

基于密集高效 Transformer 的图像超分辨率重建

傅慧滢, 王 瑜, 杨高明

安徽理工大学 计算机科学与工程学院, 安徽 淮南 232001

摘要:目的 针对超分辨率网络存在参数量大和图像特征提取能力不足导致重建图片质量不佳的问题,提出了一种新的图像超分辨率模型 LHNNet。方法 在深层特征提取模块部分,设计对称密集 CNN,以对称结构连接卷积模块,加强模型对图像局部信息的提取能力;引入双注意力模块,使网络关注像素和空间信息;此外,使用密集连接方式连接高效 Transformer 模块,增强模型全局特征的提取能力,从而更好地恢复图像纹理细节。结果 实验结果在 Set5、Set14、BSD100、Urban100 和 Manga109 等五个方法数据集进行测试。结果表明:在 Set5 数据集上,放大倍数为 3 时,LHNNet 方法相对于 IMDN 方法减少了 99 K 的参数量,同时 PSNR 值提高了 0.13 dB,SSIM 提高了 0.000 9。与当前其他方法,如 LAPAR-A、ShuffleMixer 等方法相比,LHNNet 方法也表现出优越的性能。结论 LHNNet 方法可以在使用相对较少参数的同时提高重建图片的质量,从而实现参数量和性能之间的平衡。

关键词:超分辨率重建;对称结构;双注意力模块;密集连接;高效 Transformer

中图分类号:TP391 文献标识码:A doi:10.16055/j.issn.1672-058X.2026.0001.009

Image Super-Resolution Reconstruction Based on Dense and Efficient Transformer

FU Huiying, WANG Yu, YANG Gaoming

School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 232001, Anhui, China

Abstract: Objective To address the problems of a large number of parameters and insufficient image feature extraction ability in super-resolution networks, which lead to poor quality of reconstructed images, a new image super-resolution model, LHNNet, is proposed. **Methods** In the deep feature extraction module, a symmetric dense CNN was designed. The convolution module was connected in a symmetric structure to enhance the model's ability to extract local image information. A dual attention module was introduced to make the network focus on pixel and spatial information. In addition, the efficient Transformer modules were connected using a dense connection method to strengthen the model's global feature extraction ability, thus better restoring the image texture details. **Results** The experimental results were tested on five datasets, namely Set5, Set14, BSD100, Urban100, and Manga109. The results showed that on the Set5 dataset, when the magnification factor was 3, the LHNNet method reduced the number of parameters by 99 K compared with the IMDN method. Meanwhile, the PSNR value increased by 0.13 dB, and the SSIM increased by 0.000 9. Compared with other current methods, such as LAPAR-A and ShuffleMixer, the LHNNet method also showed superior performance. **Conclusion** The LHNNet method can improve the quality of reconstructed images while using a relatively small number of parameters, thereby achieving a balance between the number of parameters and performance.

Keywords: super-resolution reconstruction; symmetric structure; dual attention module; dense connection; efficient Transformer

收稿日期:2024-01-16 修回日期:2024-03-18 文章编号:1672-058X(2026)01-0072-08

基金项目:国家自然科学基金(52374155);安徽省自然科学基金(2308085MF218);安徽省高等学校科学研究项目(2022AH040113)。

作者简介:傅慧滢(2001—),女,安徽宣城人,硕士研究生,从事计算机视觉研究。Email:fhuiying2022@aust.edu.cn。

引用格式:傅慧滢,王瑜,杨高明.基于密集高效 Transformer 的图像超分辨率重建[J].重庆工商大学学报(自然科学版),2026,43(1):72-79.

FU Huiying, WANG Yu, YANG Gaoming. Image super-resolution reconstruction based on dense and efficient Transformer[J].

Journal of Chongqing Technology and Business University (Natural Science Edition), 2026, 43(1): 72-79.

图像超分辨率(Super Resolution, SR)旨在从低分辨率图像中重建出纹理细节更加丰富的、高分辨率的图像。这项技术被广泛应用于遥感图像^[1]、图像修复^[2]和医学图像处理^[3]等领域。

随着深度学习技术的不断发展,基于卷积神经网络(Convolutional Neural Network, CNN)的方法逐渐应用在图像超分辨率任务中,并取得了令人印象深刻的性能。Dong 等^[4]将卷积神经网络引入 SR 重建中,采用三层卷积结构设计获得了比传统方法更好的性能。Lim 等^[5]提出用于图像超分辨率的增强深度残差网络,通过残差设计^[6]提取更多特征信息。Zhang 等^[7]提出非常深的残差通道注意力网络,使用通道注意力机制专注学习重要信息。然而,虽然使用 CNN 方法相较于传统方法在重建图片上表现更为出色,但是仅仅使用卷积结构设计,网络关注于局部图片信息而无法有效学习全局特征。

随着自然语言处理^[8]和高级视觉任务^[9]的提出,Transformer 技术展现出巨大潜力。Transformer 能够有效地建模远距离依赖关系,关注图像全局特征,因此可以弥补卷积提取全局信息不足的问题。近年来,研究者将 Transformer 技术引入图像超分辨率领域,用来增强网络全局信息提取能力。Liang 等^[10]提出一种基于 Swin Transformer 的网络,使用一堆残差 Swin Transformer 块进行深度特征提取,取得了良好的效果。Chen 等^[11]设计融合快速傅里叶卷积的空间频率块,并提出全新的特征集成训练策略,以提高网络全局信息提取能力。Chen 等^[12]在连续的 Transformer 块中交替应用空间和通道自注意力,实现强大的特征提取能力。

虽然这些方法有效提升了模型的性能,但同时也使得网络的参数量过大,导致运算复杂度增加。为了应对这一问题,轻量级超分辨率网络引起了广泛关注。这类网络通过精心设计结构和减少参数量,旨在保持模型性能的同时降低计算成本,更适用于资源受限的

场景。Yu 等^[13]采用结合蒸馏、渐进学习和修剪的多阶段轻量级训练策略,有效地实现单幅图像超分辨率重建。Wang 等^[14]提出高效的特征去冗余和自校准超分辨率网络,通过设计特征去冗余和自校准块,减少模型提取重复特征。此外,设计局部特征融合模块充分利用和融合提取到的特征信息,确保在减少不必要的计算量的同时保证模型性能不受影响。Zhang 等^[15]提出组卷积特征增强蒸馏网络,多尺度群卷积块的设计可以在保持 SR 性能的同时大大减少了参数量。

综上所述,虽然现有的一些 SR 方法使用 Transformer 有效解决了 CNN 不能提取全局特征的问题,但是 Transformer 的应用也带来了较大的参数量。当前的模型很难充分发挥 CNN 和 Transformer 各自的优势,并在模型大小和性能之间取得最佳平衡。针对上述问题,提出 LHNet 模型,该模型由对称密集 CNN (Symmetric Dense Convolutional Neural Network, SDCNN)和密集高效 Transformer(Dense Efficient Transformer Network, DETN)构成。SDCNN 采用对称结构设计,融合双注意力模块,用于关注像素和空间信息,加强网络局部特征的提取能力。此外,为了加强网络全局特征的提取能力,加入 DETN,通过跨通道应用自注意机制(Self-Attention, SA),即计算跨通道的交叉协方差,以生成隐式编码全局上下文意图的方式,在关注上下文信息的同时减少网络的参数量。实验结果表明:本文的模型在峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)和结构相似性(Structural Similarity, SSIM)指标评估方面优于现有的图像超分辨率重建方法。重建图像更加清晰,细节更加丰富,在模型性能和参数量之间达到了较好的平衡。

1 图像超分辨率重建方法

设计的图像超分辨率重建网络整体框架如图 1 所示。本文提出的 LHNet 方法主要分为三个部分:浅层特征提取模块、深层特征提取模块和重建模块。

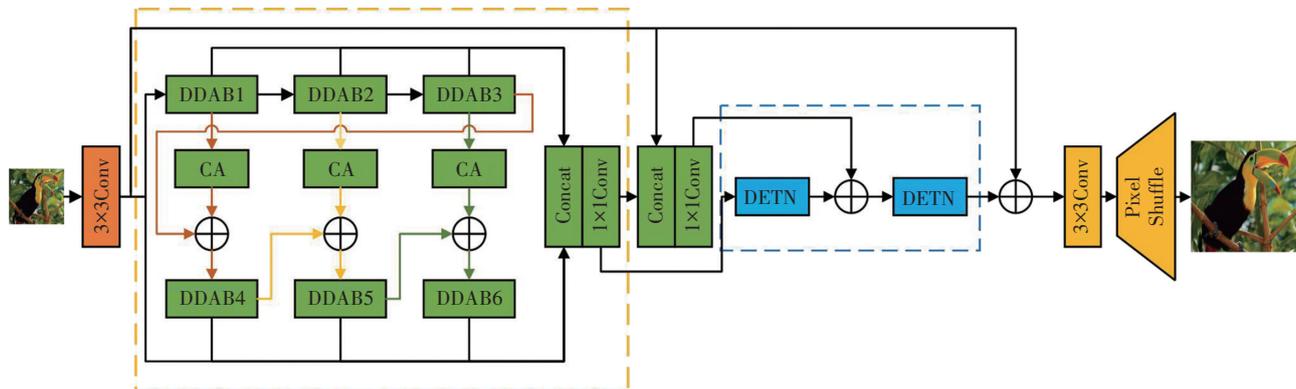


图 1 基于对称密集 CNN 和密集高效 Transformer 的超分辨率网络整体框架

Fig. 1 Overall framework of a super-resolution network based on symmetric dense CNN and dense efficient Transformer

将输入图像和输出图像分别用 I_{SR} 和 I_{LR} 来表示。首先使用一个 3×3 的卷积来提取输入图片的浅层特征信息 f_{sf} , 表示为

$$f_{sf} = F_{Conv3 \times 3}(I_{LR}) \quad (1)$$

式(1)中, $F_{Conv3 \times 3}(\cdot)$ 表示卷积层, 得到的浅层特征 f_{sf} 被传递到下一网络。

在深度特征提取部分主要由 SDCNN 和 DETN 组成。SDCNN 采用对称结构设计, 主要用于提取局部特征, 该过程为

$$f_{cnn} = F_{cnn}(f_{sf}) \quad (2)$$

式(2)中, $F_{cnn}(\cdot)$ 表示 SDCNN 操作, f_{cnn} 表示提取到的局部特征。这一局部特征继续送入 DETN 中提取全局信息。该过程为

$$f_{trans} = F_{trans}(f_{cnn}) \quad (3)$$

式(3)中, $F_{trans}(\cdot)$ 表示 DETN 操作。 f_{trans} 表示提取的全局信息特征。最后, 全局信息特征 f_{trans} 和浅层特征 f_{sf} 相加, 重新送入重建模块中生成超分辨率图像。该过程为

$$I_{SR} = F_{re}(f_{sf} + f_{trans}) \quad (4)$$

式(4)中, $F_{re}(\cdot)$ 是由 1×1 的卷积和亚像素卷积层组成的重建模块。 I_{SR} 为网络重建的高分辨率图像。

1.1 对称密集 CNN 模块

受到文献[16]的启发, 本文提出 SDCNN 进行局部特征提取。如图 1 黄色框区域所示, 该模块由密集双注意力模块(Dense Double Attention Block, DDAB)和通道注意(Channel Attention, CA)组成。整个网络分为两个分支, 网络的浅层特征首先传递到顶部分支中, 顶部分支中每个 DDAB 的输出将作为下分支中相应 DDAB 的输入部分。采用对称的结构设计使得 DDAB 模块之间参数共享, 从而提取图像更多局部特征。

如图 2 所示, DDAB 由三个双注意力模块(Double Attention Block, DAB)密集连接而成。每一层提取的特征信息都作为后续层的输入。密集连接设计很好地缓解了梯度消失问题, 通过特征信息的重用, 以重复利用提取的特征信息, 进而提高图像超分辨率模型的性能。

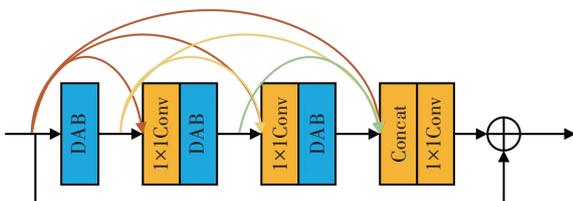


图 2 密集双注意力模块

Fig. 2 Dense dual attention block

DAB 模块设计如图 3 所示。首先, 使用 1×1 的卷积将输入特征分成两个分支, 顶部分支用于保留输入特征的原始信息。该过程为

$$x_{top_i} = F_{conv3 \times 3}(x_i) \quad (5)$$

式(5)中, x_i 表示第 i 个 DAB 模块的输入信息。 $F_{conv3 \times 3}(\cdot)$ 是标准的卷积模块, 用于学习压缩特征。 x_{top_i} 是顶部分支的输出。底部分支使用像素注意力(Pixel Attention, PA)和增强的空间注意力(Enhanced Spatial Attention, EPA)来调整 DAB 的接受域并捕获有用的特征区域。该过程为

$$x_{bot_i} = F_{bot}(x_i) \quad (6)$$

式(6)中, x_i 表示第 i 个 DAB 模块的输入信息, F_{bot} 描述第 i 个 DAB 模块底部分支的运算过程, x_{bot_i} 表示底部分支的输出。之后使用通道连接的方式连接顶部分支和底部分支得到的特征, 并使用 1×1 的卷积融合特征。但是特征融合操作却削弱了像素和空间信息。为了解决这一问题, 设计 ESA 模块用于重校准输出特征的权重, 对特征信息进行加强。该过程为

$$y_i = F_{ESA}\{F_{conv1 \times 1}[F_{cat}(x_{top_i}, x_{bot_i})]\} + x_i \quad (7)$$

式(7)中, $F_{cat}(\cdot)$ 表示通道连接, $F_{conv1 \times 1}(\cdot)$ 表示卷积核为 1×1 的标准卷积, $F_{ESA}(\cdot)$ 表示增强的空间注意力。 x_i 表示第 i 个 DAB 模块的输入信息, y_i 表示第 i 个 DAB 模块的输出信息。

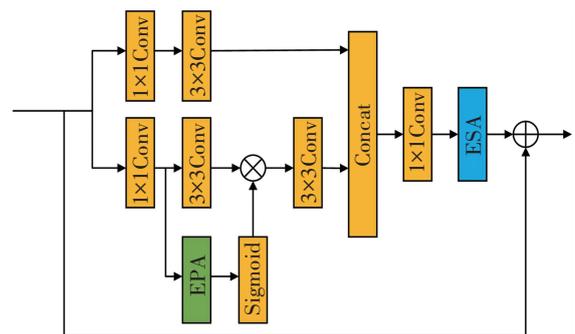


图 3 双注意力模块

Fig. 3 Dual attention block

增强空间和像素注意力模块(Enhanced Spacial and Pixel Attention, EPA)如图 4 所示。由于像素注意力和增强的空间注意力占据两个不同的空间, 因此采用并联机制进行注意力操作, 使用 1×1 的卷积和 Sigmoid 激活函数融合特征信息。

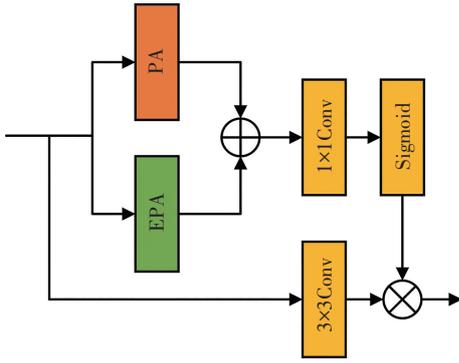


图 4 增强空间和像素注意力

Fig. 4 Enhanced spatial and pixel attention

1.2 密集高效 Transformer

SDCNN 专注于局部特征提取,但是仅仅依赖局部特征不足以实现高质量图像的重建。Tranformer 可以学习图像的长期依赖关系,加深对全局信息的关注。CNN 和 Transformer 的有效结合,可以让模型既关注局部信息也关注全局信息,加强网络特征的提取能力,重建出高频的纹理细节。

受到 Restormer^[17] 和 DenseNet^[18] 的启发,核心模块 DETN 由多卷积头转置注意力 (Multi-Dconv Head Transposed Attention, MDTA) 和门控卷积前馈网络 (Gated-dconv Feed-forward Network, GDFN) 组成。每三个高效 Transformer 为一组,共两组。每组采用同 DDAB 相似的密集连接方式连接高效 Transformer。密集结构设计使信息在网络中的所有层中传递最大化,缓解梯度消失问题,加强网络的特征提取能力。

由于 Transformer 的主要开销在于注意力层,并且随着输入大小呈二次增长。为了解决这一问题,这里使用 MDTA,不是显示的对成对像素的交互建模,而是计算跨特征通道的交叉协方差。MDTA 结构如图 5 所示,首先将输入特征分为三个分支,每个分支都使用 1x1 的卷积进行跨通道上下文像素级整合,使用 3x3 的深度卷积进行局部上下文的通道级聚合,得到键 K 、查询 Q 和值 V 。之后执行重塑操作,使得 $\hat{Q} \in \mathbf{R}^{H \times W \times C}$, $\hat{K} \in \mathbf{R}^{C \times H \times W}$, $\hat{V} \in \mathbf{R}^{H \times W \times C}$ 。这样键 \hat{K} 和查询 \hat{Q} 点积交互会生成大小为 $C \times C$ 的转置注意力图 A 。该过程为

$$\text{Attention}(\hat{Q}, \hat{K}, \hat{V}) = \hat{V} \cdot \text{Softmax}(\hat{K} \cdot \hat{Q} / \alpha) \quad (8)$$

式(8)中, $\text{Attention}(\hat{Q}, \hat{K}, \hat{V})$ 为使用 Transformer 提取到的特征信息, $\text{Softmax}(\cdot)$ 表示 Softmax 函数, α 为可学习的缩放参数。之后,将提取的特征信息与原始输入相加作为输出,该过程为

$$y_i = F_{\text{conv}1 \times 1} [\text{Attention}(\hat{Q}, \hat{K}, \hat{V})] + x_i \quad (9)$$

式(9)中, $F_{\text{conv}1 \times 1}$ 表示标准 1x1 卷积, x_i 表示第 i 个 MDTA 模块的输入特征信息, y_i 表示第 i 个 MDTA 模块的输出特征信息。

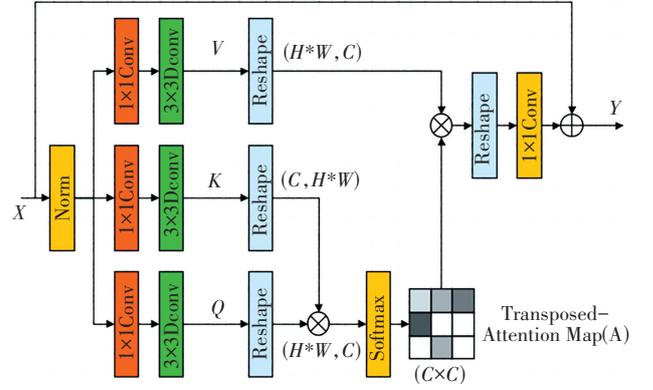


图 5 多卷积头转置注意力

Fig. 5 Multi-convolution head transposed attention

为了获得更加准确的特征信息,引入门控卷积前馈网络。如图 6 所示。该过程为

$$y_{G,2} = F_{\text{conv}3 \times 3} \{ F_{\text{conv}1 \times 1} [\text{LN}(y_i)] \} \quad (10)$$

$$y_{G,1} = \text{GELU} \{ F_{\text{conv}3 \times 3} \{ F_{\text{conv}1 \times 1} [\text{LN}(y_i)] \} \} \quad (11)$$

$$Y_i = F_{\text{conv}1 \times 1} (y_{G,1} \odot y_{G,2}) \quad (12)$$

式(10)、式(11)中, $\text{LN}(\cdot)$ 表示归一化层,式(11)中 $\text{GELU}(\cdot)$ 表示 GELU 激活函数,式(12)中 \odot 代表元素乘法。 y_i 表示第 i 个 GDFN 模块的输入特征信息, Y_i 表示第 i 个 GDFN 模块的输出特征信息。GDFN 在网络中通过门控机制控制信息流,从而抑制信息量较小的特征,只允许有用的信息通过网络进一步传递。

总体而言, DETN 的设计能够有效地帮助模型提取更多的全局信息来恢复高频纹理细节。

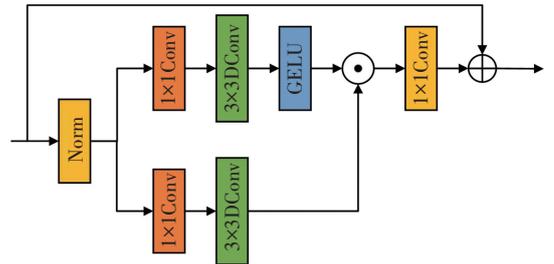


图 6 门控卷积前馈网络

Fig. 6 Gated-convolution feed-forward network

2 实验与结果分析

2.1 数据集和评价指标

本次实验使用 DIV2K 的 800 张高分辨率图像来构建训练集。采用随机旋转 90°、180°、270° 和水平翻转进行数据增强,得到 8 000 张图片用于训练。测试集采用 5 个通用的标准图像数据集: Set5、Set14、BSD100、Urban100 和 Manga109。

本文模型的性能评估与大多数研究一样,采用 PSNR 和 SSIM 两个指标评估模型的性能。其中,PSNR 和 SSIM 的计算是通过将 RGB 通道转换为 YCbCr 通道中的 Y 通道得到。

2.2 实验参数设置

实验设备为 Intel Core i5-12400 处理器,12 GB 内存,NVIDIA GeForce RTX 3060 GPU 显卡。在 Pytorch 框架上实现模型训练。实验设置训练 1 000 轮,训练批次设置为 16。在训练过程中,采用从 DIV2K 数据集中随机裁剪的 48×48 大小补丁作为输入。通道数设置为 48。初始学习率设置为 2×10^{-4} ,最后通过余弦退火降低到 6.25×10^{-6} 。整个模型采用 L1 损失函数进行训练。

2.3 消融实验

为了验证提出方法的有效性,使用 Set5 数据集作为测试集,在放大倍数为 3 的条件下,对设计的模块和模型框架进行消融实验。采用级联密集 CNN(Cascade Dense Convolutional Neural Network, CDCNN)代替 SDCNN;设计对称密集非注意力模块(Symmetric Without Double Attention Convolutional Neural Network, SWDCNN),去除 DAB,将注意力块全部替换成标准 3×3 卷积;去除 DETN 模块;使用级联高效 Transformer(Cascade Efficient Transformer Network, CETN)代替 DETN。

如表 1 所示,通过将“SDCNN+DETN”和“CDCNN+DETN”进行比较,在 Set5 数据集上,采用对称结构设计使 PSNR 提高了 0.04 dB,SSIM 提高了 0.001,体现了对称设计的有效性。将“SDCNN+DETN”和“SWDCNN+DETN”相比,PSNR 提高了 0.01 dB。说明双注意力模块可以关注图像中更多的细节信息,促进模型性能的提高。表 1 结果显示在 Set5 测试集下,“SDCNN+DETN”比 SDCNN 的 PSNR 值提高了 0.05 dB,SSIM 提

高了 0.001。证明 Transformer 的加入可以提高网络性能。“SDCNN+DETN”与“SDCNN+CETN”相比,在 Set5 数据集上,PSNR 提高了 0.05 dB,SSIM 提高了 0.000 2。证明密集连接的有效性。

表 1 不同模块对网络模型的影响

模块	倍数	PSNR/dB	
		Set5	
		PSNR/SSIM	
CDCNN+DETN		34.45/0.927 8	
SWDCNN+DETN		34.48/0.927 9	
SDCNN	3	34.34/0.926 8	
SDCNN+CETN		34.44/0.927 7	
SDCNN+DETN(LHNet)		34.49/0.927 9	

2.4 其他算法对比分析

为了验证本文方法的有效性,与 SRCNN^[4]、FSRCNN^[19]、IMDN^[20]、RFDN^[21]、LAPAR-A^[22]、ShuffleMixer^[23] 进行比较。如表 2、表 3、表 4 所示,分别展示了 7 种轻量级算法在测试集 Set5、Set14、BSD100、Urban100 和 Manga109 上分别放大 2、3、4 倍的客观指标的比较结果。在模型复杂度方面,假设模型输出为 720×1 280,使用参数数量和浮点运算数(Floating Point Operations, FLOPs)用于衡量模型的复杂性。观察发现,LHNet 模型的 PSNR 和 SSIM 值优于参与比较的其它大多数方法。与 IMDN 方法比较中,在放大倍数为 3 的情况下,LHNet 模型的参数量比 IMDN 模型少 99 K,PSNR 和 SSIM 指标明显高于 IMDN。在与 LAPAR-A 比较中,在放大倍数为 4 的情况下,LHNet 的 FLOPs 比 LAPAR-A 少 34 G,PSNR 和 SSIM 指标也明显高于 LAPAR-A。综上所述,可以证明所提出算法的有效性。

表 2 2 倍尺度下不同算法的 PSNR/SSIM 对比

方法	倍数	参数 /K	FLOPs /G	PSNR/dB				
				Set5	Set14	Urban100	BSD100	Manga109
				PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
SRCNN	2	57	53	36.66/0.954 2	32.45/0.906 7	29.50/0.894 6	31.36/0.887 9	35.60/0.966 3
FSRCNN	2	12	6	37.00/0.955 8	32.63/0.908 8	29.88/0.902 0	31.53/0.892 0	36.67/0.971 0
IMDN	2	694	159	38.00/0.960 5	33.63/0.917 7	32.17/0.928 3	32.19/0.899 6	38.88/0.977 4
RFDN	2	534	123	38.05/0.960 6	33.68/0.918 4	32.12/0.927 8	32.16/0.899 4	38.88/0.977 3
LAPAR-A	2	548	171	38.01/0.960 5	33.62/0.918 3	32.10/0.928 3	32.19/0.899 9	38.67/0.977 2
ShuffleMixer	2	394	91	38.01/0.960 6	33.63/0.918 0	31.89/0.925 7	32.17/0.899 5	38.83/0.977 4
This work	2	597	237	38.05/0.960 8	33.62/0.918 1	32.21/0.928 7	32.20/0.900 1	38.81/0.977 7

表 3 3 倍尺度下不同算法的 PSNR/SSIM 对比

Table 3 Comparison of PSNR/SSIM of different algorithms at 3 scale

PSNR/dB

方 法	倍 数	参 数 /K	FLOPs /G	Set5	Set14	Urban100	BSD100	Manga109
				PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
SRCNN	3	57	53	32.75/0.909 0	29.30/0.821 5	26.24/0.798 9	28.41/0.786 3	30.48/0.911 7
FSRCNN	3	12	6	33.18/0.914 0	29.37/0.824 0	26.43/0.808 0	28.53/0.791 0	31.10/0.921 0
IMDN	3	703	72	34.36/0.927 0	30.32/0.841 7	28.17/0.851 9	29.09/0.804 6	33.61/0.944 5
RFDN	3	541	55	34.41/0.927 3	30.34/0.842 0	28.21/0.852 5	29.09/0.805 0	33.67/0.944 9
LAPAR-A	3	594	114	34.36/0.926 7	30.34/0.842 1	28.15/0.852 3	29.11/0.805 4	33.51/0.944 1
ShuffleMixer	3	415	43	34.40/0.927 2	30.37/0.842 3	28.08/0.849 8	29.12/0.805 1	33.69/0.944 8
This work	3	604	106	34.49/0.927 9	30.35/0.842 6	28.24/0.854 0	29.12/0.806 1	33.66/0.945 3

表 4 4 倍尺度下不同算法的 PSNR/SSIM 对比

Table 4 Comparison of PSNR/SSIM of different algorithms at 4 scale

PSNR/dB

方 法	倍 数	参 数 /K	FLOPs /G	Set5	Set14	Urban100	BSD100	Manga109
				PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
SRCNN	4	57	53	30.48/0.862 6	27.50/0.751 3	24.52/0.722 1	26.90/0.710 1	27.58/0.855 5
FSRCNN	4	12	5	30.72/0.866 0	27.61/0.755 0	24.62/0.728 0	26.98/0.715 0	27.90/0.861 0
IMDN	4	715	41	32.21/0.894 8	28.58/0.781 1	26.04/0.783 8	27.56/0.735 3	30.45/0.907 5
RFDN	4	550	32	32.24/0.895 2	28.61/0.781 9	26.11/0.785 8	27.57/0.736 0	30.58/0.908 9
LAPAR-A	4	659	94	32.15/0.894 4	28.61/0.781 8	26.14/0.787 1	27.61/0.736 6	30.42/0.907 4
ShuffleMixer	4	411	28	32.21/0.895 3	28.66/0.782 7	26.08/0.783 5	27.61/0.736 6	30.65/0.909 3
This work	4	613	60	32.19/0.895 1	28.68/0.783 1	26.17/0.788 1	27.60/0.737 3	30.64/0.910 2

2.5 主观视觉分析

为了更加深入地验证提出的 LHNet 方法的有效性,提供 LHNet 和其他轻量级模型的可视化比较。如图 7 所示,展示了不同算法在 Set14 数据集下,将“face.

png”图像放大 4 倍重建的可视化结果。相比于其他方法生成的图片,本文提出的算法重建出的图片脸部雀斑更加清晰,图片色泽也更加鲜艳。这证明了本文算法的优越性。

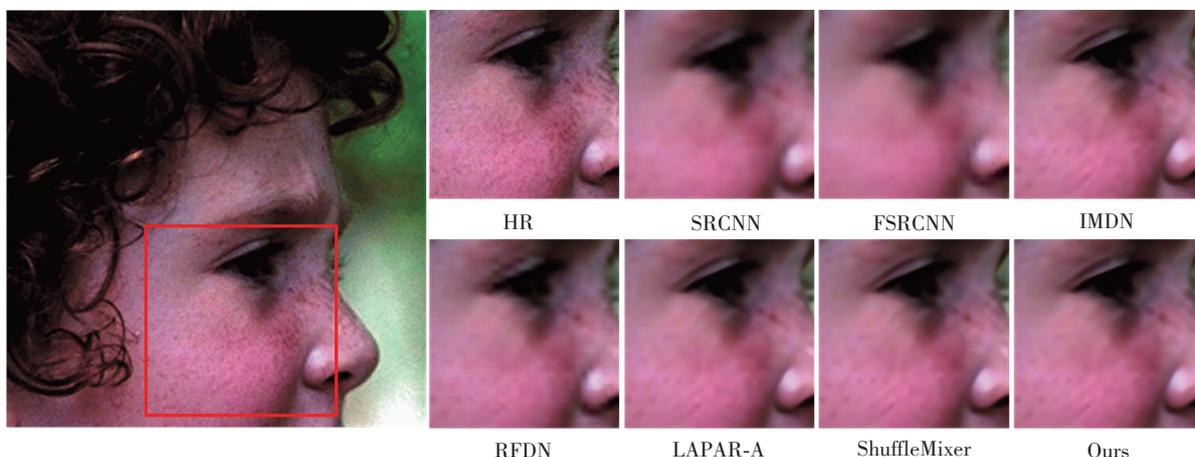


图 7 face.png 4 倍放大重建结果

Fig. 7 Reconstruction results of face.png at 4 magnification

如图 8 所示,为“baboon. png”在放大倍数为 4 的条件下各种算法重建的可视化图片。相比于这些方法, LHNNet 方法重建的图片更加清晰,能够很好地看到 baboon 的胡须。如图 9 所示,为使用“monarch. png”图片在放大倍数为 4 的条件下,各个算法生成的可视化图片。

可以看到其他算法重建的算法边缘不清晰,细节纹理模糊。本文算法重建的图片在边缘部分能够很好地看见蝴蝶翅膀的细节纹理,重建效果更准确,图片质量更佳。

通过一系列对比, LHNNet 方法在重建的清晰度、纹理细节、色泽等方面优于其他算法。

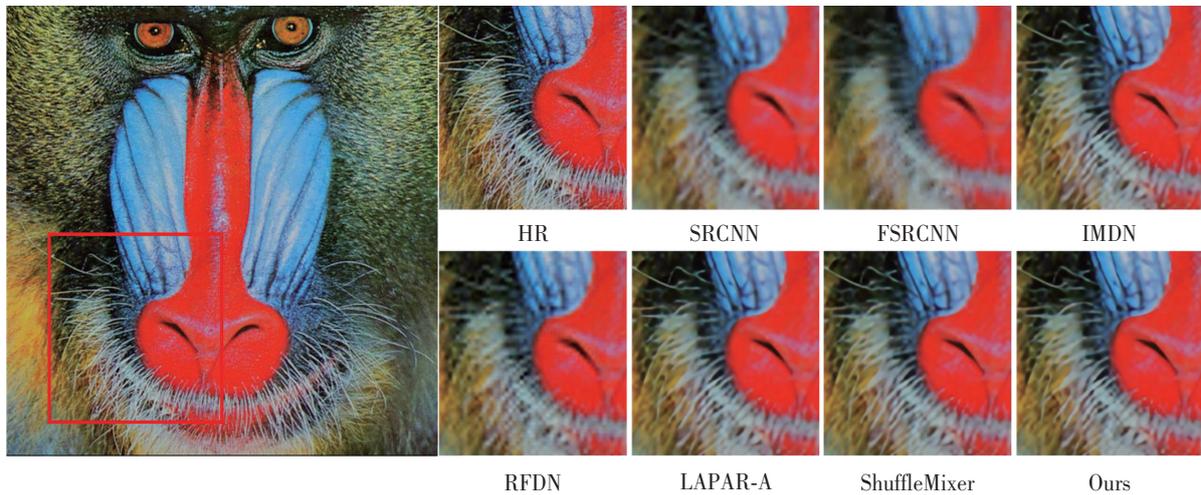


图 8 baboon. png 4 倍放大重建结果

Fig. 8 Reconstruction results of baboon. png at 4 magnification

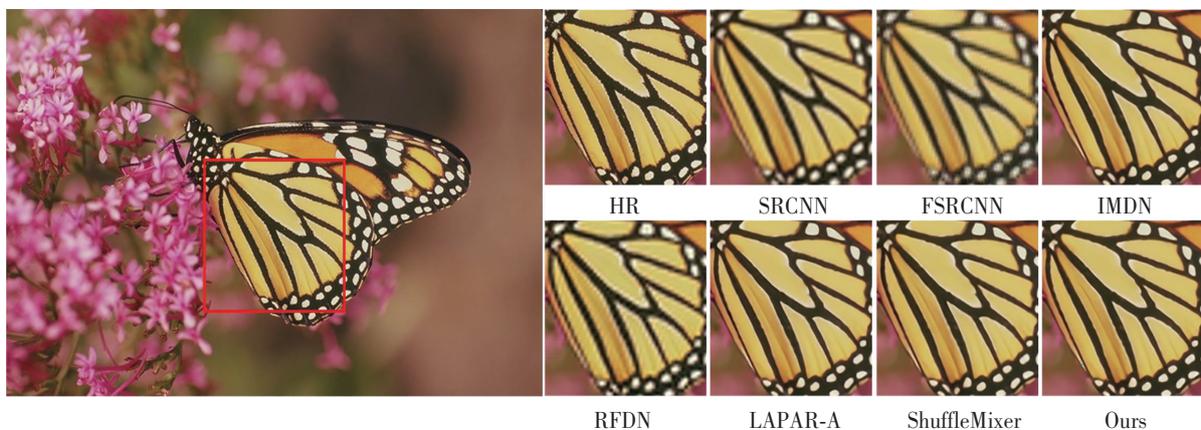


图 9 monarch. png 4 倍放大重建结果

Fig. 9 Reconstruction results of monarch. png at 4 magnification

3 结 论

针对模型参数量和性能之间不平衡问题,本文提出了基于 SDCNN 和 DETN 的混合网络 LHNNet, LHNNet 方法用于图像超分辨率重建任务。为了更加有效地提取局部特征,使用对称方式连接卷积结构,以增强网络特征的提取能力。通过双注意力模块的设计加强网络对像素信息和空间信息的关注。此外,通过密集连接高效 Transformer,尽可能使用较少的参数量进行全局特征学习。实验结果表明:LHNNet 集成 CNN 和 Transformer 的优势在于与同量级的模型相比,拥有更加出色的性能,在模型参数量和性能之间取得了较好的平衡。

参考文献(References):

- [1] FANG J, XIAO J, WANG X, et al. Arbitrary scale super resolution network for satellite imagery[J]. China Communications, 2022, 19(8): 234-246.
- [2] 王梅, 李云红, 李丽敏, 等. 像素补偿的轻量级图像超分辨率重建[J]. 西安工业大学学报, 2023, 43(4): 383-392.
WANG Mei, LI Yun-hong, LI Li-min, et al. Lightweight super-resolution network based on pixel compensation[J]. Journal of Xi'an Technological University, 2023, 43(4): 383-392.
- [3] AHMAD W, ALI H, SHAH Z, et al. A new generative adversarial network for medical images super resolution [J].

- Scientific Reports, 2022, 12: 9533.
- [4] 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.
- [5] LIM B, SON S, KIM H, et al. Enhanced deep residual networks for single image super-resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2017: 1132–1140.
- [6] 任宇, 陈新泉, 王岱嵘, 等. 改进残差网络与峰值帧的微表情识别[J]. 重庆工商大学学报(自然科学版), 2024, 41(1): 21–29.
- REN Yu, CHEN Xin-quan, WANG Dai-rong, et al. Micro-expression recognition based on improved residual network and apex frame[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2024, 41(1): 21–29.
- [7] ZHANG Y, LI K, LI K, et al. Image super-resolution using very deep residual channel attention networks[C]//Proceedings of the Computer Vision-ECCV 2018: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings Part VII, 2018: 294–310.
- [8] WOLF T, DEBUT L, SANH V, et al. Transformers: State-of-the-art natural language processing[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. 2020: 38–45.
- [9] LIU Z, LIN Y, CAO Y, et al. Swin Transformer: Hierarchical vision transformer using shifted windows [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 10012–10022.
- [10] LIANG J, CAO J, SUN G, et al. SwinIR: Image restoration using swin transformer[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2021: 1833–1844.
- [11] CHEN K, LI L, LIU H, et al. SwinFSR: Stereo image super-resolution using SwinIR and frequency domain knowledge[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2023: 1764–1774.
- [12] CHEN Z, ZHANG Y, GU J, et al. Dual aggregation transformer for image super-resolution[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 12278–12287.
- [13] YU L, LI X, LI Y, et al. DIPNet: Efficiency distillation and iterative pruning for image super-resolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2023: 1692–1701.
- [14] WANG Z, GAO G, LI J, et al. Lightweight feature de-redundancy and self-calibration network for efficient image super-resolution[J]. ACM Transactions on Multimedia Computing Communications and Applications, 19(3): 110–119.
- [15] ZHANG W, FAN Z, SONG Y, et al. Lightweight image super-resolution with group-convolutional feature enhanced distillation network[J]. International Journal of Machine Learning and Cybernetics, 2023, 14(7): 2467–2482.
- [16] NGUYEN Q T, QUANG HIEU T. Enhancing single-image super-resolution using patch-mosaic data augmentation on lightweight bimodal network [J]. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems, 2023, 10(2): 47–59.
- [17] ZAMIR S W, ARORA A, KHAN S, et al. Restormer: Efficient transformer for high-resolution image restoration[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 5718–5729.
- [18] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2261–2269.
- [19] DONG C, LOY C C, TANG X. Accelerating the super-resolution convolutional neural network[C]//European Conference on Computer Vision. Cham: Springer, 2016: 391–407.
- [20] HUI Z, GAO X, YANG Y, et al. Lightweight image super-resolution with information multi-distillation network[C]//Proceedings of the 27th ACM International Conference on Multimedia. 2019: 2024–2032.
- [21] LIU J, TANG J, WU G. Residual feature distillation network for lightweight image super-resolution[C]//Bartoli A, Fusiello A. European Conference on Computer Vision. Cham: Springer, 2020: 41–55.
- [22] XU S, ZHANG L. Densely multi-path network for single image super-resolution[C]//Yang H, Pasupa K, Leung ACS, et al. International Conference on Neural Information Processing. Cham: Springer, 2020: 253–265.
- [23] SUN L, PAN J, TANG J. ShuffleMixer: An efficient ConvNet for image super-resolution [EB/OL]. 2022: arXiv: 2205.15175. <http://arxiv.org/abs/2205.15175>.

责任编辑:代小红