

采用感受野优化与渐进特征融合的 图像超分辨率算法

吴洪伍^{1,2}, 盖绍彦^{1,2}, 达飞鹏^{1,2}

(1. 东南大学自动化学院, 210096, 南京; 2. 东南大学复杂工程系统测量与控制教育部重点实验室, 210096, 南京)

摘要: 针对现有基于深度学习的超分辨率方法存在参数冗余以及难以学习到全局上下文信息、重建图像高频特征能力欠佳的问题,提出一种基于感受野优化与渐进特征融合的超分辨率网络(RPSRnet),其在单幅图像重建方面实现了较高的性能。该网络采用像素注意力机制与大感受野卷积相结合的方式,设计两条渐进路径将输入表征为不同层次的特征抽象,增强网络捕获上下文信息的能力,同时减少了网络参数冗余。通过分层卷积和多重感受野分支,在保持轻量卷积的前提下,于分层的多路径上学习不同尺度的融合特征,增强网络重建边缘细节和复杂纹理特征的能力。实验结果表明:相比于先进算法,所提算法在基准测试集 Set5 上的峰值信噪比达到 32.47 dB,在测试集 Set14 上达到 28.81 dB,优于现有的先进算法,且网络参数更少,实现了 9% 的参数缩减,从而验证了算法的有效性。

关键词: 超分辨率; 注意力机制; 感受野优化; 特征融合

中图分类号: TP391 **文献标志码:** A

DOI: 10.7652/xjtub202501013 **文章编号:** 0253-987X(2025)01-0136-12

Image Super Resolution Algorithm Based on Receptive Field Optimization and Progressive Feature Fusion

WU Hongwu^{1,2}, GAI Shaoyan^{1,2}, DA Feipeng^{1,2}

(1. School of Automation, Southeast University, Nanjing 210096, China; 2. Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Southeast University, Nanjing 210096, China)

Abstract: To tackle the issue of parameter redundancy and the challenge of learning comprehensive global contextual information in current deep learning-based super-resolution methods, as well as the limited capacity to reconstruct high-frequency image features, a super-resolution network (RPSRnet) based on receptive field optimization and progressive feature fusion is proposed, delivering exceptional performance in single image reconstruction. By integrating pixel attention mechanisms and large receptive field convolutions, two progressive pathways are devised to interpret inputs as features at varying levels of abstraction. This design enriches the network's capability to grasp contextual information while streamlining parameter redundancy. Through the utilization of hierarchical convolutions and multiple receptive field branches, the network sustains lightweight convolutions and acquires fused features from diverse scales on hierarchical pathways. This process enhances the network's proficiency in reconstructing edge details and intricate texture

收稿日期: 2024-04-07。 作者简介: 吴洪伍(1998—),男,硕士生;盖绍彦(1979—),男,副教授,博士生导师。 基金

项目: 江苏省前沿引领技术基础研究专项资助项目(BK20192004C)。

网络出版时间: 2024-07-18

网络出版地址: <https://link.cnki.net/urlid/61.1069.T.20240717.1734.004>

features. The experimental results reveal that the proposed algorithm achieves a peak signal-to-noise ratio of 32.47 dB on the Set5 benchmark test set and 28.81 dB on the Set14 test set. The algorithm surpasses existing advanced algorithms, utilizing fewer parameters and resulting in a 9% reduction in parameters, effectively validating the algorithm's efficacy.

Keywords: image super resolution; attention mechanism; receptive field optimization; feature fusion

单图像的超分辨率是一项经典的计算机视觉任务,目标是从低分辨率图像中重建出视觉质量更高的高分辨率图像。它在公共安全、目标检测和医疗诊断等图像分析领域有着广泛的应用^[1-4]。通常,可以对高分辨率图像添加退化类型以获得相应的低分辨率图像。然而,这一过程是不可逆的,高分辨率图像与低分辨率图像之间并非一一对应关系。现有的超分辨率方法从低分辨率图像中恢复对应的高分辨率图像时仍然受到阻碍,因此,图像超分辨率技术具有一定的研究价值。传统的超分辨率方法包括机器学习和插值算法。

与传统算法相比,卷积神经网络在单图像超分辨率视觉任务中表现出优良的性能,归因于卷积网络拟合复杂非线性函数的强大能力^[5-7]。Dong 等^[8]提出的超分辨率卷积神经网络(SRCNN)是卷积神经网络在图像超分辨率领域的首个应用。该网络先通过插值算法将输入的低分辨率图像上采样至目标分辨率,再利用三层卷积提取特征图以提升图像重建质量。作为改进,加速超分辨率卷积神经网络(FSRCNN)^[9]将特征图的上采样操作移至网络末端而非网络入口,并采用转置卷积调整特征图的大小,以此降低计算开销。为提取更深层的高频特征,增加卷积层数的方法得到广泛采用。Kim 等^[10]利用残差连接构建了一个含 20 个卷积层的深度网络。尽管众多研究结果表明,增加网络深度能够提升性能,然而深层网络和类似递归结构会引入大量参数和庞大的计算需求,这促使研究人员去设计更为轻量化的网络。

近年来,针对便携设备以及需要实时反馈结果的场景,提出了许多轻量级的算法。这些算法的特点是轻量、高效。级联残差网络(CARN)^[11]通过在残差结构中采用级联的方式实现轻量化。一些方法^[12-13]通过改进多重蒸馏网络在性能和参数平衡方面取得了更优的效果。然而在这些基于卷积神经网络的方法中,仅是捕捉单一尺度内的高频细节。此外,这些方法很少利用大的感受野来捕捉上下文内容,单一尺度特征提取或仅使用堆叠的注意力块

会导致在恢复高频图像细节方面表现不佳。

为了解决上述问题,本文提出了一种基于感受野优化与渐进特征融合的图像超分辨率网络(RPSRnet),该网络能够有效地跨不同尺度提取丰富的上下文信息。所提出的 RPSRnet 网络中,设计了一个局部多尺度特征提取模块(LMFE),增强网络的多尺度特征提取。基于注意力机制和感受野优化,提出了一个渐进式的像素注意力模块(PPAB),通过渐进式的分层提取,结合感受野优化有效地提升网络对上下文信息的理解,加强对高频特征的捕获能力。与先前的方法相比,本文提出的算法主要有以下改进:①设计了渐进式的像素注意力模块,基于感受野优化和渐进式的特征提取提升网络对特征长程关系的理解能力,更有效地捕获图像的高频细节;②提出的局部多尺度特征提取块,能够从多路径获取不同感受野范围的特征信息,增强网络对图像更大范围特征的提取能力,提升网络的性能;③在性能和参数方面,所提网络以更少的参数实现了最优的性能。在对 4 个基准数据集进行评估的结果中,RPSRnet 获得了相比于先进算法更优的性能,涵盖峰值信噪比、结构相似性、网络尺寸和主观视觉质量。

1 超分辨率相关研究

1.1 基于卷积神经网络和 Transformer 的超分辨率

卷积神经网络(CNN)在各类计算机视觉任务中已取得显著进展^[14-15],近期许多方法尝试构建深层卷积网络以提升重建性能。文献[10]中的甚深超分辨率神经网络(VDSR)采用 20 层卷积来提取深层高频特征。文献[16]中的增强超分辨率神经网络(EDSR)通过引入残差缩放并改进深度残差模块,提出一个包含 32 个残差模块的深度网络。为使超分辨率网络实现轻量化,信息蒸馏神经网络(IMDN)^[17]提出了信息多重蒸馏方案,结合通道拆分以减少参数的使用。特征蒸馏神经网络(RFDN)^[12]改进了多重蒸馏,提出了特征蒸馏连接方案,特征残差蒸馏神经网络(RLFN)^[13]在特征蒸

馏基础上提出了残差特征蒸馏,性能和参数得到进一步优化。晶格神经网络(LatticeNet)^[18]利用格子块和反向特征融合来平衡性能和参数。蓝图可分离神经网络(BSRN)^[19]引入了蓝图可分离卷积,分离逐通道和逐点卷积以提高参数利用率。

Liang 等^[20]在超分辨率任务中引入视觉滑动窗口变换器(Swin Transformer),借助移位窗口机制进行长距离依赖建模以提升性能。Lu 等^[21]将 CNN 与 Transformer 相结合,提出一种高效的超分辨率变换器,利用图像内部的自相似性建立长程依赖关系。Chen 等^[22]在连续的 Transformer 块中交替运用空间和通道自注意力,聚合块间特征以提升高频细节的提取能力。Wang 等^[23]对视觉 Transformer 进行改进,提出了多维聚合网络以及同质聚合方案,能够在不同维度上聚合多层次信息。Chen 等^[24]改进了基于窗口的自注意力并结合通道注意力,引入交叉注意力模块增强窗口特征间的交互。

1.2 基于感受野优化的卷积方法

近年来,Transformer 在视觉任务中表现卓越,其出色的性能得益于 Transformer 所提供的广阔感受野。扩大感受野的一种替代方法是采用更大的卷积核,这个特性启发了研究人员利用大卷积核设计神经网络。新一代卷积神经网络(ConvNeXt)^[25]通过运用大卷积核来提升更广阔的感受野,从而达到

了与基于 Transformer 的方法具有相同竞争力的性能。Hesse 等^[26]设计了新的自适应下采样策略,通过增加感受野来提升采样的性能。视觉注意力网络(VAN)^[27]探究了大卷积核注意力在视觉任务中的性能,提出了大卷积核注意力模块。大感受野神经网络(VapSR)^[28]拓展了此项工作,提出了一个具有大感受野的像素注意力模块。基于此理论基础,本文提出了一个基于感受野优化的渐进式注意力模块。

2 RPSRnet 网络结构设计

在本节中,首先阐述了如图 1 所示的 RPSRnet 网络结构,接着详细阐释网络内部的细节。

2.1 主体网络结构

本文提出的 RPSRnet 由 3 个主要部分组成:浅层特征提取、基于渐进局部多尺度注意力块(PLMA)的深层特征提取以及图像重建层。由图 1 可以看出,初始阶段将输入的低分辨率图像 I_{LR} 转换为浅层特征,以便为后续的深层特征提取作准备。然后,提取出的浅层特征被输送到堆叠的 PLMA 中,用于提取并融合更深层次的特征。最后,利用像素清洗层将图像大小调整至预定分辨率并生成重建图像。设计的 PLMA 包括 4 个部分:通道注意力模块(CAB)、局部多尺度特征提取模块、渐进式的像素注意力模块和增强型空间注意力模块(ESA)。

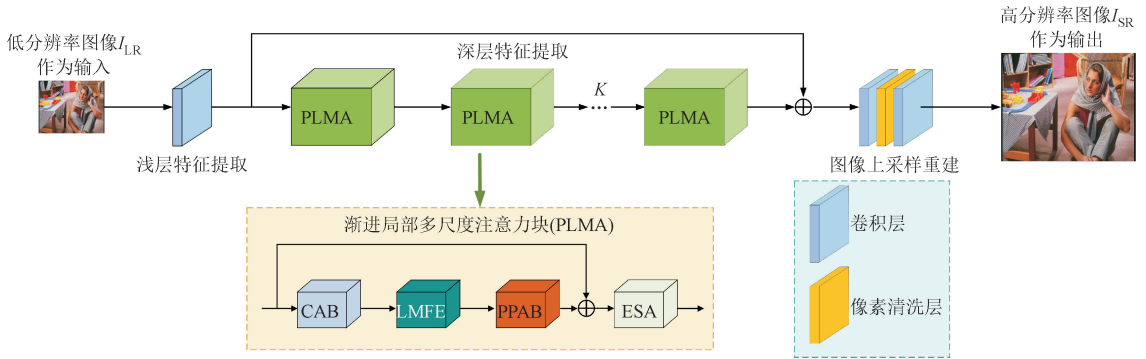


图 1 RPSRnet 的网络结构图

Fig. 1 Architecture of the RPSRnet framework

下面对 RPSRnet 的工作流程进行解释。首先,在浅层特征提取阶段,输入图像 I_{LR} 经过卷积层被提取为高维度的浅层特征,如下所示

$$X_0 = f_{3 \times 3}(I_{LR}) \quad (1)$$

式中: $f_{3 \times 3}$ 表示 3×3 的卷积层; X_0 表示提取到的浅层特征。在深层特征提取阶段,将获得的高维度浅层特征输送到连续堆叠的 PLMA 中进行深层特征提取

$$X_k = f_{PLMA}^k(f_{PLMA}^{k-1}(\cdots f_{PLMA}^1(X_0)\cdots)) \quad (2)$$

$$X_{k_0} = f_{ESA}^k(f_{PPAB}^k(f_{LMFE}^k(f_{CAB}^k(X_{k_i})))) + X_{k_i} \quad (3)$$

式中: f_{PLMA} 表示渐进式的局部特征提取模块; X_k 表示第 k 个 PLMA 输出的特征图;在 PLMA 中,通过级联 4 个模块来实现特征提取和融合,其中 X_{k_0} 、 X_{k_i} 分别为第 k 个 PLMA 内的输出特征和输入特征, f_{ESA}^k 、 f_{PPAB}^k 、 f_{LMFE}^k 、 f_{CAB}^k 分别为第 k 个 PLMA 内的 ESA、PPAB、LMFE 和 CAB,在这之后,通过残差连

接将特征 X_k 和 X_0 加和,并使用像素清洗层将特征图调整为高分辨率图像的像素尺寸

$$I_{SR} = f_{rec}(X_k + X_0) \quad (4)$$

式中: I_{SR} 是超分辨率网络最终生成的重建图像; f_{rec} 是上采样层。特别地,此处采用了两个 2 倍缩放因子的像素清洗操作来实现 4 倍或 3 倍的上采样。

关于损失函数的选择,由于 L1 目标函数具有倾向于逼近真值图像的参考像素值的特性,有助于网络的特征选择并提高网络可解释性,此外,L1 目标函数针对绝对值施加惩罚,减少了异常量的影响从而增强所训练网络的鲁棒性。预处理的数据集 $\{I_{LR}, I_{HR}\}_{i=1}^N$ 中 $\{I_{LR}^i, I_{HR}^i\}$ 为第 i 对低分辨率图像和高分辨率图像, I_{LR}^i, I_{HR}^i 分别为数据集中第 i 幅低、高分辨率图像。最小化 L1 目标函数训练 RPSRnet,其目标函数如下式

$$L(\theta) = \sum_{i=1}^N |I_{SR}^i - I_{HR}^i| \quad (5)$$

式中: θ 是网络学习到的参数; I_{SR}^i 为从第 i 幅低分辨率图像 I_{LR}^i 中复原得到的超分辨率图像。

2.2 局部多尺度特征提取模块

在 PLMA 中,设计了两个模块,分别为 LMFE 和 PPAB。首先,介绍局部多尺度特征提取块 LMFE,如图 2 所示。为了减少参数并提高参数采用效率,LMFE 采用具有不同感受野的卷积层从不同通道提取特征。与之前的方法不同,该多尺度特征提取器采用 1×1 卷积、 3×3 卷积和带有扩张度的 3×3 卷积来捕获不同尺度的特征。扩张卷积采用组卷积,组数等于输入通道数,扩张值设置为 3。通过 3 种不同感受野的并行特征提取结构,LMFE 有效地利用参数并扩展感受野。与仅使用 3×3 卷积进行多尺度特征提取相比,分层的多路径提取方式可有效融合来自不同层次的特征。在提取各种尺度的特征之后,采用级联融合和压缩多尺度特征,再使用大感受野注意力卷积模块 VAB 从融合特征中进一步逐像素分配权重。LMFE 的工作流程如下

$$X_{Lm} = f_c(f_{1 \times 1}(X_{Li}), f_{3 \times 3}(X_{Li}), f_{3 \times 3 \text{dil}}(X_{Li})) \quad (6)$$

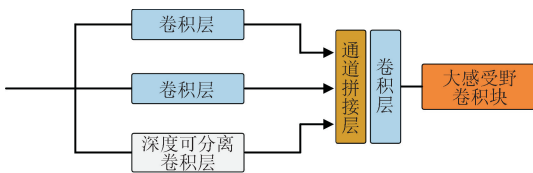


图 2 LMFE 的内部细节

Fig. 2 Interior particulars about LMFE

$$X_{Lo} = f_{VAB}(f_{1 \times 1}(X_{Lm})) \quad (7)$$

式中: $f_{1 \times 1}, f_{3 \times 3}, f_{3 \times 3 \text{dil}}$ 分别是 $1 \times 1, 3 \times 3$ 及扩张值为 3 的 3×3 卷积层运算; f_c 是对卷积的通道进行拼接运算; X_{Li}, X_{Lm}, X_{Lo} 分别是 LMFE 的输入特征、分层融合特征和输出特征; f_{VAB} 是大感受野像素注意力模块。

2.3 渐进式像素注意力模块

像素注意力神经网络(PAN)^[29]中提出的像素注意力模块,可以生成比普通注意力更高维的注意力图,相比于其他的注意力模块,它能够对图像的细粒度特征进行精细的学习与调整,有效帮助网络理解图像的局部信息,而更大的感受野有助于网络理解图像的长程关系,这对于恢复图像的高频细节十分重要。PPAB 是 RPSRnet 中设计的第二个模块,相比于 VapSR, PPAB 为了改进 VAB 中的像素注意力,采用了渐进式的特征提取方案。通过设置两个不同路径的注意力分支,在不同层次上逐步对输入提取特征,帮助模型理解图像长程关系。此外,渐进式的特征提取将输入表征为不同层次的特征抽象,每个路径上提取的特征代表不同级别的语义信息,有利于网络对高频特征的提取。

PPAB 的内部细节如图 3 所示,PPAB 首先对输入进行一次点卷积(Point-wise)和深度卷积(Depth-wise)组合的蓝图卷积,并在点卷积时调整特征通道数,随后将大卷积核像素注意力模块独立应用于两条不同路径上。在第一条路径中,通过一个像素注意力分支提取特征的高频信息,在另一条路径上,采取先前 Point-wise 卷积和 Depth-wise 卷积级联的蓝图卷积与像素注意力模块组合。之后将两个分支的输出结果相加融合,使用 1×1 的卷积调整特征通道数后与最初的输入执行残差相加,特别的是,在每个大感受野卷积块之后,对输出进行了像素归一化。具体地,首先将特征图大小 $B \times C \times H \times W$ 转换到 $B \times H \times W \times C$ 维度,对特征图执行卷积层归一化后,再次转换回原来的特征图大小 $B \times C \times H \times W$,这样就可以对每个像素值在通道维度上进行归一化,从而加速训练并提升性能。PPAB 的工作流程如下

$$X_{mid} = f_{bs}(X_{Pin}) \quad (8)$$

$$X_{1st} = f_{VAB}(X_{mid}) \quad (9)$$

$$X_{2nd} = f_{VAB}(f_{bs}(X_{mid})) \quad (10)$$

式中: $X_{Pin}, X_{mid}, X_{1st}, X_{2nd}$ 分别表示 PPAB 的输入特征、中间特征、第一条路径的输出特征以及第二条路径的输出特征; f_{bs} 是蓝图可分离卷积运算。

VAB 内的运算如下

$$X_{Vo} = X_{Vi} \otimes f_{5 \times 5 \text{dil}}(f_{5 \times 5}(f_{1 \times 1}(X_{Vi}))) \quad (11)$$

式中： X_{Vi} 、 X_{Vo} 、 $\otimes$ 表示 VAB 的输入和输出以及按像素作乘法运算； $f_{5 \times 5 \text{dil}}$ 和 $f_{5 \times 5}$ 为含扩张值的 5×5 卷积层和不含扩张值的 5×5 卷积运算。运算得到两

个不同层次的输出特征图之后，调整特征图通道并输出结果

$$X_{Pout} = \text{Norm}(X_{Pin} + f_{1 \times 1}(X_{1st} + X_{2nd})) \quad (12)$$

式中： X_{Pout} 是 PPAB 的输出； $\text{Norm}(\cdot)$ 表示像素归一化运算。

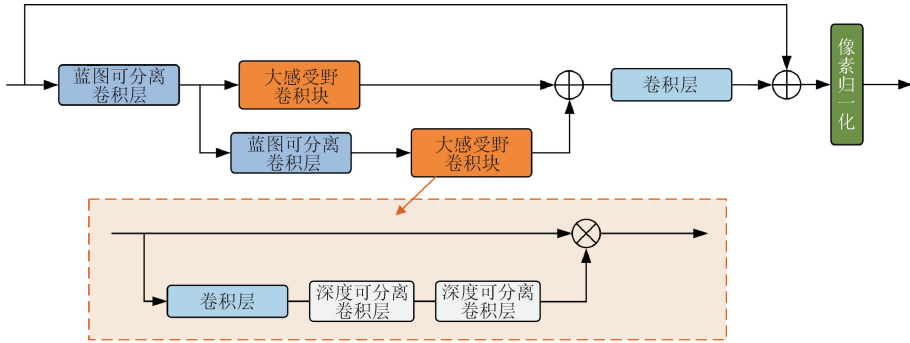
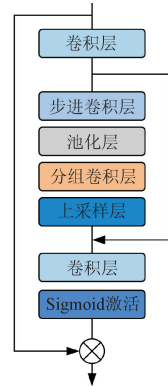


图 3 PPAB 的内部细节
Fig. 3 Interior particulars about PPAB

2.4 通道注意力模块和增强型空间注意力模块

通道注意力模块在卷积神经网络中的有效性已经在之前的研究中得到证实^[30-31]。在本文提出的网络中，将通道注意力模块融入到 PLMA 的前端，如图 4(a)所示。与普通的通道注意力不同的是，该网络在通道注意力层之前使用了蓝图可分离卷积 (BSConv)，取代了 3×3 卷积。

除了通道注意力模块以外，空间注意力模块^[32-33]可以显著提高网络对输入特征图中不同位置关系的捕捉能力。经过多次局部特征提取，网络在重建高频特征细节方面表现了足够的性能，但多次的残差连接会导致网络分层特征的利用率下降。将 ESA 置于 PLMA 的残差连接之后，聚合先前级联模块输出的残差特征，使网络更加关注空间维度上的关键内容，提高 RPSRnet 的分层特征利用率。ESA 的网络结构如图 4(b)所示。



(b)增强型空间注意力模块的内部细节

图 4 CAB 和 ESA 的网络结构

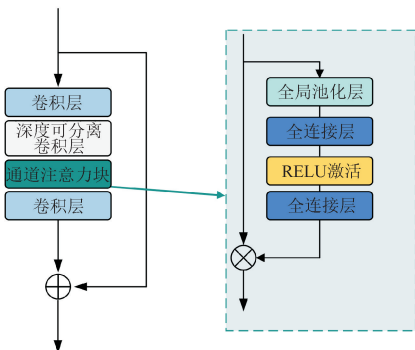
Fig. 4 Structure of CAB and ESA

3 实验结果与分析

本节介绍一系列用于验证所提出 RPSRnet 网络有效性的实验。

3.1 数据集与评价指标

所提出的 RPSRnet 网络使用 DIV2K 数据集进行训练和验证，该数据集包括 800 张训练图像对和 100 张验证图像对。同时在 4 个广泛使用的基准数据集 Set5、Set14、BSD100 和 Urban100 上进行测试。在评估过程中，采用了常规数据增强技术，包括随机水平翻转和 $90^\circ/270^\circ$ 旋转。低分辨率图像是通过原始图像进行双三次插值下采样得到的。此外，这些用于训练的图像在进行数据增强之前被转



(a)通道注意力模块的内部细节

换为 lmdb 格式,以加速设备的读取和模型的训练。需要指出的是,输出图像的 RGB 通道被转换为 YCbCr 通道。评估指标包括图像的峰值信噪比 (PSNR, 设为 P_{SNR}) 和结构相似度 (SSIM, 设为 S_{SIM}),在图像通道 Y 上计算这些评估指标。

峰值信噪比基于均方误差计算复原图像与原图像之间在像素级别上的差异,衡量了两者之间的像素相近程度;结构相似度不同于前者的像素级比较,它更注重人眼对图像结构和细节的感知,通过计算图像的亮度相似度、对比度相似度、结构相似度得到评估结果。利用峰值信噪比和结构相似度可以客观地对比不同方法复原出的图像的质量。

3.2 实验细节

关于实验的各项配置列举在表 1 中,此外,在制作 lmdb 格式数据前,训练数据集还被裁剪成 480×480 像素的子图像,需要指出的是,所有实验训练时不使用任何预训练模型和热启动策略。

表 1 RPSRnet 网络训练时的各项实验细节
Table 1 Experiment details when training RPSRnet

参数	数值或项目
PLMA 块数	8
网络通道数	48
PPAB 内通道数	64
优化器	AdamW
AdamW 参数	$\beta_1=0.9, \beta_2=0.999$
	权重衰减率为 0.000 1
批量大小	64
迭代次数	1 000 000
学习率	初始为 0.001,每 200 000 次迭代后减半
训练数据格式	lmdb
输入图像尺寸	256×256
Pytorch 版本	2.2.2
操作系统	Ubuntu 20.04
显示芯片	Nvidia RTX4090(24 GB)

3.3 与先进方法的对比实验

在本小节中,为了验证所提 RPSRnet 重建算法的有效性,在 4 个广泛使用的基准数据集上将 RPSRnet 与近年来主流的先进方法进行了对比,这些方法都在一定的参数量级之内有着优秀的性能。对比实验所采用的放大因子为 2、3、4,由于上采样的差异,同一模型在不同放大比例下参数略有不同。

这些方法已经在互联网上公开共享了实验结果以及开源代码,方便在相同的评估数据集上进行公正的比较。

表 2 列举了各种先进方法以及所提 RPSRnet 网络在 4 个数据集上的定量评估和参数量。以 4 倍放大因子为例,RPSRnet 以更少的参数取得了优越的性能,对比结果表明,所提出的方法在综合性能上更出色,VDSR 在重建性能方面表现较差,像 EDSR 这种与 RPSRnet 参数相当或更多的方法重建性能也只是次优,VapSR、BSRN 的重建性能更低。值得注意的是,相比于基于 Transformer 的重建方法 SwinIR,所提方法不仅在性能上表现更好 (Set5、Set14 数据集中达到了 0.03、0.04 dB 的提升),而且使用的参数缩减了 19.28%,这主要是因为像素注意力模块内的大感受野以及渐进式的特征提取,增强了网络对长程上下文关系的理解,有效地帮助网络恢复图像的高频细节,这也表明在纯卷积神经网络中使用大感受野的机制可提升性能。此外,与最新发表的 CVPR2023 论文中的 Omni-SR 方法相比,RPSRnet 同样具备很强的竞争力,特别是在 Set14 和 BSD100 数据集上,取得了更优秀的评估结果。尽管在 Urban100 数据集上,RPSRnet 的重建性能低于 Omni-SR,这是由于 Urban100 中存在大量相似的斑块,而 Omni-SR 中的全尺度聚合模块引入的长程依赖可以很好地恢复这些细节。尽管如此,RPSRnet 仍然优于其他方法,且参数更少 (相比 Omni-SR,参数缩减了约 9%),网络更加紧凑。此外,对于 2 倍和 3 倍的放大实验,RPSRnet 在评估指标以及参数量方面也表现优秀。综上所述,RPSRnet 的有效性以及重建性能得到了充分验证,尤其适用于 4 倍分辨率放大的任务。

除了对不同方法的定量评估分析,在运行效率上,以 4 倍放大因子为例,还进一步评估了 RPSRnet 以及 VDSR、EDSR 实现分辨率放大所需的平均运行时间,如表 3 所示,这 3 个方法都属于轻量模型,所提方法在运行效率上表现更出色,不论是参数量较多的 EDSR,还是参数量较少的 VDSR,其推理速度都更慢,且实验平台使用了高性能 CPU 以及图形处理器,运行时间的差异仍然明显,充分证明了 RPSRnet 方法的高效性。

表 2 在 4 个基准数据集上使用不同方法得到的峰值信噪比和结构相似度的定量结果

Table 2 Quantitative results (PSNR/SSIM) with different state of the art methods on four benchmark datasets

网络模型	放大 倍数	参数量/ 10^3	P_{SNR}/dB				S_{SIM}			
			Set5	Set14	BSD100	Urban100	Set5	Set14	BSD100	Urban100
VDSR ^[10]	2	666	36.66	33.05	31.90	30.76	0.954 2	0.912 7	0.896 0	0.914 0
EDSR ^[16]		1 335	37.99	33.57	32.16	31.98	0.960 4	0.917 5	0.899 4	0.927 2
IMDN ^[17]		694	38.00	33.63	32.19	32.17	0.960 5	0.917 7	0.899 6	0.928 3
RFDN ^[12]		626	38.08	33.67	32.18	32.24	0.960 6	0.919 0	0.899 6	0.929 0
LatticeNet ^[18]		756	38.15	33.78	32.25	32.43	0.961 0	0.919 3	0.900 5	0.930 2
RLFN ^[13]		527	38.07	33.72	32.22	32.33	0.960 7	0.918 7	0.900 0	0.929 9
SwinIR ^[20]		878	38.14	33.86	32.31	32.76	0.961 1	0.920 6	0.901 2	0.934 0
BSRN ^[19]		332	38.10	33.74	32.24	32.34	0.961 0	0.919 3	0.900 6	0.930 3
VapSR ^[28]		329	38.08	33.77	32.27	32.45	0.961 2	0.919 5	0.901 1	0.931 6
Omni-SR ^[23]	3	772	38.22	33.98	32.36	33.05	0.961 3	0.921 0	0.902 0	0.936 3
RPSRnet(本文)		709	38.12	33.85	32.31	32.76	0.961 5	0.920 3	0.902 3	0.934 2
VDSR ^[10]		666	33.66	29.77	28.82	27.14	0.921 3	0.831 4	0.797 6	0.827 9
EDSR ^[16]		1 555	34.37	30.28	29.09	28.15	0.927 0	0.841 7	0.805 2	0.852 7
IMDN ^[17]		703	34.36	30.32	29.09	28.17	0.927 0	0.841 7	0.804 6	0.851 9
RFDN ^[12]		633	34.47	30.35	29.11	28.32	0.928 0	0.842 1	0.805 3	0.854 7
LatticeNet ^[18]		765	34.53	30.39	29.15	28.33	0.928 1	0.842 4	0.805 9	0.853 8
SwinIR ^[20]		886	34.62	30.54	29.20	28.66	0.928 9	0.846 3	0.808 2	0.862 4
BSRN ^[19]		340	34.46	30.47	29.18	28.39	0.927 7	0.844 9	0.806 8	0.856 7
VapSR ^[28]	4	337	34.52	30.53	29.19	28.43	0.928 4	0.845 2	0.807 7	0.858 3
Omni-SR ^[23]		780	34.70	30.57	29.28	28.84	0.929 4	0.846 9	0.809 4	0.865 6
RPSRnet(本文)		720	34.59	30.50	29.19	28.47	0.929 1	0.845 4	0.808 6	0.859 1
VDSR ^[10]		666	31.35	28.01	27.29	25.18	0.883 8	0.767 4	0.725 1	0.752 4
EDSR ^[16]		1 518	32.09	28.58	27.57	26.04	0.893 8	0.781 3	0.735 7	0.784 9
IMDN ^[17]		715	32.21	28.58	27.56	26.04	0.894 8	0.781 1	0.735 3	0.783 8
RFDN ^[12]		643	32.28	28.61	27.58	26.20	0.895 7	0.781 8	0.736 3	0.788 3
LatticeNet ^[18]		777	32.30	28.68	27.62	26.25	0.896 2	0.783 0	0.736 7	0.787 3
RLFN ^[13]		543	32.24	28.62	27.60	26.17	0.895 2	0.781 3	0.736 4	0.787 7
SwinIR ^[20]	5	897	32.44	28.77	27.69	26.47	0.897 6	0.785 8	0.740 6	0.798 0
BSRN ^[19]		352	32.35	28.73	27.65	26.27	0.896 6	0.784 7	0.738 7	0.790 8
VapSR ^[28]		342	32.38	28.77	27.68	26.35	0.897 8	0.785 2	0.739 8	0.794 1
Omni-SR ^[23]		792	32.49	28.78	27.71	26.64	0.898 8	0.785 9	0.741 5	0.801 8
RPSRnet(本文)		724	32.47	28.81	27.71	26.49	0.898 5	0.786 5	0.741 7	0.797 8

表 3 在 4 倍放大因子下的平均运行时间对比

Table 3 Average running time required in scale 4

网络 模型	参数量/ 10^3	平均运行时间/s			
		Set5	Set14	BSD100	Urban100
EDSR	1 518	0.19	0.24	0.12	0.39
VDSR	666	0.12	0.18	0.08	0.38
RPSRnet	724	0.09	0.16	0.05	0.31

在主观视觉效果上,采用本文 RPSRnet 方法及其他超分辨率方法 (Bicubic、SRCNN、FSRCNN、VDSR、IMDN、RLFN) 进行对比,选取 2 个数据集中 4 幅图像进行重建,并将局部放大图进行对比,实验结果如图 5 所示。由图 5 可见,所选择的图像有明显的共同特点:图像中存在许多边缘或非光滑的纹理,此类区域包含灰度值变化剧烈的图像结构,在复原过程中的难度包含这类区域的图像相对于图像

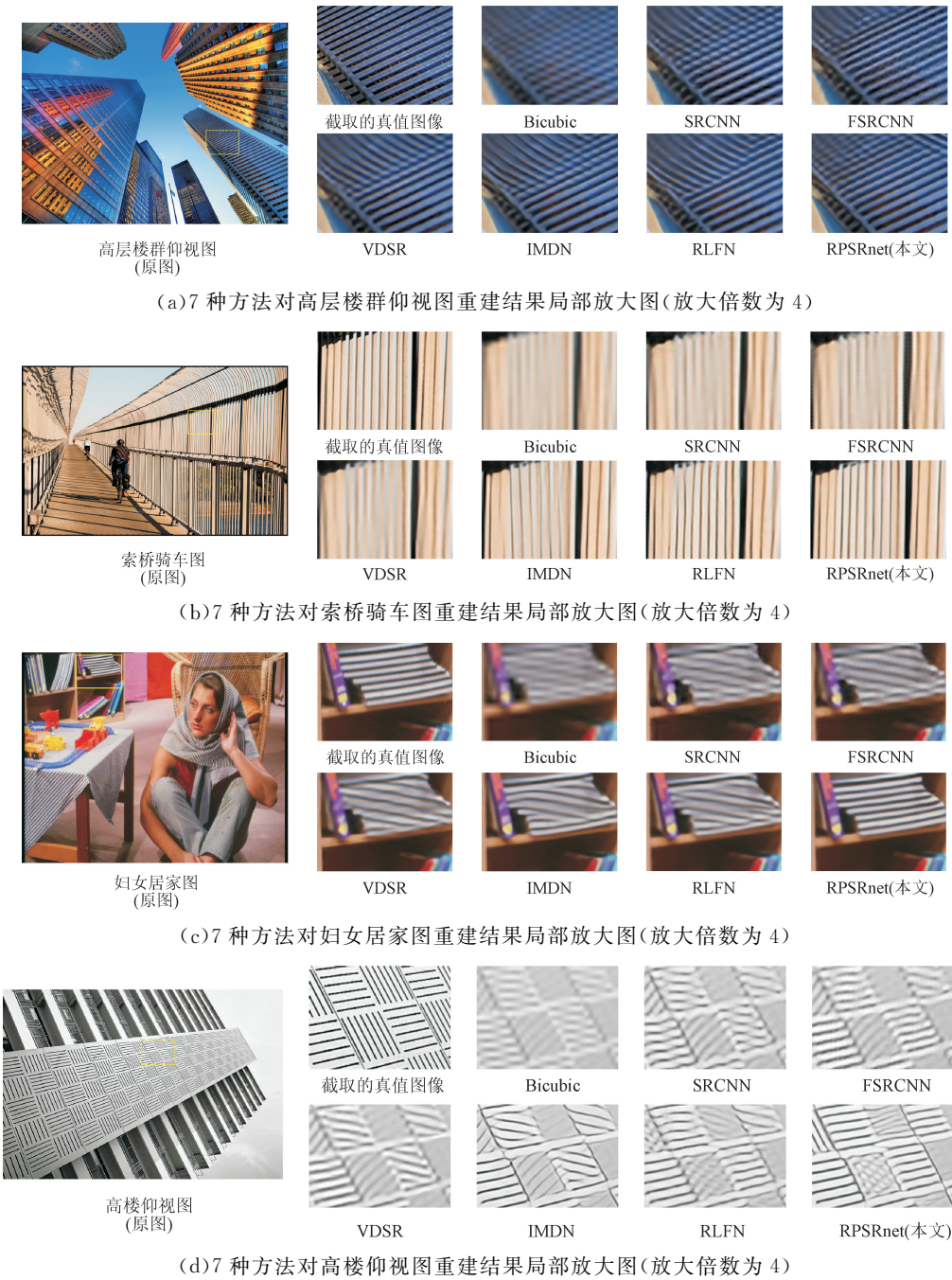


图 5 放大倍数为 4 时采用 7 种方法对 2 个测试集中 4 幅图像重建结果的比较

Fig. 5 Reconstruction results comparison with four images in two test sets using seven methods in scale 4

平滑区域大得多,8 幅低分辨率的复原图是不同超分辨率方法在同一幅图像上的复原结果,为了便于呈现复原的结果,这里仅截取复原图像中的一部分区域作展示,截取的真值图像区域与原图中黄色矩形框内的部分相对应。就恢复的图像细节而言,Bicubic 的复原结果相比于深度学习方法逊色得多,而 SRCNN 与 FSRCNN 的效果相对较好,能够恢复一定的图像细节和纹理,如图 5(a),视觉效果有明显改善,但在图 5(b)中则出现了严重的模糊并且发

生结构偏移,VDSR 在这些图像上的模糊效果相对少了许多,表现出更好的重建效果,但仍然存在结构偏移,IMDN 和 RLFN 在处理模糊和结构偏移方面表现更优,例如图 5(b)和图 5(d),视觉效果比 VDSR 好很多,也不会发生太严重的结构偏移,但在复原图 5(c)上出现了严重的偏移,特别是遇到像素灰度值发生剧烈变化的图像边缘时,其重建结果发生了畸变并伴随严重的模糊,可以明显看到的是,在这幅图像上,所提出的 RPSRnet 能够精准地恢复其

中的边缘细节,尤其是黑白相间的过渡区域,复原图像的这部分区域没有出现明显的模糊和偏移,对于其他图像也是如此,在图 5(d)中,对于横线和纵线密集分布的区域,能够更好地修正横线的弯曲程度,而图 5(b)图像中,栏杆与背景的分离区域也能恢复得很好,没有造成结构偏移。

通过主观视觉分析得出,经过 RPSRnet 复原的图像呈现出更令人满意的视觉效果,包括图像的边缘纹理、图像辨识度等,从而表明了所提方法对高频细节有更高敏感性以及对复杂边缘有精细恢复能力,证明了所提方法不仅在参数上足够轻量,复原的图像整体视觉效果也更好。

3.4 不同像素注意力模块的对比实验

为了验证所提出的渐进式注意力模块 PPAB 的有效性,进一步设置了像素注意力模块的对比实验,比较 PPAB 与 VapSR 中提出的单分支像素注意力模块 VAB 的复原性能。在该实验中,为了尽可能保持公平,以 RPSRnet 为主体网络结构,将 PPAB 的内部从渐进式的注意力模块修改为仅含单像素注意力模块的级联形式。同时为了尽可能保持参数一致,实验中保持了 PLMA 数量相同,并微调 PPAB 的结构和卷积层,以确保它们都只在像素注意力模块部分有所不同。准确地说,将单分支的像素注意力模块与 1×1 卷积层结合起来,并通过卷积内的通道数增加以实现参数调整,对比结果见表 4,可以看出,在提出的 RPSRnet 中,PPAB 对性能的改善的贡献更加显著,而使用 VAB 的 RPSRnet 虽然相比于同量级的其他方法也表现出了更优秀的性能,但是相较于使用 PPAB 仍然有一定性能劣势,这也验证了在像素注意力模块中,使用渐进式结构逐渐提取多尺度特征是有有效的。

表 4 4 倍放大因子下的像素注意力模块峰值信噪比对比

Table 4 PSNR comparison of different pixel attention block in scale 4

网络模型	峰值信噪比/dB			
	Set5	Set14	BSDS100	Urban100
使用渐进像素注意力块的 RPSRnet	32.47	28.81	27.71	26.49
使用大感受野卷积块的 RPSRnet	32.41	28.77	27.68	26.43

3.5 不同数量的 PLMA 对比实验

针对不同数量的深度特征提取模块 PLMA 进行了探讨。由于增加模块数可以帮助网络提升对不同数据分布的泛化能力,在一定程度上增加模块可以有效减少欠拟合现象,也可以帮助网络更好地学习复杂的特征和模式;然而模块过多会伴随一定的风险,过度添加模块会增加梯度消失和梯度爆炸的可能性。此外,层数的增加不仅导致训练时的难度加大,也会增加实际部署时计算量和内存消耗,在内存和计算资源受限的便携设备上表现得尤为明显。所以,合理选择网络的模块数有助于平衡性能和计算需求,这也与 RPSRnet 更侧重于轻量化网络的设计初衷相符。为了和近期提出的轻量化方案在相似的参数量级作比较,对不同数量的 PLMA 模块进行了评估实验,相应的结果见表 5。从表 5 中可以看出,当 PLMA 的数量 K 设置为 8 时,所提方法在性能和参数量之间实现了最合适的平衡,虽然模块数量更少时也能实现近似的平衡,但会损失更多的复原精度。此外,在 $K=9$ 时模型的性能相比于 Omni-SR 有所提升,但是提升并不明显,在 Set5 数据集中,当 $K=9$ 时 RPSRnet 额外使用的 8.5×10^4 参数量相比于 Omni-SR 仅带来了 0.03 dB 性能的提升,相比于使用 8 个 PLMA,增加模块所提升的性能无法很好地平衡额外参数的引入,因此,考虑到模型轻量化的需求,所提方法将 $K=8$ 作为最终的配置。

表 5 在 4 倍放大因子下不同数量 PLMA 的 PSNR 对比

Table 5 PSNR comparison of different progressive local-multi attention block numbers in scale 4

模块数 K	参数量	峰值信噪比/dB			
		Set5	Set14	BSDS100	Urban100
6	556	32.31	28.68	27.62	26.24
7	640	32.38	28.72	27.65	26.32
8	724	32.47	28.81	27.71	26.49
9	809	32.50	28.85	27.74	26.53

3.6 不同结构的 LMFE 对比实验

除了对所提出的 PPAB 进行对比外,本小节针对提出的局部多尺度特征提取模块 LMFE 设计了对比实验,验证该模块的性能提升效果。在该实验中,额外训练了仅含扩张卷积的多尺度分支对比模块(称为基础块 basic block),其结构如图 6 所示,对比模块使用的扩张卷积与 LMFE 内的扩张卷积配置相同,由于 3 个路径的卷积层都一致,感受野一致,这有利于在保持参数量级近似的实验条件下对

比不同感受野的多尺度特征提取模块的性能提升,同之前的对比实验类似,为了保持相近的参数量级,采取了级联 1×1 卷积层,并在卷积层上调整通道数的方法微调对比模块,除了该模块有所不同,其他训练配置均保持一致。

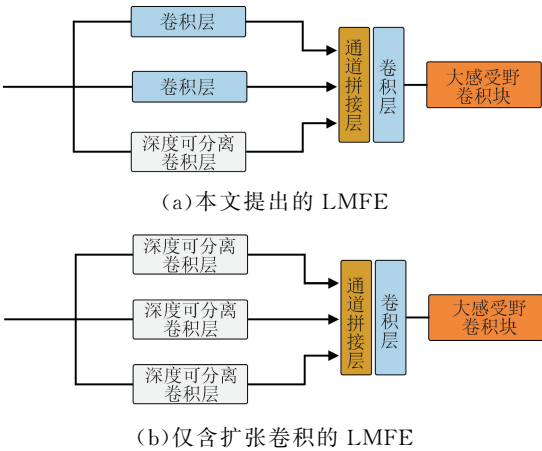


图 6 不同的 LMFE 网络结构

Fig. 6 Different LMFE network structures

图 7 给出了两种不同多尺度特征提取模块结构的 RPSRnet 训练轨迹对比(在 DIV2K 的验证数据集上进行验证,包含 100 张图片,实验使用 tensorboard 库记录实验结果),和所提出的原始 LMFE 对比,可以明显看到,所提出的 LMFE 在各个训练阶段的峰值信噪比相比基础模块都有显著提升,验证了对于混合不同感受野的多尺度特征提取方法可以提升性能。

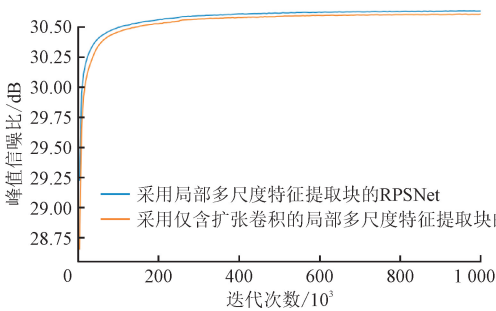


图 7 两种多尺度特征提取块在 DIV2K 验证集上的训练轨迹(放大因子为 4)

Fig. 7 Training trajectories of two multi-scale feature extraction blocks on the DIV2K validation set in scale 4

4 结 论

本文提出了一种基于感受野优化和渐进特征融合的轻量化单图像超分辨率网络。提出了局部多尺

度特征提取块 LMFE,它通过多尺度的分层特征提取,有效融合了来自不同感受野范围内的特征信息,同时重新审视并分析了大感受野在注意力机制下的表现,提出了渐进式的像素注意力模块 PPAB,改进了像素注意力的性能。实验结果表明,渐进式的大卷积核注意力模块增强了网络捕捉高频细节的能力,渐进式特征提取的设计与大感受野相结合,使提出的方法能够有效地利用更广泛的图像上下文信息。大量实验结果表明,本文的方法相对于现有方法取得了更好的性能。

同时,从实验结果可知,在 Set5、Set14 数据集上 RPSRnet 与最新的方法相比性能更好,但在 BSD100、Urban 数据集上有些差距,RPSRnet 在重建图像时并没有考虑到不同复杂退化的影响,导致网络在部分测试集图像上的重建性能受限,这也是目前超分辨率领域内亟待解决的问题,还需要更有效的解决方案。

参考文献:

[1] AAKERBERG A, NASROLLAHI K, MOESLUND T B. Real-world super-resolution of face-images from surveillance cameras [J]. IET Image Processing, 2022, 16(2): 442-452.

[2] HARIS M, SHAKHNAROVICH G, UKITA N. Task-driven super resolution: object detection in low-resolution images [C]//Neural Information Processing. Cham, Switzerland: Springer International Publishing, 2021: 387-395.

[3] 李嫣,任文琦,张长青,等. 基于真实退化估计与高频引导的内窥镜图像超分辨率重建 [J]. 自动化学报, 2024, 50(2): 334-347.

LI Yan, REN Wenqi, ZHANG Changqing, et al. Super-resolution of endoscopic images based on real degradation estimation and high-frequency guidance [J]. Acta Automatica Sinica, 2024, 50(2): 334-347.

[4] 徐胜军,杜森,段中兴,等. 多尺度区域注意力 InfoGAN 车牌识别网络 [J]. 西安交通大学学报, 2023, 57(8): 206-218.

XU Shengjun, DU Miao, DUAN Zhongxing, et al. Multi-scale regional attention InfoGAN license plate recognition network [J]. Journal of Xi'an Jiaotong University, 2023, 57(8): 206-218.

[5] CHEN Wenlong, YAO Pengcheng, GAI Shaoyan, et al. Multi-scale feature aggregation network for image super-resolution [J]. Applied Intelligence, 2022, 52(4): 3577-3586.

- [6] 韩玉兰, 顾琴, 崔玉杰, 等. 基于树状多分支残差注意力网络的真实场景图像超分辨率重构 [J]. 控制与决策, 2024, 39(10): 3442-3450.
HAN Yulan, GU Qin, CUI Yujie, et al. Real-world super-resolution based on residual attention network with tree based multi-branch structure [J]. Control and Decision, 2024, 39(10): 3442-3450.
- [7] KIM J H, CHOI J H, CHEON M, et al. MAMNet: multi-path adaptive modulation network for image super-resolution [J]. Neurocomputing, 2020, 402: 38-49.
- [8] DONG Chao, LOY C C, HE Kaiming, et al. Learning a deep convolutional network for image super-resolution [C]//Computer Vision-ECCV 2014. Cham, Switzerland: Springer International Publishing, 2014: 184-199.
- [9] DONG Chao, LOY C C, TANG Xiaoou. Accelerating the super-resolution convolutional neural network [C]//Computer Vision-ECCV 2016. Cham, Switzerland: Springer International Publishing, 2016: 391-407.
- [10] 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). Piscataway, NJ, USA: IEEE, 2016: 1646-1654.
- [11] AHN N, KANG B, SOHN K A. Fast, accurate, and lightweight super-resolution with cascading residual network [C]//Computer Vision-ECCV 2018. Cham, Switzerland: Springer International Publishing, 2018: 256-272.
- [12] LIU Jie, TANG Jie, WU Gangshan. Residual feature distillation network for lightweight image super-resolution [C]//Computer Vision-ECCV 2020 Workshops. Cham, Switzerland: Springer International Publishing, 2020: 41-55.
- [13] KONG Fangyuan, LI Mingxi, LIU Songwei, et al. Residual local feature network for efficient super-resolution [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ, USA: IEEE, 2022: 765-775.
- [14] 梁敏, 汪西莉. 结合超分辨率和域适应的遥感图像语义分割方法 [J]. 计算机学报, 2022, 45(12): 2619-2636.
LIANG Min, WANG Xili. Semantic segmentation model for remote sensing images combining super resolution and domain adaption [J]. Chinese Journal of Computers, 2022, 45(12): 2619-2636.
- [15] CHEN Haoyu, GU Jinjin, LIU Yihao, et al. Masked image training for generalizable deep image denoising [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2023: 1692-1703.
- [16] 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). Piscataway, NJ, USA: IEEE, 2017: 1132-1140.
- [17] HUI Zheng, GAO Xinbo, YANG Yunchu, et al. Lightweight image super-resolution with information multi-distillation network [C]//Proceedings of the 27th ACM International Conference on Multimedia. New York, NY, USA: ACM, 2019: 2024-2032.
- [18] LUO Xiaotong, XIE Yuan, ZHANG Yulun, et al. LatticeNet: towards lightweight image super-resolution with lattice block [C]//Computer Vision-ECCV 2020. Cham, Switzerland: Springer International Publishing, 2020: 272-289.
- [19] LI Zheyuan, LIU Yingqi, CHEN Xiangyu, et al. Blueprint separable residual network for efficient image super-resolution [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ, USA: IEEE, 2022: 832-842.
- [20] LIANG Jingyun, CAO Jiezhong, SUN Guolei, et al. SwinIR: image restoration using Swin transformer [C]//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Piscataway, NJ, USA: IEEE, 2021: 1833-1844.
- [21] LU Zhisheng, LI Juncheng, LIU Hong, et al. Transformer for single image super-resolution [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ, USA: IEEE, 2022: 456-465.
- [22] CHEN Zheng, ZHANG Yulun, GU Jinjin, et al. Dual aggregation transformer for image super-resolution [C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2023: 12278-12287.
- [23] WANG Hang, CHEN Xuanhong, NI Bingbing, et al. Omni aggregation networks for lightweight image super-resolution [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2023: 22378-22387.
- [24] CHEN Xiangyu, WANG Xintao, ZHOU Jiantao, et al. Activating more pixels in image super-resolution transformer [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

Piscataway, NJ, USA: IEEE, 2023: 22367-22377.

[25] LIU Zhuang, MAO Hanzi, WU Chaoyuan, et al. A ConvNet for the 2020s [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2022: 11966-11976.

[26] HESSE R, SCHAUB-MEYER S, ROTH S. Content-adaptive downsampling in convolutional neural networks [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ, USA: IEEE, 2023: 4544-4553.

[27] GUO Menghao, LU Chengze, LIU Zhengning, et al. Visual attention network [J]. Computational Visual Media, 2023, 9(4): 733-752.

[28] ZHOU Lin, CAI Haoming, GU Jinjin, et al. Efficient image super-resolution using vast-receptive-field attention [C]//Computer Vision-ECCV 2022 Workshops. Cham, Switzerland: Springer Nature Switzerland, 2023: 256-272.

[29] ZHAO Hengyuan, KONG Xiangtao, HE Jingwen, et al. Efficient image super-resolution using pixel attention [C]//Computer Vision-ECCV 2020 Workshops. Cham, Switzerland: Springer International Publishing, 2020: 56-72.

[30] HU Jie, SHEN Li, ALBANIE S, et al. Squeeze-and-excitation networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.

[31] SHANG Jianrun, ZHANG Xue, ZHANG Guisheng, et al. Gated multi-attention feedback network for medical image super-resolution [J]. Electronics, 2022, 11(21): 3554.

[32] ZONG Zhikai, ZHA Lin, JIANG Jiande, et al. Asymmetric information distillation network for lightweight super resolution [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ, USA: IEEE, 2022: 1248-1257.

[33] LIU Jie, ZHANG Wenjie, TANG Yuting, et al. Residual feature aggregation network for image super-resolution [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 2356-2365.

(编辑 武红江)



(上接第 135 页)

JIA Hua, LIU Yanjun, WANG Yu, et al. Adaptive-neural-network-based terminal sliding mode control for six-axis robotic manipulators [J]. Journal of Xi'an Jiaotong University, 2022, 56(11): 21-30.

[22] 殷春武, 甘婷, 楚天乐, 等. 基于时变终端滑模的任意迭代初态抑制控制 [J]. 控制理论与应用, 2023, 40(6): 1105-1112.

YIN Chunwu, GAN Ting, CHU Tianle, et al. Arbitrary iterative initial value suppression control based on time-varying terminal sliding mode [J]. Control Theory & Applications, 2023, 40(6): 1105-1112.

[23] WU Shaohua, WU Aiguo, DONG Na, et al. Adaptive iterative learning control of robotic manipulator with second-order terminal sliding mode method [C]//2018 13th World Congress on Intelligent Control and Automation (WCICA). Piscataway, NJ, USA: IEEE, 2018: 1488-1493.

[24] 张程琳, 桑文闯, 孙宁, 等. 多自由度机器人自适应滑模迭代学习跟踪控制 [J]. 控制与决策, 2024, 39(6): 1819-1828.

ZHANG Chenglin, SANG Wenchuang, SUN Ning, et al. Adaptive sliding mode iterative learning tracking control of multi-degree-of-freedom robots [J]. Control and Decision, 2024, 39(6): 1819-1828.

[25] 刘金琨. 机器人控制系统的设计与 MATLAB 仿真: 基本设计方法 [M]. 北京: 清华大学出版社, 2016.

[26] 朱玉强, 刘振, 朱全民, 等. 不确定机械臂系统的模糊滑模有限时间容错控制 [J/OL]. 控制理论与应用. (2024-06-11) [2024-07-02]. <http://kns.cnki.net/kcms/detail/44.1240.TP.20240607.1011.012.html>.

ZHU Yuqiang, LIU Zhen, ZHU Quanmin, et al. Fuzzy sliding mode finite-time fault-tolerant control for uncertain manipulator systems [J/OL]. Control Theory & Applications. (2024-06-11) [2024-07-02]. <http://kns.cnki.net/kcms/detail/44.1240.TP.20240607.1011.012.html>.

(编辑 杜秀杰)