jiang 等<sup>[95]</sup>提出的重参数化视觉 Mamba 轻量遥感图像超分辨率网络(Re-Parameterization in Vision Mamba for Lightweight Remote Sensing Image Super-Resolution, Rep-Mamba)。Rep-Mamba 通过重参数化技术和渐进式多尺度融合设计,打造一个兼顾效率与精度的 Mamba 轻量网络。

首先,Rep-Mamba 没有直接采用标准 Mamba 模块,而是设计了跨尺度状态传播机制(Cross-Scale State Propagation, CSSP),构建一个轻量级渐进融合模块(Lightweight Progressive Fusion Module, LPFM)用于特征提取。从网络结构来看,LPFM 包含多个并行分支来处理不同尺度的特征,并在每个阶段通过跨尺度残差连接和级联,进一步动态融合各分支,从而鼓励网络捕获遥感图像中层次丰富的空间依赖关系,同时避免单一路径下信息丢失。其次,为提升推理效率,Rep-Mamba 在 LPFM 中引入可微重参数卷积(RepConv)结构。在训练阶段,每个 RepConv 模块由卷积核组合构成更宽的结构,以增强局部结构和全局上下文的联合建模能力;而在推理部署时,这些卷积分支可等价重组为单一卷积核,从而在不影响性能的前提下大幅简化模型计算。

图 13 为 Rep-Mamba 所设计的 RepConv-Mamba 双分支块结构,一支通过池化和 2D-SSM 实现全局细节提取;另一支通过 RepConv 保留局部细节,再将两支特征融合重构输出。该结构在训练时优化了全局与局部建模的协同,在推理时则简化为高效单元,实现了性能与复杂度的优良折中。

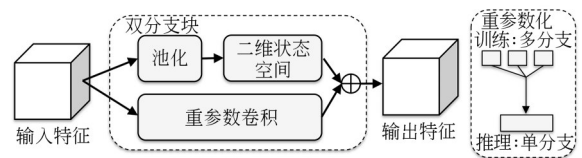


图 13 轻量化重参数化双分支块示意图

Fig. 13 Schematic of lightweight dual-branch block with re-parameterization

除了在通用超分领域的压缩优化,Mamba 架构在特定领域的轻量化改造也有探索。Liu 等<sup>[96]</sup>提出的近地面气象场下尺度化模型(Near-Surface Meteorological Field Downscaling with Topography Constrained Selective State Space Modeling, MambaDS)。将 Mamba 用于气象场景下的下采样重建任务,并针对多变量、物理先验强的特点进行了定制优化。一方面,MambaDS 在模型结构上引入多变量相关增强的视觉状态空间模块,通过增加平行通道注意力分支来捕获不同气象变量之间的相关关系。进而将各气象要素

映射到 Mamba 的不同通道,并通过通道注意力在状态更新时建模它们的相关性,让模型同时学习多变量之间的耦合规律,而非简单地将多通道当作独立的 RGB 通道处理。另一方面, MambaDS 在重建末段设计了一个高效的地形约束层。利用数字高程模型等地形先验,以极低计算开销将地形影响融合到输出中,从物理一致性上约束下采样结果。这一设计避免了以往方法需要将地形信息嵌入中间特征所带来的额外计算,同时确保了生成的高分气象场符合实际地形分布。实验结果显示,在中国及北美区域的多个气象下采样基准上, MambaDS 在温度、降水、风场这 3 种典型气象要素的重建中均取得当前最佳表现。与传统 CNN 和 Transformer 下采样方法相比精度全面提升,并超越了直接应用原始 Mamba 架构的效果。这证明了 Mamba 模型在融合物理先验、适应领域需求方面具有很高的灵活性和潜力。通过针对性地增强多源信息利用并加入物理约束, MambaDS 成功将 SSM 方法扩展到了遥感气象数据重建这一全新领域。

最后, Zhang 等<sup>[97]</sup>提出了面向遥感图像超分辨率的可学习 TransMamba 混合网络(Learnable TransMamba Hybrid Network for Remote Sensing Image Superresolution, LTMHN)对 Mamba 和 Transformer 的融合轻量化探索。该方法提出可学习 TransMamba 块,以双分支结构同时包含 Mamba 状态空间分支和 Transformer 注意力分支,并通过可训练的适配器实现二者特征动态融合。这种设计思想与前述 MambaFormerSR 相似但更进一步。

LTMHN 在每个基本块中引入可学习融合参数,使网络可以自行调整 Mamba 分支与 Transformer 分支的输出比例,从而在训练过程中找到最优的融合策略。此外, LTMHN 引入了自适应通道增强模块,结合传统通道注意力和可学习门控机制,按任务需求突出重要通道、抑制冗余特征,经统计该模块使有效通道利用率提高了约 60%。在细节处理上, LTMHN 还设计了多尺度重叠交叉注意力机制,融合相邻窗口的多尺度信息,以增强图像边缘和纹理等精细细节的恢复。

凭借上述设计, LTMHN 在整体复杂度显著低于纯 Transformer 的同时,重建精度优于大部

分 Transformer 和 Mamba 模型。主观结果也显示, LTMHN 生成的影像在条带、建筑边缘等局部纹理上最为清晰逼真,几乎看不到过度平滑或伪影,与原始高分辨率影像的细节分布高度一致。这表明,通过双架构取长补短和模块级重参数化、注意力优化,可以在很大程度上突破各自单一架构的局限,实现远优于单模型的效果。LTMHN 为 Mamba 与 Transformer 的深度融合提供了借鉴,也印证了轻量化设计在保持甚至提升模型性能方面的巨大潜力。

轻量化与重参数化方法的目标在于将长程建模能力保留在可部署的复杂度预算内<sup>[98-100]</sup>。这类方法通常通过训练期增强表征、推理期压缩结构的方式,在兼顾重建性能的同时降低模型复杂度与运行开销,从而提升其在遥感超分任务中的工程可用性。重参数化卷积通过在训练阶段引入多分支结构增强局部表征能力,并在推理阶段等价折叠为单分支算子,从而在不增加额外运行时开销的前提下获得更强的细节建模能力;多尺度渐进融合则通过跨尺度状态传播增强特征复用,减少重复计算,并缓解信息在单一路径传播中的损失。此类策略对于大幅面遥感图像处理具有较高的工程价值,但其性能仍需在压缩比例与重建精度之间建立合理权衡;当模型规模被压缩到极限时,细节恢复能力往往首先受到影响,因此更有效的设计思路通常不是单纯裁剪参数,而是将轻量化策略与结构先验进行协同考虑。

相较于参数规模较大的 CNN 或 Transformer 方法,轻量化与重参数化类方法在模型复杂度、推理速度和部署灵活性方面更具优势,因此更适用于大幅面遥感图像处理、边缘设备部署及对实时性要求较高的应用场景。

#### 2.4.6 共性局限与适用性分析

综上所述,现有研究一方面不断拓展了 Mamba 在遥感图像超分辨率重建中的应用边界,另一方面也从侧面揭示了该架构在实际任务中的若干共性局限。Mamba 虽然在长程依赖建模和计算效率方面具有优势,但其序列扫描机制对局部高频纹理、弱边缘与精细结构的直接表征能力仍相对有限,因此在复杂地物边界、重复纹理和细粒度目标恢复中,往往仍需借助频域补偿、

局部增强或注意力交互机制加以弥补。其次,状态空间建模效果较大程度依赖于序列组织方式、扫描路径和特征交互策略,不同空间位置之间的信息可见性并不完全均衡,这使其在大尺度结构保持、跨区域一致性恢复以及多尺度特征协同方面仍存在进一步优化空间。此外,当模型被拓展到更复杂的数据条件或部署场景时,其性能还会受到模态对齐质量、退化建模准确性及复杂度约束等因素影响。由此可见,Mamba并非遥感超分问题中的统一最优解,其实际表现仍与任务类型、数据特性及网络设计方式密切相关。

从适用性角度看,上述四类改进路径虽然均以提升Mamba在遥感超分中的重建能力为目标,但在性能表现、适用场景与计算代价等方面仍呈现出较为明显的差异。频域增强类方法在纹理细节补偿方面更具优势,但对复杂退化条件的适应性相对有限;结构融合与注意力增强类方法通常具有更好的综合重建精度和结构一致性,但计算与显存开销相对较高;跨模态与多源建模类方法在辅助信息充分且配准可靠时潜力更大,但适用范围受数据条件限制;轻量化与重参数化类方法更适合部署效率敏感场景,但在极限压缩条件下可能牺牲部分细节恢复能力。因此,不同改进路径并不存在绝对优劣,而更适合根据任务目标、数据条件和算力约束进行针对性选择。

## 3 实验

### 3.1 实验数据、环境与参数配置

本实验遵循统一的遥感超分辨率评测协议<sup>[101]</sup>。训练数据选取AID数据集<sup>[102]</sup>,AID包含30类典型航拍场景,样本规模约为10 000幅,原始图像分辨率为600×600。训练集构建时每类随机抽取100幅图像,并统一裁剪为512×512,最终得到3 000幅训练样本用于网络参数学习。AID测试集从未参与训练的样本中按类别随机抽取,每类选取30幅图像,最终得到900幅512×512尺寸的测试样本,与训练集不存在重叠。为检验模型的跨域泛化能力,测试阶段除采用AID测试集进行域内评测外,还进一步引入DIOR与UCAS-AOD两个公开遥感数据集构建跨域测试集<sup>[103]</sup>。DIOR与UCAS-AOD两者均随

机选取900幅分辨率为800×800图像作为评测样本,且不参与模型训练<sup>[104]</sup>。退化模型仅采用双三次插值下采样构造×4倍率的LR输入,并以原始图像作为HR参考<sup>[105]</sup>。该设置主要用于保证不同方法在统一退化假设下的横向可比性,便于考察各模型在受控条件下的重建能力。由于真实遥感成像过程常同时受到模糊、噪声及传感器响应差异等因素影响,其退化形式通常较双三次下采样更为复杂,已有研究也开始从真实退化建模、实用退化模型和多退化重建等角度开展探索。因此,本文实验结果主要反映各方法在统一合成退化设置下的相对性能。

在此基础上,训练阶段从LR图像中随机裁剪64×64的图像块作为网络输入,其对应的HR图像块尺寸为256×256,以提升训练效率并增加样本多样性,同时引入随机旋转与水平、垂直翻转等数据增强操作以改善泛化能力<sup>[106]</sup>。训练与测试均在单张NVIDIA RTX 3090 GPU上完成。优化器采用Adam( $\beta_1=0.9, \beta_2=0.99$ )<sup>[107]</sup>,初始学习率设置为 $1 \times 10^{-4}$ ,并在250 k,400 k,450 k,475 k迭代处按比例衰减,总训练迭代次数设为500 k;batch size设置为4,损失函数采用L1损失以约束重建图像与参考图像之间的像素差异<sup>[108]</sup>。各对比方法均采用其公开实现中对应×4超分任务的标准网络配置进行重新训练,不额外统一网络深度或参数规模,而是在统一实验协议下比较其重建性能。测试阶段采用端到端推理,不引入额外后处理,从而保证重建精度与运行效率的可比性。在性能评价时,按照超分辨率领域的通用设置,将彩色图像转换到YCbCr空间并仅在亮度Y通道上计算指标,同时为避免边界效应,对重建结果与参考图像均裁剪4个边缘像素后再进行PSNR与SSIM统计<sup>[109]</sup>。此外,本文进一步引入学习感知图像块相似度(Learned Perceptual Image Patch Similarity, LPIPS)指标,用于补充评价重建结果的感知质量<sup>[6]</sup>。

### 3.2 评价指标

为全面衡量遥感超分辨率重建结果的保真度与结构一致性,本文采用PSNR与SSIM作为客观评价指标<sup>[110]</sup>。PSNR通过均方误差(Mean Squared Error, MSE)刻画重建图像与参考图像之间的像素级偏差,能够直接反映细节恢复的精

确程度<sup>[111]</sup>;SSIM则从亮度、对比度与结构信息三方面综合度量两幅图像的结构一致性,对遥感图像中边缘、纹理与几何形态的保持更为敏感<sup>[112]</sup>。因此,两者联合使用可以从像素误差和结构相似两个互补维度表征模型的重建能力,为不同方法在复杂遥感场景下的性能比较提供可靠依据<sup>[113-117]</sup>。

在计算上,首先由重建图像与参考图像的像素差得到MSE:

$$MSE = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W (I_{SR}(i,j) - I_{HR}(i,j))^2, \quad (8)$$

其中: $H$ 与 $W$ 分别为图像高度与宽度,MSE越小表示像素误差越低。基于MSE可进一步得到PSNR:

$$PSNR = 10 \log_{10} \left( \frac{MAX^2}{MSE} \right), \quad (9)$$

其中: $MAX$ 为像素最大取值,PSNR越大意味着重建结果与参考图像越接近、像素级保真度越高。对于结构一致性的度量,SSIM通常在局部窗口内计算,其表达式为:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (10)$$

其中: $x$ 与 $y$ 分别表示两幅待比较图像在局部窗口内的像素集合, $\mu_x, \mu_y$ 为均值, $\sigma_x^2, \sigma_y^2$ 为方差, $\sigma_{xy}$ 为协方差, $C_1, C_2$ 为稳定常数以避免分母为零。SSIM取值为 $[0, 1]$ ,越接近1表示结构一致性越强,说明重建图像在边缘形状与纹理模式等结构信息上更接近参考图像。

除PSNR和SSIM外,本文进一步引入LPIPS作为感知质量评价指标。LPIPS通过比较重建图像与参考图像在深层特征空间中的差异来衡量视觉感知相似性,其值越小表示重建结果在感知质量上越接近参考图像。与PSNR和SSIM侧重失真度和结构相似性的评价特点不同,LPIPS能够在一定程度上补充反映纹理细节和视觉真实感恢复效果。

### 3.3 实验结果

表1和表2汇总了一些代表性模型在AID, DIOR与UCAS\_AOD 3个遥感超分基准数据集上的定量对比结果<sup>[118-120]</sup>。整体来看,传统双三次插值作为非学习基线,在3个数据集上的平均

PSNR和SSIM分别为28.60 dB, 0.752 8,重建结果普遍存在明显平滑与细节缺失现象,难以满足遥感图像对边缘结构与纹理真实性的要求。随着深度学习方法的引入,CNN代表方法EDSR与RCAN在3个数据集平均PSNR上分别提升至30.77 dB与31.08 dB,同时SSIM分别达到0.825 8与0.831 2,说明残差学习与通道注意力对遥感场景的纹理恢复具有稳定增益。进一步,Transformer系列方法在该任务上同样具备竞争力,其中SwinIR与HAT的平均PSNR分别达到30.93 dB与31.12 dB,体现出基于窗口注意力的长程依赖建模对遥感图像多尺度结构重建的促进作用。上述结果表明,遥感超分性能的提升不仅依赖局部卷积表征能力,也逐渐受益于更强的全局建模机制<sup>[121-123]</sup>。

在以Mamba为主体的对比中,Mamba系列模型在3个数据集上呈现出更突出的整体优势。MambaIR在AID, DIOR, UCAS\_AOD上的PSNR分别达到30.90 dB, 31.67 dB与30.87 dB,3数据集平均PSNR为31.15dB,说明基于状态空间模型的序列建模方式能够在遥感图像重建中有效提升全局一致性与结构表达。进一步,MambaIRv2在3个数据集上继续取得提升,其AID数据集中的PSNR达到31.01 dB,DIOR数据集中的PSNR达到31.78 dB,UCAS\_AOD数据集中的PSNR达到31.03 dB,平均PSNR达到31.27 dB;相应SSIM也稳定提升至0.836 6。与SwinIR和HAT等Transformer代表方法相比,MambaIRv2在平均PSNR上保持领先,同时在DIOR与UCAS\_AOD上也位于最优水平梯队,表明其在跨数据集的泛化能力方面具有较强稳定性。从补充的感知质量指标LPIPS来看,MambaIRv2在AID, DIOR和UCAS\_AOD数据集上分别取得0.305 6, 0.313 3和0.273 9的最优值,在平均指标上也以0.297 6达到最佳水平;MambaIR在平均指标上也与Transformer代表方法HAT保持相近水平。说明Mamba系列在保持较优PSNR和SSIM表现的同时,在感知质量与纹理细节恢复方面同样具有较强竞争力。

从更细粒度的类别层面观察,MambaIRv2在AID多类典型场景中仍表现出稳定优势。例如在机场、裸地、棒球场、海滩、桥梁等类别中,

表 1 在 AID 数据集上对 30 个场景类别的各种 4 倍超分辨率方法定量比较

Tab. 1 Quantitative comparison of various 4× super-resolution methods on 30 scene categories in AID dataset

类别	双三次插值		EDSR		RCAN		SwinIR		HAT		MambaIR		MambaRv2	
	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑
机场	27.44	0.7517	29.42	0.8294	29.55	0.8314	29.50	0.8320	30.15	0.8319	29.63	0.8338	29.77	0.8374
裸地	35.36	0.8397	36.91	0.8756	36.94	0.8761	36.95	0.8763	36.88	0.8841	36.99	0.8766	37.02	0.8773
棒球场	30.72	0.8318	33.22	0.8862	33.45	0.8888	33.34	0.8881	33.25	0.8789	33.53	0.8902	33.61	0.8915
海滩	34.07	0.8446	35.59	0.8763	35.61	0.8767	35.64	0.8775	34.34	0.8756	35.69	0.8775	35.72	0.8785
桥梁	28.92	0.7770	31.30	0.8379	31.59	0.8409	31.50	0.8404	33.04	0.8809	31.66	0.8423	31.89	0.8459
中心区	26.22	0.7089	28.64	0.8057	28.97	0.8124	28.84	0.8113	28.92	0.7956	29.05	0.8160	29.36	0.8239
教堂	22.98	0.6028	25.05	0.7309	25.40	0.7440	25.16	0.7349	26.56	0.7532	25.49	0.7493	25.62	0.7546
商业区	27.34	0.7358	29.11	0.8172	29.32	0.8227	29.19	0.8200	29.21	0.8007	29.38	0.8256	29.47	0.8281
密集住宅区	24.00	0.6171	25.66	0.7287	25.89	0.7410	25.80	0.7345	24.67	0.6936	25.97	0.7465	26.11	0.7518
沙漠	40.87	0.9292	42.09	0.9442	41.87	0.9457	42.19	0.9459	40.37	0.9278	42.19	0.9461	42.25	0.9471
农田	33.03	0.8237	35.05	0.8743	35.18	0.8760	35.16	0.8761	35.03	0.8691	35.25	0.8776	35.33	0.8788
森林	29.73	0.6914	30.72	0.7606	30.78	0.7638	30.80	0.7647	30.01	0.7363	30.85	0.7682	30.89	0.7702
工业区	25.13	0.6812	27.42	0.7950	27.74	0.8019	27.48	0.7975	29.04	0.7980	27.80	0.8061	27.94	0.8113
草地	33.65	0.7365	34.28	0.7663	34.30	0.7671	34.27	0.7672	34.70	0.7815	34.33	0.7685	34.34	0.7699
中等密度住宅区	25.47	0.6230	27.40	0.7317	27.75	0.7438	27.52	0.7359	28.46	0.7408	27.81	0.7483	27.91	0.7510
山地	29.02	0.7362	30.28	0.7996	30.32	0.8002	30.34	0.8018	30.78	0.7923	30.34	0.8014	30.39	0.8036
公园	28.51	0.7456	30.17	0.8179	30.34	0.8222	30.24	0.8207	30.71	0.8189	30.39	0.8247	30.46	0.8273
停车场	24.10	0.7378	26.82	0.8431	28.05	0.8667	27.28	0.8529	27.56	0.8405	28.31	0.8723	28.47	0.8771
操场	31.12	0.8189	33.94	0.8798	34.28	0.8845	34.13	0.8826	35.49	0.8959	34.38	0.8866	34.50	0.8891
池塘	30.64	0.8151	32.11	0.8549	32.22	0.8564	32.17	0.8561	32.18	0.8555	32.29	0.8579	32.28	0.8581
港口	26.07	0.8030	28.45	0.8768	28.69	0.8805	28.63	0.8800	28.81	0.8638	28.82	0.8831	28.99	0.8862
火车站	27.80	0.7239	29.95	0.8211	30.35	0.8290	30.16	0.8260	28.88	0.7780	30.34	0.8305	30.60	0.8357
度假区	25.80	0.6786	27.56	0.7710	27.73	0.7752	27.67	0.7748	28.71	0.7849	27.81	0.7794	27.90	0.7829
河流	29.96	0.7493	31.40	0.8091	31.44	0.8096	31.47	0.8118	31.63	0.7909	31.50	0.8120	31.53	0.8132
学校	26.86	0.7137	28.84	0.8052	29.07	0.8098	28.97	0.8095	29.54	0.8104	29.16	0.8137	29.26	0.8166
稀疏住宅区	26.56	0.5975	27.59	0.6784	27.78	0.6854	27.65	0.6818	27.88	0.6759	27.80	0.6879	27.84	0.6895
广场	28.86	0.7699	31.37	0.8491	31.62	0.8540	31.52	0.8531	31.00	0.8251	31.71	0.8566	31.80	0.8598
体育场	27.15	0.7561	29.91	0.8455	30.16	0.8494	30.08	0.8506	29.77	0.8422	30.27	0.8537	30.51	0.8603
储油罐	26.39	0.7100	28.45	0.8009	28.60	0.8039	28.56	0.8039	27.60	0.7698	28.65	0.8070	28.76	0.8093
高架桥	27.38	0.6982	29.51	0.7967	29.73	0.8018	29.65	0.8005	29.11	0.7794	29.77	0.8051	29.89	0.8082
平均值	28.71	0.7416	30.61	0.8170	30.82	0.8220	30.73	0.8203	30.81	0.8124	30.90	0.8248	31.01	0.8278

MambaIRv2 的 PSNR 与 SSIM 均处于较高水平, 说明其对规则线性结构、强对比边缘以及大面积平滑背景与细小纹理并存的遥感场景具备较好的统一建模能力; 这类场景往往同时包含长直边

界、重复纹理与局部高频细节, 对网络的全局依赖建模与局部细节保真提出了更高要求。尤其在具有细线条和高对比特征的目标(如跑道标线、停车位线、道路边界)上, 能够保持线条连续

表2 在AID, DIOR和UCAS\_AOD测试数据集上4×超分辨率方法的量化比较

Tab. 2 Quantitative comparison of 4× super-resolution methods on AID, DIOR, and UCAS\_AOD test datasets

方 法	AID			DIOR			UCAS_AOD			平均值		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
双三次插值	28.71	0.7416	0.4830	29.17	0.7606	0.4617	27.91	0.7563	0.4433	28.60	0.7528	0.4627
EDSR	30.60	0.8170	0.3141	31.28	0.8264	0.3195	30.43	0.8339	0.2787	30.77	0.8258	0.3041
RCAN	30.82	0.8220	0.3731	31.61	0.8321	0.3187	30.81	0.8395	0.2777	31.08	0.8312	0.3034
SwinIR	30.73	0.8203	0.3130	31.44	0.8290	0.3159	30.62	0.8370	0.2756	30.93	0.8288	0.3015
HAT	30.85	0.8137	0.3121	31.62	0.8332	0.3150	30.88	0.8412	0.2750	31.12	0.8294	0.3007
MambaIR	30.90	0.8248	0.3097	31.67	0.8342	0.3163	30.87	0.8419	0.2770	31.15	0.8336	0.3010
MambaIRv2	31.01	0.8278	0.3056	31.78	0.8378	0.3133	31.03	0.8444	0.2739	31.27	0.8366	0.2976

性与几何一致性的模型往往在视觉质量上更占优势。

在定性分析上,本文选取若干典型场景片段进行主观视觉对比,方法包括双三次插值、Transformer代表方法HAT与SwinIR,以及Mamba系列MambaIR与MambaIRv2,对比结果如图14所示。图14左侧给出了选取的典型场景及其红框标注重点观察的局部区域,右侧展示了对应区域在不同方法下的放大重建结果,其中真实图像用于作为视觉参照,以判断各方法在结构恢复、边缘保持和细节重建方面与真实结果的接近程度。整体来看,双三次插值普遍存在过度平滑现象,细线条、窄边界与弱纹理容易被明显钝化;HAT与SwinIR能够提升局部清晰度,但在规则重复结构、细长边界以及局部几何一致性保持方面仍存在一定不足。相比之下,Mamba系列方法在边缘清晰度、线条连续性、重复纹理稳定性以及背景伪纹理抑制方面整体更为稳定,重建结果也更接近真实图像。

以“体育场\_213”和“N0243”为例,局部区域分别包含平行条带结构以及道路标线与斜向重复纹理,这类目标具有明显的周期性细节和较强的几何约束。由图14可见,双三次插值难以保持条带与标线之间的清晰分离,容易表现为边缘模糊、间隔不清和局部灰度过渡过宽。HAT与SwinIR虽能一定程度恢复结构轮廓,但在条带密集或方向变化较明显的区域,仍会出现线宽不一致、局部粘连或重复结构排列不够稳定的问题。相比之下,MambaIR与MambaIRv2能够更好地

维持条带间隔、边界方向和重复结构的规整性,其中MambaIRv2在边缘收敛性和局部结构可分性方面更接近真实图像。

在“操场\_270”和“00121”样例中,场景同时包含长边界、细线结构以及大面积相对平滑的背景区域,是检验边缘保持能力与背景抑制能力的典型区域。双三次插值容易造成边界过渡带变宽、细线钝化以及交汇处定位不清。HAT与SwinIR虽然增强了边界对比,但在细线利落度、长边界连续性及角点定位方面仍可见轻微模糊、线条偏粗或局部不均匀锐化。相比之下,Mamba系列能够在保持背景较为平滑的同时,更准确地恢复长边界与细线的几何关系,使边界更集中、交汇位置更清晰,整体结构也更接近真实图像。

在“N0131”和“棒球场\_80”样例中,局部区域分别表现为规则的重复几何单元以及细小的暗色线状目标,这类区域对细节辨识能力和伪纹理控制能力要求较高。可以看到,双三次插值难以恢复重复单元之间的边界分隔,细小目标也容易被模糊为不清晰的暗斑。HAT与SwinIR虽能提升局部对比度,但在重复单元的形状一致性以及细小目标的轮廓稳定性方面仍与真实图像存在差距,部分区域还会伴随轻微伪纹理增强。相比之下,MambaIR与MambaIRv2对重复结构的形状保持更稳定,对细小线状目标的边界恢复也更充分,尤其MambaIRv2在抑制背景无关纹理放大方面表现更优,目标轮廓与周围背景之间的分离更加自然。

结合主观对比结果可知,Mamba系列方法在

规则结构、重复纹理和长距离线性要素重建中具有更明显的优势。这主要是因为其基于状态空间的序列建模能够在较大空间范围内实现信息传递与上下文聚合,使具有强几何一致性的边界、条带和轮廓在重建过程中受到更稳定的全局约束,从而有效减少断裂、粘连、错位与线宽漂移等问题。同时,选择性门控机制能够对不同位置 and 不同通道的特征响应进行自适应调制,在强化关键结构信息的同时抑制无关噪声成分和非结

构化纹理响应,降低伪纹理被过度放大或错误迁移的风险。进一步,MambaIRv2在此基础上强化了全局特征交互与信息筛选能力,使跨区域语义关联的聚合更加充分,结构约束也更加稳定,从而在细节可辨性、结构连续性与自然观感之间取得了更可靠的平衡。因此在图 14 所示的多类典型遥感场景中,MambaIRv2 整体表现更优,也更能体现 Mamba 系列方法在复杂结构重建任务中的潜力。

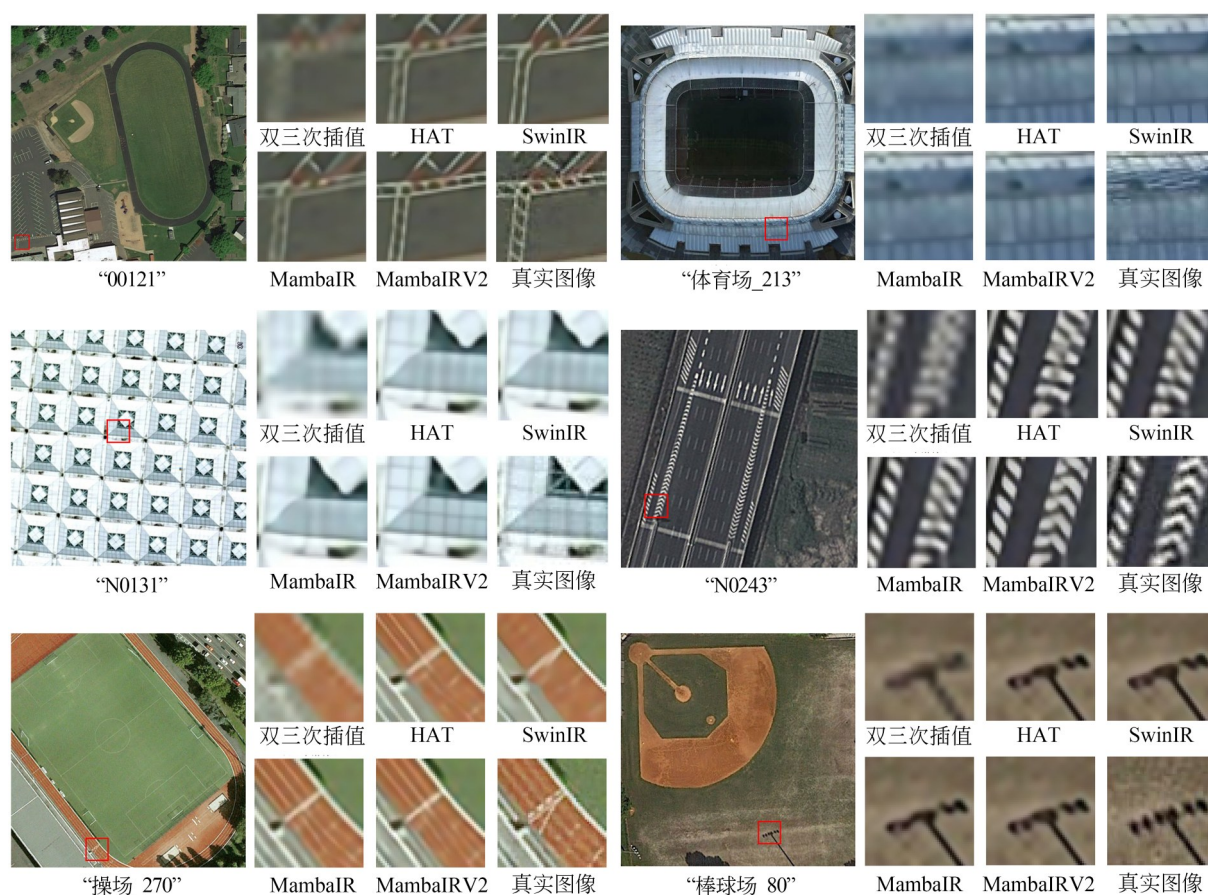


图 14 在  $4\times$  超分辨率任务下 AID、DIOR 和 UCAS\_AOD 数据集上不同方法的视觉对比,每幅原始图像下方的编号或名称表示该样本在相应数据集中的标识

Fig. 14 Visual comparison results of different methods on AID, DIOR, and UCAS\_AOD datasets for  $4\times$  super-resolution task, where the code or name below each original image denotes sample identifier in the corresponding dataset

## 4 结论及展望

### 4.1 结论

本文围绕遥感图像超分辨率中的 Mamba 重建方法进行了系统综述与实验对比,从状态空间

建模视角梳理了其通过选择性扫描与状态更新实现长程依赖聚合、兼顾全局表征与计算效率的基本特征,并将现有研究归纳为频域建模与频谱增强、结构融合与注意力增强、跨模态与多源遥感建模、轻量化与重参数化四类典型改进路径。

结合统一合成退化评测协议下的定量与主观对比结果可以看出,Mamba系列方法在多数数据集上表现出较好的稳定性与泛化一致性,在细长结构连续性保持、跨区域几何一致性恢复以及背景伪纹理抑制等方面展现出较强潜力。总体而言,Mamba通过选择性状态空间建模在近线性复杂度下实现长程依赖聚合,为遥感图像大幅面、强异质、多尺度细节并存条件下的高效高质量重建提供了新的技术路径,并进一步拓展了轻量化部署与实时应用的可行边界。

Mamba系列方法已成为遥感超分辨率重建的重要新方向,兼具全局建模能力与计算效率。相较于CNN依赖局部感受野、Transformer在高分辨率下计算与显存开销显著增加,Mamba通过选择性扫描与状态更新以更复杂度实现跨区域信息传递,有利于保持长边界结构的连续性与全局一致性,适用于道路、跑道、建筑群边缘等长程结构的恢复。面向遥感超分任务的适配改造是Mamba性能提升的核心。现有研究已形成较清晰的改进脉络:以频域建模与频谱增强补偿高频纹理,以结构融合与注意力增强缓解因果扫描的信息可见性限制,以跨模态与多源融合引入光学、SAR、红外与高光谱等互补信息提升重建质量,并通过轻量化与重参数化结构实现高效部署与可用性提升。这些方向从结构、信息建模与训练策略层面持续强化其适配性与泛化能力,推动Mamba由通用视觉SSM骨干演化为面向遥感场景的高效重建框架,并为模块复用与工程落地奠定基础。实验对比表明,Mamba系列在多数数据集上表现稳定,具备作为主干框架的可靠性与可扩展性。定量结果显示,MambaIRv2在AID、DIOR与UCAS\_AOD等基准上整体处于领先地位且平均指标占优,说明状态空间建模不仅降低了模型复杂度,也能在遥感纹理恢复中取得较高精度。主观对比亦显示其在细长结构连续保持、跨区域几何一致性与背景纹理抑制方面更具优势,这与其长程上下文聚合及选择性信息调制机制相关,可在增强细节的同时降低噪声放大或错误纹理迁移导致的伪影风险。Mamba方法仍有待完善。首先,遥感图像真实退化常含复合噪声、非均匀模糊与传感器响应差异,现有方法在真实

退化稳定性与跨域泛化方面仍有提升空间;其次,内部扫描方式与特征组织对重建质量影响显著,如何在保持线性复杂度的同时进一步增强对局部细节与多尺度结构的表达能力,仍是后续研究的重要议题。

## 4.2 展望

面向遥感图像场景的复杂性与工程部署需求,Mamba超分辨率重建方法未来仍具有较大的拓展空间。后续研究可围绕遥感特性建模、真实退化稳定性、多源融合稳定性、轻量化部署与评价体系完善等方面持续推进,以形成更具通用性与实用价值的技术体系。

面向遥感图像特性构建谱到空到尺度协同的状态空间建模机制。遥感图像多尺度结构突出、场景异质性强且多谱段响应差异明显。未来可将光谱一致性、辐射保真约束与几何先验融入状态更新过程,并在门控调制、序列组织与跨尺度交互层面强化适配性,以提升复杂纹理与边缘几何的重建可靠性。

加强真实退化建模与盲超分以提升跨域泛化。遥感成像链路中的复合噪声、压缩伪影、非均匀模糊与配准误差会削弱重建稳定性。未来可从退化估计、退化条件化恢复与自监督学习出发,构建面向真实退化的Mamba超分框架,将退化信息作为提示或调制信号引入状态更新,提高跨区域、跨传感器与跨时相数据的稳定性与迁移能力。

推动多源跨模态融合由可融合走向可控融合。光学到SAR、可见光到红外、HSI到MSI等组合可提供互补细节,但也带来噪声差异与配准偏差的不确定性。未来可研究基于选择性机制的可信度建模与动态融合,使模型按区域、按尺度自适应分配引导权重,并在频域与空间域联合建模中保持结构一致性,减少引导失配导致的伪影与纹理错配。

以轻量化、低延迟与可部署为目标推进算法和系统协同优化。虽有近线性复杂度优势,但在超高分辨率输入与多分支结构下仍受算力、显存与带宽限制。后续可围绕重参数化、扫描并行、量化剪枝蒸馏一体化训练等开展系统研究,形成面向端侧、机载与星载平台的高效推理方案,提

升业务链条可用性。

完善面向遥感应用的基准数据集与评价体系。当前评价仍主要依赖 PSNR, SSIM 及 LPIPS 等指标,难以全面刻画光谱一致性、辐射保真与结构可用性。未来需构建覆盖多场景、多退化、多传感器的数据集与统一评测协议,引入结构一致性、光谱一致性与任务驱动指标,建立重建质量到应用收益的闭环评价框架,为 Mamba

方法迭代提供更准确的目标牵引。

#### 作者贡献声明:

李秉豪:论文构思和撰写;  
姜肖楠:论文中心的提出;  
傅 瑶:论文审核;  
王亚楠、万龙腾:文献调研;  
吴凡路:论文架构构思及搭建。

#### 参考文献:

- [1] LIU B D, ZHAO L F, SHAO S, *et al.* RAN: region-aware network for remote sensing image super-resolution [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 5408113.
- [2] MITTAL A, SOUNDARARAJAN R, BOVIK A C. Making a “completely blind” image quality analyzer[J]. *IEEE Signal Processing Letters*, 2013, 20(3): 209-212.
- [3] SHEIKH H R, BOVIK A C. Image information and visual quality[J]. *IEEE Transactions on Image Processing*, 2006, 15(2): 430-444.
- [4] ZHANG L, ZHANG L, MOU X Q, *et al.* FSIM: a feature similarity index for image quality assessment[J]. *IEEE Transactions on Image Processing*, 2011, 20(8): 2378-2386.
- [5] ZHANG J Q, LEI J, XIE W Y, *et al.* SuperY-OLO: super resolution assisted object detection in multimodal remote sensing imagery [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 5605415.
- [6] ZHANG R, ISOLA P, EFROS A A, *et al.* The unreasonable effectiveness of deep features as a perceptual metric[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 586-595.
- [7] PRASHNANI E, CAI H, MOSTOFI Y, *et al.* PieAPP: perceptual image-error assessment through pairwise preference[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 1808-1817.
- [8] WANG J, XIANG L Y, LIU L, *et al.* Toward real-world remote sensing image super-resolution: a new benchmark and an efficient model [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 1-13.
- [9] 孙巍, 王乾宙, 陈雪凌, 等. 真实场景下图像超分辨率技术现状与趋势[J]. *中国图象图形学报*, 2025, 30(6): 1576-1592.
- SUN W, WANG Q ZH, CHEN X L, *et al.* Development of real-world image super-resolution [J]. *Journal of Image and Graphics*, 2025, 30(6): 1576-1592. (in Chinese)
- [10] DING K Y, MA K D, WANG S Q, *et al.* Image quality assessment: unifying structure and texture similarity[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(5): 2567-2581.
- [11] WANG Y, BASHIR S M A, KHAN M, *et al.* Remote sensing image super-resolution and object detection: Benchmark and state of the art[J]. *Expert Systems with Applications*, 2022, 197: 116793.
- [12] 赵晓, 杜光磊. 基于条件先验增强和扩散模型的遥感图像超分辨重建算法[J]. *液晶与显示*, 2025, 40(7): 1067-1079.
- ZHAO X, DU G L. Super-resolution of remote sensing images based on conditional prior enhancement and diffusion models[J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(7): 1067-1079. (in Chinese)
- [13] DONG C, LOY C C, HE K M, *et al.* Learning a deep convolutional network for image super-resolution [C]. *Computer Vision-ECCV 2014*. Cham: Springer, 2014: 184-199.
- [14] LIANG J Y, CAO J Z, SUN G L, *et al.* Swin-IR: image restoration using swin transformer[C].

- 2021 *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. October 11-17, 2021, Montreal, BC, Canada. IEEE, 2021: 1833-1844.
- [15] CHEN X Y, WANG X T, ZHOU J T, *et al.* Activating more pixels in image super-resolution transformer[C]. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 17-24, 2023, Vancouver, BC, Canada. IEEE, 2023: 22367-22377.
- [16] GU A, DAO T. Mamba: linear-time sequence modeling with selective state spaces [EB/OL]. 2023: *arXiv*: 2312.00752. <https://arxiv.org/abs/2312.00752>.
- [17] XIAO Y, YUAN Q Q, JIANG K, *et al.* Frequency-assisted mamba for remote sensing image super-resolution[J]. *IEEE Transactions on Multimedia*, 2025, 27: 1783-1796.
- [18] QI Y L, LOU M, LIU Y M, *et al.* Advancing image super-resolution techniques in remote sensing: a comprehensive survey [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2026, 231: 68-100.
- [19] VU-DUC A, NGUYEN-VI K, BUI-QUOC B, *et al.* A comprehensive survey of super-resolution remote sensing image datasets: evolution, challenges, and future directions [J]. *IEEE Access*, 2025, 13: 145350-145372.
- [20] BAO M Y, LYU S C, XU Z Y, *et al.* Vision mamba in remote sensing: a comprehensive survey of techniques, applications and outlook [J]. *Remote Sensing*, 2026, 18(4): 594.
- [21] 周冬梅, 贾琴琴. 融合Mamba的多尺度遥感图像实例分割网络[J]. *光学精密工程*, 2025, 33(24): 3940-3955.
- ZHOU D M, JIA Q Q. Multi-scale remote sensing image instance segmentation network with mamba integration[J]. *Optics and Precision Engineering*, 2025, 33(24): 3940-3955. (in Chinese)
- [22] CARRATO S, RAMPONI G, MARSIS. A simple edge-sensitive image interpolation filter [C]. *Proceedings of 3rd IEEE International Conference on Image Processing*. September 19-19, 1996, Lausanne, Switzerland. IEEE, 2002: 711-714.
- [23] LI X, ORCHARD M T. New edge-directed interpolation[J]. *IEEE Transactions on Image Processing*, 2001, 10(10): 1521-1527.
- [24] 刘宇豪, 吴福培, 吴树壮, 等. 基于插值超分辨率的双目三维重建方法[J]. *中国光学(中英文)*, 2024, 17(4): 862-874.
- LIU Y H, WU F P, WU SH ZH, *et al.* Binocular 3D reconstruction method based on interpolation super-resolution[J]. *Chinese Optics*, 2024, 17(4): 862-874. (in Chinese)
- [25] YANG J C, WRIGHT J, HUANG T S, *et al.* Image super-resolution via sparse representation [J]. *IEEE Transactions on Image Processing*, 2010, 19(11): 2861-2873.
- [26] ZHENG Z H, WANG B, SUN K. Single remote sensing image super-resolution and denoising via sparse representation [C]. *2011 International Workshop on Multi-Platform/Multi-Sensor Remote Sensing and Mapping*. January 10-12, 2011, Xiamen, China. IEEE, 2011: 1-5.
- [27] ZHANG Y Y, WU W, DAI Y, *et al.* Remote sensing images super-resolution based on sparse dictionaries and residual dictionaries [C]. *2013 IEEE 11th International Conference on Dependable, Autonomic and Secure Computing*. December 21-22, 2013, Chengdu, China. IEEE, 2014: 318-323.
- [28] GOU S P, LIU S Z, YANG S Y, *et al.* Remote sensing image super-resolution reconstruction based on nonlocal pairwise dictionaries and double regularization[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, 7(12): 4784-4792.
- [29] SHAO Z F, WANG L, WANG Z Y, *et al.* Remote sensing image super-resolution using sparse representation and coupled sparse autoencoder[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019, 12(8): 2663-2674.
- [30] DEKA B, MULLAH H U, BARMAN T, *et al.* Joint sparse representation-based single image super-resolution for remote sensing applications[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023, 16: 2352-2365.
- [31] WANG X L, LIU N F. Super-resolution of re-

- remote sensing images *via* sparse structural manifold embedding [J]. *Neurocomputing*, 2016, 173: 1402-1411.
- [32] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks [C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 1646-1654.
- [33] LEDIG C, THEIS L, HUSZÁR F, *et al.* Photo-realistic single image super-resolution using a generative adversarial network [C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 105-114.
- [34] 卜丽静, 陈香雪, 张正鹏, 等. 基于两阶段残差条件扩散网络的遥感图像超分辨重建 [J]. *液晶与显示*, 2025, 40(11): 1647-1660.
- BU L J, CHEN X X, ZHANG ZH P, *et al.* Two-stage residual conditional diffusion network for super-resolution reconstruction of remote sensing images [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(11): 1647-1660. (in Chinese)
- [35] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need [EB/OL]. 2017: *arXiv*: 1706.03762. <https://arxiv.org/abs/1706.03762>.
- [36] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al.* An image is worth 16x16 words: transformers for image recognition at scale [EB/OL]. 2020: *arXiv*: 2010.11929. <https://arxiv.org/abs/2010.11929>.
- [37] LU Y T, MIN L T, WANG B L, *et al.* Cross-spatial pixel integration and cross-stage feature fusion-based transformer network for remote sensing image super-resolution [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 5625616.
- [38] MAO Y J, HE G J, WANG G Z, *et al.* DE-SAT: a distance-enhanced strip attention transformer for remote sensing image super-resolution [J]. *Remote Sensing*, 2024, 16(22): 4251.
- [39] XIAO Y, YUAN Q Q, JIANG K, *et al.* TTST: a top-k token selective transformer for remote sensing image super-resolution [J]. *IEEE Transactions on Image Processing*, 2024, 33: 738-752.
- [40] CHEN H T, WANG Y H, GUO T Y, *et al.* Pre-trained image processing transformer [C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 20-25, 2021, Nashville, TN, USA. IEEE, 2021: 12294-12305.
- [41] YANG F Z, YANG H, FU J L, *et al.* Learning texture transformer network for image super-resolution [C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 13-19, 2020, Seattle, WA, USA. IEEE, 2020: 5790-5799.
- [42] CAI Y H, LIN J, LIN Z D, *et al.* MST++: multi-stage spectral-wise transformer for efficient spectral reconstruction [C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. June 19-20, 2022, New Orleans, LA, USA. IEEE, 2022: 744-754.
- [43] DENG S Q, DENG L J, WU X, *et al.* PSRT: pyramid shuffle-and-reshuffle transformer for multi-spectral and hyperspectral image fusion [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 5503715.
- [44] CHEN S, ZHANG L F, ZHANG L P. MSD-former: multiscale deformable transformer for hyperspectral image super-resolution [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 5525614.
- [45] SUN W W, WANG Y, LIU W W, *et al.* STANet: a hybrid spectral and texture attention pyramid network for spectral super-resolution of remote sensing images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5525915.
- [46] CHEN Z, ZHANG Y L, GU J J, *et al.* Dual aggregation transformer for image super-resolution [C]. 2023 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 1-6, 2023, Paris, France. IEEE, 2024: 12278-12287.
- [47] WANG Z D, CUN X D, BAO J M, *et al.* Uformer: a general U-shaped transformer for image restoration [C]. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 18-24, 2022, New Orleans, LA, USA. IEEE, 2022: 17662-17672.

- [48] ZAMIR S W, ARORA A, KHAN S, *et al.* Restormer: efficient transformer for high-resolution image restoration [C]. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 18-24, 2022, New Orleans, LA, USA. IEEE, 2022: 5718-5729.
- [49] 王军, 陈莹莹, 程勇. 基于多重蒸馏与Transformer的遥感图像超分辨率重建[J]. *计算机系统应用*, 2025, 34(2): 225-236.
- WANG J, CHEN Y Y, CHENG Y. Remote sensing image super-resolution reconstruction based on multi-distillation and transformer [J]. *Computer Systems and Applications*, 2025, 34(2): 225-236. (in Chinese)
- [50] LIU Z, LIN Y T, CAO Y, *et al.* Swin transformer: hierarchical vision transformer using shifted windows[C]. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. October 10-17, 2021, Montreal, QC, Canada. IEEE, 2022: 9992-10002.
- [51] ZHANG W J, TAN Z, LV Q B, *et al.* An efficient hybrid CNN-transformer approach for remote sensing super-resolution [J]. *Remote Sensing*, 2024, 16(5): 880.
- [52] HU T, CHEN Z J, WANG M Y, *et al.* Global sparse attention network for remote sensing image super-resolution [J]. *Knowledge-Based Systems*, 2024, 304: 112448.
- [53] HAO J L, LI W K, LU Y T, *et al.* Scale-aware backprojection transformer for single remote sensing image super-resolution[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5649013.
- [54] GU A, GOEL K, RÉ C. Efficiently modeling long sequences with structured state spaces [EB/OL]. 2021: *arXiv*: 2111.00396. <https://arxiv.org/abs/2111.00396>.
- [55] ZHU L H, LIAO B C, ZHANG Q, *et al.* Vision mamba: efficient visual representation learning with bidirectional state space model [EB/OL]. 2024: *arXiv*: 2401.09417. <https://arxiv.org/abs/2401.09417>.
- [56] JIAO J B, LIU Y, LIU Y F, *et al.* VMamba: visual state space model[C]. *Advances in Neural Information Processing Systems 37*. December 10-15, 2024. Vancouver, BC, Canada. Neural Information Processing Systems Foundation, Inc. (NeurIPS), 2024: 103031-103063.
- [57] HUANG T, PEI X H, YOU S, *et al.* LocalMamba: visual state space model with Windowed selective scan [C]. *Computer Vision-ECCV 2024 Workshops*. Cham: Springer, 2025: 12-22.
- [58] GUO H, LI J M, DAI T, *et al.* MambaIR: a simple baseline for Image restoration with State-space model [C]. *Computer Vision-ECCV 2024*. Cham: Springer, 2025: 222-241.
- [59] LI Z Y, GAO J Y, YUAN Y. Dual-stage prior-driven diffusion model for remote sensing spectral super-resolution[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 5525414.
- [60] CHEN W B, JIANG S K, WANG F H, *et al.* Infrared remote-sensing image super-resolution based on physical characteristic deduction[J]. *Results in Physics*, 2024, 64: 107897.
- [61] REN K, SUN W W, MENG X C, *et al.* A locally optimized model for hyperspectral and multispectral images fusion[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5519015.
- [62] SUN W W, REN K, MENG X C, *et al.* MLR-DBPFN: a multi-scale low rank deep back projection fusion network for anti-noise hyperspectral and multispectral image fusion[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5522914.
- [63] LI Y S, ZHOU L, XU F, *et al.* OGSRN: Optical-guided super-resolution network for SAR image [J]. *Chinese Journal of Aeronautics*, 2022, 35(5): 204-219.
- [64] LIU J J, WU Z B, XIAO L. A spectral diffusion prior for unsupervised hyperspectral image super-resolution [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5528613.
- [65] NGUYEN H V, ULFARSSON M O, SVEINSSON J R, *et al.* Deep SURE for unsupervised remote sensing image fusion[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5412613.
- [66] WANG Z, ZHAO Y W, CHEN J C. Multi-scale fast Fourier transform based attention network for remote-sensing image super-resolution [J]. *IEEE*

- Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023, 16: 2728-2740.
- [67] ALI A M, BENJDIRA B, KOUBAA A, *et al.* TESR: two-stage approach for enhancement and super-resolution of remote sensing images[J]. *Remote Sensing*, 2023, 15(9): 2346.
- [68] WANG J R, LU Y T, WANG S Z, *et al.* Two-stage spatial-frequency joint learning for large-factor remote sensing image super-resolution [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-13.
- [69] XIAO Y, YUAN Q Q, JIANG K, *et al.* EDiffSR: an efficient diffusion probabilistic model for remote sensing image super-resolution [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-14.
- [70] ZHANG Z H, HU Z S, CAO B H, *et al.* Wiener filter-based mamba for remote sensing image super-resolution with novel degradation[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025, 18: 26295-26308.
- [71] CHEN W Y, LUO L G, QU S M, *et al.* MFEM: multiscale frequency-enhanced mamba for lightweight remote sensing image super-resolution [J]. *IEEE Geoscience and Remote Sensing Letters*, 2025, 22: 1-5.
- [72] CHEN W K, GAO F, GAN Y H, *et al.* Wavelet-assisted mamba for satellite-derived sea surface temperature super-resolution [EB/OL]. 2025: *arXiv*: 2509.24334. <https://arxiv.org/abs/2509.24334>.
- [73] GUO H, GUO Y, ZHA Y H, *et al.* MambaIRv2: attentive state space restoration[C]. 2025 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 10-17, 2025. Nashville, TN, USA. IEEE, 2025: 28124-28133.
- [74] LI M Y, XIONG C Y, GAO Z R, *et al.* HAM: hierarchical attention mamba with spatial-frequency fusion for remote sensing image super-resolution [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 1-14.
- [75] ZHI R C, FAN X P, SHI J Y. MambaFormerSR: a lightweight model for remote-sensing image super-resolution[J]. *IEEE Geoscience and Remote Sensing Letters*, 2024, 21: 1-5.
- [76] ZHANG Z Q, SUN Q, YE Z W, *et al.* Adaptive multi-scale feature aggregation transformer network for single remote sensing image super-resolution [J]. *Geo-spatial Information Science*, 2025: 1-22.
- [77] SHANG J R, GAO M L, LI Q L, *et al.* Hybrid-scale hierarchical transformer for remote sensing image super-resolution [J]. *Remote Sensing*, 2023, 15(13): 3442.
- [78] XIAO Y, YUAN Q Q, HE J, *et al.* Remote sensing image super-resolution via cross-scale hierarchical transformer [J]. *Geo-spatial Information Science*, 2024, 27(6): 1914-1930.
- [79] LIN M Y, MO Z Q, ZHANG H T, *et al.* Spectral reconstruction via dual cross-scanning and cross-attention mechanisms[C]. 2024 *14th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. December 9-11, 2024. Helsinki, Finland. IEEE, 2024: 1-5.
- [80] ZHAO Z C, GAO Q, YAN J Q, *et al.* HSFMBamba: hierarchical selective fusion mamba network for optics-guided joint super-resolution and denoising of noise-corrupted SAR images[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025, 18: 16445-16461.
- [81] XU Y, DONG W Q, QU J H, *et al.* CTMEMdiff: a cross-modal texture matching and enhancement mamba diffusion model for unpaired hyperspectral image super-resolution[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 1-13.
- [82] 张兵, 高连如, 李嘉鑫, 等. 高/多光谱遥感图像超分辨率融合研究进展与展望[J]. *测绘学报*, 2023, 52(7): 1074-1089.
- ZHANG B, GAO L R, LI J X, *et al.* Advances and prospects in hyperspectral and multispectral remote sensing image super-resolution fusion[J]. *Acta Geodaetica et Cartographica Sinica*, 2023, 52(7): 1074-1089. (in Chinese)
- [83] WU Y J, DAI J G, MA Z, *et al.* A diffusion model for hyperspectral and multispectral fusion guided by prior knowledge [J]. *International Journal of*

- Applied Earth Observation and Geoinformation*, 2025, 144: 104923.
- [84] ZHANG M L, ZHENG G Z, JIANG Z B, *et al.* Local-aware coupled network for hyperspectral image super-resolution[J]. *GIScience & Remote Sensing*, 2023, 60(1): 2233725.
- [85] CHEN B W, LIU L Q, LIU C Y, *et al.* Spectral-cascaded diffusion model for remote sensing image spectral super-resolution [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-14.
- [86] SUI J L, WU Q Q, PUN M O. Denoising diffusion probabilistic model with adversarial learning for remote sensing super-resolution [J]. *Remote Sensing*, 2024, 16(7): 1219.
- [87] ZHU C, LIU Y, HUANG S, *et al.* Taming a diffusion model to revitalize remote sensing image super-resolution[J]. *Remote Sensing*, 2025, 17(8): 1348.
- [88] XIAO Y, YUAN Q Q, JIANG K, *et al.* From degrade to upgrade: Learning a self-supervised degradation guided adaptive network for blind remote sensing image super-resolution [J]. *Information Fusion*, 2023, 96: 297-311.
- [89] SI P J, JIA M, WANG H, *et al.* DC-mamba: a degradation-aware cross-modality framework for blind super-resolution of thermal UAV images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 5005815.
- [90] AN T, XUE B, HUO C L, *et al.* Efficient remote sensing image super-resolution via lightweight diffusion models[J]. *IEEE Geoscience and Remote Sensing Letters*, 2024, 21: 6000105.
- [91] HU T, CHEN Z J, ZENG Q Q, *et al.* Multi-phase Transformer for remote sensing image super-resolution[J]. *Expert Systems with Applications*, 2026, 295: 128853.
- [92] LIU B D, ZHAO L F, LIU W F, *et al.* MWLN: multilevel wavelet learning network for continuous-scale remote-sensing image super-resolution [J]. *IEEE Geoscience and Remote Sensing Letters*, 2024, 21: 5000505.
- [93] 易见兵, 陈俊宽, 曹锋, 等. 轻量级重参数化的遥感图像超分辨率重建网络设计[J]. *光学精密工程*, 2024, 32(2): 268-285.
- YI J B, CHEN J K, CAO F, *et al.* Design of lightweight re-parameterized remote sensing image super-resolution network[J]. *Optics and Precision Engineering*, 2024, 32(2): 268-285. (in Chinese)
- [94] 阎刚, 宋子怡, 耿树泽. 基于全方位状态空间模型的轻量化图像超分辨率重建[J]. *液晶与显示*, 2025, 40(4): 642-654.
- YAN G, SONG Z Y, GENG SH Z. PMambaIR: panoramic vision state space model for lightweight image super-resolution[J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(4): 642-654. (in Chinese)
- [95] JIANG K, YANG M R, XIAO Y, *et al.* Rep-mamba: re-parameterization in vision mamba for lightweight remote sensing image super-resolution [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 5637012.
- [96] LIU Z L, CHEN H, BAI L, *et al.* MambaDS: near-surface meteorological field downscaling with topography constrained selective state-space modeling[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 4112615.
- [97] ZHANG M M, LI J L, JING H Y, *et al.* LT-MHN: learnable TransMamba hybrid network for remote sensing image superresolution [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025, 18: 23548-23562.
- [98] CHENG H R, ZHANG M, SHI J Q. A survey on deep neural network pruning: taxonomy, comparison, analysis, and recommendations[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(12): 10558-10578.
- [99] CHENG Y, WANG D, ZHOU P, *et al.* Model compression and acceleration for deep neural networks: the principles, progress, and challenges [J]. *IEEE Signal Processing Magazine*, 2018, 35(1): 126-136.
- [100] SELVARAJU R R, COGSWELL M, DAS A, *et al.* Grad-CAM: visual explanations from deep networks via gradient-based localization[J]. *International Journal of Computer Vision*, 2020, 128(2): 336-359.
- [101] WANG P J, BAYRAM B, SERTEL E. A comprehensive review on deep learning based remote

- sensing image super-resolution methods [J]. *Earth-Science Reviews*, 2022, 232: 104110.
- [102] XIA G S, HU J W, HU F, *et al.* AID: a benchmark data set for performance evaluation of aerial scene classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55 (7) : 3965-3981.
- [103] LI K, WAN G, CHENG G, *et al.* Object detection in optical remote sensing images: a survey and a new benchmark[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, 159: 296-307.
- [104] ZHU H G, CHEN X G, DAI W Q, *et al.* Orientation robust object detection in aerial images using deep convolutional neural network[C]. 2015 *IEEE International Conference on Image Processing (ICIP)*. September 27-30, 2015, Quebec City, QC, Canada. IEEE, 2015: 3735-3739.
- [105] KEYS R. Cubic convolution interpolation for digital image processing [J]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1981, 29(6): 1153-1160.
- [106] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. *Journal of Big Data*, 2019, 6(1): 60.
- [107] KINGMA D P, BA J. Adam: a method for stochastic optimization [EB/OL]. 2014: *arXiv*: 1412.6980. <https://arxiv.org/abs/1412.6980>.
- [108] SU J W, XU B Y, YIN H J. A survey of deep learning approaches to image restoration[J]. *Neurocomputing*, 2022, 487: 46-65.
- [109] LEE I S, NOH J H, OH S J. A Survey and analysis on a troposcatter propagation model based on ITU-R recommendations [J]. *ICT Express*, 2023, 9(3): 507-516.
- [110] TEKALP A M. Deep learning for image/video restoration and super-resolution[J]. *Foundations and Trends® in Computer Graphics and Vision*, 2022, 13(1): 1-110.
- [111] HUYNH-THU Q, GHANBARI M. Scope of validity of PSNR in image/video quality assessment [J]. *Electronics Letters*, 2008, 44 (13) : 800-801.
- [112] WANG Z, BOVIK A C, SHEIKH H R, *et al.* Image quality assessment: from error visibility to structural similarity [J]. *IEEE Transactions on Image Processing*, 2004, 13(4): 600-612.
- [113] ARABBOEV M, BEGMATOV S, RIKHSI-VOEV M, *et al.* Comprehensive review of image super-resolution metrics: classical and AI-based approaches [J]. *Acta IMEKO*, 2024, 13 (1) : 1-8.
- [114] HORÉ A, ZIOU D. Is there a relationship between peak-signal-to-noise ratio and structural similarity index measure?[J]. *IET Image Processing*, 2013, 7(1): 12-24.
- [115] DOHMEN M, KLEMENS M A, BALTRUSCHAT I M, *et al.* Similarity and quality metrics for MR image-to-image translation [J]. *Scientific Reports*, 2025, 15: 3853.
- [116] AL-MEKHLAFI H, LIU S G. Single image super-resolution: a comprehensive review and recent insight [J]. *Frontiers of Computer Science*, 2023, 18(1): 181702.
- [117] AFNAN, ULLAH F, YASEEN, *et al.* Subjective assessment of objective image quality metrics range guaranteeing visually lossless compression [J]. *Sensors*, 2023, 23(3): 1297.
- [118] LIM B, SON S, KIM H, *et al.* Enhanced deep residual networks for single image super-resolution [C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. July 21-26, 2017. Honolulu, HI, USA. IEEE, 2017: 1132-1140.
- [119] ZHANG Y L, TIAN Y P, KONG Y, *et al.* Residual dense network for image super-resolution [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 2472-2481.
- [120] ZHANG Y L, LI K P, LI K, *et al.* Image super-resolution using very deep residual channel attention networks [C]. *Computer Vision-ECCV* 2018. Cham: Springer, 2018: 294-310.
- [121] SUI J L, MA X P, ZHANG X K, *et al.* GCRDN: global context-driven residual dense network for remote sensing image superresolution [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023, 16: 4457-4468.
- [122] LIU T T, PU X T, SHI Y, *et al.* Hyperspectral

image super-resolution based on Mamba and bidirectional feature fusion network[J]. *Expert Systems with Applications*, 2025, 285: 127905.

[123] CHEN B W, CHEN K Y, YANG M H, *et al.*

SeG-SR: integrating semantic knowledge into remote sensing image super-resolution *via* vision-language model[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 5643915.

#### 作者简介:



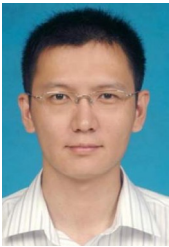
李秉豪(2002—),男,山东济南人,硕士研究生,2024年于山东理工大学获得学士学位,主要从事遥感图像处理与计算机视觉的研究。E-mail:libinghao24@mails.ucas.ac.cn

#### 通讯作者:



吴凡路(1988—),男,安徽宿州人,博士,副研究员,2011年于长春理工大学获得学士学位,2014年于中国科学院大学获得硕士学位,2018年于天津大学获得博士学位,主要从事空间光学遥感、遥感图像智能解译、深空探测具身智能、新体制成像等方面的研究。E-mail:flwu@ciomp.ac.cn

#### 通讯作者:



姜肖楠(1981—),男,吉林省吉林市人,博士,研究员,2004年、2010年于哈尔滨工业大学分别获得学士和博士学位,主要从事遥感图像智能处理和空间遥感载荷总体论证的研究。E-mail:jiangxn@ciomp.ac.cn