

基于注意力机制的残差特征聚合网络超分辨率图像重建研究

孙阳, 丁建伟, 张琪, 魏慧雯, 田博文

引用本文

孙阳, 丁建伟, 张琪, 魏慧雯, 田博文. 基于注意力机制的残差特征聚合网络超分辨率图像重建研究[J]. 计算机科学, 2024, 51(6A): 230600039-6.

SUN Yang, DING Jianwei, ZHANG Qi, WEI Huiwen, TIAN Bowen. [Study on Super-resolution Image Reconstruction Using Residual Feature Aggregation NetworkBased on Attention Mechanism \[J\].](#)

Computer Science, 2024, 51(6A): 230600039-6.

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)

[面向产线AI质检的少样本评测方法研究和验证](#)

Study and Verification on Few-shot Evaluation Methods for AI-based Quality Inspection in Production Lines

计算机科学, 2024, 51(6A): 230700086-8. <https://doi.org/10.11896/jsjcx.230700086>

[基于SAMNV3的滚动轴承智能故障诊断方法](#)

Intelligent Fault Diagnosis Method for Rolling Bearing Based on SAMNV3

计算机科学, 2024, 51(6A): 230700167-6. <https://doi.org/10.11896/jsjcx.230700167>

[基于BERT和CNN的药物不良反应个案报道文献分类方法](#)

Literature Classification of Individual Reports of Adverse Drug Reactions Based on BERT and CNN

计算机科学, 2024, 51(6A): 230400049-6. <https://doi.org/10.11896/jsjcx.230400049>

[基于LSTM和注意力机制的远程会诊需求预测](#)

Forecasting Teleconsultation Demand Based on LSTM and Attention Mechanism

计算机科学, 2024, 51(6A): 230800119-7. <https://doi.org/10.11896/jsjcx.230800119>

[结合图卷积神经网络和集成方法的推荐系统恶意攻击检测](#)

Malicious Attack Detection in Recommendation Systems Combining Graph Convolutional Neural Networks and Ensemble Methods

计算机科学, 2024, 51(6A): 230700003-9. <https://doi.org/10.11896/jsjcx.230700003>

基于注意力机制的残差特征聚合网络超分辨率图像重建研究

孙 阳 丁建伟 张 琪 魏慧雯 田博文

中国人民公安大学信息网络安全学院 北京 100038

摘 要 针对单图像超分辨率算法级联残差块的输出特征仅在局部作用的问题,提出了一种结合注意力机制的残差特征聚合网络。该网络通过跳跃连接将各残差块输出不同层次的特征聚合到残差组的尾部,实现特征的充分提取与复用,扩大网络的感受野并增强特征的表达能力,使得不同层次的特征图更充分地参与到图像重建中。同时,为增强特征信息空间上的相关性,引入增强空间注意力机制以改善残差块的性能。大量实验表明,此模型可以获得良好的超分辨率性能。在 $\times 4$ 倍SR任务中与RCAN,SAN和HAN等主流方法相比,在5个基准测试集上取得的峰值信噪比平均提升0.07 dB,0.06 dB,0.006 dB,结构相似度平均提升0.0012,0.0011,0.0008,重建图像质量明显提高,细节更加丰富,充分说明了所提方法的有效性与先进性。

关键词: 图像超分辨率重建;深度学习;注意力机制;特征聚合;卷积神经网络

中图分类号 TP391.4

Study on Super-resolution Image Reconstruction Using Residual Feature Aggregation Network Based on Attention Mechanism

SUN Yang,DING Jianwei,ZHANG Qi,WEI Huiwen and TIAN Bowen

School of Information Network Security,People's Public Security University of China,Beijing 100038,China

Abstract To address the problem of the local effect of the output features of cascaded residual blocks in single image super-resolution algorithms,a residual feature aggregation network combined with attention mechanism is proposed. The network aggregates the features of different levels output by each residual block through skip connections to the end of the residual group, achieves sufficient feature extraction and reuse,expands the receptive field of the network and enhances the expression ability of features. Meanwhile,to improve the spatial correlation of feature information,an enhanced spatial attention mechanism is introduced to improve the performance of the residual blocks. Extensive experiments demonstrate that the proposed model achieves good super-resolution performance. Compared with state-of-the-art methods such as RCAN,SAN,and HAN,the proposed method demonstrates significant effectiveness and advancement in the task of $\times 4$ super-resolution. On five benchmark datasets,our method achieves an average improvement of 0.07 dB,0.06 dB,and 0.006 dB in peak signal-to-noise ratio,as well as an average improvement of 0.0012,0.0011,and 0.0008 in structural similarity index. The reconstructed images exhibit a notable increase in quality,with more abundant details. These results verifies he efficacy and advancement of the proposed method.

Keywords Image super-resolution reconstruction,Deep learning,Attention mechanisms,Feature aggregation,Convolutional neural networks

1 引言

单图像超分辨率(Single Image Super-Resolution,SISR)技术指将低分辨率(Low Resolution,LR)图像通过相应的算法将其重建成高分辨率(High Resolution,HR)图像。相对于LR图像,HR图像拥有更大的像素密度和更丰富的细节信息,在刑侦、国防、医学等领域有广泛应用,是图像复原领域的热点研究问题之一。且SISR具有不适定(ill-posed)特性,即一个LR输入可以对应许多可能的HR输出,映射空间太大,其仍然具有挑战性。

超分辨率技术包括基于插值的方法^[1]、基于重建的方法^[2]和基于学习的方法。但通过基于插值与基于重建的方法

得到的图像质量较差且效率极低。Dong等^[3]于2014年提出的SRCNN开创了基于学习的超分辨率技术的先河。近年来,得益于神经网络的残差结构和深层架构,基于卷积神经网络(Convolutional Neural Network,CNN)的图像超分辨率重建方法获得了重大的进展。VDSR^[4],SRResNet^[5],EDSR^[6]等方法将残差结构引入图像超分辨率方法中增加网络的深度,从而增强图像重建质量。Zhang等^[7]将通道注意力(Channel Attention,CA)机制引入到超分网络中构建残差通道注意力网络RCAN,采用嵌套残差结构堆叠多个残差通道注意力块(Residual Channel Attention Block,RCAB),在单图像超分辨率任务中实现了良好的性能。但作为目前卷积神经网络在超分辨率任务中的主流模型,其还有很多不足。首先,

基金项目:中国人民公安大学安全防范工程双一流专项(2023SYL08)

This work was supported by the People's Public Security University of China Double First-class Project on Security and Prevention Engineering (2023SYL08).

通信作者:丁建伟(jwding@ppsuc.edu.cn)

经典卷积结构都是通过残差块的连接构成,当前残差块的输出只能作为下一残差块的输入,不同残差块输出的 LR 图像分层特征没有得到充分利用;其次,不同的残差块可以反映空间方面的层次特征,但是使用 CA 机制捕捉的残差特征还不够突出,没有充分利用 LR 图像各分层特征图的信息,并且特征提取残差块的网络容量较低,限制了超分网络的学习能力。针对目前卷积神经网络在超分任务中存在的一系列问题,提出了一种全新的卷积神经网络,其中最主要的贡献点包括:

1) 提出了一种基于注意力机制的残差特征聚合密集网络 (Residual Feature Aggregation Dense Network, RFADN), 该网络在超分辨率任务中展现出了极大的优势。

2) 关注残差块输出的 LR 图像分层特征, 搭建残差特征聚合框架, 充分复用图像特征信息, 提高卷积神经网络的图像重建能力。

3) 引入增强空间注意力 (Enhanced Spatial Attention, ESA) 机制构建密集空间注意力块 (Dense Spatial Attention Block, DSAB), 增强网络容量和学习能力, 使网络能够恢复 LR 图像更多的高频细节, 大幅提高了图像的重建质量。

2 相关工作

2.1 基于卷积神经网络的超分辨率重建

SRCNN^[3] 采用经典的端到端结构, 通过简单的三层卷积结构在保证图像质量的同时又可以实现高分辨率图像的快速重建。Shi 等^[8] 通过对像素进行重新排列提出了 ESPCN 算法, 降低了网络的计算量, 提高了模型的上采样效率, 增强了图像的重建效果。Kim 等^[4] 将残差网络引入单图像超分辨率方法中, 构建了一个极深的残差网络 VDSR, 同时加速了网络的收敛速度。Lim 等^[6] 在 SRResNet^[5] 的基础上舍弃了批归一化 (Batch Normalization, BN) 层, 在网络层数加深的同时计算量也大大得到简化, 图像重建效果也有所提高。Zhang

等^[7] 创新性地提出嵌套残差结构堆叠多个 RCAB, 区分不同通道特征对超分图像重建的重要程度并进行赋权, 重建图像质量得到极大提升。

以上方法虽然取得了非常不错效果, 但是这些方法都是通过简单堆叠残差块实现的, 没有利用分层特征信息帮助重建图像, 造成了特征浪费, 因此本文采用特征聚合框架充分利用各残差块输出的分层特征, 获取更多包含细节、纹理的特征, 实现特征的复用。

2.2 基于注意力机制的超分辨率重建

注意力机制指导网络关注输入特征图信息量较大的部分, 从而产生更具分辨性的特征表示, 被广泛地应用于计算机视觉任务中。Hu 等^[9] 提出的 SENet 通过对特征图进行处理变换, 建立通道依赖关系, 极大地改善了图像分类任务的性能。RCAN^[7] 首次将注意力机制引入到超分辨率任务中, CA 机制通过考虑各通道间的相互依赖性, 自适应调整每个特征通道的权重。SAN^[10] 通过引入高阶注意力机制捕获长距离信息相关性, 从而获得更好的特征表达。HAN^[11] 采用整体注意力网络模拟层、位置、通道之间的相互依存关系, 使网络聚焦于信息量更大的特征中, 提升了模型的重建能力。

以上方法通过使用注意力机制使重建图像达到了良好的视觉效果, 且超分任务需要残差特征包含更多的空间结构信息来建立特征长距离依赖关系, 以提高残差特征聚合网络的有效性, 因此需要具有较大感受野且轻量级的注意力模块来增强模型的表现能力。

3 本文方法

3.1 网络总体结构

本文所提出的 RFADN 网络结构如图 1 所示, 包括浅层特征提取、深层特征提取、上采样模块、重建模块 4 部分。

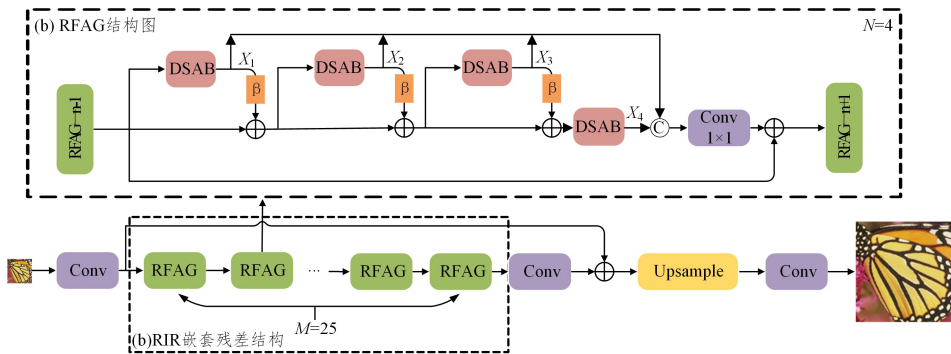


图 1 基于注意力机制的残差特征聚合网络结构

Fig. 1 Residual feature aggregation network structure based on attention mechanism

首先, 给定输入图像 $I_{LR} \in R^{H \times W \times C}$, 使用一个 3×3 卷积层获得图像浅层特征 F_0 :

$$F_0 = H_{\text{conv}}(I_{LR}) \quad (1)$$

其中, $H_{\text{conv}}(\cdot)$ 表示卷积核大小为 3×3 的卷积层。然后将 F_0 输入到通过嵌套残差 (Residual in residual, RIR) 结构堆叠的 M 个残差特征聚合组 (Residual Feature Aggregation Group, RFAG) 深层特征提取模块中, 以获得图像的深层特征 F_{DF} 。另外, 采用嵌套残差结构可以堆叠多个残差块, 形成深度的神经网络。每个残差块由多个卷积层和激活函数组成, 允许网络逐渐提取更高级别的特征。同时残差连接允许网络

在学习过程中跳过一些层的操作, 从而减轻了梯度消失问题, 有助于更深层次的网络训练。RFAG 由 N 个 DSAB 块构成, M 和 N 的具体参数请见 4.1 节。

$$F_{DF} = H_{RIR}(F_0) \quad (2)$$

其中, $H_{RIR}(\cdot)$ 表示嵌套残差结构, 能够达到很深的深度, 提供非常大的感受野。深层特征 F_{DF} 通过上采样模块进行上采样。

$$F_{UP} = H_{UP}(F_{DF} + F_0) \quad (3)$$

其中, $H_{UP}(\cdot)$ 采用 ESPCN^[8] 所提出的亚像素卷积。最后通过一个卷积层将放大后的特征重建至 SR 图像。

$$I_{SR} = HR(FUP) = HRFADN(I_{LR}) \quad (4)$$

其中, $HR(\cdot)$ 表示卷积核大小为 3×3 的卷积层, I_{SR} 表示重建得到的高分辨率图像。

尽管 L_1 损失函数^[12]、 L_2 损失函数^[13]、感知损失^[14]、对抗损失^[15]等多种损失函数的组合会为重建图像带来视觉效果的提升,但是为了验证 RFADN 网络的有效性,本文采用与前人工作相同的损失函数(L_1 损失函数)进行优化。 L_1 损失函数可以提供稳定的梯度,使得模型可以快速收敛。给定训练集 $\{I_{LR}, I_{HR}\}_{i=1}^N$, 包括 N 对相应的 LR 与 HR 图像, 训练 RFADN 网络的目标是最小化 L_1 损失函数:

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N \| H_{RFADN}(I_{LR}^i) - I_{HR}^i \|_1 \quad (5)$$

其中, θ 表示 RFADN 网络模型的参数, $H_{RFADN}(\cdot)$ 表示 RFADN 网络模型函数。

3.2 残差特征聚合

残差网络^[16]通过引入跳跃连接的方式来缓解神经网络深度增加可能会产生的梯度消失或梯度爆炸问题,使网络学习难度减小且更容易优化。在 RCAN 和 SAN 网络中,残差块按序堆叠在网络中形成残差组(Residual Group, RG)的主干部分。在超任务中,由于各残差块的输入不同,其输出也会包含 LR 图像的不同层次特征。一个残差块输出的特征通常要经过许多残差块进行不同运算才能被输入到下一模块中,特征利用率偏低。这样就导致了 LR 图像不同层次的特征很难被充分利用,从而限制了卷积神经网络的学习与表达能力。

利用各层次特征实现高精度图像重建是非常必要的, Zhang 等^[17]在堆叠的密集块(Dense Block, DB)中提取层次特征进行整合,取得了良好的效果。同样, Li 等^[18]将每个多尺度残差块(Multi-Scale Residual Blocks, MSRBs)的输出发送到网络尾部进行图像重建。但是这些方法都未考虑到浅层残差块与深层残差块之间所输出特征的相似性与差异性,同时当堆叠残差块数量过多时,多通道聚合特征将大大影响计算效率。

因此,采用残差特征聚合组^[19]实现特征复用,其结构如图 1(b)所示。不同于 SR 任务中的其他网络,本文所使用的网络结构可以有效地利用各残差块输出的 LR 图像不同的层次特征。前 3 个 DSAB 输出的残差特征直接与第四个 DSAB 输出的残差特征相连,然后通过一个 1×1 卷积对特征进行通道维度的融合。与简单堆叠多个残差块的 SR 网络相比,残差特征聚合组可以使 LR 图像不同层次的特征更大程度地参与到图像重建中,并且通过这种方式,不同残差块输出的不同特征不会相互干扰。

Wang 等^[20]结合密集块和 RIR 结构构建了嵌套残差密集块(Residualin Residual Dense Block, RRDB),其兼具残差连接和密集连接的优点,拥有更大的容量、更低的内存占用、更小的计算成本等优点,更易于训练与优化。受 RRDB 的启发,在将密集空间注意力块输出的特征输入到主干网络时,将各密集空间注意力块的输出乘以一个缩放权重 β 以调整残差,从而提高网络的稳定性。残差特征聚合组的整个过程可以表示为:

$$F_{RFAG} = Conv(cat(x_1, x_2, x_3, x_4); \theta_1) \oplus x \quad (6)$$

其中, F_{RFAG} 表示残差特征聚合组输出的特征, (x_1, x_2, x_3, x_4)

分别表示 4 个 DSAB 块的输出, $Conv(\cdot)$ 表示卷积核大小为 1×1 大小的卷积层, θ_1 表示卷积层的超参数, $cat(\cdot)$ 表示对 4 个残差输出按通道维度进行拼接, \oplus 表示逐像素相加操作。

3.3 密集空间注意力块

RCAN^[6]在其残差块中引入注意力机制,使得网络聚焦于信息量更大的通道上。但是通过 CA 机制捕获的残差特征还不够突出,同时 RCAB 的网络容量较低,限制了其在超任务中的学习能力。SAN^[10]在其残差块中引入高阶注意力机制提取长距离信息依赖性,但是高阶注意力机制十分消耗计算资源,限制了模型的应用场景。

为了使每个残差块输出的特征更具有辨别性,也为了使网络结构能够最大限度地发挥作用,构建了 DSAB 模块作为网络的基础块,使网络拥有更高的容量与更强的学习能力,其结构如图 2 所示。

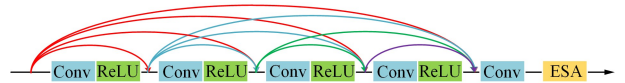


图 2 密集空间注意力块

Fig. 2 Dense space attention block(DSAB)

为了堆叠多个残差块,模型需要的注意力机制必须是轻量级的,同时为了更好地完成 SR 任务,还需要较大的感受野。受 Liu 等^[19]的启发,采用 ESA 机制增强密集块的表达能力,其结构如图 3 所示。ESA 模块使用残差结构将输入特征直接发送到模块的末端,为了使整个模块可以轻量化,首先使用 1×1 卷积层压缩通道维度,有助于提取更具表达力的特征。然后使用一个步长为 2 的跨步卷积以及窗口大小为 7 和步长为 3 的最大池化层快速降低空间维度,使模型可以更集中地关注特征图中的重要信息。与下采样层相对应,通过上采样层恢复空间维度,然后使用一个 1×1 卷积层恢复通道维度,保持特征图的通道数不变,以便进行后续的特征融合和加权,最后使用 Sigmoid 函数生成注意力掩码,以指示特征图中的重要区域。

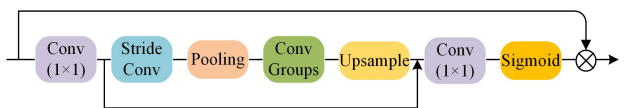


图 3 增强空间注意力机制

Fig. 3 Enhanced spatial attention mechanism(ESA)

4 实验

4.1 数据集及训练细节

本文使用 DIV2K 数据集^[22]的 800 张训练图像对 RFADN 网络进行训练,为了扩充数据集,对 800 张训练图像随机旋转 $90^\circ, 180^\circ, 270^\circ$ 或者水平翻转,在每次迭代中使用 16 个大小为 48×48 的 LR 图像块作为输入。同时使用拥有不同特征的 5 个标准测试集 Set5, Set14, BSD100, Urban100, Manga109 进行测试,并与其他方法进行对比。使用 Adam 优化器进行训练,其中, $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 1 \times 10^{-8}$, 初始学习率 $lr = 1 \times 10^{-4}$, 之后每 1.5×10^5 次反向传播迭代后学习率降为之前的一半。

使用 PyTorch 框架在 NVIDIA GeForce RTX 3080 显卡上实现。在 RIR 结构中,设置 RFAG 为 $M = 25$, 每个 RFAG 中 DSAB 数为 $N = 4$, 缩放权重 $\beta = 0.2$ 。

4.2 实验结果分析

为了验证 RFADN 网络的有效性,在输出图像的 YCbCr 空间 Y 通道上计算峰值信噪比 (Peak signal-to-noise ratio, PSNR) 和结构相似度 (Structural similarity, SSIM) 指标评价 SR 结果。本文通过与 Bicubic, SRCNN, VDSR, EDSR, RDN,

DBPN, RCAN, SAN 以及 HAN 进行视觉与指标对比来进行验证。表 1—表 3 列出了以上经典算法与 RFADN 网络重建效果的 PSNR 值和 SSIM 值。其中, PSNR 值越高表示重建效果越好, SSIM 值越接近 1 表示 SR 图像与 HR 图像越接近。

表 1 各模型在不同数据集上的表现($\times 2$)

Table 1 Performance of each model on different datasets($\times 2$)

Method	Scale	Set5		Set14		BSD100		Urban100		Manga109	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	$\times 2$	33.66	0.9299	30.24	0.8688	29.56	0.8431	26.88	0.8403	30.80	0.9339
SRCNN ^[3]	$\times 2$	36.66	0.9542	32.45	0.9067	31.36	0.8879	29.50	0.8946	35.60	0.9663
VDSR ^[4]	$\times 2$	37.05	0.9560	33.05	0.9130	31.90	0.8960	30.77	0.9140	37.22	0.9750
EDSR ^[6]	$\times 2$	38.11	0.9602	33.92	0.9195	32.32	0.9013	32.93	0.9351	39.10	0.9773
RDN ^[15]	$\times 2$	38.24	0.9614	34.01	0.9212	32.34	0.9017	32.89	0.9353	39.18	0.9780
DBPN ^[23]	$\times 2$	38.09	0.9600	33.85	0.9190	32.27	0.9000	32.55	0.9324	38.89	0.9775
RCAN ^[6]	$\times 2$	38.27	0.9614	34.11	0.9216	32.41	0.9026	33.34	0.9384	39.43	0.9786
SAN ^[10]	$\times 2$	38.31	0.9620	34.07	0.9213	32.42	0.9028	33.10	0.9370	39.32	0.9792
HAN ^[11]	$\times 2$	38.27	0.9614	34.16	0.9217	32.41	0.9027	33.35	0.9385	39.46	0.9785
MCFSR ^[24]	$\times 2$	<u>38.29</u>	0.9618	34.15	0.9212	32.35	0.9030	<u>33.35</u>	0.9389	39.47	<u>0.9788</u>
RFADN(ours)	$\times 2$	38.26	0.9622	34.11	0.9217	<u>32.41</u>	0.9031	33.39	0.9389	39.41	0.9788

表 2 各模型在不同数据集上的表现($\times 3$)

Table 2 Performance of each model on different datasets($\times 3$)

Method	Scale	Set5		Set14		BSD100		Urban100		Manga109	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	$\times 3$	30.39	0.8682	27.55	0.7742	27.21	0.7385	24.46	0.7349	26.95	0.8556
SRCNN ^[3]	$\times 3$	32.75	0.9090	29.30	0.8215	28.41	0.7863	26.24	0.7989	30.48	0.9117
VDSR ^[4]	$\times 3$	33.67	0.9210	29.78	0.8320	28.83	0.7990	27.14	0.8290	32.01	0.9340
EDSR ^[6]	$\times 3$	34.65	0.9280	30.52	0.8462	29.25	0.8093	28.80	0.8653	34.17	0.9476
RDN ^[15]	$\times 3$	34.71	0.9296	30.57	0.8468	29.26	0.8093	28.80	0.8653	34.13	0.9484
RCAN ^[6]	$\times 3$	34.74	0.9299	30.64	0.8481	<u>29.32</u>	0.8111	29.08	0.8702	34.43	0.9498
SAN ^[10]	$\times 3$	34.75	<u>0.9300</u>	30.59	0.8476	29.33	0.8112	28.93	0.8671	34.30	0.9494
HAN ^[11]	$\times 3$	34.75	0.9299	<u>30.67</u>	<u>0.8483</u>	<u>29.32</u>	0.8110	29.10	<u>0.8705</u>	34.48	<u>0.9500</u>
MCFSR ^[24]	$\times 3$	<u>34.77</u>	0.9298	<u>30.67</u>	0.8477	29.26	0.8123	29.14	0.8716	34.45	0.9497
RFADN(ours)	$\times 3$	34.79	0.9306	30.69	0.8491	29.33	0.8112	29.13	0.8716	34.44	0.9505

注:DBPN 公布数据无 $\times 3$ 倍 SR 结果。

表 3 各模型在不同数据集上的表现($\times 4$)

Table 3 Performance of each model on different datasets($\times 4$)

Method	Scale	Set5		Set14		BSD100		Urban100		Manga109	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	$\times 4$	28.42	0.8104	26.00	0.7027	25.96	0.6675	23.14	0.6577	24.89	0.7866
SRCNN ^[3]	$\times 4$	30.48	0.8628	27.50	0.7513	26.90	0.7101	24.52	0.7221	27.58	0.8555
VDSR ^[4]	$\times 4$	31.35	0.8830	28.02	0.7680	27.29	0.7251	25.18	0.7540	28.83	0.8870
EDSR ^[6]	$\times 4$	32.46	0.8968	28.80	0.7876	27.71	0.7420	26.64	0.8033	31.02	0.9148
RDN ^[15]	$\times 4$	32.47	0.8990	28.81	0.7871	27.72	0.7419	26.61	0.8028	31.00	0.9151
DBPN ^[23]	$\times 4$	32.47	0.8980	28.82	0.7860	27.72	0.7400	26.38	0.7946	30.91	0.9137
RCAN ^[6]	$\times 4$	32.63	0.9002	28.87	0.7889	27.77	0.7436	26.82	0.8087	31.22	0.9173
SAN ^[10]	$\times 4$	32.64	0.9003	28.92	0.7888	27.78	0.7436	26.79	0.8068	31.18	0.9169
HAN ^[11]	$\times 4$	32.64	0.9002	28.90	0.7890	27.80	0.7442	26.85	<u>0.8094</u>	31.42	<u>0.9177</u>
MCFSR ^[24]	$\times 4$	32.65	<u>0.9013</u>	28.88	<u>0.7891</u>	27.75	<u>0.7441</u>	<u>26.88</u>	0.8061	31.27	0.9175
ERAN ^[25]	$\times 4$	<u>32.66</u>	0.8999	28.92	<u>0.7891</u>	<u>27.79</u>	0.7429	26.86	0.8073	<u>31.39</u>	0.9172
RFADN(ours)	$\times 4$	32.69	0.9016	<u>28.91</u>	0.7896	27.80	0.7442	26.90	0.8101	31.34	0.9190

注:ERAN 公布数据只含 $\times 4$ 倍 SR 结果。

为了展示 RFADN 网络的有效性与先进性,表 1—表 3 列出了不同方法在 5 个标准测试集上进行不同尺度图像超分的对比结果,加粗数值表示为最佳结果,下划线数值表示次佳结果。在 $\times 4$ 倍的 SR 任务中,RFADN 网络与其他主流模型在 Set5, Urban100, BSD100 测试集上相比, PSNR 和 SSIM 指标均取得了最好的表现,在 Set14 和 Manga109 测试集中, SSIM 取得了最优的表现, PSNR 指标都十分接近最佳结果。在 $\times 3$ 倍的 SR 任务中,RFADN 网络与其他主流模型在 5 个基准测试集上相比, PSNR 均获得了最佳指标或与最佳指标非常接近的结果。同时 SSIM 均取得了最优或次优的结果。另外在 $\times 2$ 倍的 SR 任务中,RFADN 网络的表现虽然不如在 $\times 3$ 和 $\times 4$ 倍 SR 任务中具有统治性,但综合衡量指标,其也

远远优于之前的超分模型。从另一种角度来说,当进行 $\times 4$ 或 $\times 3$ 倍 SR 任务时,信息丢失比 $\times 2$ 倍更严重。这意味着从 LR 到 HR 的重建更具挑战性,需要模型能够更好地推断丢失的细节。因此,RFADN 网络在更高的缩放因子下会表现更优秀,其强大的学习能力可以还原更多的细节。

值得强调的是,RFADN 网络在 Urban100 测试集上表现良好,这是因为其包含了丰富的结构信息,RFADN 网络可以分层聚合这些信息,从而产生更具有代表性的特征和更为丰富的细节信息。另外,需要注意到,RFADN 网络在 5 个不同测试集上得到的 SSIM 结果均达到了最优或非常接近最优的结果,这说明 RFADN 网络重建得到的图像更接近原图像,也说明了 RFADN 网络的先进性与鲁棒性。

图 4、图 5 给出了不同模型在 $\times 4$ 倍 SR 任务中的主观视觉感知评价,即不同网络所重建图像之间的对比图,并对局部区域进行标注与放大,以便直观地感受不同模型所重建图像的细节与纹理。

图 4 给出了 BSD100 数据集 86000.png 图像的测试结果,传统 Bicubic 插值法重建得到的图像十分模糊,图像的细节与边缘基本丢失,图像内容无法恢复甚至丢失。VDSR,EDSR 等模型虽然较传统插值法有了很大的提升,但会出现细节模糊、纹理丢失等问题。RCAN,SAN,HAN 模型基本可以重建图像内容,但其分别在细节、清晰度、边缘方面还可以进一步改善。综合观察,RFADN 网络所重建的图像内容更清晰,边缘更锐化,细节更丰富,也与 HR 图像更为接近。

图 5 给出了 Urban100 数据集中 img_092.png 图像的 $\times 4$ 倍 SR 结果。Bicubic 插值法重建结果十分模糊,图像细节几乎无法展现。基于深度学习的其他方法重建图像可以恢复主要轮廓,但是在纹理结构、细节内容等方面都呈现不同程度的扭曲与伪影。相比之下,RFADN 网络所重建的图像具有轮廓清晰、边缘锐化、细节丰富等优点,通过恢复更多的信息获得更好的结果。这正好展现了 RFADN 网络对于 Urban100 这类具有丰富细节纹理与层次结构图像的重建优势。并且相比其他网络,RFADN 网络生成的 HR 图像在一定程度上缓解了 ill-posed 问题,更加展现了 RFADN 网络的强大能力。

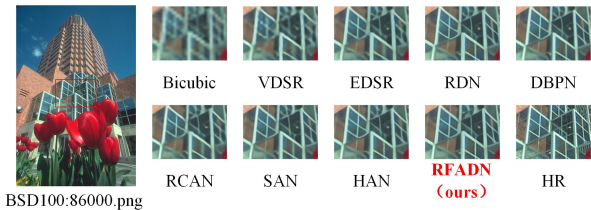


图 4 BSD100 数据集中 86000.png 的测试结果 ($\times 4$)

Fig. 4 Test results of 86000.png on BSD100 dataset ($\times 4$)

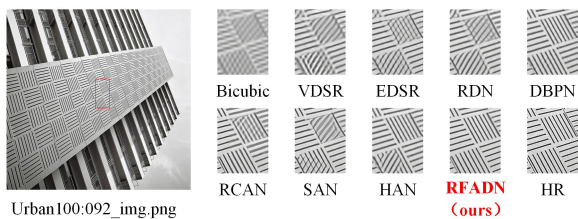


图 5 Urban100 数据集中 img_092.png 的测试结果 ($\times 4$)

Fig. 5 Test results of img_092.png on Urban100 dataset ($\times 4$)

4.3 消融实验

为了验证 RFADN 网络中 DSAB 模块与特征聚合结构的有效性,对各结构进行消融实验,并对实验结果进行对比与分析。为了更清晰地展示不同模块的有效性,只在 Set5 数据集上测试 PSNR 和 SSIM 两个指标。在不同模块的 $\times 4$ 倍 SR 任务中,遵循本文 4.1 节的训练细节,具体性能如表 4 所列。

表 4 在 Set5 测试集上的消融实验 ($\times 4$)

Table 4 Ablation study on Set5 ($\times 4$)

	×	√	×	√
RFA	×	√	×	√
DSAB	×	×	√	√
PSNR	32.54	32.58	3.63	32.69
SSIM	0.9000	0.9003	0.9005	0.9016

表 4 中,RFA 表示残差特征聚合框架,DSAB 表示密集空间注意力块。

由表 4 中的第 3、4 列可知,相比 Baseline,在分别引入

RFA 和 DSAB 之后,PSNR 值分别提升了 0.04 dB,0.09 dB,SSIM 分别提升了 0.0003,0.0005。需要强调的是,为了使消融实验各网络卷积层数目大致持平,使用密集块 (Dense Block,DB) 替换 DSAB。由表 4 中的第 5 列可见,在整合了 RFA 和 DSAB 的 RFADN 网络两项指标达到了最大值,相比 Baseline,PSNR 提高了 0.15 dB,SSIM 提高了 0.0016。

为了验证 RFAG 模块数量 M 对网络性能的影响,本文构建不同 M 的 RFADN 网络进行消融实验。具体来说,分别使 $M=10, M=20, M=25, M=30$ 搭建 4 个不同参数数量的 RFADN 网络,并在 5 个标准数据集上测试 PSNR 值评估模型,实验结果如表 5 所列。

由表 5 可知,当 $M=25$ 时,RFADN 网络各项指标最优。参数数量过小,神经网络无法学习到图像深层特征;参数数量过大,会导致网络难以训练,容易出现过拟合现象,这说明神经网络的复杂程度与所求解问题相匹配才能获得最好的效果。

表 5 不同 RFAG 数量的消融实验 ($\times 4$)

Table 5 Ablation study with different numbers of RFAG ($\times 4$)

M	Set5	Set14	B100	Urban100	Manga109
10	32.47	28.80	27.71	26.59	31.02
20	32.55	28.86	27.77	26.75	31.22
25	32.69	28.91	27.80	26.90	31.34
30	32.64	28.88	27.79	26.84	31.30

为了验证控制缩放权重 β 对训练过程中网络稳定性的影响,以 $\beta=0.1, \beta=0.2, \beta=0.5, \beta=1.0$ 分别进行消融实验,并在标准测试集上测试 PSNR 指标以获得最佳缩放权重 β ,实验结果如表 6 所列。

表 6 不同缩放参数 β 的消融实验 ($\times 4$)

Table 6 Ablation study with different parameters β ($\times 4$)

β	Set5	Set14	B100	Urban100	Manga109
0.1	32.69	28.89	27.78	26.84	31.32
0.2	32.69	28.91	27.80	26.90	31.34
0.5	32.66	28.88	27.82	26.82	31.28
1.0	32.65	28.84	27.72	26.78	31.25

由表 6 可知,当 $\beta=1.0$ 时,即对于 DSAB 的输出不进行控制缩放时,模型的性能远低于采用缩放参数。因为缩放参数 β 可以避免梯度消失或梯度爆炸问题,有助于稳定训练过程,使模型能够更有效地更新参数。且当 $\beta=0.2$ 时,RFADN 获得的 PSNR 指标在 5 个基准测试集上的表现更为优异。

结束语 本文针对目前超分辨率卷积神经网络中各残差块输出的分层特征未充分利用且模型容量小、学习能力较弱的问题,提出了基于注意力机制的特征聚合密集网络 RFADN。该网络模型首先构建密集空间注意力块提升网络的容量与残差块的感受野,然后使用特征聚合框架聚合来自不同残差块输出的层次特征信息,同时利用 RIR 结构堆叠多个 DSAB 加深网络的深度,增强模型的学习能力。与 SR 任务中主流的卷积神经网络相比,RFADN 网络模型能够重建出内容更丰富、纹理更清晰、边缘更锐利的 SR 图像,在得到良好视觉效果的同时,也获得了最为优异的客观评价指标。综合来看,RFADN 网络在图像超分辨率技术中具有先进性与优异性。

虽然 RFADN 网络在 SR 任务中的表现较主流网络有了长足的提升,但是模型在某类数据集上的表现欠佳,后续将通过分析这类数据集的特点对 RFADN 网络进行改进,从而实现泛化性能更佳的超分辨率模型。

参考文献

- [1] ZHANG L, WU X. An edge-guided image interpolation algorithm via directional filtering and data fusion[J]. *IEEE Transactions on Image Processing*, 2006, 15(8): 2226-2238.
- [2] ZHANG K, GAO X, TAO D, et al. Single image super-resolution with non-local means and steering kernel regression[J]. *IEEE Transactions on Image Processing*, 2012, 21(11): 4544-4556.
- [3] DONG C, LOY C C, HE K, et al. Learning a deep convolutional network for image super-resolution[C]// *Computer Vision-ECVCV 2014: 13th European Conference*. 2014: 184-199.
- [4] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016: 1646-1654.
- [5] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realistic single image super-resolution using a generative adversarial network [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017: 4681-4690.
- [6] 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: 136-144.
- [7] ZHANG Y, LI K, LI K, et al. Image super-resolution using very deep residual channel attention networks[C]// *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018: 286-301.
- [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]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016: 1874-1883.
- [9] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018: 7132-7141.
- [10] DAI T, CAI J, ZHANG Y, et al. Second-order Attention Network for Single Image Super-Resolution[C]// *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2019.
- [11] NIU B, WEN W, REN W, et al. Single image super-resolution via a holistic attention network[C]// *Computer Vision-ECCV 2020: 16th European Conference*. 2020: 191-207.
- [12] LAI W S, HUANG J B, AHUJA N, et al. Deep laplacian pyramid networks for fast and accurate super-resolution[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017: 624-632.
- [13] 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*, 2015, 38(2): 295-307.
- [14] JOHNSON J, ALAHI A, FEI-FEI L. Perceptual losses for real-time style transfer and super-resolution[C]// *Computer Vision-ECCV 2016: 14th European Conference*. 2016: 694-711.
- [15] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative Adversarial Nets[J]. *arXiv:1406.2661*, 2014.
- [16] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016: 770-778.
- [17] ZHANG Y, TIAN Y, KONG Y, et al. Residual dense network for image super-resolution[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018: 2472-2481.
- [18] LI J, FANG F, MEI K, et al. Multi-scale residual network for image super-resolution[C]// *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018: 517-532.
- [19] LIU J, ZHANG W, TANG Y, et al. Residual feature aggregation network for image super-resolution [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020: 2359-2368.
- [20] WANG X, YU K, WU S, et al. Esrgan: Enhanced super-resolution generative adversarial networks [C] // *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. 2018.
- [21] LIU J, TANG J, WU G. Residual feature distillation network for lightweight image super-resolution [C] // *Computer Vision-ECVCV 2020 Workshops*. 2020: 41-55.
- [22] AGUSTSSON E, TIMOFTE R. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study[C]// *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2017.
- [23] HARIS M, SHAKHAROVICH G, UKITA N. Deep back-projection networks for super-resolution[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018: 1664-1673.
- [24] YANG X, ZHU Y, GUO Y, et al. An image super-resolution network based on multi-scale convolution fusion[J]. *The Visual Computer*, 2022, 38(12): 4307-4317.
- [25] HAO F, ZHANG T, ZHAO L, et al. Efficient residual attention network for single image super-resolution[J]. *Applied Intelligence*, 2022, 52(1): 652-661.
- [26] LIU Z X, ZHU C J, HUANG J, et al. Image Super-resolution by Residual Attention Network with Multi-skip Connection [J]. *Chinese Journal of Computers*, 2021, 48(11): 258-267.
- [27] GUO L, LI C, CHEN C, et al. Image Super-Resolution Reconstruction Using Recursive Residual Network Based on Channel Attention[J]. *Chinese Journal of Computers*, 2021, 48(8): 139-144.
- [28] LIU Z X, ZHU C J, HUANG J et al. Image Super-resolution by Residual Attention Network with Multi-skip Connection [J]. *Chinese Journal of Computers*, 2021, 48(11): 258-267.



SUN Yang, born in 1998, postgraduate. His main research interests include image super-resolution and computer vision.



DING Jianwei, born in 1984, Ph.D, associate professor. His main research interests include computer vision and artificial intelligence security.