

基于跨尺度Transformer全局-局部交互的遥感图像 连续超分辨率

熊承义^{ab}, 王薇^{ab*}, 高志荣^c

(中南民族大学 a. 电子信息工程学院; b. 智能无线通信湖北省重点实验室; c. 计算机学院, 武汉 430074)

摘要 遥感图像连续超分辨率技术对多尺度地物识别、变化检测与语义分析等任务至关重要。然而, 现有方法在复杂背景干扰和大尺度跨度场景下, 难以兼顾局部细节重建与全局语义一致性。为此, 提出了一种跨尺度Transformer全局-局部交互的遥感图像连续超分辨率方法。设计了结合上下文注意力机制(CAM)的多尺度参数生成器, 使之选择性地增强不同尺度下的局部高频特征; 构建了一种跨尺度Transformer交互模块, 利用自注意力机制实现全局语义建模与局部特征融合; 提出双分支全局-局部解析器, 联合优化坐标相关的位置编码与上下文依赖的语义解码, 提升不同缩放倍数下的重建精度。实验结果表明: 所提出方法相比先进的连续超分辨率方法可获得近0.17dB的PSNR增益。

关键词 连续超分辨率; 遥感图像; 自注意力机制; Transformer网络; 全局-局部交互

中图分类号 TP391.4 文献标志码 A 文章编号 1672-4321(2026)03-0353-11

doi: 10.20056/j.cnki.ZNMDZK.20250841

Continuous super-resolution for remote sensing image via cross-scale Transformer with global-local interaction

XIONG Chengyi^{ab}, WANG Wei^{ab*}, GAO Zhirong^c

(South-Central Minzu University, a. College of Electronic and Information Engineering; b. Hubei Key Lab of Intelligent Wireless Communication; c. College of Computer Science, Wuhan 430074, China)

Abstract Continuous super-resolution technology for remote sensing images is critical for tasks such as multi-scale ground object recognition, change detection and semantic analysis. However, existing methods struggle to balance local detail reconstruction with global semantic consistency under complex background interference and large-scale variations. To address this problem, a continuous super-resolution method for remote sensing image via cross-scale Transformer with global-local interaction is proposed. A multi-scale parameter generator integrated with Contextual Attention Mechanism (CAM) is designed, which selectively enhances local high-frequency features at different scales, builds a cross-scale Transformer interaction module that leverages self-attention mechanism to achieve global semantic modeling and local feature fusion. A dual-branch global-local parser that jointly optimizes coordinate-aware positional encoding and context-dependent semantic decoding to ensure reconstruction accuracy at different scaling factors is proposed. Experimental results demonstrate that the proposed method can achieve a gain of 0.17 dB in PSNR compared to state-of-the-art continuous super-resolution approaches.

Keywords continuous super-resolution; remote sensing image; self-attention mechanism; Transformer network; global-local interaction

收稿日期 2025-05-15 * 通信作者 王薇, 研究方向: 图像处理, E-mail: 2674775692@qq.com

作者简介 熊承义(1969-), 男, 教授, 研究方向: 图像处理与模式识别, E-mail: xiongcy@mail.scuec.edu.cn

基金项目 多谱信息处理技术国家重点实验室基金资助项目(6142113210303); 中央高校基本科研业务费专项资金资助项目(CZZ25010)

图像超分辨率(Super-Resolution, SR)是计算机视觉领域中的一项重要任务,其目标是从低分辨率(Low-Resolution, LR)图像中重建对应的高分辨率(High-Resolution, HR)图像,广泛应用于医学成像^[1]、视频监控^[2]、遥感图像分析^[3-4]等多个领域.特别是在遥感领域,高分辨率图像包含丰富的纹理细节与空间结构信息,对于地物识别、变化检测等任务具有重要价值^[5].然而,由于遥感图像的获取通常受限于卫星传感器分辨率、传输带宽等多重因素,高质量遥感图像的获取成本仍然较高.因此,利用超分辨率技术对遥感图像进行重建,是提升图像感知能力、增强下游任务性能的关键手段.

近年来,深度学习方法显著推动了SR技术的发展.与早期的基于边缘检测、稀疏编码等方法相比,CNN、GAN以及Transformer等网络架构在重建精度与表达能力方面均取得了显著提升.早期方法如SRCNN^[6]采用预先采样策略,先通过插值扩大图像再进行处理,存在效率低下的问题.后来的后上采样方法,如反卷积^[7]和亚像素重排^[8],虽然提升了效率,但通常依赖于具体的放大倍数,难以实现对连续尺度的统一建模.在遥感图像应用中,具备多尺度图像重建能力尤为重要.一方面,不同遥感场景和任务对图像分辨率存在差异化需求;另一方面,图像在不同尺度下呈现的空间结构和层次信息也有显著差异.然而,传统的超分辨率模型大多为固定尺度设计,仅支持预定义的放大因子.这种模式不仅限制了模型的适用范围,而且在面对有多分辨率需求的实际应用中,需要分别训练多个模型,带来额外的计算与存储开销,难以满足实际部署的灵活性要求.为解决这一问题,研究者开始探索任意放大倍数的连续超分方法.Meta-SR^[9]首次将尺度因子作为模型输入,实现了对非整数倍图像的连续放大,但在实际应用中仍受到网络结构的限制,出现尺度选择策略不灵活等问题.

随着隐式神经表示(Implicit Neural Representation, INR)的兴起,连续超分辨率技术进入了新的阶段.INR将图像视作连续函数,使用多层感知器(Multi-Layer Perceptron, MLP)将空间坐标映射到像素值,结合深度特征编码实现坐标到图像的函数拟合.形式上,设 I 为连续图像, x 为图像中的任意二维坐标,结合由深度网络提取的潜在特征向量 Z ,图像像素值可通过一个隐式函数进行预测,其形式定义为:

$$I(x) = f(Z, x), \quad (1)$$

随后,局部隐式图像函数(Local Implicit Image

Function, LIIF^[10])采用了基于隐式神经表示的方法,把LR图像编码为特征,再将任意查询坐标输入多层感知器,生成对应的RGB值,有效摆脱了固定上采样结构的限制.后续工作如UltraSR^[11]、IPE^[12]和LTE^[13]等进一步引入坐标编码、频域信息和局部纹理估计器,显著提升了连续超分任务中的重建精度与表达能力.此外,注意力机制被引入到连续超分辨率技术中,提升了网络的上下文建模能力.CHEN等^[14]提出了一个级联局部隐式变换器,将注意力机制和频率编码技术集成到局部隐式图像函数中,在任意尺度超分任务中取得了良好效果.CAO等设计了一种隐式注意力网络学习局部特征集合权重并嵌入尺度感知注意力利用非局部信息^[15].

尽管现有的连续超分方法在自然图像领域取得了显著成效,但在遥感图像场景下仍面临诸多挑战.遥感图像通常具有场景种类繁多、尺度跨度大、背景复杂度高特性,对模型在不同尺度下的泛化能力、全局结构建模能力与局部细节还原能力提出了更高要求.WU等^[16]引入了动态尺度感知模块与全局隐式函数,以增强模型对多尺度信息的适应性.CHEN等^[5]提出基于上下文交互的连续超分框架,从全局语义层面提升了重建的一致性.这些方法多以全局建模为主,缺乏对局部细节的表达能力.因此,如何设计出一种兼顾全局语义建模与局部高频细节表达的连续超分网络,已成为实现高质量遥感图像重建的关键所在.

基于以上背景,本文提出了一种跨尺度Transformer全局-局部交互的遥感图像连续超分辨率方法CTGI-ConSR(Cross-Scale Transformer with Global-Local Interaction for Continuous Super-Resolution of RSI).受CoTNet^[17]中上下文注意力思想的启发,本文设计了一种多尺度参数生成器(Multi-Scale Parameter Map Generator, MPMG),通过引入上下文注意力机制(Contextual Attention Mechanism, CAM)以充分挖掘输入特征中相邻位置之间的上下文信息,从而生成多尺度的特征参数图,增强模型对局部高频细节的表达能力.这些参数图被编码为一组局部特征单元(Local Token, T_{local}),并结合一个可学习的全局特征单元(Global Token, T_{global}),通过Transformer编码器对所有特征单元进行自注意力交互,实现局部细节与全局结构之间的协同表达.所提出的全局-局部解析器(Global-Local Parser, GP)分别从全局与局部两个层面对任意输入坐标进行解析,并通过 1×1 卷积与 3×3 深度卷积进行有效融

合,输出高质量的连续放大图像.

1 提出的方法

1.1 模型整体结构

本文模型整体结构如图1所示,包括3个阶段:多尺度特征参数图提取、全局-局部上下文交互以及全局-局部解析.第一阶段通过编码器提取初始低分

辨率特征,并构建多尺度参数生成器,生成多个不同尺度的特征参数图;第二阶段将这些多尺度特征参数图编码为一组 T_{local} ,并引入一个可学习的 T_{global} ,通过Transformer编码器实现 T_{global} 与 T_{local} 之间的自注意力交互;第三阶段以任意坐标点为输入,利用双分支全局-局部解析器分别从全局语义和局部细节两个层面解析像素值并融合.三者协同工作,实现从低分辨率输入到任意尺度图像输出的连续重建过程.

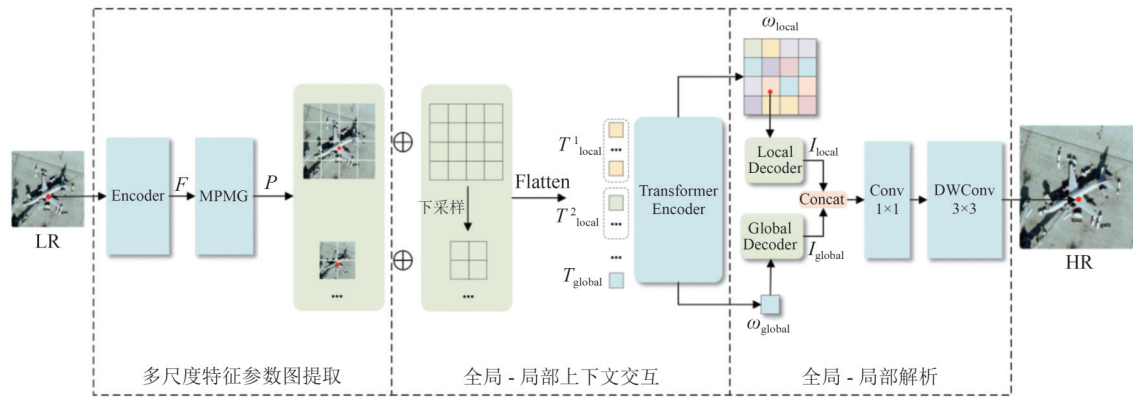


图1 本文方法的整体结构

Fig. 1 Overall structure of the proposed method

1.2 多尺度特征参数图提取

该阶段首先通过编码器提取初始低分辨率特征 F ,这里采用之前工作中的模块作为编码器,包括EDSR、RDN.接着经过多尺度参数生成器得到多个不同尺度的特征参数图 $P = \{p^1, p^2, p^3, p^4\}$:

$$F = \varphi_{\text{encoder}}(I_{\text{LR}}), \quad (2)$$

$$P = \{p^1, p^2, p^3, p^4\} = \varphi_{\text{MPMG}}(F), \quad (3)$$

其中, $\varphi_{\text{encoder}}(\cdot)$ 表示编码器提取初始低分辨率特征的操作, I_{LR} 表示输入的低分辨率图像, $\varphi_{\text{MPMG}}(\cdot)$ 表示多尺度参数生成器用来提取多尺度特征参数图的操作.

1.2.1 多尺度参数生成器

为了增强对局部高频细节的恢复能力,受CSPNet^[18]启发,本文基于跨阶段部分层(Cross-Stage Partial Layer, CSPLayer)设计了一种多尺度参数生成器.CSPLayer将输入特征沿通道划分为两部分,一部分特征经过深度处理提取高级语义特征,另一部分则直接作为残差连接跨过以保留原始信息,这种结构既降低了计算量,又有效兼顾了低层细节和高层语义表达.

MPMG具体结构如图2(a)所示,该结构由3个CSPLayer堆叠而成,每个CSPLayer由多个Darknet瓶颈块(Darknet Bottleneck Block, DBB)、两个1×1卷积以及一个3×3卷积组成,其中CSPLayer1、

CSPLayer2和CSPLayer3依次设置DBB数量为2、4、6.

以CSPLayer2为例,其工作流程如下.输入 $p^2 \in \mathbb{R}^{(H/2) \times (W/2) \times C}$ 首先经过一个1×1卷积操作将维度降为原始的一半,并沿着通道维度划分为两部分 $\{p_1^2, p_2^2 \in \mathbb{R}^{(H/2) \times (W/2) \times (C/4)}\}$;接着 p_2^2 经过4个DBB操作提取更高层次的特征 $F_n \in \mathbb{R}^{(H/2) \times (W/2) \times (C/4)}$, F_n 再与 p_1^2 沿着通道维度进行拼接得到 $H \in \mathbb{R}^{(H/2) \times (W/2) \times (C/2)}$;最后 H 经过一个1×1卷积操作将维度扩展2倍,再经过一个3×3卷积操作得到下采样两倍的特征参数图 $p^3 \in \mathbb{R}^{(H/4) \times (W/4) \times C}$.该过程可表示为:

$$p_1^2, p_2^2 = \text{Split}(\text{Conv}_{1 \times 1}(p^2)), \quad (4)$$

$$H = \text{Concat}(\text{DBB}_4(p_2^2), p_1^2), \quad (5)$$

$$p^3 = \text{Conv}_{3 \times 3}(\text{Conv}_{1 \times 1}(H)), \quad (6)$$

其中, $\text{Split}(\cdot)$ 表示通道划分操作.最终输入经过3层CSPLayer的逐级下采样处理,生成4组空间分辨率不同的特征参数图 $P \in \mathbb{R}^{(H/2^{i-1}) \times (W/2^{i-1}) \times C}$, $i = 1, 2, 3, 4$.

1.2.2 上下文注意力机制

DBB由两个“卷积层——批量归一化层——激活函数”基本卷积单元和上下文注意力机制(CAM)组成.CAM作为DBB的核心部分,能够利用输入特征中相邻位置之间的上下文信息指导注意力权重的学习,从而提升特征的表达能力.

图2(b)是CAM的具体结构.输入特征图

$X \in \mathbb{R}^{H \times W \times C}$, 其对应的键(key, K)、查询(query, Q)和值(value, V)分别定义为 $K=X$ 、 $Q=X$ 以及 $V=XW_v$, 其中 W_v 表示通过 1×1 卷积实现的线性映射. CAM 首先对 K 进行 3×3 卷积操作, 得到包含局部上下文信息的特征 $K^1 \in \mathbb{R}^{H \times W \times C}$; 随后将 K^1 与 Q 进行通道拼接, 通过两个连续的 1×1 卷积生成权重 $A' \in \mathbb{R}^{H \times W \times (C \cdot k^2)}$:

$$K^1 = \text{Conv}_{3 \times 3}(K), \quad (7)$$

$$A' = \text{Conv}_{1 \times 1}(\text{ReLU}(\text{Conv}_{1 \times 1}(K^1, Q))), \quad (8)$$

其中, 设置局部感受野大小 $k=3$, 定义注意力权重的空间范围. $\text{ReLU}(\cdot)$ 表示非线性激活函数. 将 A' 重塑为 $\mathbb{R}^{C \times k^2 \times HW}$, 使得每个空间位置的每个通道都关联

一个 $k \times k$ 局部注意力核, 从而有效捕捉该位置的邻域上下文信息. 随后沿邻域维度执行均值操作, 得到每个空间位置上每个通道的注意力权重 $A \in \mathbb{R}^{C \times HW}$. 将 softmax 归一化后的注意力权重 A 与 V 表示进行逐元素相乘, 得到注意力特征 $K^2 \in \mathbb{R}^{C \times HW}$:

$$K^2 = \text{Softmax}(A) \odot V, \quad (9)$$

其中, \odot 表示逐元素相乘. 将 K^2 重塑回 $\mathbb{R}^{H \times W \times C}$, K^1 和 K^2 进行逐元素相加得到最终输出 Y :

$$Y = K^1 \oplus K^2, \quad (10)$$

其中, \oplus 表示逐元素相加. 最终得到具有丰富上下文信息的增强特征表示.

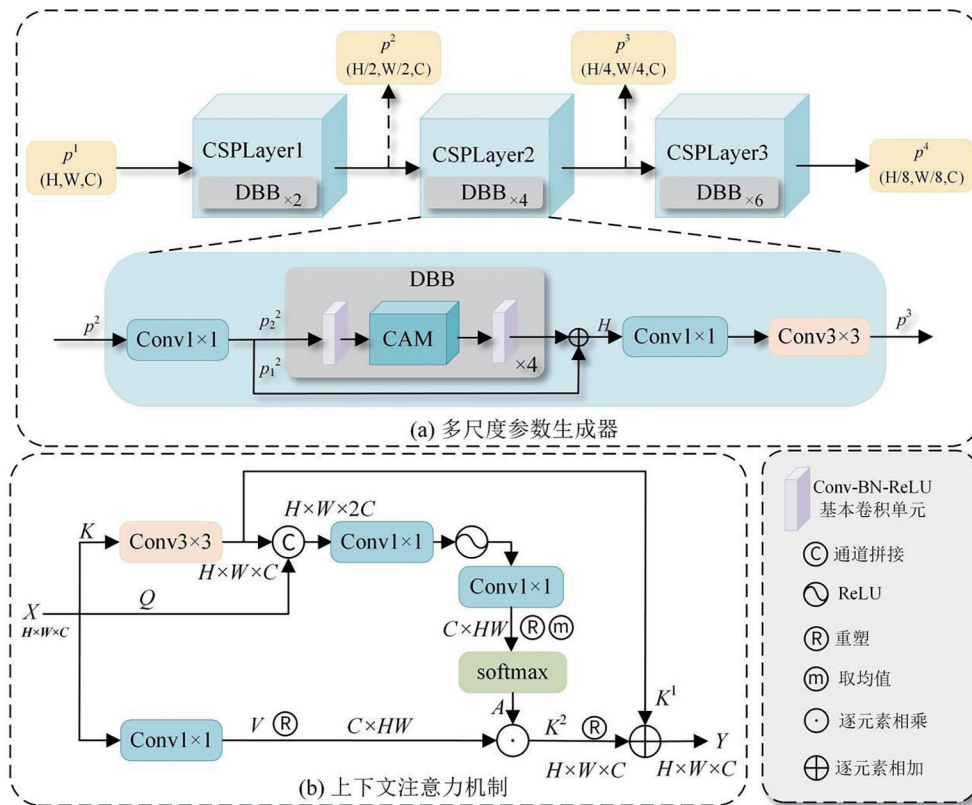


图2 多尺度参数生成器的内部结构图

Fig. 2 Internal structural of the multi-scale parameter generator

1.3 全局-局部上下文交互

本阶段采用三层 Transformer 编码器架构实现全局与局部信息的交互. 具体来说, 首先, 对 4 组不同尺度的特征参数图 P^i ($i = 1, 2, 3, 4$) 分别添加可学习的位置编码(Positional Encoding, PE), 并通过双三次下采样将其适配到其余尺度, 确保各尺度特征都保留了精确的空间位置信息; 接着将增强位置信息的特征参数图展平为多组 T_{local}^i ($i = 1, 2, 3, 4$), 并引入一个可训练的 $T_{\text{global}} \in \mathbb{R}^{1 \times C}$. 该 T_{global} 在模型初始化时随机生成, 其维度与 T_{local} 保持一致, 并在训练过程中持续优化; 随后, 将两者沿特征维度进行拼

接得到完整的特征表示 T , 再通过 Transformer 编码器的多头自注意力机制进行交互. 该过程可表示为:

$$T_{\text{local}}^i = \text{Flatten}(P^i + \varphi_{\text{bicubic}}(PE)), i = 1, 2, 3, 4, \quad (11)$$

$$T = \text{Concat}([T_{\text{global}}, T_{\text{local}}^1, T_{\text{local}}^2, T_{\text{local}}^3, T_{\text{local}}^4]), \quad (12)$$

$$(\omega_{\text{global}}, \omega_{\text{local}}) = \varphi_{\text{transformer}}(T), \quad (13)$$

其中, $\varphi_{\text{bicubic}}(\cdot)$ 表示双三次下采样操作, $\text{Flatten}(\cdot)$ 表示将多维张量转换为一维向量的过程. $\omega_{\text{global}} \in \mathbb{R}^{1 \times C}$ 表示全局参数向量, 该向量从输出的 T_{global} 中提取, 整合了图像的全局信息. $\omega_{\text{local}} \in \mathbb{R}^{H \times W \times C}$ 表示局部参数图, 该参数图由输出的 T_{local} 重塑而得, 与最高多尺度

参数图尺度保持一致,负责在局部像素层面精细化地表达特征.

1.4 全局-局部解析

本阶段通过双分支全局-局部解析器的协同作用预测查询坐标的像素值.给定查询坐标 (x_i, y_i) ,全局解析器直接基于全局参数向量 ω_{global} 预测像素值:

$$I_{\text{global}} = \varphi_{\text{global}}((x_i, y_i); \omega_{\text{global}}), \quad (14)$$

其中, $\varphi_{\text{global}}(\cdot)$ 表示全局解析操作, I_{global} 表示由全局解析器处理后得到的HR图像.接着通过插值操作从局部参数图 ω_{local} 中提取查询坐标对应的局部参数向量 $\omega_{\text{local}}^{(x_i, y_i)}$,局部解析器利用该参数预测像素值,该过程可表示为:

$$\omega_{\text{local}}^{(x_i, y_i)} = \varphi_{\text{interp}}(\omega_{\text{local}}, (x_i, y_i)), \quad (15)$$

$$I_{\text{local}} = \varphi_{\text{local}}((x_i, y_i); \omega_{\text{local}}^{(x_i, y_i)}), \quad (16)$$

其中, $\varphi_{\text{interp}}(\cdot)$ 表示插值操作, $\varphi_{\text{local}}(\cdot)$ 表示局部解析操作, I_{local} 表示由局部解析器处理后得到的HR图像.在特征融合阶段,首先采用通道拼接的方式整合 I_{global} 和 I_{local} ,再通过 1×1 卷积进行跨通道特征重组以建立全局特征与局部特征的关联,最后利用 3×3 深度可分离卷积在局部邻域内进行空间特征细化,最终实现多层次特征的有效融合.该过程可表示为:

$$I_{\text{HR}} = \text{DWConv}(\text{Conv}(\text{Concat}(I_{\text{global}}, I_{\text{local}}))), \quad (17)$$

其中,DWConv表示 3×3 深度可分离卷积, I_{HR} 表示最终重建的高分辨率图像.

双分支全局-局部解析器中全局解析器与局部解析器的网络结构完全相同,均采用了 N 层MLP结构,并通过正弦激活函数来捕捉图像的高频细节.具体来说,对于查询坐标 s ,全局解析器结合全局参数向量 ω_{global} 解析对应坐标的像素值;局部解析器结

合局部参数向量 $\omega_{\text{local}}^{(x_i, y_i)}$ 解析对应坐标的像素值.因此,本文以全局解析器为例介绍具体工作原理.

全局解析器的原理框图如图3所示.首先,第一层MLP将坐标转化为特征.具体来说,将坐标 s 通过线性变换映射到高维空间,通过正弦激活函数引入高频表达能力.其中对输入的全局参数向量 ω_{global} 做线性变换,并通过ReLU激活函数生成调制参数 m_1 ,用来调整正弦激活函数的相位,使网络能适应不同位置的信号特性.该过程可表示为:

$$m_1 = \text{ReLU}(w_1^m \omega_{\text{global}} + b_1^m), \quad (18)$$

$$d_1 = \sin(w_1^d s + m_1), \quad (19)$$

其中, w_1^m 、 w_1^d 和 b_1^m 表示第一层线性变换的权重和偏置. $\sin(\cdot)$ 表示正弦激活函数操作, d_1 表示初始输出特征.其次,再通过 $(N-1)$ 层MLP重复细化特征.经过第 i 层MLP时,对前一层的输出特征 d_{i-1} 进行线性变换,加上调制参数 m_i 后输入正弦激活函数,并残差连接前一层的输出特征 d_{i-1} 保留原始信息,得到该层的输出特征 d_i .其中将 ω_{global} 、前一层的调制参数 m_{i-1} 和输出特征 d_{i-1} 拼接作为输入,通过线性变换和ReLU激活函数,生成当前层调制参数 m_i .该过程可表示为:

$$m_i = \text{ReLU}(w_i^m \text{Concat}([\omega_{\text{global}}, m_{i-1}, d_{i-1}]) + b_i^m), \quad (20)$$

$$d_i = \sin(w_i^d d_{i-1} + m_i) + d_{i-1}, \quad (21)$$

其中, w_i^m 、 w_i^d 和 b_i^m 表示第 i 层线性变换的权重和偏置.后续层重复上述增强特征,最后通过线性变换将最后一层的输出特征映射到坐标 s 对应的像素值 y :

$$y = w_y^d d_N + b_y^d, \quad (22)$$

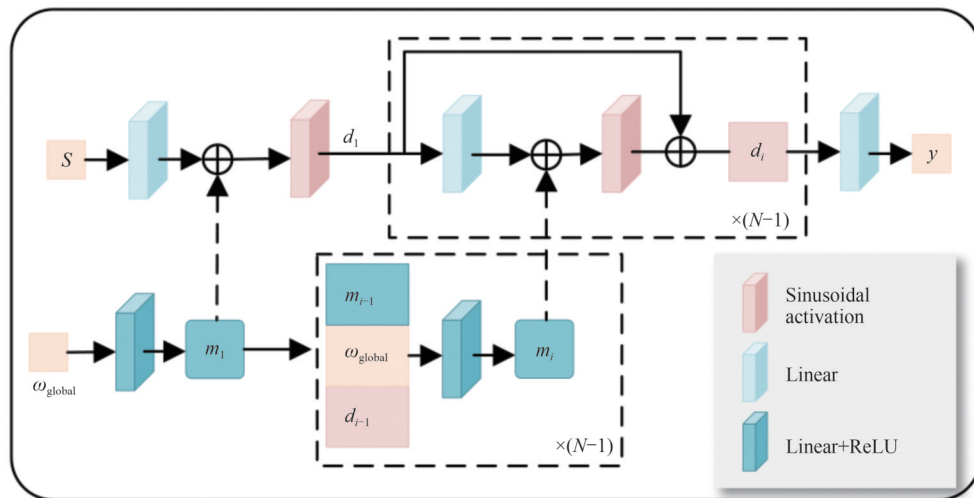


图3 全局解析器的整体结构

Fig. 3 Overall structure of global parser

其中, w_y^d, b_y^d 表示线性变换的权重和偏置, d_N 表示最后一层的输出特征.

1.5 损失函数

本文使用 L1 损失函数, 结合训练数据集对模型进行优化, 具体的损失函数表达式如下:

$$L = \frac{1}{N \cdot M} \sum_{i=0}^N \sum_{j=0}^M \left\| \hat{I}_{\text{HR}}^{(x_i, y_j)} - I_{\text{HR}}^{(x_i, y_j)} \right\|, \quad (23)$$

其中, \hat{I}_{HR} 表示模型生成的高分辨率图像, I_{HR} 表示相应的原始高分辨率图像. (x_i, y_j) 表示查询像素点的坐标, N 和 M 分别表示图像的宽度和高度.

2 实验结果与分析

2.1 实验设置

2.1.1 数据集

采用 UCMered^[19] 和 AID^[20] 两个公开遥感数据集进行性能验证. UCMered 数据集包含 21 类遥感场景, 每类 100 张 256×256 像素图像, 按 6:2:2 比例将每类图像划分为训练集、验证集和测试集. AID 数据集包含 30 类场景, 共 10000 张 600×600 像素图像, 从中随机抽取 2000 张作为验证集.

2.1.2 实现细节

将 Transformer 编码器的输入输出通道维度设为 256 维. 全局与局部解析器均采用 5 层 MLP 结构. 训练开始前, 对原始 HR 图像进行双三次下采样生成 LR 图像. 在训练阶段, 采用随机裁剪策略获取 48×48 的 LR-HR 图像块对, 并通过随机旋转和水平

翻转进行数据增强. 网络优化采用 AdamW 优化器, 并将损失函数设置为 L1 函数, 初始学习率设置为 1×10^{-4} , 并采用余弦退火策略进行动态调整. 批量大小设置为 8, 共训练 4000 个 epoch. 实验在 PyTorch 框架下实现, 使用两块 NVIDIA GTX 1080Ti GPU, 模型性能通过 PSNR 和 SSIM 指标进行量化评估.

2.2 对比实验

2.2.1 定量分析

为了验证本文方法的有效性, 将其在不同放大倍数 ($2 \times$ - $8 \times$) 下与现有先进连续放大 SR 方法作对比. 为全面评估模型在连续尺度下的重建能力, 本文遵循已有工作中常用的实验设置, 将测试放大倍数划分为“分布内”与“分布外”两类; 分布内是指训练阶段使用的放大倍数范围, 而分布外是指训练中未出现的放大倍数, 用于检验模型在未见尺度上的泛化能力.

如表 1 所示, 本文方法在 UCMered 数据集的多数放大倍数下均实现了最佳性能. 以 $4 \times$ 放大为例, 采用 EDSR 编码器时, PSNR 指标相较于 DIINN、SADN 和 FunSR 分别提升了 0.27 dB、0.32 dB 和 0.17 dB, 较其他对比方法有 0.4 dB 以上的性能优势. 此外, 尽管与次优方法 FunSR 在参数量上相当, 本文方法的重建效果仍显著更优. 从表 2 可以看出, 在更大、场景更复杂的 AID 数据集上, 本文方法同样在大多数放大倍数下保持最优性能. 以 $2 \times$ 放大为例, 采用 EDSR 和 RDN 编码器时, PSNR 分别较 FunSR 提升了 0.03 dB 和 0.04 dB, 进一步验证了其在复杂场景下的重建能力.

表 1 不同方法在 UCMered 数据集上重构性能比较 (PSNR/SSIM)

Tab. 1 Quantitative comparison of reconstruction performance (PSNR/SSIM) among different methods on the UCMered dataset

方法	参数量 /M	分布内					分布外		
		2×	2.5×	3×	3.5×	4×	6×	8×	
EDSR	MetaSR	<u>1.7</u>	33.17/0.9175	29.88/0.8561	27.72/0.7958	26.56/0.7539	25.55/0.7138	22.81/0.5854	21.48/0.5167
	LIIF	1.6	33.83/0.9244	30.53/0.8700	28.35/0.8149	27.21/0.7779	26.25/0.7433	23.20/0.6097	21.73/0.5362
	A-LIIF	<u>1.7</u>	33.84/0.9254	30.58/0.8707	28.45/0.8186	27.35/0.7824	26.37/0.7488	23.29/0.6135	21.74/0.5389
	DIINN	2.2	34.29/0.9298	30.87/0.8767	28.70/0.8260	27.54/0.7905	26.50/0.7546	23.26/0.6152	21.73/0.5388
	SADN	1.9	34.17/0.9288	30.77/0.8739	28.63/0.8248	27.49/0.7895	26.45/0.7540	23.09/0.6085	21.54/0.5253
	FunSR	8.7	<u>34.39/0.9311</u>	<u>31.00/0.8799</u>	<u>28.79/0.8304</u>	<u>27.67/0.7971</u>	<u>26.60/0.7604</u>	<u>23.30/0.6167</u>	<u>21.75/0.5399</u>
	本文	8.7	<u>34.53/0.9321</u>	<u>31.17/0.8834</u>	<u>28.95/0.8338</u>	<u>27.83/0.8013</u>	<u>26.77/0.7667</u>	<u>23.37/0.6207</u>	<u>21.80/0.5417</u>
RDN	MetaSR	22.4	34.25/0.9290	30.99/0.8782	28.88/0.8310	27.81/0.7977	26.80/0.7636	<u>23.51/0.6244</u>	21.87/0.5411
	LIIF	22.3	34.37/0.9309	31.14/0.8832	28.98/0.8340	27.91/0.8019	26.93/0.7692	<u>23.60/0.6318</u>	<u>21.92/0.5488</u>
	A-LIIF	<u>22.4</u>	34.15/0.9292	30.90/0.8789	28.73/0.8267	27.67/0.7922	26.72/0.7606	23.47/0.6248	21.80/0.5423
	DIINN	23.0	34.70/0.9336	31.32/0.8849	29.16/0.8368	28.04/0.8047	26.99/0.7695	23.47/0.6264	21.85/0.5454
	SADN	22.7	34.64/0.9331	31.31/0.8854	29.14/0.8378	28.05/0.8049	27.01/0.7718	23.40/0.6255	21.81/0.5458
	FunSR	29.4	<u>34.82/0.9357</u>	<u>31.49/0.8890</u>	<u>29.33/0.8428</u>	<u>28.22/0.8107</u>	<u>27.17/0.7784</u>	23.45/0.6274	21.81/0.5430
	本文	29.4	<u>34.90/0.9362</u>	<u>31.53/0.8896</u>	<u>29.37/0.8437</u>	<u>28.26/0.8127</u>	<u>27.21/0.7798</u>	23.48/0.6256	<u>21.89/0.5480</u>

注: 蓝色字体表示最优结果, 下划线表示次优结果.

表 2 不同方法在 AID 数据集上重构性能比较 (PSNR/SSIM)

Tab. 2 Quantitative comparison of reconstruction performance (PSNR/SSIM) among different methods on the AID dataset

方法	分布内					分布外		
	2×	2.5×	3×	3.5×	4×	6×	8×	
MetaSR	35.63/0.9252	32.46/0.8702	30.42/0.8190	29.29/0.7833	28.32/0.7511	25.45/0.6500	23.90/0.5906	
LHIF	36.02/0.9292	32.82/0.8778	30.80/0.8297	29.67/0.7958	28.70/0.7656	25.79/0.6667	24.15/0.6089	
A-LHIF	35.78/0.9289	32.71/0.8778	30.73/0.8298	29.63/0.7962	28.68/0.7661	25.80/0.6678	24.17/0.6094	
EDSR	DIINN	36.27/0.9325	32.96/0.8795	30.90/0.8304	29.76/0.7969	28.84/0.7688	<u>25.94/0.6741</u>	<u>24.28/0.6138</u>
	SADN	35.96/0.9316	32.80/0.8791	30.78/0.8306	29.71/0.7983	28.78/0.7697	25.80/0.6700	24.10/0.6018
	FunSR	<u>36.35/0.9334</u>	<u>33.05/0.8825</u>	<u>30.96/0.8340</u>	<u>29.84/0.8018</u>	<u>28.90/0.7737</u>	25.93/0.6750	24.27/0.6136
	本文	<u>36.38/0.9338</u>	<u>33.07/0.8831</u>	<u>30.97/0.8346</u>	<u>29.87/0.8031</u>	<u>28.94/0.7753</u>	<u>25.97/0.6760</u>	<u>24.32/0.6146</u>
MetaSR	36.00/0.9310	32.87/0.8804	30.86/0.8330	29.75/0.8000	28.80/0.7710	25.80/0.6714	<u>24.25/0.6099</u>	
LHIF	36.03/0.9313	32.87/0.8806	30.84/0.8331	29.73/0.8006	28.79/0.7717	25.87/0.6720	24.23/0.6098	
A-LHIF	35.61/0.9300	32.63/0.8793	30.69/0.8320	29.62/0.7993	28.68/0.7701	25.82/0.6726	24.19/0.6093	
RDN	DIINN	36.25/0.9325	32.96/0.8808	30.73/0.8301	29.60/0.7989	28.69/0.7708	<u>25.88/0.6741</u>	24.20/0.6099
	SADN	36.14/0.9321	32.92/0.8803	30.83/0.8311	29.73/0.7989	28.78/0.7704	25.86/0.6734	<u>24.25/0.6125</u>
	FunSR	<u>36.33/0.9338</u>	<u>32.95/0.8819</u>	<u>30.86/0.8332</u>	<u>29.75/0.8013</u>	<u>28.81/0.7735</u>	25.85/0.6721	24.23/0.6110
	本文	<u>36.37/0.9339</u>	<u>33.05/0.8828</u>	<u>30.94/0.8336</u>	<u>29.83/0.8021</u>	<u>28.91/0.7750</u>	<u>25.88/0.6738</u>	24.22/0.6112

注:蓝色字体表示最优结果,下划线表示次优结果.

为进一步验证所提方法的有效性,分别对两个数据集上不同场景类别进行评估.如表 3 所示,在 UC Merced 数据集的 21 个类别中,4×放大任务下,本

方法在所有类别上均取得了最优的 PSNR/SSIM 值.根据表 4 所示,针对更具挑战性的 AID 数据集,本方法在 30 个类别中的 22 个类别上表现最佳.

表 3 不同方法在 UC Merced 数据集各类别 4×放大倍数下的重构性能对比 (PSNR/SSIM)

Tab. 3 Comparison of reconstruction performance for different methods on the UC Merced dataset at 4× scaling factor (PSNR/SSIM)

类别	MetaSR	LHIF	A-LHIF	DIINN	SADN	FunSR	本文
农业	25.84/0.5435	28.08/0.6743	<u>28.35/0.7101</u>	28.13/0.6836	27.98/0.6901	28.14/0.6850	<u>28.58/0.7124</u>
飞机	24.67/0.7299	25.19/0.7470	25.24/0.7484	25.37/0.7547	25.37/0.7535	<u>25.47/0.7582</u>	<u>25.68/0.7623</u>
棒球场	32.57/0.8191	32.98/0.8256	33.03/0.8267	<u>33.18/0.8294</u>	33.09/0.8284	33.14/0.8303	<u>33.26/0.8319</u>
海滩	32.91/0.8333	33.03/0.8368	33.01/0.8380	33.02/0.8377	<u>33.10/0.8408</u>	33.07/0.8418	<u>33.24/0.8469</u>
建筑物	24.30/0.7588	25.07/0.7839	25.30/0.7892	25.44/0.7946	25.33/0.7932	<u>25.46/0.7989</u>	<u>25.63/0.8025</u>
灌木丛	23.48/0.6134	23.72/0.6291	23.76/0.6332	23.78/0.6404	23.79/0.6382	<u>23.79/0.6425</u>	<u>23.86/0.6481</u>
密集住宅	24.14/0.7594	24.81/0.7845	24.90/0.7870	25.04/0.7935	24.98/0.7916	<u>25.13/0.7978</u>	<u>25.32/0.8037</u>
森林	25.50/0.5821	25.88/0.6072	25.94/0.6110	26.03/0.6192	25.99/0.6217	<u>26.02/0.6266</u>	<u>26.09/0.6305</u>
高速公路	26.08/0.7471	27.39/0.7944	27.59/0.7959	27.86/0.8169	27.86/0.8185	<u>28.10/0.8290</u>	<u>28.34/0.8343</u>
高尔夫球场	30.69/0.7676	31.04/0.7756	31.07/0.7765	31.12/0.7780	31.09/0.7781	<u>31.13/0.7797</u>	<u>31.22/0.7810</u>
港口	21.69/0.7646	22.36/0.7965	22.51/0.8009	22.61/0.8078	22.58/0.8038	<u>22.77/0.8140</u>	<u>23.00/0.8221</u>
交叉路口	24.10/0.7049	24.56/0.7272	24.61/0.7305	24.80/0.7379	24.70/0.7352	<u>24.88/0.7471</u>	<u>25.11/0.7562</u>
中心住宅	23.76/0.6876	24.31/0.7093	24.40/0.7128	24.46/0.7162	24.44/0.7165	<u>24.50/0.7209</u>	<u>24.64/0.7257</u>
移动住宅	21.89/0.6895	22.68/0.7267	22.84/0.7315	22.99/0.7410	22.95/0.7395	<u>23.12/0.7473</u>	<u>23.31/0.7534</u>
立交桥	22.81/0.6855	24.32/0.7429	24.57/0.7535	25.00/0.7679	24.76/0.7592	<u>25.21/0.7786</u>	<u>25.49/0.7883</u>
停车场	19.61/0.6691	20.26/0.7059	20.43/0.7151	20.71/0.7339	20.67/0.7298	<u>20.96/0.7451</u>	<u>21.26/0.7563</u>
河流	25.81/0.6754	26.20/0.6920	26.25/0.6939	<u>26.31/0.6985</u>	26.26/0.6991	26.28/0.7033	<u>26.34/0.7058</u>
跑道	27.94/0.7414	28.88/0.7634	29.10/0.7692	29.37/0.7775	29.28/0.7755	<u>29.68/0.7842</u>	<u>29.83/0.7886</u>
稀疏住宅	25.79/0.7014	26.19/0.7161	26.26/0.7181	26.30/0.7214	26.27/0.7212	<u>26.33/0.7251</u>	<u>26.48/0.7282</u>
储油罐	26.33/0.7662	26.87/0.7867	26.95/0.7881	27.05/0.7922	27.04/0.7923	<u>27.21/0.7997</u>	<u>27.32/0.8031</u>
网球场	26.66/0.7501	27.51/0.7837	27.73/0.7960	27.87/0.8049	27.87/0.8072	<u>28.08/0.8142</u>	<u>28.21/0.8188</u>
平均	25.55/0.7138	26.25/0.7433	26.37/0.7488	26.50/0.7546	26.45/0.7540	<u>26.60/0.7604</u>	<u>26.77/0.7667</u>

注:蓝色字体表示最优结果,下划线表示次优结果.

表 4 不同方法在 AID 数据集各类别 4×放大倍数下的重构性能对比(PSNR/SSIM)

Tab. 4 Comparison of reconstruction performance for different methods on the AID dataset at 4× scaling factor (PSNR/SSIM)

类别	MetaSR	LIIF	A-LIIF	DIINN	SADN	FunSR	本文
飞机	27.63/0.7695	27.94/0.7823	27.93/0.7830	28.00/0.7850	27.98/0.7846	<u>28.04/0.7875</u>	<u>28.09/0.7892</u>
裸地	<u>35.13/0.8297</u>	35.54/0.8334	35.24/0.8302	35.14/0.8237	35.23/0.8270	35.22/0.8270	35.18/0.8263
棒球场	34.05/0.8487	34.54/0.8567	34.41/0.8563	34.57/0.8575	<u>34.41/0.8558</u>	34.58/0.8589	34.57/0.8592
海滩	33.12/0.8311	<u>33.23/0.8350</u>	33.17/0.8339	33.22/0.8329	33.18/0.8347	33.21/0.8341	<u>33.26/0.8363</u>
桥梁	30.05/0.8335	30.90/0.8550	30.91/0.8558	30.83/0.8571	30.84/0.8573	31.12/0.8623	<u>31.16/0.8635</u>
中心	28.63/0.7892	29.28/0.8086	29.21/0.8095	29.31/0.8112	29.15/0.8105	29.32/0.8134	<u>29.36/0.8154</u>
教堂	22.11/0.6214	22.59/0.6471	22.67/0.6507	22.77/0.6562	<u>22.69/0.6540</u>	22.78/0.6622	22.77/0.6638
商业	25.51/0.7183	25.68/0.7282	25.67/0.7262	25.61/0.7253	25.65/0.7283	25.70/0.7332	<u>25.75/0.7357</u>
密集住宅	21.78/0.6105	22.05/0.6307	22.07/0.6320	22.05/0.6316	22.07/0.6339	22.10/0.6394	<u>22.11/0.6423</u>
沙漠	<u>37.27/0.8881</u>	37.64/0.8915	37.45/0.8910	37.55/0.8881	37.34/0.8873	37.61/0.8906	37.50/0.8884
农田	33.15/0.8394	33.80/0.8552	33.87/0.8568	33.94/0.8592	33.85/0.8565	33.98/0.8616	<u>34.07/0.8637</u>
森林	27.55/0.6437	<u>27.55/0.6519</u>	27.73/0.6555	27.42/0.6394	27.53/0.6470	27.47/0.6460	27.57/0.6522
工业	25.14/0.7242	25.61/0.7477	25.67/0.7502	25.74/0.7528	25.69/0.7512	25.79/0.7582	<u>25.82/0.7608</u>
草地	<u>32.27/0.7166</u>	32.53/0.7214	32.50/0.7192	32.40/0.7135	32.36/0.7182	32.51/0.7217	32.45/0.7174
中型住宅	25.41/0.6676	25.86/0.6879	25.87/0.6896	25.87/0.6903	<u>25.85/0.6918</u>	25.90/0.6958	25.89/0.6972
山地	28.62/0.7285	28.62/0.7328	28.67/0.7311	28.58/0.7272	28.60/0.7306	28.60/0.7316	<u>28.64/0.7331</u>
公园	27.62/0.7333	27.80/0.7412	27.78/0.7413	27.73/0.7392	27.77/0.7417	27.79/0.7445	<u>27.84/0.7471</u>
停车场	22.22/0.7521	22.63/0.7725	22.68/0.7749	22.36/0.7679	22.54/0.7728	22.70/0.7801	<u>22.79/0.7835</u>
游乐场	35.06/0.8745	35.78/0.8852	35.91/0.8878	35.95/0.8875	35.66/0.8860	35.96/0.8887	<u>36.08/0.8905</u>
池塘	37.49/0.8724	<u>37.69/0.8753</u>	37.49/0.8762	37.73/0.8772	37.09/0.8764	37.65/0.8776	<u>37.75/0.8783</u>
港口	24.59/0.7863	24.89/0.8015	24.91/0.8027	24.87/0.8037	24.87/0.8049	24.91/0.8078	<u>24.95/0.8090</u>
火车站	26.61/0.7334	27.33/0.7717	27.33/0.7737	27.37/0.7760	27.35/0.7752	27.56/0.7872	<u>27.62/0.7909</u>
度假村	24.94/0.7048	25.23/0.7198	25.26/0.7205	25.27/0.7204	25.24/0.7214	25.28/0.7249	<u>25.29/0.7266</u>
河流	29.00/0.7165	29.20/0.7244	29.21/0.7227	29.17/0.7202	29.19/0.7236	29.22/0.7260	<u>29.23/0.7262</u>
学校	24.71/0.7151	25.01/0.7322	24.99/0.7317	24.92/0.7313	24.95/0.7313	25.03/0.7369	<u>25.07/0.7394</u>
稀疏住宅	23.77/0.6261	24.11/0.6411	24.15/0.6424	24.10/0.6384	24.12/0.6429	<u>24.19/0.6475</u>	24.17/0.6479
广场	28.16/0.7877	28.59/0.8014	<u>28.59/0.8021</u>	28.56/0.8021	28.51/0.8013	28.57/0.8035	<u>28.62/0.8051</u>
体育场	33.90/0.8674	34.24/0.8730	34.30/0.8736	34.33/0.8742	34.12/0.8733	<u>34.34/0.8750</u>	<u>34.36/0.8752</u>
储油罐	23.52/0.6775	23.95/0.6985	24.01/0.6998	24.05/0.7029	24.00/0.7013	<u>24.06/0.7072</u>	<u>24.10/0.7108</u>
高架桥	25.72/0.6842	26.38/0.7194	26.40/0.7215	26.49/0.7265	26.45/0.7255	<u>26.59/0.7335</u>	<u>26.66/0.7384</u>
平均	28.32/0.7511	28.70/0.7656	28.68/0.7661	28.84/0.7688	28.78/0.7697	<u>28.90/0.7737</u>	<u>28.94/0.7753</u>

注:蓝色字体表示最优结果,下划线表示次优结果。

2.2.2 可视化结果分析

本小节通过视觉对比实验,将本文提出的方法与其他方法的重建结果进行比较,以进一步验证其有效性.图4展示了本文方法采用EDSR编码器时在

不同放大倍数下的重建效果.第一行选用UCMerced测试集中包含大量密集车辆的“parkinglot_29”图像,具有典型的复杂背景干扰特征;第二行为AID测试集中的“school_47”图像,涵盖大型校园建筑与微

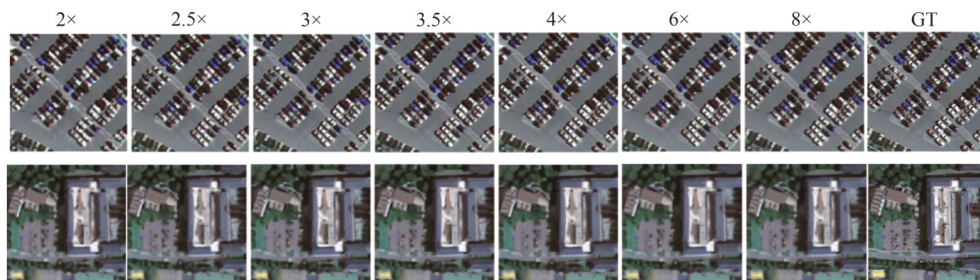


图4 本文方法在连续放大倍数下的重构结果

Fig. 4 Reconstruction results of the proposed method at continuous scaling factors

小目标,具有大尺度跨度特性.结果表明,本文方法在各放大倍数下均保持优异的视觉质量,验证了其在复杂背景与大尺度场景中的重建能力.

图 5 和图 6 分别展示了 UCMerced 与 AID 数据集在 4×放大任务下的定性对比结果.其中,airplane_11 和 bridge_252 具有大尺度跨度特征,tenniscourt_93

和 square_19 则包含复杂背景干扰.通过与现有方法的视觉比较,本文方法在重建质量上表现更优.

2.3 消融实验

2.3.1 不同组件的有效性

本节通过一系列消融分析来验证不同组件的有效性.所有实验都是在 UCMerced 数据集上开展,

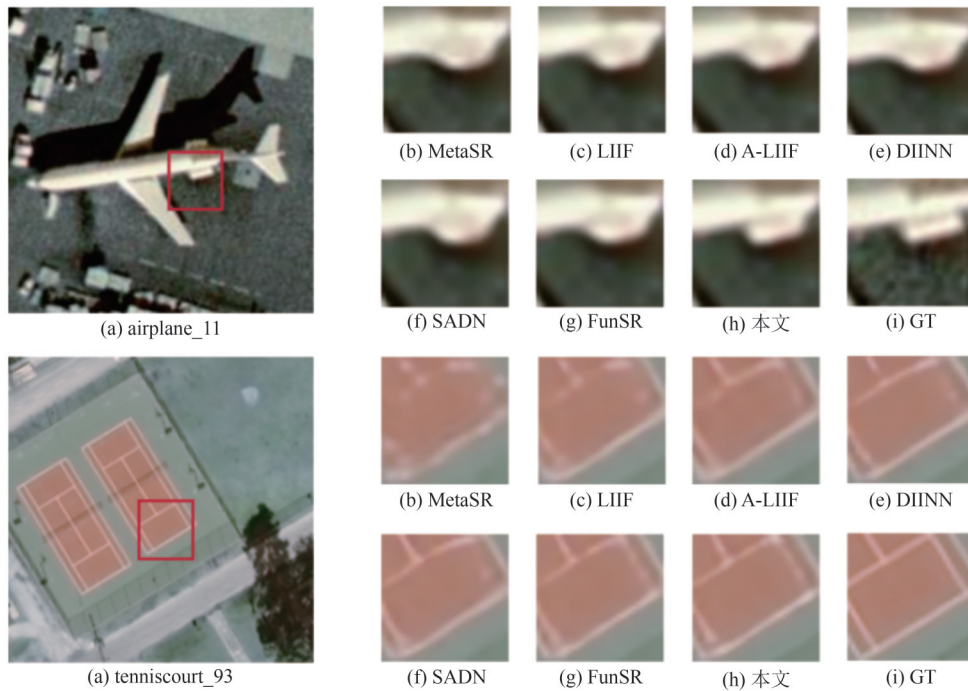


图 5 不同方法在图像 airplane_11 和 tenniscourt_93 上放大因子为 4×的重构结果

Fig. 5 Reconstruction results of airplane_11 and tenniscourt_93 for different methods at 4× scaling factor

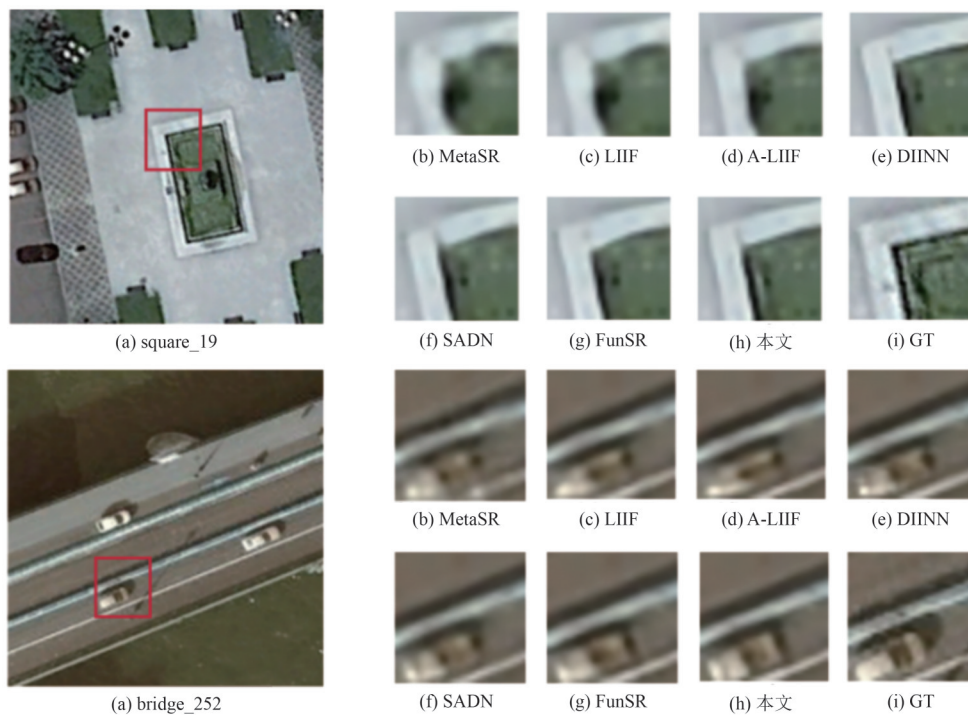


图 6 不同方法在图像 bridge_252 和 square_19 上放大因子为 4×的重构结果

Fig. 6 Reconstruction results of bridge_252 和 square_19 for different methods at 4× scaling factor

采用EDSR编码器架构,并保持一致的训练参数配置.设计了4个模型变体:仅保留基础特征提取模块,移除上下文注意力机制(CAM)、Transformer编码器和全局解析器,将此变体记为Base;在Base基础上增加CAM,将此变体记为Base+CAM;进一步集成Transformer编码器,将此变体记为Base+CAM+Trans;再引入全局解析器得到完整模型.

如表5所示,各组件对性能提升均具有显著贡献.在4×放大任务中,Base模型的PSNR为26.16 dB,引入CAM后PSNR提升至26.51 dB,说明该机制能有效增强局部特征提取能力;当加入Transformer编码器时,性能进一步提升至26.65 dB,验证了全局-局部上下文建模的重要性;加入全局

解析器后,最终CTGI-ConSR达到26.77dB的最佳性能,表明全局-局部解析器策略能实现更优的特征融合.实验结果充分验证了各组件设计的合理性和必要性.

2.3.2 采用不同编码器的效果

如表1和表2所示,本文方法在采用不同编码器架构时均取得优异性能,且多数放大倍数下优于其他方法,验证了所提结构的鲁棒性.图7进一步展示了UCMerced数据集中图像building_65和AID数据集中图像railwaystation_12在不同编码器架构下的视觉重构效果.结果表明,编码器结构对重建图像的细节恢复能力和边缘清晰度具有显著影响,为实际应用中编码器的选择提供了参考.

表5 不同组件对模型重建性能的有效性比较

Tab. 5 Comparison of the effects of different variants on model reconstruction performance

方法	2×	2.5×	3×	3.5×	4×
Base	33.90/0.9240	30.45/0.8683	28.34/0.8155	27.37/0.7877	26.16/0.7487
Base+CAM	34.30/0.9270	30.84/0.8761	28.75/0.8286	27.62/0.7960	26.51/0.7576
Base+CAM+Trans	34.42/0.9315	31.05/0.8817	28.82/0.8308	27.70/0.7988	26.65/0.7632
本文	34.53/0.9321	31.17/0.8834	28.95/0.8338	27.83/0.8013	26.77/0.7667

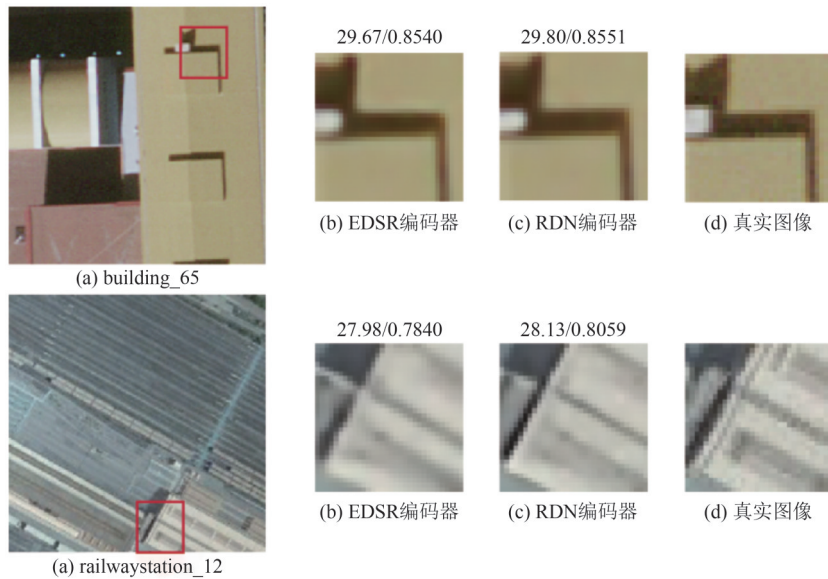


图7 选用不同编码器在放大因子为4×的重构图像

Fig. 7 The reconstructed images at 4× scaling factor using different encoders

3 结论

本文提出了一种基于全局-局部上下文交互的遥感图像连续超分辨率方法(CTGI-ConSR),通过协同优化全局语义一致性与局部细节保真度,显著提升了任意缩放倍数下的重建质量.通过引入动态注意力机制,自适应地挖掘局部邻域的多尺度纹理与

结构特征,增强了高频信息的建模能力;基于Transformer编码器构建全局语义表征Token,通过跨尺度长程依赖建模捕捉全局上下文关联,建立空间-语义协同优化机制;设计双分支全局-局部解析器,分别从语义约束与细节回归维度实现坐标到像素值的精细化映射,兼顾结构一致性与高频细节的精准重建.在UCMerced和AID遥感数据集上的实验

结果验证了本文方法在改善重构性能方面的有效性.

参 考 文 献

- [1] QIU D, CHENG Y, WANG X. Medical image super-resolution reconstruction algorithms based on deep learning: A survey[J]. *Computer Methods and Programs in Biomedicine*, 2023, 238: 107590.
- [2] HSU W Y, YANG P Y. Pedestrian detection using multi-scale structure-enhanced super-resolution [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2023, 24(11): 12312-12322.
- [3] JIANG H, PENG M, ZHONG Y, et al. A survey on deep learning-based change detection from high-resolution remote sensing images[J]. *Remote Sensing*, 2022, 14(7): 1552.
- [4] WANG P, BAYRAM B, SERTEL E. A comprehensive review on deep learning based remote sensing image super-resolution methods [J]. *Earth-Science Reviews*, 2022, 232: 104110.
- [5] CHEN K, LI W, LEI S, et al. Continuous remote sensing image super-resolution based on context interaction in implicit function space [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 4702216.
- [6] DONG C, LOY C C, HE K, et al. Image super-resolution using deep convolutional networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(2): 295-307.
- [7] ZEILER M D, KRISHNAN D, TAYLOR G W, et al. Deconvolutional networks [C]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco: IEEE, 2010: 2528-2535.
- [8] SHI W, CABALLERO J, HUSZÁR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 1874-1883.
- [9] HU X, MU H, ZHANG X, et al. Meta-SR: A magnification-arbitrary network for super-resolution [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2020: 1575-1584.
- [10] CHEN Y, LIU S, WANG X. Learning continuous image representation with local implicit image function [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021: 8624-8634.
- [11] XU X, WANG Z, SHI H. UltraSR: Spatial encoding is a missing key for implicit image function-based arbitrary-scale super-resolution[J]. arXiv: 2103.12716.
- [12] LIU Y T, GUO Y C, ZHANG S H. Enhancing multi-scale implicit learning in image super-resolution with integrated positional encoding[J]. arXiv: 2112.05756.
- [13] LEE J, HWAN J. Local texture estimator for implicit representation function [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022: 1919-1928.
- [14] CHEN H W, XU Y S, HONG M F, et al. Cascaded local implicit transformer for arbitrary-scale super-resolution [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver: IEEE, 2023: 18257-18267.
- [15] CAO J, WANG Q, XIAN Y, et al. CiaoSR: Continuous implicit attention-in-attention network for arbitrary-scale image super-resolution [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver: IEEE, 2023: 1796-1807.
- [16] WU H, NI N, ZHANG L. Learning dynamic scale awareness and global implicit functions for continuous-scale super-resolution of remote sensing images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 5602315.
- [17] LI Y, YAO T, PAN Y, et al. Contextual transformer networks for visual recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(2): 1489-1500.
- [18] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle: IEEE, 2020: 1571-1580.
- [19] YANG Y, NEWSAM S. Bag-of-visual-words and spatial extensions for land-use classification [C]//Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. San Jose California: ACM, 2010: 270-279.
- [20] XIA G S, HU J, 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.

(责编 曹东,校对 雷建云)