



计算机科学

COMPUTER SCIENCE

基于递归门控卷积的遥感图像超分辨率研究

刘长新, 吴宁, 胡俐蕊, 高霸, 高学山

引用本文

刘长新, 吴宁, 胡俐蕊, 高霸, 高学山. 基于递归门控卷积的遥感图像超分辨率研究[J]. 计算机科学, 2024, 51(2): 205-216.

LIU Changxin, WU Ning, HU Lirui, GAO Ba, GAO Xueshan. Recursive Gated Convolution Based Super-resolution Network for Remote Sensing Images [J]. Computer Science, 2024, 51(2): 205-216.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

基于语义注意力的医学图像超分辨率方法

Medical Image Super-resolution Method Based on Semantic Attention

计算机科学, 2023, 50(11A): 221200107-6. <https://doi.org/10.11896/jsjcx.221200107>

基于潜在注意力的高性能视频超分辨率技术

Efficient Video Super-Resolution with Latent Attention

计算机科学, 2023, 50(11A): 221100156-10. <https://doi.org/10.11896/jsjcx.221100156>

基于特征权重感知的VNF资源需求预测方法

Feature Weight Perception-based Prediction of Virtual Network Function Resource Demands

计算机科学, 2023, 50(9): 331-336. <https://doi.org/10.11896/jsjcx.221000012>

基于生成对抗网络的遥感图像锐化方法

Remote Sensing Image Pan-sharpening Method Based on Generative Adversarial Network

计算机科学, 2023, 50(8): 133-141. <https://doi.org/10.11896/jsjcx.220600065>

基于视频多帧融合的医学超声图像超分辨率重建方法

Medical Ultrasound Image Super-resolution Reconstruction Based on Video Multi-frame Fusion

计算机科学, 2023, 50(7): 143-151. <https://doi.org/10.11896/jsjcx.220700232>

基于递归门控卷积的遥感图像超分辨率研究

刘长新¹ 吴宁² 胡俐蕊³ 高霸¹ 高学山⁴

1 广西大学计算机与电子信息学院 南宁 530004

2 广西海洋工程装备与技术重点实验室(北部湾大学) 广西 钦州 535011

3 北部湾大学电子与信息工程学院 广西 钦州 535011

4 北部湾大学机械与船舶海洋工程学院 广西 钦州 535011

(m18278591041@163.com)

摘要 由于受到硬件条件的限制,通常难以获得具有高分辨率(HR)的遥感图像。利用单幅图像超分辨率(SISR)技术对低分辨率(LR)遥感图像进行超分辨率重建是获取高分辨率遥感图像的常用方法。近年来,在图像超分辨率领域引入的卷积神经网络(CNN)改进了图像重建性能。然而,现有的基于CNN的超分辨率模型通常使用低阶注意力机制提取深层特征,其表征能力有待提高,且常规卷积的感受野有限,缺乏对远距离依赖关系的学习。为了解决以上问题,提出了一种基于递归门控卷积的遥感图像超分辨率方法RGCSR。该方法引入递归门控卷积 g^n Conv学习全局依赖和局部细节,通过高阶空间交互来获取高阶特征。首先,使用由高阶交互子模块(HorBlock)和前馈神经网络(FFN)组成的高阶交互——前馈神经网络模块(HFB)提取高阶特征。其次,利用由通道注意力(CA)和 g^n Conv构建的特征优化模块(FOB)优化各个中间模块的输出特征。最后,在多个数据集上的对比结果表明,RGCSR比现有的基于CNN的超分辨率方法具备更好的重建性能和视觉效果。

关键词: 递归门控卷积;高阶空间交互;通道注意力;遥感图像;超分辨率

中图分类号 TP751.1

Recursive Gated Convolution Based Super-resolution Network for Remote Sensing Images

LIU Changxin¹, WU Ning², HU Lirui³, GAO Ba¹ and GAO Xueshan⁴

1 School of Computer, Electronics and Information, Guangxi University, Nanning, 530004, China

2 Guangxi Key Laboratory of Marine Engineering Equipment and Technology, Beibu Gulf University, Qinzhou, Guangxi 535011, China

3 College of Electronics and Information Engineering, Beibu Gulf University, Qinzhou, Guangxi 535011, China

4 College of Mechanical, Naval Architecture and Ocean Engineering, Beibu Gulf University, Qinzhou, Guangxi 535011, China

Abstract Due to hardware manufacturing constraints, it is usually difficult to obtain high-resolution(HR) images in the area of remote sensing. From low resolution remote-sensing image to reconstruct high-resolution(HR) image via single-image super-resolution(SISR) technique is a common method. Recently, the convolutional neural network(CNN) was introduced to the field of super-resolution image reconstruction, and it effectively improved the image reconstruction performance. However, the classic CNN-based approaches typically use low-order attention to extract deep features, which limits its reconstructing ability. Moreover, the receptive field is limited, which lacks the ability to learn long-range dependency. To solve the above problems, a recursive gated convolution-based super-resolution method for remote sensing images(RGCSR) is proposed. The RGCSR introduces recursive gated convolution(g^n Conv) to learn global dependencies and local details, and high-order features are acquired by high-order spatial interactions. Firstly, a high-order interaction—feedforward neural network(HFB) consisting of a high-order interaction sub-module(HorBlock) and a feedforward neural network(FFN) is applied to extract high-order features. Then, a feature optimization module(FOB) contains channel attention(CA) and g^n Conv is used to optimize the output features of each intermediate module. Finally, the comparison results on multiple datasets show that RGCSR has better reconstruction and visualization performances than existing CNN-based solutions.

Keywords Recursive gated convolution, High-order spatial interaction, Channel attention, Remote sensing images, Super-resolution

到稿日期:2023-08-04 返修日期:2023-11-27

基金项目:国家自然科学基金(61961004);广西重点研发计划(2021AB10030)

This work was supported by the National Natural Science Foundation of China(61961004) and Guangxi Key Project of Research and Development Plan(2021AB10030).

通信作者:胡俐蕊(hulr163@163.com)

1 引言

遥感图像携带着大量地表信息,人类通过对遥感图像的研究能够获得大量参考信息,并以此指导人类改进自身的生存和发展。目前,遥感图像已在许多实际工作中起到重要作用,如环境评估^[1]、植被检测^[2]、城市规划^[3]等。与低分辨率(Low-Resolution, LR)遥感图像相比,高分辨率(High-Resolution, HR)遥感图像能够提供大量的细节信息。然而,由于受到传输噪声、技术限制和运动模糊等因素的影响,所获取的遥感图像的分辨率较低,阻碍了对遥感图像的进一步应用。最直接的办法是对物理设备更新换代,但需要耗费巨大的成本。因此,通过软件实现的图像超分辨率(Super-Resolution, SR)技术是一种可行的方案。图像超分辨率能够将图像从 LR 恢复到 HR,是低级图像处理任务^[4-5]中的一种。其在遥感图像处理^[6]、医学成像^[7]和安全监控^[8]等领域有着广泛的应用。

在图像 SR 领域中,目前的主流方法是基于卷积神经网络(Convolutional Neural Networks, CNN)搭建网络模型。CNN 能够实现端到端的训练,自动学习 LR 和 HR 之间的映射关系,并基于学习到的特征对 LR 图像进行 SR 重建。从最早的 SRCNN^[9](Super-Resolution Convolutional Neural Network)开始,基于 CNN 的 SR 模型所重建的图像,无论是在数值指标还是视觉表现方面,都展现出了优于传统方法^[10-11]的效果。此后,图像 SR 研究的重点是利用 CNN 网络从 LR 和 HR 之间的映射关系中学习更多有效特征。文献^[12-15]使用残差学习、密集连接等操作以堆叠更多的卷积层到网络中,并构建越来越深的网络以提升 SR 重建效果。在构建深层网络的同时,文献^[16-20]将注意力机制结合到图像 SR 中,以强化网络的学习能力。

虽然 CNN 的引入推进了图像 SR 的发展,但仍存在一些缺陷。首先,现有 CNN 工作通常使用低阶注意力^[16-17],难以学习到更精炼的特征,重建复杂图像时不能有效重构纹理细节。而在具有大量地面设施的低空间分辨率遥感图像中,建筑表面的复杂纹理存在耦合问题,需要超分辨率模型具备强大的纹理重建能力。目前,还没有工作对高阶注意力在图像超分辨率领域的表现进行探究。此外,卷积的局部感受野虽然可以有效学习局部特征,但是缺乏对远距离依赖关系的学习能力。文献^[21]通过多尺度大核注意力(Multi-Scale Large Kernel Attention, MLKA)来学习不同距离的依赖关系,并取得了一定效果,表明加入对远距离信息的学习有利于提升 SR 模型的性能。

而最近的计算机视觉上游工作中,文献^[22]所提出的递归门控 g^n Conv,为使用高阶注意力机制和学习远距离特征提供了一个可探索的解决方案。 g^n Conv(n 表示高阶交互的阶数)基于门控卷积和递归设计实现高阶空间交互,能够作为 SR 网络的组件从 LR 图像中提取有效的高阶特征。另外,在进行高阶空间交互之前, g^n Conv 使用深度卷积对输入特征进行处理,以捕捉远距离依赖关系。此外,文献^[16]提出的通道

注意力(Channel Attention, CA)能够加入通道层面的优化能力。

本文提出了一种基于递归门控卷积 g^n Conv 的遥感图像超分辨率网络 RGCSR(Recursive Gated Convolution Super-Resolution network, RGCSR)。首先,以 g^n Conv 为基础的高阶交互——前馈神经网络模块(HorBlock-FFN Block, HFB)是 RGCSR 的基本模块。通过对多个 HFB 串联并引入残差连接组成一个特征提取残差组(Residual Feature Extraction Group, RFEG),HFB 在残差组中可用于执行高阶空间交互。其次,为了强化对中间模块所提取的深度特征的提炼,构建了一个由通道注意力(CA)和 g^n Conv 组成的特征优化模块(Feature Optimization Block, FOB)。在 RFEG 和主干网络的末端放置 FOB,以增强辨别学习能力。实验结果表明,与其他基于 CNN 的方法相比, RGCSR 在自然图像和遥感图像的测试集上具有更好的性能。

综上所述,本文的主要工作和创新点如下:

(1)为解决目前基于 CNN 的 SR 方法难以学习到远距离知识的问题和探究高阶注意力在该领域的表现,在遥感图像 SR 领域引入了递归门控卷积 g^n Conv,利用递归门控卷积学习图像的远距离依赖关系和提取高阶特征。

(2)为了强化对中间模块所提取的深度特征的提炼,加强模型对通道层面的辨识能力,提出了一个由通道注意力(CA)和递归门控卷积构建的特征优化模块 FOB。

(3)提出了一种基于递归门控卷积的遥感图像超分辨率方法 RGCSR。在多个数据集上进行了广泛实验,结果表明 RGCSR 相比现有基于 CNN 的图像 SR 方法具有更好的重建结果。

2 相关工作

2.1 基于 CNN 的图像超分辨率网络

自 CNN 被广泛用于图像超分辨率研究以来,RCAN^[16]是最常见的 CNN 图像 SR 方法之一。该方法提出了通道注意力(CA)机制,其通过单次元素乘法进行一阶特征交互,实现了自适应调整通道特征。之后,Dai 等^[17]使用二阶通道注意模块(Second-Order Channel Attention module, SOCA)获取二阶特征分布,以此学习网络中间层的特征相关性。相比简单加深网络^[18],这些注意力模块所获取的特征能够恢复更多细节,在一定程度上提升了模型的表征能力。但在应对存在边缘耦合、复杂纹理部分缺失以及细节过于模糊等问题的低分辨率遥感图像时,低阶特征难以为图像重建提供更多正确信息。以上信息表明超分辨率模型需要具备高阶特征提取能力以应对低分辨率遥感图像的挑战,但目前未有研究尝试通过提取高阶特征来改进模型重建效果。

此外,下游图像任务的文献^[23]表明,强化 CNN 模型对远距离依赖关系的学习能力有助于性能的提升。在图像超分辨率领域,个别工作^[21]开始在其 CNN 模型中加入对远距离依赖的学习能力。Wang 等^[21]将多尺度机制和大核注意力(Large Kernel Attention, LKA)相结合,提出了 MLKA 来

同时关注局部信息和远程依赖,通过多个一阶注意力模块 LKA 来获取不同距离的依赖关系并取得了性能提升。该工作表明增强对远距离依赖关系的学习有助于提升模型性能。

现有方法^[16-18]缺乏对中间模块结果特征的重视,通常使用单一 3×3 卷积层进行简单整合,未充分利用中间网络知识。

综上所述,目前还没有图像超分辨率方法将高阶特征提取与远距离依赖学习相结合进行研究,且缺乏对中间特征的优化。因此,本文引入递归门控卷积实现对特征远距离依赖关系的学习,并通过高阶空间交互操作提取高阶特征。同时,本文基于通道注意力和递归门控卷积构造一个特征优化模块以优化中间模块的结果特征。

2.2 递归门控卷积

Transformer^[24]被引入计算机视觉领域后,在多个图像任务中均取得出色表现^[25-26]。Transformer 所使用的自注意力机制是一种二阶交互的注意力机制,能够学习远距离依赖关系。为了进一步提高模型效果,一些工作^[22-23,27-28]利用 Transformer 的优势改进 CNN。其中,文献[22]首次从高阶空间

相互作用的角度分析自注意在计算机视觉任务中的有效性,提出了基于 CNN 框架的有效实现——递归门控卷积 g^r Conv。通过门控卷积和递归设计执行高阶空间相互作用, g^r Conv 能够有效结合 Vision Transformer 和 CNN 的优点。Rao 等^[22]以 g^r Conv 为基础构建的 HorNet 在架构和配置相似的条件,在 ImageNet 分类等多个下游图像任务上取得了优于 Swin Transformer^[29]和 ConvNeXt^[23]的效果。

鉴于 g^r Conv 能够通过高阶交互提炼高阶特征和学习远距离依赖,且在下游图像工作上具有一定效果,本文将其应用于上游图像处理任务——遥感图像 SR 上,以探索具有学习远距离依赖能力的高阶特征提取模块在该任务上的有效性。

3 方法描述

3.1 主体结构

如图 1(a)所示,本文所提出的遥感图像超分辨率方法 RGCSR 由浅层特征提取、高阶空间交互和高质量图像重建 3 个阶段组成。

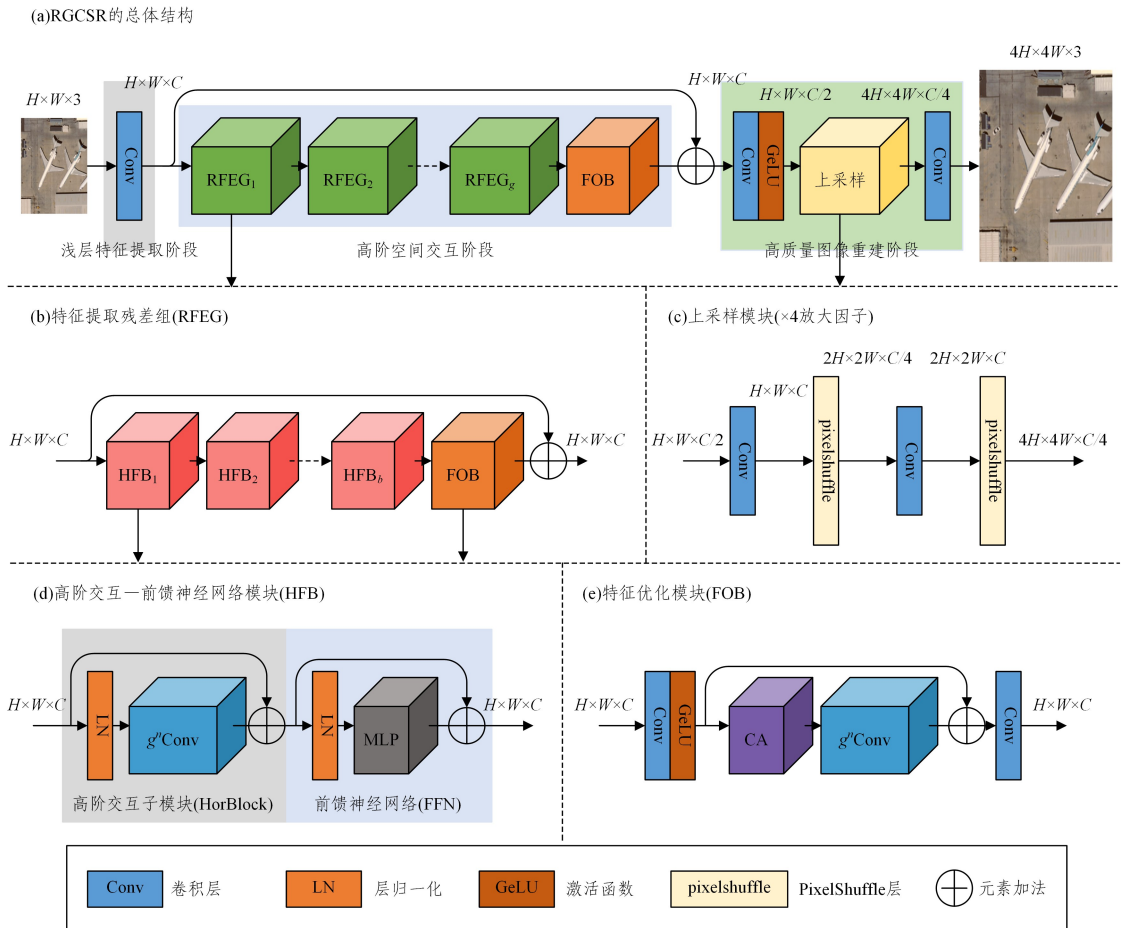


图 1 RGCSR 的总体架构和各模块结构

Fig. 1 Overall architecture of RGCSR and the structure of each module

浅层提取阶段使用一个 3×3 卷积层从低分辨率遥感图像中提取浅层特征。高阶空间交互阶段由 g 个特征提取残差组 (RFEG) 和一个特征优化模块 (FOB) 组成。RFEG 能够对特征进行深度提取,每个 REFG 有 b 个高阶

交互——前馈神经网络模块 (HFB) 和一个 FOB。FOB 能够对深度特征进行自适应调整和进一步的高阶空间交互。在高阶空间交互模块外部使用长跳转连接,以提高训练稳定性并绕过丰富的低频信息。高质量图像重建阶段首先

通过一个 3×3 卷积层进行特征降维,再通过一个上采样模块进行特征放大,最后通过一个 3×3 卷积层进行特征重构。

表 1 列出了当输入一张尺寸为 $H \times W \times 3$ 的 RGB 图像

时, RGCSR 的各阶段网络参数。Conv 表示卷积层,其内部参数分别表示卷积核大小和输出通道数。本章的后续小节将以图 1 和表 1 为基础,对递归门控卷积 $g^n \text{Conv}$ 和 RGCSR 各组成模块进行详细介绍。

表 1 RGCSR 的网络总体架构(高阶交互阶数 $n=4$)

Table 1 Overall network architecture of RGCSR(high-order interaction order $n=4$)

阶段	模块名	子模块	输出尺寸	网络参数设置
浅层特征提取	—	—	$H \times W \times C$	Conv($3 \times 3, C$)
				LayerNorm
		HFB _{1~b}	$H \times W \times C$	$g^4 \text{Conv}$: Proj-in; Conv($1 \times 1, 2C$), 深度卷积层, ⊗1, Proj-1; Conv($1 \times 1, C/4$), ⊗2, Proj-2; Conv($1 \times 1, C/2$), ⊗3, Proj-3; Conv($1 \times 1, C$), ⊗4, Proj-out; Conv($1 \times 1, C$)
				LayerNorm
				MLP: 全连接层, GeLU, 全连接层
				Conv($1 \times 1, C$), GeLU
	REFG _{1~g}			CA: 全局平均池化, Conv($1 \times 1, 4$), ReLU, Conv($1 \times 1, C$), Sigmoid
高阶空间交互		FOB	$H \times W \times C$	$g^4 \text{Conv}$: Proj-in; Conv($1 \times 1, 2C$), 深度卷积层, ⊗1, Proj-1; Conv($1 \times 1, C/4$), ⊗2, Proj-2; Conv($1 \times 1, C/2$), ⊗3, Proj-3; Conv($1 \times 1, C$), ⊗4, Proj-out; Conv($1 \times 1, C$)
				Conv($1 \times 1, C$)
				Conv($1 \times 1, C$), GeLU
				CA: 全局平均池化, Conv($1 \times 1, 4$), ReLU, Conv($1 \times 1, C$), Sigmoid
		FOB	$H \times W \times C$	$g^4 \text{Conv}$: Proj-in; Conv($1 \times 1, 2C$), 深度卷积层, ⊗1, Proj-1; Conv($1 \times 1, C/4$), ⊗2, Proj-2; Conv($1 \times 1, C/2$), ⊗3, Proj-3; Conv($1 \times 1, C$), ⊗4, Proj-out; Conv($1 \times 1, C$)
				Conv($1 \times 1, C$)
				Conv($3 \times 3, C/2$), GeLU
高质量 图像重建	上采样	—	$4H \times 4W \times C/4$	Conv($3 \times 3, C$), PixelShuffle(2), Conv($3 \times 3, C$), PixelShuffle(2)
	—	—	$4H \times 4W \times 3$	Conv($3 \times 3, 3$)

3.2 递归门控卷积

$g \text{Conv}$ 是 $g^n \text{Conv}$ ^[22] 的基础操作,它被视为一阶交互。给定输入特征 $X_{g \text{ in}} \in \mathbb{R}^{H \times W \times C}$, 此时 $g \text{Conv}$ 的结果特征 $Y_{g \text{ out}}$ 可由式(1)计算:

$$\begin{aligned} [P_0, Q_0] &= F_{\text{Proj-in}}(X_{g \text{ in}}) \\ P_1 &= DWConv(Q_0) \otimes P_0 \\ Y_{g \text{ out}} &= F_{\text{Proj-out}}(P_1) \end{aligned} \quad (1)$$

其中, $F_{\text{Proj-in}}(\cdot)$ 和 $F_{\text{Proj-out}}(\cdot)$ 分别表示 $g \text{Conv}$ 的输入线性投影层和输出线性投影层, $DWConv(\cdot)$ 表示深度卷积层操作。在 $g \text{Conv}$ 中, 输入特征 $X_{g \text{ in}}$ 被投影为 P_0 和 Q_0 ($P_0, Q_0 \in \mathbb{R}^{H \times W \times C}$)。深度卷积被用于捕捉远距离依赖关系, 并通过元素乘法实现相邻特征的单次相互作用。

文献[22]在 $g \text{Conv}$ 的基础上设计递归门控卷积 $g^n \text{Conv}$, 以引入高阶交互。为防止高阶交互引入过多计算开销, 参与每阶计算的通道数被设为 $C_k = \frac{C}{2^{n-k-1}}$ ($k=0, 1, \dots, n-1$),

较低阶交互使用的特征具有较少的通道数。给定输入特征 $X_{g \text{ in}} \in \mathbb{R}^{H \times W \times C}$, 此时 $g^n \text{Conv}$ 的结果特征 $Y_{g \text{ out}}$ 可由式(2)计算:

$$\begin{aligned} [P_0, Q_0, Q_1, \dots, Q_{n-1}] &= F_{\text{Proj-in}}(X_{g \text{ in}}) \\ P_1 &= DWConv(Q_0) \otimes P_0 / \alpha \\ P_{k+1} &= DWConv(Q_k) \otimes F_{\text{Proj-k}}(P_k) / \alpha, \\ & (Q_k \in Q_0, Q_1, \dots, Q_{n-1}) \end{aligned} \quad (2)$$

$$Y_{g \text{ out}} = F_{\text{Proj-out}}(P_n)$$

其中, $F_{\text{Proj-k}}(\cdot)$ 表示第 k ($k=1, 2, \dots, n-1$) 次阶乘后用于调整 P_k 通道数的线性投影层, 系数 $1/\alpha$ 用于缩放输出。

表 1 列出了 $g^4 \text{Conv}$ 的网络参数设置, 且图 2(a) 举例了输入特征 $X_{g \text{ in}}$ 的通道数 $C=192$ 时的 $g^4 \text{Conv}$ 网络结构。对于深度卷积层, 本文使用与文献[22]一致的实现, 使用单个深度卷积处理所有 Q_k ($k=0, 1, \dots, n-1$)。通过深度卷积获取远距离依赖关系和使用递归门控机制实现高阶交互, 能够更好地学习特征表示。

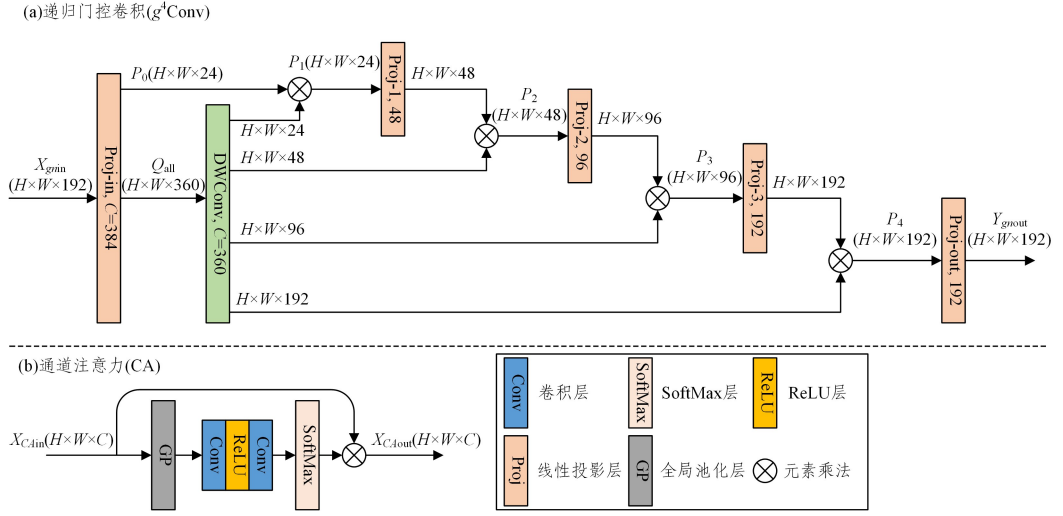


图2 各关键子模块的结构
Fig. 2 Structure of key sub-modules

为了令 $g^n \text{Conv}$ 捕获远距离依赖,Rao 等^[22]采用两种方式实现深度卷积。一种使用 7×7 深度卷积 ($g^n \text{Conv}_{7 \times 7}$),另一种使用全局滤波器层 (Global Filter, GF)^[30]的修改版本 ($g^n \text{Conv}_{\text{GF}}$,输入特征的一半通道由 GF 层处理,另一半由 3×3 深度卷积处理)。由于 $g^n \text{Conv}_{\text{GF}}$ 通过由 GF 层和 3×3 组成的深度卷积层能够同时考虑全局和局部信息,因此 RGCSR 以 $g^n \text{Conv}_{\text{GF}}$ 为基础构建。4.2.1 节将对这两种 $g^n \text{Conv}$ 在超分辨率任务上的表现进行验证。

3.3 高阶交互——前馈神经网络模块

为发挥 $g^n \text{Conv}$ 的性能,我们将 Transformer 风格的 $g^n \text{Conv}$ 基本构建块^[22] (HFB)作为 RFEG 的基本组成单元。HFB 由高阶交互子模块 (HorBlock) 和前馈神经网络 (Feed-Forward neural Network, FFN) 组成。HorBlock 是一个层归一化 (LayerNorm)—— $g^n \text{Conv}$ 结构,FFN 由层归一化 (LayerNorm) 和多层感知机 (Multi-Layer Perceptron, MLP) 组成。HFB 结构如图 1(d) 所示,第 i 个 RFEG 中的第 j 个 HFB 的计算过程如式 (3) 所示:

$$\begin{aligned} X_{\text{HorBlock}} &= g^n \text{Conv}(\text{LN}(X_{i,j-1})) + X_{i,j-1} \\ X_{i,j} &= \text{MLP}(\text{LN}(X_{\text{HorBlock}})) + X_{\text{HorBlock}} \end{aligned} \quad (3)$$

其中, X_{HorBlock} , $g^n \text{Conv}(\cdot)$, $\text{MLP}(\cdot)$ 和 $\text{LN}(\cdot)$ 分别表示当前 HFB 的 HorBlock 输出、递归门控卷积操作、多层感知机和层归一化操作; $X_{i,j-1}$ 和 $X_{i,j}$ 分别表示当前 HFB 的输入和输出。

3.4 特征优化模块

文献[16-18]在中间模块的末端放置单个 3×3 卷积以优化模块所提取的特征。然而,感受野较小的单一卷积层难以对中间模块提取的深度特征进行有效整合。为此,本文使用一个特征优化模块 (FOB) 替换 3×3 卷积,以优化中间模块提取的特征,该模块由 $g^n \text{Conv}$ 和通道注意力^[16] (CA) 组成。 $g^n \text{Conv}$ 能够引入更多高阶空间交互操作,但缺乏对通道层面依赖关系的关注。通道注意力 (CA) 则可以弥补这一缺陷,其能够学习通道间的依赖关系的能力,并以此调整通道特征,从而实现在通道层面优化特征提取。

特征优化模块 (FOB) 的结构如图 1(e) 所示。FOB 使用

残差连接来稳定训练,并在模块前后各加入一个 1×1 卷积,以调节 FOB 内部特征的维度。此时,第 i 个 RFEG 中的 FOB 的结果特征可由式 (4) 计算:

$$\begin{aligned} X_{\text{GeLU}} &= \text{GeLU}(\text{Conv}_{1 \times 1}(X_{i,b})) \\ X_{g^n \text{Conv}} &= g^n \text{Conv}(\text{CA}(X_{\text{GeLU}})) \\ X_{\text{FOB}} &= \text{Conv}_{1 \times 1}(X_{\text{GeLU}} + X_{g^n \text{Conv}}) \end{aligned} \quad (4)$$

其中, $\text{GeLU}(\cdot)$, $\text{Conv}_{1 \times 1}(\cdot)$ 和 $\text{CA}(\cdot)$ 分别表示 GeLU 激活函数、 1×1 卷积操作和通道注意力 (CA); X_{GeLU} 和 $X_{g^n \text{Conv}}$ 分别表示 GeLU 激活函数和 $g^n \text{Conv}$ 的输出特征; $X_{i,b}$ 和 X_{FOB} 分别表示当前 FOB 的输入和输出。如图 1(a) 和 (b) 所示,在 RGCSR 中,高阶空间交互模块和每个 RFEG 的末端均放置一个 FOB,以优化中间模块所提取的深度特征。

标准通道注意力^[16] (CA) 的结构见图 2(b)。通道注意力模块的计算式如式 (5) 所示 (设输入特征 $X_{\text{Ca}^{\text{in}}} \in \mathbb{R}^{H \times W \times C}$):

$$\begin{aligned} X_{\text{Conv1}} &= \text{Conv}_{1 \times 1}(F_{\text{GP}}(X_{\text{Ca}^{\text{in}}})) \\ X_{\text{Conv2}} &= \text{Conv}_{1 \times 1}(\text{ReLU}(X_{\text{Conv1}})) \\ Y_{\text{Ca}^{\text{out}}} &= \text{Sigmoid}(X_{\text{Conv2}}) \otimes X_{\text{Conv2}} \end{aligned} \quad (5)$$

其中, $F_{\text{GP}}(\cdot)$, $\text{ReLU}(\cdot)$, $\text{Sigmoid}(\cdot)$ 和 \otimes 分别表示全局池化层、ReLU 激活函数、Sigmoid 激活函数和元素乘法; X_{Conv1} 和 X_{Conv2} 表示两个 1×1 卷积操作的输出; $Y_{\text{Ca}^{\text{out}}}$ 表示通道注意力模块的输出。

3.5 上采样模块

上采样技术 PixelShuffle^[31] 能够将特征尺寸为 $H \times W \times C \cdot \text{factor}^2$ (factor 为放大因子) 的特征图上采样为 $H \cdot \text{factor} \times W \cdot \text{factor} \times C$,常用于图像超分辨率的上采样过程^[16,21]。

RGCSR 的上采样模块以 PixelShuffle 为主体构建。对于 $\times 2$ 和 $\times 3$ 放大因子,上采样模块由一个 3×3 卷积层和一个 PixelShuffle 层组成。卷积层先将通道数上调,再通过 PixelShuffle 层按倍数进行上采样放大。对于 $\times 4$ 放大因子,上采样模块由两个 3×3 卷积层和两个 PixelShuffle 层组成。如图 1(c) 所示, $\times 4$ 放大因子采用逐步上采样方式实现 4 倍放大 (依次进行两次 $\times 2$ 放大)。此外,所有放大因子的第一个卷积层仅调整为 $\text{factor}^2/2$ 倍,以减少上采样过程的计算

开销。详细参数设置见表 1 上采样部分。

3.6 损失函数

在经典工作^[16-17,20,32]中,L1 loss 是使用最广泛的损失函数。为验证 RGCSR 的有效性,本文同样使用 L1 loss 来训练 RGCSR。将 N 个 LR-HR 图像对作为训练集,则对 L1 loss 最小化以优化 RGCSR 的参数可由式(6)计算:

$$L(\Theta) = \frac{1}{N} \sum_{i=1}^N \| F_{\text{RGCSR}}(I_{\text{LR}}) - I_{\text{HR}} \|_1 \quad (6)$$

其中, Θ 是 RGCSR 的训练参数; $F_{\text{RGCSR}}(\cdot)$ 表示 RGCSR 超分辨率重建; I_{LR} 和 I_{HR} 分别表示在 N 个图像对中,第 i 对图像的 LR 和 HR 图像。

4 实验验证

4.1 实验设置

4.1.1 数据集说明

本文使用 DIV2K^[33] 的 800 张训练图像和 Flickr2K^[13] 的 2650 张图像共 3450 张图片作为训练集。DIV2K 和 Flickr2K 的图像取自多种不同场景,图像中的纹理细节清晰,有利于模型对纹理细节特征的学习。对于训练图像对的获取,本文通过对高分辨率(HR)图像进行双三次下采样(Bicubic)操作以获取对应低分辨率(LR)图像,从而得到成对的 HR-LR 图像。对于效果对比,RGCSR 将依次在自然图像基准数据集和遥感图像数据集上与基于 CNN 的方法进行

比较。自然图像基准数据集的对比工作有助于反映 RGCSR 的泛化能力,而遥感图像数据集上的对比则能有效验证 RGCSR 在遥感图像上的超分辨率重建有效性。

自然图像基准数据集包括 Set5^[34],Set14^[35],BSD100^[36],Urban100^[37]和 Manga109^[38]。其中,Urban100 包含 100 张城市景观图像,这些城市场景中存在大量复杂的建筑纹理和城市景观。Manga109 由 109 张漫画图像组成,这些漫画图像具有多种类型和风格。超分辨率模型在这两个数据集上的性能通常反映出自身细节特征挖掘能力的优劣。

用于测试的遥感图像从 5 个常见的遥感图像数据集获取:UCMerced LandUse data set^[39](简称 UCMerced),RSS-CN7^[40],AID^[41],WHU-RS19^[42]和 NWPU-RESISC45^[43]。具体地,从每个数据集的各类别中随机抽取 10% 的图片构成每个数据集对应的测试集,并依次编号。表 2 列出了各遥感图像数据集的基本信息和参与测试的图片数量。UCMerced 具有较高的空间分辨率,地面物体边缘耦合程度较低,类别较为丰富,易于观察模型对地面建筑边缘纹理的重建效果;RSS-CN7 的场景类别较少,但图像根据 4 种不同的比例采样得到;AID 的数据的获取条件多样,同一类别图像的多样性显著;WHU-RS19 中同类别图像来自不同分辨率的卫星图像;NWPU-RESISC45 具有多达 45 个类别的场景图像,图像的空间分辨率跨度大。

表 2 各遥感图像数据集的相关属性

Table 2 Relevant attributes of each remote sensing image dataset

数据集	类别数	图片数量	空间分辨率/m	尺寸	测试数量	测试编号
UCMerced	21	2100	0.3	256×256	210	Test-A
RSSCN7	7	2800	—	400×400	280	Test-B
AID	30	10000	0.5 to 8	600×600	1000	Test-C
WHU-RS19	19	1005	up to 0.5	600×600	96	Test-D
NWPU-RESISC45	45	31500	0.2 to 30	256×256	3150	Test-E

4.1.2 评估指标

根据图像超分辨率领域的通用设置,将峰值信噪比(Peak Signal-to-Noise Ratio,PSNR)和结构相似性指数度量(Structural Similarity Index Measure,SSIM)^[44]作为本文图像重建质量的评估指标。PSNR 可由式(7)计算得到:

$$\text{PSNR}(H,S) = 10 \cdot \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right) \quad (7)$$

其中, H 和 S 分别为原始高清图像和超分辨率重建图像;MAX 为图像像素最大值;MSE 为均方误差。PSNR 值越大,图像重建质量越好。SSIM 可由式(8)计算得到:

$$\text{SSIM}(H,S) = \frac{(2\mu_H\mu_S + C_1)(2\sigma_{HS} + C_2)}{(\mu_H^2 + \mu_S^2 + C_1)(\sigma_H^2 + \sigma_S^2 + C_2)} \quad (8)$$

其中, (μ_H, μ_S) 和 (σ_H, σ_S) 分别表示原始高清图像 H 和超分辨率重建图像 S 的平均值和标准差; σ_{HS} 为 H 和 S 的协方差; C_1 和 C_2 均表示正值常数。SSIM 值越大,图像纹理的重构效果越好。计算 PSNR 和 SSIM 时,RGB 图像先被转换到 YCbCr 空间的 Y 通道上进行评估。

此外,同样常用于遥感图像质量评估的视觉信息保真度(Visual Information Fidelity,VIF)^[45] 和学习感知图像块相似度(Learned Perceptual Image Patch Similarity,LPIPS)^[46] 被用于进一步评估各方法的遥感图像重建性能,辅助验证模型

对遥感图像超分辨率重建的有效性和优越性。

4.1.3 实验环境和参数设定

最终的 RGCSR 中,RFEG 数 g 、HFB 数 b 和中间特征通道数分别设置为 6,6 和 192。 g^n Conv 均为 g^n Conv_{GF} 版本,高阶交互阶数均为 4。

4.2 节展示了最终 RGCSR 参数设置的探索过程。所有 RGCSR 模型的训练均在 8 张 NVIDIA GeForce RTX 3090 GPU(单张 GPU 显存大小为 24GB)上使用 Pytorch1.8 框架完成。CPU 为 Intel(R) Xeon(R) Platinum 8338C。

在训练阶段,LR 图像被随机裁剪为 64×64 输入到网络中,批量大小为 32。并使用水平翻转和随机旋转 90°,180°,270°对 LR-HR 图像对进行增强。对于×2 放大因子的 RGCSR,学习率初始化为 2×10^{-4} ,总迭代次数为 7.5×10^5 。而×3 和×4 放大因子的 RGCSR,则使用×2 放大因子的结果模型来初始化模型参数,以缩短训练时间。学习率初始化为 1×10^{-4} ,总迭代次数为 3×10^5 。使用 Adam 优化器来训练模型, β_1 和 β_2 分别设置为 0.9 和 0.99。学习率在训练过程中使用多间隔策略进行减半处理,调整次数为 4 次。

4.2 消融实验

一个 RGCSR 的初始版本(见表 3(a))被用于逐步验证各组成模块的有效性,并探索最佳 RGCSR 设置。该 RGCSR 的

RFEG 数量 g 、HFB 数量 b 、通道数和高阶交互的阶数 n 被分别设置为 4, 4, 192 和 3, HFB 中不设置 HorBlock, 末端结构设置为 3×3 卷积。浅层特征提取阶段和浅层特征提取阶

段的设置与表 1 一致。使用 DIV2K 和 Flicker2K 数据集训练 $\times 2$ 放大因子的模型, 所有消融模型的训练迭代次数均为 2.5×10^5 , 并观察其在 Urban100 数据集上的表现。

表 3 模块结构的消融研究

Table 3 Ablation studies of modular structures

序号	高阶交互子模块(HorBlock)			主干和 REFG 的末端结构			参数量	计算量	PSNR
	无 HorBlock	$g^n \text{Conv}_{7 \times 7}$	$g^n \text{Conv}_{GF}$	3×3 卷积	$g^n \text{Conv}_{GF}$	FOB			
(a)	✓			✓			4.4×10^6	17.9×10^9	32.14
(b)		✓		✓			6.8×10^6	27.6×10^9	32.92
(c)			✓	✓			7.3×10^6	27.1×10^9	32.97
(d)			✓		✓		6.5×10^6	23.1×10^9	33.34
(e)			✓			✓	6.9×10^6	24.6×10^9	33.40

4.2.1 HorBlock 的消融实验

HFB 中的 HorBlock 是使用 $g^n \text{Conv}$ 进行高阶空间交互的基本单位。如表 3 所列, 与不使用 HorBlock 的实验(a)相比, 实验(b)和(c)使用 HorBlock 取得了 0.78~0.83 dB 的提升, 表明了 $g^n \text{Conv}$ 提取高阶特征的有效性。相对于初始实验(a), 实验(b)加入 $g^n \text{Conv}_{7 \times 7}$ 作为 HorBlock 后, 模型参数量和计算量大幅增加, 但模型性能的提升达到 0.78 dB。实验(c)使用 $g^n \text{Conv}_{GF}$ 替换 $g^n \text{Conv}_{7 \times 7}$ 后, 由于深度卷积的实现方式不同, 替换后参数量增加 0.5×10^6 、计算量减少 0.5×10^9 。实验(c)的性能优于实验(b), 表明与使用有限感受野的 7×7 深度卷积相比, GF 能够为后续高阶空间交互提供更有效的深度特征。此外, $g^n \text{Conv}_{GF}$ 对输入特征的另一半通道使用 3×3 深度卷积处理, 具备对局部细节的学习能力。因此, 后续实验在使用 $g^n \text{Conv}_{GF}$ 的实验(c)的基础上继续进行消融实验。

4.2.2 中间模块的末端结构的消融实验

为探究中间模块的末端结构对模型性能的影响, 本小节讨论在模块末端使用 $3 \times 3 \text{Conv}$, $g^n \text{Conv}_{GF}$ 和 FOB 时的模型表现。对比实验(c)和实验(d), 实验(d)使用 $g^n \text{Conv}_{GF}$ 替换 $3 \times 3 \text{Conv}$ 后取得了更优表现, 并且由于 $g^n \text{Conv}$ 对参与交互计算的特征通道数进行限制, 模型参数量和计算量分别降低了 0.8×10^6 和 0.4×10^9 。而在实验(d)的基础上, 引入通道注意力(CA)搭建的 FOB 在略微增加模型复杂度的情况下取得进一步提升, 由此证明了所提出的 FOB 的有效性。

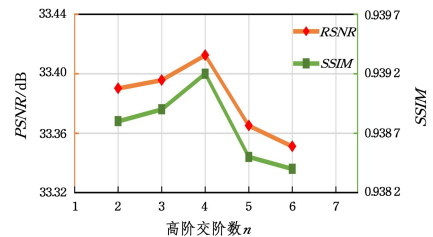
按实验(e)的设置, 此时 RGCSR 的参数量和计算量分别为 6.9×10^6 和 24.6×10^9 。相对于初始实验(a), 模型参数量增大 56.8% (2.5×10^6), 计算量仅增大 37.4% (6.7×10^9), 但 PSNR 的提升高达 1.26 dB。而以往 CNN 方法的参数量和计算量通常更大。以 RCAN^[16]($\times 2$ 放大因子)为例, 其模型参数量和计算量分别为 15.4×10^6 和 62.8×10^9 , 远大于当前 RGCSR 模型规模。以上结果表明 RGCSR 具备较大的模型规模扩展空间, 后续实验在实验(e)的基础上继续进行。

4.2.3 高阶交互的阶数 n 的消融实验

为探究高阶交互的阶数 n 对 SR 重建性能的影响和确定最优的阶数, 本小节讨论不同阶数 n 设置下的模型表现。

在实验(e)的基础上, 图 3 展示了不同阶数 n 时的模型表现, 此时 $C=192$ 。随着阶数递增, 模型性能在 $n=4$ 时达到最优, 此后出现下降趋势。这是 n 较大时, 参与空间交互的原始特征量过度减少所导致的。在式(2)中, P_0 被作为原始特征与通过深度卷积后的 Q_0 进行交互。 n 被设置为 2, 3 和 4 时, 对应通道数 C_0 分别为 96, 48 和 24。而 n 被设置为 5 或 6 时,

C_0 仅为 12 和 6, 此时被用于执行更高阶交互的原始特征较少, 加大了 $g^n \text{Conv}$ 提取高阶特征的难度。因此, 本文使用 $C=192$ 时表现最佳的 $n=4$ 作为最终高阶交互阶数设置。

图 3 对高阶交互阶数 n 的消融研究Fig. 3 Ablation study of order n for high-order interactions

4.2.4 模块数量的消融实验

本小节通过堆叠更多子模块来加深模型的深度, 以进一步证明所使用模块的有效性, 并确定最优模块设置。

表 4 列出了硬件条件满足的前提下, 采用不同数量的各模块构建 RGCSR 时的模型性能。可以发现, 当 RFEG 数量 g 以及每个 RFEG 中 HFB 数量 b 逐渐增长时, 模型性能取得稳定提升。当 $g=6, b=6$ 时, 模型性能达到最优值, 此时模型参数 (14.3×10^6) 仍少于同一尺度下的 RCAN (15.4×10^6)。以上结果表明, 堆叠更多子模块能够更好地学习特征, 进一步验证构成 RGCSR 的各模块的有效性。因此, 在最终的 RGCSR 模型中, RFEG 数量 g 和 HFB 数量 b 均设置为 6。

表 4 对模块数量的消融研究

Table 4 Ablation study for the number of modules

模块数量	参数量	PSNR
$g4b4$	7.0×10^6	33.41
$g4b6$	9.6×10^6	33.48
$g4b8$	12.3×10^6	33.56
$g6b4$	10.2×10^6	33.46
$g6b6$	14.3×10^6	33.61
$g8b4$	13.4×10^6	33.58

4.3 在自然图像基准测试集上的比较

为验证 RGCSR 的有效性, 首先在 5 个自然图像基准测试集上将 RGCSR 与其他方法进行对比。这些方法包括双线性插值(Bicubic)、EDSR^[13]、RCAN^[16]、NLSN^[47] 和 MAN^[21]。

(1) 定量比较

表 5 列出了在 $\times 2$ 、 $\times 3$ 和 $\times 4$ 放大因子下的定量结果, 黑色粗体数字和下划线数字分别表示最优和次优结果。其中, MAN 的指标数据引用自原始论文, EDSR, RCAN 和 NLSN 的指标数据使用 DIV2K 和 Flicker2K 重新训练得到。

表5 在自然图像基准数据集上与基于 CNN 方法的定量比较

Table 5 Quantitative comparison with CNN-based methods on natural image benchmark datasets

Method	Scale	Set5		Set14		BSD100		Urban100		Manga109	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	×2	33.66	0.9299	30.24	0.8688	29.56	0.8431	26.88	0.8403	30.80	0.9339
EDSR	×2	38.26	0.9616	34.01	0.9213	32.38	0.9022	32.97	0.9359	39.43	0.9790
RCAN	×2	38.33	0.9619	34.37	0.9242	32.44	0.9031	33.41	0.9392	39.80	0.9796
NLSN	×2	38.40	0.9620	34.27	0.9233	32.47	0.9033	33.51	0.9399	39.82	0.9797
MAN	×2	38.42	0.9622	34.40	0.9242	32.53	0.9043	33.73	0.9422	40.02	0.9801
RGCSR	×2	38.42	0.9624	34.54	0.9254	32.53	0.9043	33.86	0.9426	40.10	0.9803
Bicubic	×3	30.39	0.8682	27.55	0.7742	27.21	0.7385	24.46	0.7349	26.95	0.8556
EDSR	×3	34.77	0.9299	30.64	0.8478	29.34	0.8111	29.05	0.8703	34.59	0.9503
RCAN	×3	34.87	0.9308	30.74	0.8496	29.37	0.8119	29.23	0.8726	34.82	0.9515
NLSN	×3	34.87	0.9308	30.77	0.8497	29.38	0.8122	29.31	0.8733	34.95	0.9520
MAN	×3	34.91	0.9312	30.88	0.8514	29.43	0.8138	29.52	0.8782	35.06	0.9526
RGCSR	×3	34.98	0.9316	30.92	0.8523	29.45	0.8142	29.63	0.8790	35.15	0.9530
Bicubic	×4	28.42	0.8104	26.00	0.7027	25.96	0.6675	23.14	0.6577	24.89	0.7866
EDSR	×4	32.64	0.9003	28.95	0.7904	27.82	0.7443	26.92	0.8111	31.46	0.9177
RCAN	×4	32.68	0.9012	28.99	0.7912	27.83	0.7452	27.04	0.8136	31.70	0.9213
NLSN	×4	32.80	0.9022	29.06	0.7927	27.89	0.7471	27.27	0.8190	31.88	0.9221
MAN	×4	32.81	0.9024	29.07	0.7934	27.90	0.7477	27.26	0.8197	31.92	0.9230
RGCSR	×4	32.80	0.9018	29.08	0.7938	27.91	0.7480	27.35	0.8215	32.04	0.9243

与其他方法相比, RGCSR 在大多数情况下取得了最佳表现。特别是对于包含有大量复杂纹理信息的 Urban100 和 Manga109 测试集, RGCSR 均取得了优于其他方法的表现。且相比表现次优的 MAN, 随着放大因子的增大, RGCSR 在 Urban100 和 Manga109 上的提升也增大。以上结果表明 g^r Conv 能够有效挖掘高阶信息, 进而提升网络应对包含复杂信息的图像时的表示能力。此外, 与 RGCSR 模型参数量 (14.2×10^6) 最相近的 RCAN (15.6×10^6) 相比, RGCSR 在所有比较条件下均有提升, 在 ×2 放大因子的 Urban100 上的 PSNR 提升达到 0.45dB。以上定量比较结果表明了 RGCSR 的有效性。

(2) 视觉比较

图 4 展示了各方法在应对具有挑战性的区域时的表现, 这些区域均来自 ×4 放大因子的 Urban100 数据集。对于

图像 img_004, 大多数方法无法对后视镜中的网格区域进行有效重建; RCAN, NLSN 和 MAN 的重建结果存在大范围模糊区域; 而 RGCSR 能够对网格区域进行精细重建。对于图像 img_044, Bicubic 和 EDSR 以及 RCAN 无法对天花板横纹进行恢复, 仅能对图像的主要信息进行重构; NLSN 通过全局相似块恢复了部分横纹; MAN 仅产生了少量的横纹伪影, 而 RGCSR 能够对当前区域的所有横纹进行清晰的重建。对于图像 img_074, EDSR 和 RCAN 以及 NLSN 均能正确地恢复纵向的建筑条纹, 但不能正确地恢复横向条纹; 使用多尺度大核注意力的 MAN 能够对所有条纹进行重建, 但部分条纹存在明显的伪影; RGCSR 能够在准确重建条纹的基础上有效减少伪影的出现。视觉结果表明, RGCSR 具有强大的表示能力, 在正确还原图像整体结构的同时, 能够对纹理细节进行精细重构。

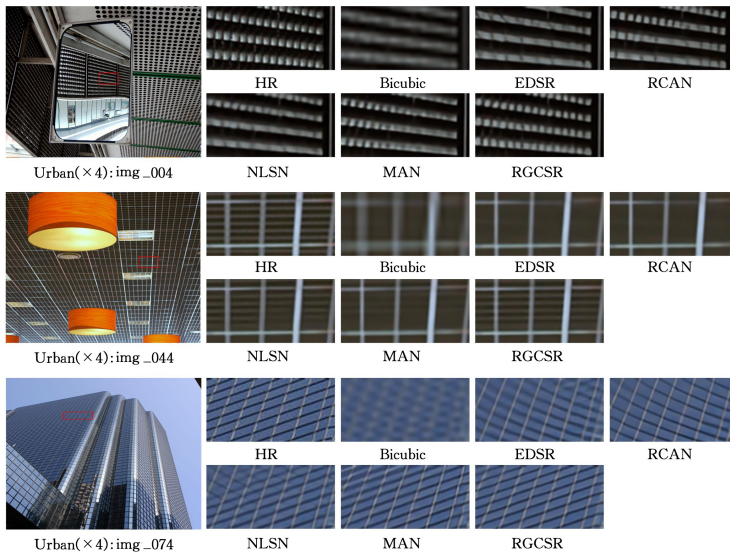


图4 在 Urban100 数据集上与其他图像 SR 方法的视觉比较(×4)

Fig. 4 Visual comparison with other image SR methods on Urban100(×4)

4.4 在常用遥感图像数据集上的比较

为验证 RGCSR 在遥感图像超分辨率重建上的有效性, 我们在 5 个遥感图像测试集上将 RGCSR 与双线性插值

(Bicubic)、EDSR^[13]、RCAN^[16]、NLSN^[47]、MAN^[21] 进行比较。这些方法的对比数据均使用由 DIV2K 和 Flicker2K 训练的模型测试得到。

(1) 定量比较

表 6 列出了遥感图像测试集上的 $\times 2$, $\times 3$ 和 $\times 4$ 放大因子的定量结果(黑色粗体数字和下划线数字分别表示最佳和次佳结果)。与其他方法相比, RGCSR 在所有放大因子的所有遥感测试集上均取得了最佳表现。特别是 RGCSR 在应对具有不同空间分辨率的遥感图像时具有强大的鲁棒性。例如, 对于空间分辨率跨度较大的测试集 Test-E, RGCSR 在 $\times 2$ 放大因子下的 PSNR 和 SSIM 提升比次优的 MAN 高 0.07 dB 和 0.0009。而对于空间分辨率

较高的 Test-A, RGCSR 同样优于其他方法。在应对具有不同空间分辨率图像的测试集 Test-B 时, RGCSR 仍展现出了最佳性能。在遥感图像测试集上的定量比较结果表明, RGCSR 在遥感图像超分辨率重建上表现优异, 同时也验证了在递归门控卷积 g^r Conv 在遥感图像超分辨率工作上的适用性。

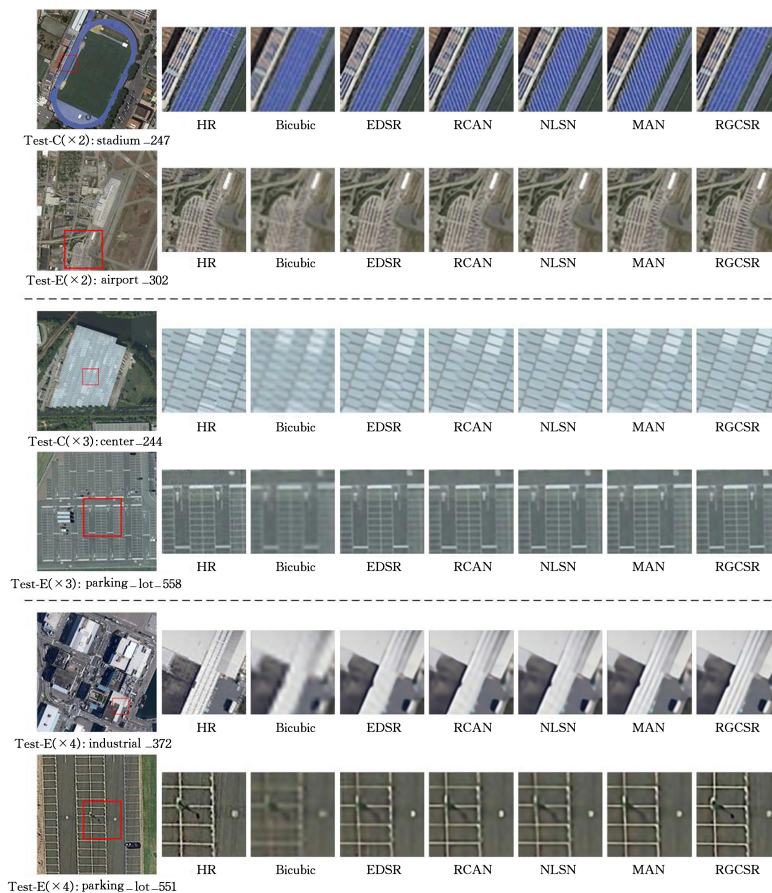
(2) 视觉比较

图 5 展示了不同方法在不同放大因子的遥感图像测试集 Test-C 和 Test-E 上的视觉结果。

表 6 在遥感图像测试集上与基于 CNN 方法的定量比较

Table 6 Quantitative comparison with CNN-based methods on remote sensing image test sets

Method	Scale	Test-A		Test-B		Test-C		Test-D		Test-E	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	$\times 2$	32.52	0.8917	31.82	0.8623	33.93	0.9026	35.28	0.9227	32.03	0.8791
EDSR	$\times 2$	36.50	0.9416	33.96	0.9110	36.99	0.9446	38.35	0.9584	34.67	0.9281
RCAN	$\times 2$	36.68	0.9427	34.05	0.9129	37.09	0.9454	38.48	0.9594	34.83	0.9302
NLSN	$\times 2$	36.71	<u>0.9432</u>	34.03	0.9126	37.07	0.9451	38.46	0.9591	34.74	0.9291
MAN	$\times 2$	36.75	0.9429	34.09	0.9137	37.14	0.9460	38.57	0.9599	34.84	0.9304
RGCSR	$\times 2$	36.77	0.9434	34.10	0.9140	37.16	0.9461	38.58	0.9601	34.91	0.9313
Bicubic	$\times 3$	29.06	0.7923	29.29	0.7586	30.56	0.8090	31.51	0.8282	29.11	0.7744
EDSR	$\times 3$	32.28	0.8746	30.82	0.8202	32.98	0.8742	33.89	0.8901	31.20	0.8463
RCAN	$\times 3$	32.39	0.8765	30.87	0.8222	33.05	0.8754	33.96	0.8916	31.25	0.8478
NLSN	$\times 3$	32.46	<u>0.8776</u>	30.87	0.8216	33.04	0.8749	33.96	0.8911	31.25	0.8476
MAN	$\times 3$	32.46	0.8773	30.89	<u>0.8232</u>	33.09	<u>0.8761</u>	34.01	0.8923	31.29	0.8488
RGCSR	$\times 3$	32.52	0.8790	30.91	0.8237	33.11	0.8765	34.04	0.8927	31.31	0.8494
Bicubic	$\times 4$	27.14	0.7104	27.96	0.6848	28.74	0.7348	29.50	0.7475	27.55	0.6946
EDSR	$\times 4$	29.83	0.8051	29.26	0.7496	30.74	0.8089	31.49	0.8225	29.29	0.7735
RCAN	$\times 4$	29.95	0.8092	29.29	0.7511	30.80	0.8105	31.56	0.8245	29.34	0.7757
NLSN	$\times 4$	30.04	<u>0.8121</u>	29.32	0.7528	30.84	0.8117	31.59	0.8252	29.37	0.7770
MAN	$\times 4$	30.00	0.8100	29.33	<u>0.7531</u>	30.86	<u>0.8121</u>	<u>31.61</u>	0.8258	<u>29.38</u>	<u>0.7773</u>
RGCSR	$\times 4$	30.06	0.8125	29.33	0.7532	30.88	0.8126	31.64	0.8262	29.39	0.7774

图 5 在遥感图像测试集 Test-C 和 Test-E 上与其他图像 SR 方法的视觉比较($\times 4$)Fig. 5 Visual comparison with other image SR methods on remote sensing image test sets Test-C and Test-E($\times 4$)

从视觉结果可知, RGCSR 在所有尺度下均取得了比其他方法更好的重建效果。可以发现, 其他方法在重建建筑表面和露天设施的纹理时, 存在无法重建纹理或者是重建错误纹理的现象。而 RGCSR 能够对操场跑道、建筑物外墙和停车场场地线的纹理进行正确重建。此外, 随着重建尺度增大, LR 图像所提供的有效信息变少, 其他方法的重建结果出现了更多模糊区域。特别是在 EDSR, RCAN, NLSN 和 MAN 的 $\times 4$ 放大因子下的重建结果中, 建筑物顶端线条没有被有效重建, 且一些小型地面物体和周边环境出现了边缘耦合现象。RGCSR 能够对建筑物顶端线条进行有效恢复, 小型地面物体也能够进行清晰重建, 证明了所提出的 RGCSR 在遥感图像超分辨率重建上的可行性和先进性。

(3) 基于其他指标的评估

表 7 列出了各方法在遥感测试集 Test-A 和 Test-B 上的 VIF 和 LPIPS 评估结果, 以辅助验证 RGCSR 在遥感图像超分辨率重建上的有效性。Test-A(取自 UCMerced)分辨率较高, 用于考验各方法对建筑纹理的重构能力。Test-B(取自 RSSCN7)的遥感图像采样自不同比例, 用于考验各方法应对不同分辨率图像的鲁棒性。在 Test-A 和 Test-B 上的评估结果能够较为全面地反映模型能力。

表 7 在 Test-A 与 Test-B 上与基于 CNN 方法的定量比较

Table 7 Quantitative comparison with CNN-based methods on Test-A and Test-B

Method	Scale	Test-A		Test-B	
		VIF \uparrow	LPIPS \downarrow	VIF \uparrow	LPIPS \downarrow
Bicubic	$\times 2$	0.6106	0.1676	0.5634	0.1959
EDSR	$\times 2$	0.7964	0.0945	0.7109	0.1429
RCAN	$\times 2$	0.8013	0.0921	0.7154	0.1391
NLSN	$\times 2$	0.8028	0.0902	0.7147	0.1403
MAN	$\times 2$	0.8038	0.0899	0.7184	0.1358
RGCSR	$\times 2$	0.8043	0.0903	0.7182	0.1358
Bicubic	$\times 3$	0.3801	0.2807	0.3400	0.3073
EDSR	$\times 3$	0.5541	0.1965	0.4558	0.2582
RCAN	$\times 3$	0.5589	0.1926	0.4589	0.2552
NLSN	$\times 3$	0.5612	0.1915	0.4585	0.2552
MAN	$\times 3$	0.5620	0.1902	0.4617	0.2518
RGCSR	$\times 3$	0.5642	0.1898	0.4622	0.2523
Bicubic	$\times 4$	0.2497	0.3642	0.2200	0.3814
EDSR	$\times 4$	0.3926	0.2823	0.3093	0.3325
RCAN	$\times 4$	0.3986	0.2763	0.3112	0.3305
NLSN	$\times 4$	0.4025	0.2714	0.3133	0.3277
MAN	$\times 4$	0.4014	0.2739	0.3144	0.3262
RGCSR	$\times 4$	0.4037	0.2717	0.3144	0.3254

与其他方法相比, RGCSR 在绝大多数情况下取得了最优 VIF, 表明 RGCSR 重建的遥感图像具有最优图像质量。对于 LPIPS, RGCSR 同样能够取得最优或次优结果, 表明 RGCSR 重建的遥感图像具有较高图像相似度。以上结果进一步验证了 RGCSR 在遥感图像超分辨率重建上的有效性和优越性。

4.5 模型尺寸与计算量对比

表 8 列出了 RGCSR 与其他方法在模型参数数量和计算量上的对比结果。这些方法中, RGCSR 的 PSNR 和 SSIM 均为最优值。与参数数量相近的 RCAN 相比, RGCSR 在计算量略多的前提下, SSIM 提升了 0.0033; 对于具有大规模参数且需要计算量较大的 EDSR 和 NLSN, RGCSR 在参数数量和计算量更少的前提下取得了最佳表现; 虽然 MAN 的参数数量和计算

量略少于 RGCSR, 但 RGCSR 在 SSIM 上的性能提升要比 MAN 高 0.0025。以上结果表明, RGCSR 在具有适当参数规模的前提下实现了最好的重建效果, 即在模型性能、参数数量和计算量之间取得了更好的平衡。

表 8 在 Test-A 上的参数数量和计算量对比($\times 4$)

Table 8 Comparison of params and FLOPs on Test-A($\times 4$)

方法	参数数量	计算量	Test-A	
			PSNR	SSIM
EDSR	43.1×10^6	206.0×10^9	29.83	0.8051
RCAN	15.6×10^6	65.3×10^9	29.95	0.8092
NLSN	44.2×10^6	221.8×10^9	30.04	0.8121
MAN	8.7×10^6	35.2×10^9	30.00	0.8100
RGCSR	14.3×10^6	51.7×10^9	30.06	0.8125

结束语 为进一步提升遥感图像超分辨率重建的效果, 引入并探索了递归门控卷积 g^n Conv 在遥感图像超分辨率领域的表现, 提出了一种基于递归门控卷积的遥感图像超分辨率方法 RGCSR。首先, 使用高阶交互—前馈神经网络模块 (HFB) 来提取高阶特征。HFB 模块中的 g^n Conv 能够兼顾全局和局部信息, 利用高阶空间交互提取高阶特征。其次, 对各个中间模块的输出特征进行优化, 提出由通道注意力 (CA) 和 g^n Conv 构建的特征优化模块 (FOB)。最后, 在多个自然图像和遥感图像数据集上的对比结果表明, RGCSR 具备比现有基于 CNN 的方法更好的重建性能表现。下一步的研究工作包括探究模型的轻量化和多尺度放大等。

参考文献

- [1] SONG W, GU H H, SONG W, et al. Environmental assessments in dense mining areas using remote sensing information over qian'an and qianxi regions china[J]. Ecological Indicators, 2023, 146: 1-13.
- [2] GARCÍA-PARDO K A, MORENO-RANGEL D, DOMÍNGUEZ-AMARILLO S, et al. Remote sensing for the assessment of ecosystem services provided by urban vegetation. a review of the methods applied[J]. Urban Forestry & Urban Greening, 2022, 74: 1-12.
- [3] XING Z, YANG S, ZAN X, et al. Flood vulnerability assessment of urban buildings based on integrating high-resolution remote sensing and street view images[J]. Sustainable Cities and Society, 2023, 92: 1-14.
- [4] CHENG L B, LI X Y, LI Z, et al. Remote sensing image denoising method based on curvelet transform and goodness-of-fit test [J]. Journal of Jilin University(Engineering and Technology Edition), 2023, 53(11): 3207-3213.
- [5] WU Y Y, WANG Y J, XIE Y Y. The panchromatic sharpening method based on multi-scale residual attention network [J]. Journal of Chongqing University of Technology(Natural Science), 2023, 37(8): 222-230.
- [6] ZHANG D Y, SHAO J, LI X Y, et al. Remote sensing image super-resolution via mixed high-order attention network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(6): 5183-5196.
- [7] ISAAC J S, KULKARNI R. Super resolution techniques for medical image processing [C] // International Conference on

- Technologies for Sustainable Development. Mumbai; IEEE, 2015:1-6.
- [8] RASTI P,UIBOUPIN T,ESCALERA S,et al. Convolutional neural network super resolution for face recognition in surveillance monitoring[C]// Articulated Motion and Deformable Objects;9th International Conference. Palma de Mallorca; Springer,2016:175-184.
- [9] DONG C,LOY C C,HE K M,et al. Learning a deep convolutional network for image super-resolution[C]// Proceedings of the European Conference on Computer Vision(ECCV). Mumbai;Springer,2014:184-199.
- [10] DUCHON C E. Lanczos Filtering in One and Two Dimensions [J]. Journal of Applied Meteorology and Climatology, 1979, 18(8):1016-1022.
- [11] SUN J,XU Z B,SHUM H Y,et al. Image super-resolution using gradient profile prior[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Anchorage;IEEE,2008:1-8.
- [12] KIM J,LEE J K,LEE K M. Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Anchorage;IEEE,2016:1646-1654.
- [13] LIM B,SON S,KIM H,et al. Enhanced deep residual networks for single image super-resolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Honolulu;IEEE,2017:136-144.
- [14] HARIS M,SHAKHNAROVICH G,UKITA N. Deep back-projection networks for super-resolution[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City;IEEE,2018:1664-1673.
- [15] ZHANG Y L,TIAN Y P,KONG Y,et al. Residual dense network for image super-resolution[C]// In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City;IEEE,2018:2473-2481.
- [16] ZHANG Y L,LI K P,LI K,et al. Image super-resolution using very deep residual channel attention networks[C]//Proceedings of the European Conference on Computer Vision(ECCV). Munich;Springer,2018:286-301.
- [17] DAI T,CAI J R,ZHANG Y B,et al. Second-order attention network for single image super-resolution[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach;IEEE,2019:11065-11074.
- [18] NIU B,WEN W L,REN W Q,et al. Single image super-resolution via a holistic attention network[C]//Proceedings of the European Conference on Computer Vision (ECCV). Glasgow; Springer,2020:191-207.
- [19] LI Q,ZHAO F Y. Super-resolution of Remote Sensing Images Based on Multi-path and Mixed Attention[J]. Journal of Chinese Computer Systems,2023,44(7):1508-1513.
- [20] MEI Y Q,FAN Y C,ZHOU Y Q,et al. Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2020; 5690-5699.
- [21] WANG Y,LI Y S,WANG G,et al. Multi-scale attention network for single image super-resolution[EB/OL]. (2022-9-29) [2023-07-23]. <https://arxiv.org/abs/2209.14145>.
- [22] RAO Y M,ZHAO W L,TANG Y S,et al. HorNet: Efficient High-Order Spatial Interactions with Recursive Gated Convolutions[EB/OL]. (2022-10-11) [2023-07-23]. <https://arxiv.org/abs/2207.14284>.
- [23] LIU Z,MAO H Z,WU C Y,et al. A convnet for the 2020s [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans; IEEE, 2022; 11976-11986.
- [24] VASWANI A,SHAZEER N,PARMAR N,et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach; MIT Press,2017:6000-6010.
- [25] DOSOVITSKIY A,BEYER L,KOLESNIKOV A,et al. An image is worth 16×16 words: Transformers for image recognition at scale [EB/OL]. (2021-06-03) [2023-07-23]. <https://arxiv.org/abs/2010.11929>.
- [26] RADFORD A,KIM J W,HALLACY C,et al. Learning Transferable Visual Models From Natural Language Supervision[EB/OL]. (2021-02-26) [2023-07-23]. <https://arxiv.org/abs/2103.00020v1>.
- [27] LI J S,XIA X,LI W,et al. Nextvit:Next generation vision transformer for efficient deployment in realistic industrial scenarios[EB/OL]. (2022-08-16) [2023-07-23]. <https://arxiv.org/abs/2207.05501>.
- [28] GUO J Y,HAN K,WU H,et al. Cmt:Convolutional neural networks meet vision transformers[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans;IEEE,2022:12175-12185.
- [29] LIU Z,LIN Y T,CAO Y,et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE,2021:10012-10022.
- [30] RAO Y M,ZHAO W L,ZHU Z,et al. Global filter networks for image classification[C]// Proceedings of the 33rd International Conference on Neural Information Processing Systems. MIT Press,2021:980-993.
- [31] SHI W Z,CABALLERO J,HUSZÁR F,et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Las Vegas;IEEE,2016:1874-1883.
- [32] ZHOU S C,ZHANG J W,ZUO W M,et al. Cross-scale internal graph neural network for image super-resolution[C]// Proceedings of the Thirty-fourth International Conference on Neural Information Processing Systems. MIT Press,2020:3499-3509.
- [33] AGUSTSSON E,TIMOFTE R. Ntire 2017 challenge on single image super-resolution: Dataset and study[C]// Proceedings of

- the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Honolulu:IEEE,2017:126-135.
- [34] BEVILACQUA M, ROUMY A, GUILLEMOT C, et al. Low-complexity singleimage super-resolution based on nonnegative neighbor embedding[C]// British Machine Vision Conference. Surrey:British Machine Vision Association,2012:135. 1-135. 10.
- [35] ZEYDE R, ELAD M, PROTTER M. On single image scale-up using sparse-representations[C]// Curves and Surfaces; 7th International Conference. Avignon;Springer,2010:711-730.
- [36] MARTIN D, FOWLKES C, TAL D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics[C]// Proceedings Eighth IEEE International Conference on Computer Vision. Vancouver:IEEE,2001:416-423.
- [37] HUANG J B, SINGH A, AHUJA N. Single image super-resolution from transformed self-exemplars[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Boston:IEEE,2015:5197-5206.
- [38] MATSUI Y, ITO K, ARAMAKI Y, et al. Sketch-based manga retrieval using manga109 dataset[J]. Multimedia Tools and Applications,2017,76:21811-21838.
- [39] 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. New York:ACM,2010:270-279.
- [40] ZOU Q, NI L Y, ZHANG T, et al. Deep learning based feature selection for remote sensing scene classification[J]. IEEE Geoscience and Remote Sensing Letters,2015,12(11):2321-2325.
- [41] 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.
- [42] DAI D X, YANG W. Satellite image classification via two-layer sparse coding with biased image representation[J]. IEEE Geoscience and Remote Sensing Letters,2010,8(1):173-176.
- [43] CHENG G, HAN J W, LU X Q, et al. Remote sensing image scene classification: Benchmark and state of the art[J]. Proceedings of the IEEE,2017,105(10):1865-1883.
- [44] 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):1-14.
- [45] SHEIKH H R, BOVIK A C. Image information and visual quality[J]. IEEE Transactions on Image Processing,2006,15(2):430-444.
- [46] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City:IEEE,2018:586-595.
- [47] MEI Y Q, FAN Y C, ZHOU Y Q. Image super-resolution with non-local sparse attention[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE,2021:3517-3526.



LIU Changxin, born in 1998, postgraduate. His main research interests include image processing and deep learning.



HU Lirui, born in 1966, Ph.D, professor, master supervisor. His main research interests include embedded system, image recognition, and machine learning.

(责任编辑:何杨)