

文章编号 1004-924X(2023)15-2273-14

## 基于注意力和宽激活密集残差网络的 图像超分辨率重建

寇旗旗<sup>1\*</sup>, 李超<sup>2</sup>, 程德强<sup>2</sup>, 陈亮亮<sup>2</sup>, 马浩辉<sup>2</sup>, 张剑英<sup>2</sup>

(1. 中国矿业大学 计算机科学与技术学院, 江苏 徐州 221116;

2. 中国矿业大学 信息与控制工程学院, 江苏 徐州 221116)

**摘要:**针对全局和局部高低频空间信息利用不足而导致重建图像纹理细节模糊的问题,提出一种基于注意力和宽激活密集残差网络的图像超分辨率重建模型。首先,四个不同尺度且平行的卷积核被用来充分提取图像低频特征作为空间特征转换的先验信息。在深层特征映射模块中构建融合注意力的宽激活残差块,并利用低频先验信息来引导高频特征的提取。该宽激活残差块通过扩大激活函数前的特征通道数来提取更深层次的特征图,且所构造的全局和局部残差连接在加强残差块和网络特征前向传播的同时,在不增加参数情况下使得所提取特征的多样性更加丰富。最后,对得到的特征图进行上采样和重建以得到清晰的高分辨率图像。实验表明,所提算法在 BSD100 数据集上 4 倍超分辨率时,相比 LatticeNet 模型的 PSNR 指标提升了 0.14 dB, SSIM 提升了 0.001,在主观视觉方面,重建出的图像局部纹理细节也更加清晰。

**关键词:**残差网络;超分辨率;宽激活;注意力机制;密集连接

中图分类号: TP391 文献标识码: A doi: 10.37188/OPE.20233115.2273

## Image super-resolution reconstruction based on attention and wide-activated dense residual network

KOU Qiqi<sup>1\*</sup>, LI Chao<sup>2</sup>, CHENG Deqiang<sup>2</sup>, CHEN Liangliang<sup>2</sup>, MA Haohui<sup>2</sup>, ZHANG Jianying<sup>2</sup>

(1. School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China;

2. School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China)

\* Corresponding author, E-mail: kouqiqi@cumt.edu.cn

**Abstract:** To address the problem of the blurring of the texture details of reconstructed images due to the insufficient utilization of global and local high- and low-frequency spatial information, this paper proposes an image super-resolution reconstruction model based on attention and a wide-activated dense residual network. First, four parallel convolution kernels with different scales are used to fully extract the low-frequency features of the image as the prior information for spatial feature transformation. Second, a wide-activated residual block fused with attention is constructed in the deep feature mapping module, and the low-fre-

收稿日期: 2022-11-01; 修订日期: 2022-11-28.

基金项目: 中央高校基本科研业务费专项资金资助项目 (No. 2020QN49)

quency prior information is used to guide the extraction of the high-frequency features. In addition, the wide-activated residual block extracts deeper feature maps by expanding the number of feature channels before the activation function. As a result, the constructed global and local residual connections not only strengthen the forward propagation of the residual blocks and network features, but also enrich the diversity of the extracted features without increasing the number of parameters. Finally, the feature map is upsampled and reconstructed to obtain a clear high-resolution image. the experimental results show that compared with the LatticeNet model, the peak signal-to-noise ratio of the proposed algorithm is improved by 0.14 dB, and the structural similarity is improved by 0.001 at  $4\times$  super resolution on the BSD100 dataset. In addition, the local texture details of the reconstructed image are also clearer in subjective visualization.

**Key words:** residual network; super-resolution; wide activation; attention mechanism; dense connectivity

## 1 引 言

超分辨率重建(Super-Resolution, SR)旨在对输入的一幅或多幅低分辨率图像(Low-Resolution, LR)进行重建,以得出细节更加丰富的高分辨率图像(High-Resolution, HR)。目前,随着计算机视觉的飞速发展,图像超分辨率重建技术已经广泛应用于夜间高噪图像重建、遥感卫星成像、医学影像分析和雾霾图像重建等领域<sup>[1-3]</sup>。对于任意的一张低分辨图像,其重建后对应无数个不同的高分辨图像,因此,图像的超分辨率重建是一个典型的病态问题。为解决这个问题,国内外学者提出了基于插值的方法<sup>[4-5]</sup>、基于重建的方法<sup>[6]</sup>和基于学习的方法。其中,基于学习的方法是通过学习低分辨率图像和高分辨图像之间的对应关系来获得图像的先验知识,进而预测目标的高分辨率图像。在基于学习的重建任务中,将高分辨率图像使用双三次差值下采样得到相应低分辨率图像,组成用于模型训练的高低分辨率图像对。与传统重建方法相比,基于学习的方法可以得到更好的重建效果,因此受到了众多学者的关注。

近些年来,随着深度学习技术的发展,涌现出大量用于解决超分辨重建问题的网络模型。2015年,Dong<sup>[7]</sup>等首次将卷积神经网络应用于图像的超分辨率重建,并提出了一种超分辨率重建卷积神经网络(Super-Resolution Convolutional Neural Network, SRCNN),该网络直接学习从低分辨率图像到高分辨率图像之间的映射关系,是一项开创性的工作。随后,Shi<sup>[8]</sup>等在 SRCNN

的基础上对反卷积层上采样操作进行了改进并提出了高效的亚像素卷积层,该层通过重新对特征图进行排列,在减少网络参数的同时,获得了更好的重建效果。然而,上述方法均采用浅层的网络来获取图像的低频特征信息,忽略了深层特征信息对图像纹理细节重建的影响。Kim<sup>[9]</sup>等通过采用更深的网络结构以及更大的感受野,同时引入了残差网络(ResNet)<sup>[10]</sup>思想,提出了一种超分辨率重建极深卷积神经网络(Super-Resolution Using Very Deep Convolutional Networks, VDSR),该网络模型不仅实现了深层特征信息的提取,且残差信息的引入也有效缓解了网络层加深导致的梯度弥散问题,提高了网络的重建性能。但随着网络深度的增加,网络参数量也在剧增,这导致网络难以训练,收敛缓慢。因而 Kim<sup>[11]</sup>等又在深度递归网络(Deeply Recursive Convolutional Network, DRCN)中结合 VDSR 模型的优点,提出了递归监督的学习策略以减少模型的参数。紧接着,Tai<sup>[12]</sup>等在深度递归卷积网络(Image Super-Resolution Via Deep Recursive Residual Network, DRRN)中提出了递归残差块来实现权重共享,以此进一步降低网络参数量。Lim<sup>[13]</sup>等通过去除批量归一化(Batch Normalization, BN)来改进残差块,提出了一种增强型深度超分辨率网络(Enhanced Deep Residual Networks for Single Image Super-Resolution, ED-SR),重建效果得到了显著的提升。Hui<sup>[14]</sup>等研究发现网络深度也是影响超分辨率重建速度的重要因素,因此提出了轻量化的多蒸馏网络(Lightweight Image Super-Resolution with Infor-

ma-tion Multi-Distillation network, IMDN) 用来提升网络的重建速度,为超分辨率重建的轻量化开启一个新思路。与 Hui 等思路不同, Luo<sup>[15]</sup> 等通过提出双支路特征提取模块和反向级联融合策略来降低计算复杂度和内存需求,然而此方法的性能提升却并不明显。Tong<sup>[16]</sup> 等设计了一种密集连接的网络以促进图像特征的信息流通,从而提出了密集跳跃连接超分辨率网络(Image Super-Resolution Using Dense Skip Connections, SRDe-nseNet)。此后, Zhang<sup>[17]</sup> 等通过将残差连接和密集连接进行融合,进一步提升了 LR 图像的超分重建效果。2018 年, Zhang<sup>[18]</sup> 等通过设计一种通道注意力模块(Channel Attention, CA) 来对不同通道的特征重新进行加权分配以增强网络的辨别学习能力,提出了基于残差注意力的超分辨率重建网络(Image Super-Resolution Using Very Deep Residual Channel Attention Networks, RCAN), 注意力机制的引入使得其重建效果相比其他 SR 方法得到了显著的提升。随后, Dai<sup>[19]</sup> 等对 RCAN 中的通道注意力进行改进,提出了基于二阶统计的注意力机制,并第一次将非局部注意力<sup>[20]</sup> 引入超分辨率重建任务中。2021 年, Chen<sup>[21]</sup> 等从底层研究注意力机制对超分辨率重建的作用,并提出了一种注意力嵌套图像超分辨率重建网络(Attention in Attention for Image Super-Resolution, A2N)。鉴于注意力机制优异的特征表达能力,研究人员开始在 SR 任务中引入各种注意力机制来提高网络的重建性能。此外, Ledig<sup>[22]</sup> 等利用感知损失和对抗损失来实现低分辨率图像的超分重建,并提出了一种超分辨率生成对抗网络模型(Super-Resolution Generative Adversial Network, SRGAN), 也取得了良好的图像重建效果。

虽然上述图像超分辨率重建网络取得了较好的重建效果,但大部分网络模型没有充分利用低分辨率图像的全局和局部信息,存在特征图太少可能导致重要信息丢失的问题。目前,虽然也有少量网络模型综合考虑了该问题,如 Yu<sup>[23]</sup> 等提出了一种宽激活深度残差网络(Wide Activation for Efficient and Accurate Image Super-resolution, WDSR) 用于图像超分辨率重建,该网络在尽量不增加参数量的前提下,通过扩大残差块

内激活函数前的特征通道数实现了深层次高频信息的提取。然而,该方法虽然提取到不同的深层特征信息,但是仍然存在特征复用和利用率较低的问题。

为了解决以上问题,本文提出了一种基于注意力和宽激活密集残差网络的图像超分辨率重建模型(Image Super Resolution Reconstruction Based on Attention and Wide-Activated Dense Residual Network, WDRN)。针对低分辨率图像浅层特征信息利用不足的问题,在浅层特征提取模块,采用 4 个不同尺度且平行的卷积核提取低分辨率图像的局部和全局信息,作为空间转换的先验信息,通过获取多层次的感受野来丰富底层关键特征的多样性,提高重建性能。为保留低分辨率图像的空间信息,通过引入空间特征转换层,将多尺度浅层特征作为该层的先验信息来引导高频特征的提取,从而为图像的重建增添更加深层次的纹理信息。为了获取丰富多样的高频特征信息,对激活函数前的特征通道数进行了扩大以获取多通道高频信息。构造全局和局部残差连接以加强残差块和网络特征前向传播,使网络在不增加参数的情况下提取多样性更加丰富的特征。在深层特征映射模块,在宽激活残差块中融入轻量级注意力机制,对高频区域信息分配更多权重,使模型的重建结果更清晰的同时仍维持较低的数量水平。

## 2 原 理

### 2.1 网络结构

本文提出融合注意力的宽激活密集残差网络模型的整体结构如图 1 所示。该网络模型有 3 个子模块:(1)多尺度浅层特征提取模块,不同尺度的卷积可以充分提取低分辨率图像全局和局部的信息,该模块提取的信息一方面作为深层特征映射模块的输入,另一方面作为空间特征转换层的先验信息来引导图像的重建;(2)深层特征映射模块,用于提取图像高频特征以及对特征重复利用,该模块由融合注意力机制的宽激活残差块(Wide Activation Residual Block, WARB)密集连接形成的密集残差组(Wide Activation Residual Block Group, WARBG)构成,空间特征转换层在密集残差组的前面,用于转换特征图的中间特

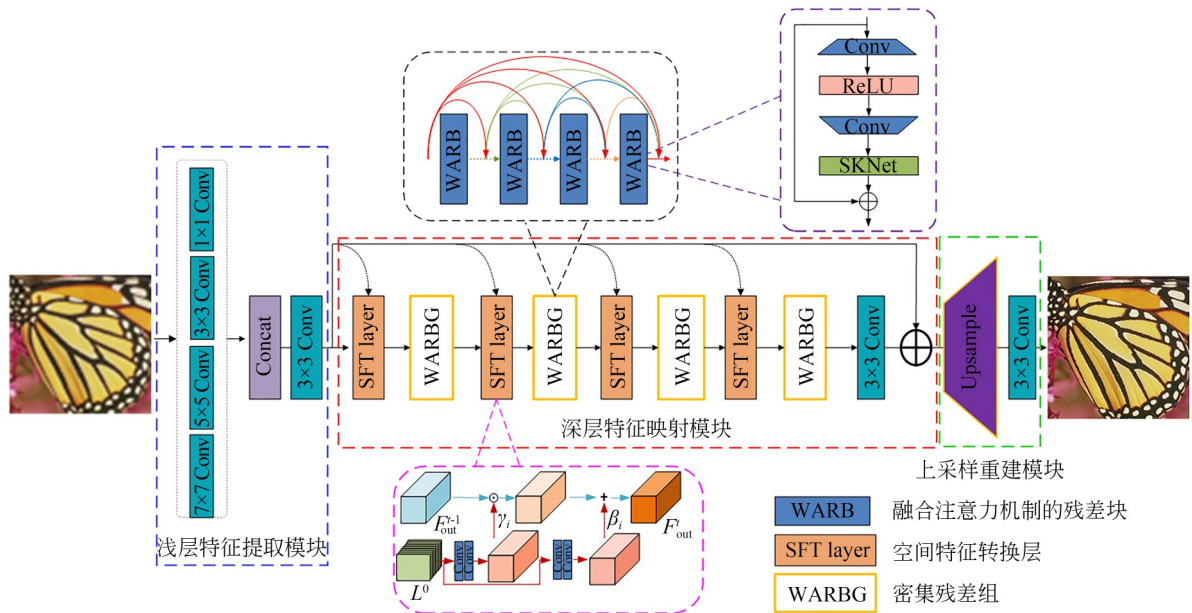


图1 融合注意力的宽激活密集残差网络结构

Fig. 1 Structure of attention and wide-activated dense residual network

征;(3)上采样重建模块,该模块主要由亚像素卷积层和一个 $3 \times 3$ 卷积层组成,用于重建出最终的高分辨率图像。

## 2.2 多尺度浅层特征提取模块

受到Szegedy<sup>[24]</sup>等提出的拓展网络宽度思想的启发,WDRN中多尺度浅层特征提取模块使用4个不同尺寸且平行的卷积核对LR图像进行卷积操作,4个卷积核的大小分别为 $1 \times 1$ , $3 \times 3$ , $5 \times 5$ , $7 \times 7$ 。小尺寸的 $1 \times 1$ 和 $3 \times 3$ 卷积核可以学习得到LR图像的局部特征, $5 \times 5$ 和 $7 \times 7$ 的较大卷积核在对LR图像进行特征提取时能获得大感受野的特征图。为获得更佳的特征映射效果,将每个卷积核输出64个特征图后进行多尺度特征级联的融合操作,进而为空间特征提取模块提供丰富的先验信息。然后,再将该先验信息融入深层特征映射模块,使低分辨率图像的信息得以充分利用,从而提高重建效果。多尺度浅层特征提取过程表示如下:

$$L_{n \times n} = F_{\text{SFE}, n \times n}(I_{\text{LR}}), \quad (1)$$

$$L^0 = f_0([L_{1 \times 1}, L_{3 \times 3}, L_{5 \times 5}, L_{7 \times 7}]), \quad (2)$$

其中: $F_{\text{SFE}, n \times n}(\cdot)$ 表示分别用 $n$ 为1,3,5,7的卷积核对低分辨率图像进行浅层特征卷积处理, $f_0$ 表示将获得的多尺度特征图经过concat后把通道数降为64的操作。

## 2.3 深层特征映射模块

### 2.3.1 注意力机制

人类视觉感知系统会有选择地聚焦于全局信息中较为显著的部分,同时忽略其他不重要的信息。而计算机视觉中的注意力机制与人类视觉感知相似,会从不同角度或在不同程度上重新为网络所提取的信息加以区分。文献[19]提出的二阶通道注意力机制是对RCAN中通道注意力机制的改进,但两者仅通过调整通道特征来获得对于信息的区别性表示,忽略了空间特征注意力信息的表示。文献[20]中的非局部注意力模块由于只涉及特征图位置信息而缺乏对于特征图通道的注意力,且其矩阵乘操作计算量较大,因此在网络模型中并不适合多次使用该注意力机制模块。综上,本文在深层特征映射部分采用轻量级注意力模块<sup>[25]</sup>,选择性内核网络(Selective Kernel Networks, SKNet),该模块可以根据不同尺度的输入信息自适应地选择合适的卷积核来关注某些特征信息,并且经过分离、融合以及选择操作后拥有对于通道和空间信息加以区分的能力,因而更适用于在超分辨率重建任务中捕捉图像的高频信息,该模块的具体结构如图2所示。

在分离操作阶段,对输入的初始特征图 $X$ 分别使用 $3 \times 3$ 和 $5 \times 5$ 的两种卷积核进行卷积操

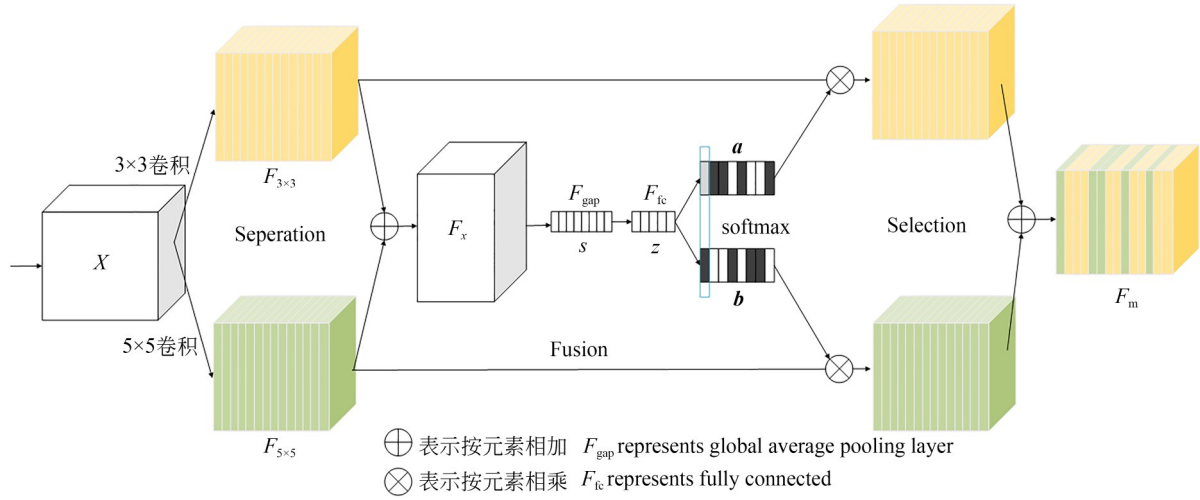


图 2 SKNet 注意力机制

Fig. 2 SKNet attention mechanism

作,得到两个输出  $F_{3 \times 3}$  和  $F_{5 \times 5}$ 。其次,通过融合操作将此两部分的特征图按元素相加的方式融合,得到新的特征图  $F_x$ :

$$F_x = F_{3 \times 3} + F_{5 \times 5}. \quad (3)$$

为获得特征图的注意力信息,特征图  $F_x$  依次经过全局平均池层、先升维再降维的全连接层,得到一个紧凑的特征向量:

$$Z = F_{\text{fc}}(F_{\text{gap}}(F_x)), \quad (4)$$

其中:  $F_{\text{gap}}$  表示全局平均池化层(global average pooling, gap),  $F_{\text{fc}}$  表示全连接层(fully connected, fc)。

最后,在选择操作阶段首先将特征向量  $Z$  重新分为  $a$  和  $b$  两个特征向量,两者为互补关系,然后再分别与  $F_{3 \times 3}$  和  $F_{5 \times 5}$  进行相应通道的加权操作,进而求和得到最终的输出向量  $F_m$ ,即:

$$a_c = \frac{A}{A+B}, b_c = \frac{B}{A+B}, \quad (5)$$

$$F_m = a_c \cdot F_{3 \times 3} + b_c \cdot F_{5 \times 5}, \quad (6)$$

其中:  $A, B$  分别表示由不同的全连接矩阵进行特征维度转化后的结果,  $a_c$  和  $b_c$  表示输出特征向量  $a$  和  $b$  的第  $c$  个元素。最终的特征映射  $F_m$  是通过各种卷积核的注意力权重获得的。

### 2.3.2 融合注意力的宽激活残差块

在本文中每个融合注意力的 WARB 包括 2 个卷积核为 1 的卷积层,1 个卷积核为 3 的卷积层与 1 个 SK 注意力模块。WARB<sup>[23]</sup> 首先在 ReLU 激活层前通过  $1 \times 1$  卷积的方式扩展通道数来获得更多特征图,利用多个通道数增大信息提取的

能力,从而避免单一通道数造成高频有效信息丢失的问题。然后,在 ReLU 激活层后分别又用卷积核大小为 1 和卷积核大小为 3 的低秩卷积核替换一个大的卷积核,以此降低其计算开销。因此,WARB 不仅有着强大的特征表达能力,且将参数量也维持在较低水平。鉴于该宽激活残差块的优异特性,为了给网络提取的高频信息自适应地分配更多的注意力,以得到更加丰富的高频特征,本文提出在 WARB 中融入轻量级的 SK 注意力模块,并将该融合结果作为提取深层特征信息的基本单元模块,具体流程如图 1 所示。由于简单地堆叠残差块不利于特征信息的传递,因此将每一个残差块提取的特征都输送给后面的残差块中来构建一个 WARBG,从而促进特征信息的有效传递。

### 2.3.3 空间特征转换层

为获取更加丰富的深层次图像空间信息,本文采用空间特征转换层(Spatial Feature Transform Layer, SFT)<sup>[26]</sup> 来对同一个特征图的不同位置进行差异化处理。该空间特征转换层通过生成仿射变换参数来实现空间特征的调制,并且可以与超分辨率重建主干网络结合在一起,使用相同的损失函数进行端到端的训练,其结构如图 1 所示。空间特征变换层包含两个输入,分别是前一个密集残差组的输出特征  $F_{\text{out}}^{\prime-1}$  和浅层特征提取模块中条件共享的先验信息  $L^0$ 。通过对  $F_{\text{out}}^{\prime-1}$  进行变换得到新的输出  $F_{\text{out}}^{\prime}$ ,并将它作为下一个

密集残差组的输入。浅层先验信息  $L^0$  分别送入到由两个  $1 \times 1$  的卷积层组成的卷积块,通过学习得到参数对  $\gamma_i$  和  $\beta_i$ ,然后再据此对深层特征映射模块网络中的每个中间特征图进行仿射变换,从而实现自适应输出。相比没有空间特征变换的网络,此模块有助于网络重建出纹理更加清晰、边缘轮廓更为明显的图像。

#### 2.4 上采样重建模块

上采样重建模块包括亚像素卷积和一个  $3 \times 3$  的卷积层,亚像素卷积旨在对特征图进行上采样放大,再通过最后的卷积层得到通道数为 3 的高分辨率重建结果,其输出如下:

$$I_{SR} = f_{up}(P_0), \quad (7)$$

其中:  $f_{up}$  表示由上采样重建模块组成的函数,  $P_0$  表示深层特征模块和浅层特征融合后的输出,  $I_{SR}$  表示最终的重建结果。

### 3 实验结果及分析

#### 3.1 数据集和实验设置

在实验训练阶段,采用 DIV2K 公共标准数据集作为本文的训练数据集,该数据集共包含 1 000 张样本图像,其中每幅图像都包含非常清晰的细节纹理信息,非常适合作为自然图像超分辨率重建的训练数据。在测试和分析阶段,为测试和验证本文所提 WDRN 网络性能的优越性,选取 Set5, Set14<sup>[27]</sup>, BSD100<sup>[28]</sup>, Urban100<sup>[29]</sup> 4 个公共标准数据集作为基准测试集来进行测验和分析。

本实验训练测试的实验硬件环境为: Intel (R) Core(TM) i9-10980XE CPU@ 3.00GHz, 18 核 36 线程;系统内存为 64 GB;GPU 为 NVIDIA RTX 3090, 24 GB 显存容量;软件环境为 Ubuntu20.04 操作系统。

将 L1 函数作为损失函数来优化重建后的高分辨率图像与真实高分辨率图像之间的差值,从而进一步优化网络,可以表示为:

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N \|I_{HR}^i - I_{SR}^i\|_1, \quad (8)$$

其中:  $\theta$  表示网络需要学习的参数集,  $I_{HR}^i$  表示原始的高分辨率图像,  $I_{SR}^i$  表示利用本文所提 WDRN 网络重建后的高分辨率图像。模型在训练阶段使用 Adam 算法优化网络参数,共训练

300 个训练周期(Epoch),初始学习率为  $1 \times 10^{-4}$ ,训练 200 个 Epoch 后的学习率衰减为  $0.5 \times 10^{-4}$ 。

#### 3.2 消融实验结果与分析

##### 3.2.1 密集残差组数和残差块的消融实验

为了测试和验证不同数目的残差组( $D$ )和残差块( $R$ )组合方式对所提出的 WDRN 网络重建性能的影响,并选取最有效的组合方式,具体测试结果如表 1 所示。从表 1 可以看出,随着残差组数目以及残差组中残差块数目的增加,各个测试集的 PSNR 指标都有所提升。当模型规模大于 R4D4 时,参数量增加、训练时间延长,性能提高不明显。由实验结果可知, R6D4 的模型在 Set5 和 Set14 的测试集上分别提高了 0.02 dB 和 0.053 dB,一方面是由于 DIV2K 训练集与 Set5 和 Set14 测试集中的图片类型比较一致,另一方面是由于 Set5, Set14 是基于非负领域嵌入的低复杂度单图像的测试集,而在包含高复杂度图像的 BSD100 和 Urban100 测试集上性能提升不明显,然而 R6D4 的模型参数量却比 R4D4 的模型参数量高 40.97%。因此,在综合考虑算法性能、网络参数量因素后,本文最终采用 R4D4 的组合方式。

表 1 残差组数量( $D$ )和残差块数量( $R$ )对 2 倍超分辨率重建性能的影响 PSNR

Tab. 1 Effect of number of residual groups ( $D$ ) and number of residual blocks ( $R$ ) on performance of  $2 \times$  super-resolution reconstruction

| 组合方式 | PSNR/dB       |               |               |               | 参数量/M |
|------|---------------|---------------|---------------|---------------|-------|
|      | Set5          | Set14         | BSD100        | Urban100      |       |
| R2D4 | 38.069        | 33.565        | 32.189        | 32.086        | 2.68  |
| R4D2 | 38.029        | 33.538        | 32.153        | 32.069        | 2.52  |
| R4D4 | 38.080        | 33.623        | <b>32.230</b> | <b>32.341</b> | 4.54  |
| R4D6 | 37.896        | 33.325        | 31.944        | 31.317        | 6.55  |
| R6D4 | <b>38.100</b> | <b>33.676</b> | 32.227        | 32.320        | 6.40  |

为进一步证明 R4D4 组合方式的优越性,选取 BSD100 数据集的 42 049 图像和 Urban100 数据集的 img002 图像在不同组合方式下进行重建图像的可视化对比,如图 3 和图 4 所示。各组合方式的 PSNR 和 SSIM 值如图下方数字所示。在主观视觉上,采用 R4D4 组合方式重建后的图像相比其他组合方式,在图 3 中树杈阴影部分的处

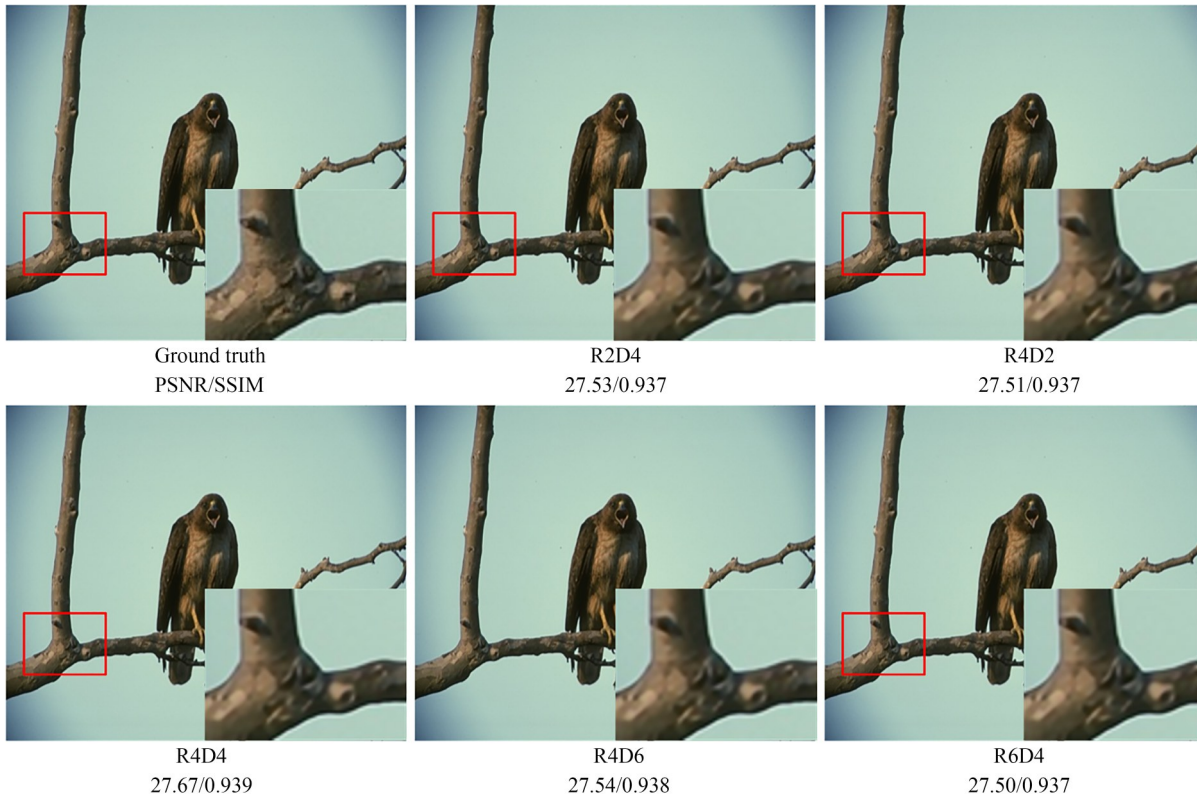


图 3 不同组合方式在 BSD100\_42049 上的重建效果对比

Fig. 3 Comparison of reconstruction effect of different combination methods on BSD100\_42049

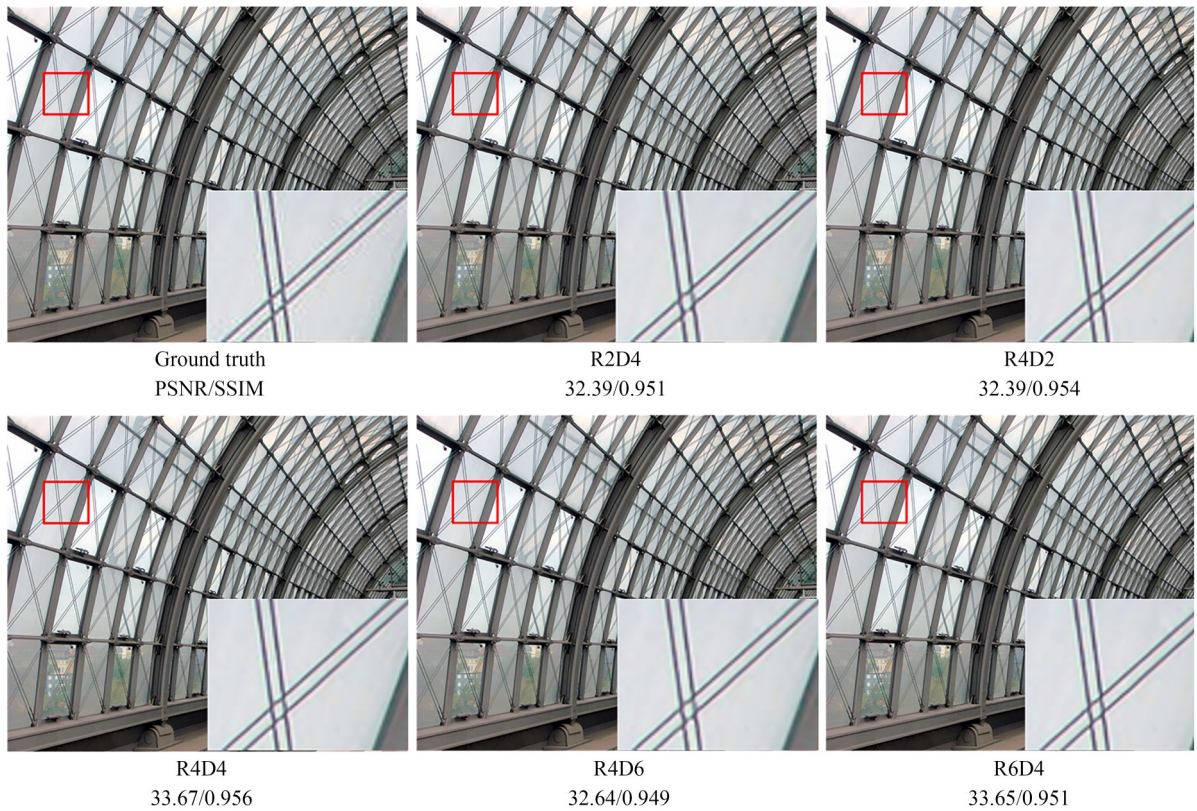


图 4 不同组合方式在 Urban100\_img002 上的重建效果对比

Fig. 4 Comparison of reconstruction effects of different combinations on Urban100\_img002

理更为明亮清晰,在图 4 中建筑物线条交汇处的像素分辨能力最强。

### 3.2.2 不同通道倍数的消融实验

本文针对 WARB 内激活函数前放大不同倍数(用  $k$  表示)的特征通道数也进行了有效性研究。在 R4D4 模型的基础上,分别采用不同扩大倍数对网络进行训练,实验结果如图 5 所示。由折线图的趋势不难看出,在 BSD100 和 Urban100 测试集上扩大倍数小于 4 时的网络性能随着  $k$  值的增大而提高,当扩大倍数为 4 时两个测试集的性能均达到最佳; $k$  值继续增大后,在 BSD100 测试集上性能表现趋于平缓,但是在图像纹理细节更加丰富的 Urban100 测试集的性能则呈现先下降后升高的现象,表明此时的网络已经出现了过拟合现象,且对应的计算量和参数量也随之增大。因此,将通道数扩大为原来的 4 倍可以达到最佳状态。

### 3.2.3 空间特征转换层的消融实验

网络结构的微小改动往往会导致结果有较大的提升,这里针对 SFT 采用不同连接方式及不同大小卷积核进行了有效性验证。图 1 中,将共享的低分辨率特征信息送入两个  $1 \times 1$  并联的卷积层学习以得到一对调制参数。通过将卷积层的并联替换成串联的方式,并探讨不同尺度卷积核对性能的影响,其测试结果如表 2 所示。由表

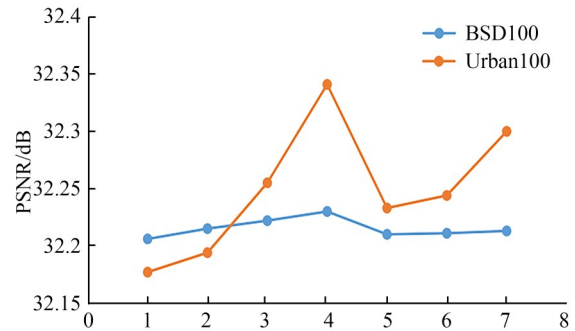


图 5 不同通道倍数下 4 倍超分辨率的 PSNR

Fig. 5 PSNR with  $4 \times$  super-resolution at different channel multiples

2 可以看出,在空间特征转换层的连接方式上,同一卷积核大小的情况下,除了在 Set5 测试集上卷积层并联的效果稍逊于串联形式,在 Set14, BSD100 和 Urban100 测试集上的效果均明显优于串联的连接方式。在卷积核尺寸选择上,串联情况下,卷积核的增大除了在 Set5 测试集上提高了 0.008 dB 之外,在其他测试集上均有所下降;并联情况下的性能对比则更为明显,卷积核为  $1 \times 1$  比  $3 \times 3$  的 PSNR 平均高出 0.03 dB。从实验结果不难看出,SFT 中两个卷积层的并联组合效果要优于串联的效果,并且卷积核的增大对性能的提升作用并不明显。因此,选用  $1 \times 1$  的卷积核效果更好。

表 2 空间特征转换层卷积核尺寸及连接方式对 2 倍超分辨率重建性能的影响

Tab. 2 Effect of convolution kernel size and connection mode of spatial feature transformation layer on performance of  $2 \times$  super resolution reconstruction

| 连接方式 | 卷积核 | PSNR/dB       |               |               |               |
|------|-----|---------------|---------------|---------------|---------------|
|      |     | Set5          | Set14         | BSD100        | Urban100      |
| 串联   | 1   | 38.075        | 33.612        | 32.220        | <u>32.283</u> |
| 串联   | 3   | <b>38.083</b> | 33.597        | 32.219        | 32.247        |
| 并联   | 1   | <u>38.080</u> | <b>33.623</b> | <b>32.230</b> | <b>32.341</b> |
| 并联   | 3   | <u>38.080</u> | <u>33.612</u> | <u>32.222</u> | 32.239        |

注:加粗字体表示该列最优值,下划线表示该列次优值。

### 3.2.4 注意力机制的消融实验

为了进一步说明采用选择性注意力机制的有效性和优越性,将它与现有的经典注意力机制进行比较。实验结果如表 3 所示,相较于没有加入注意力机制的情况,网络在 4 个测试集上的

PSNR 指标分别提升了 0.042, 0.051, 0.028, 0.283 dB,充分证明了网络融合注意力机制的有效性。加入 SK 注意力相比于加入 SE, CA, CBAM 等经典注意力机制,网络性能提升更为明显,证明了网络中融合 SK 注意力机制的优越性。

表 3 注意力机制对 2 倍超分辨率重建性能的影响

Tab. 3 Effects of attention mechanisms on  $2\times$  super-resolution reconstruction performance

| 注意力机制的选择 | PSNR/dB       |               |               |               |
|----------|---------------|---------------|---------------|---------------|
|          | Set5          | Set14         | BSD100        | Urban100      |
| ×        | 38.038        | 33.572        | <u>32.202</u> | 32.058        |
| SE       | 38.046        | 33.543        | 32.165        | 32.020        |
| CA       | <u>38.069</u> | <u>33.558</u> | 32.175        | <u>32.082</u> |
| CBAM     | 38.005        | 33.482        | 32.167        | 31.913        |
| SK       | <b>38.080</b> | <b>33.623</b> | <b>32.230</b> | <b>32.341</b> |

注:加粗字体表示该列最优值,下划线表示该列次优值。

### 3.3 算法先进性验证

本文算法训练 300 个迭代周期(Epoch),网络在不同放大倍数下的收敛曲线如图 6 所示。

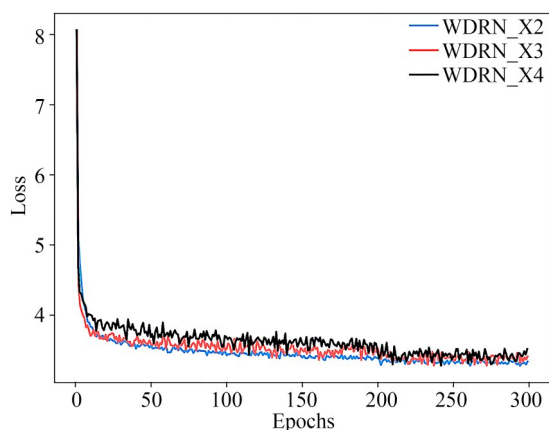


图 6 不同倍数下的损失函数曲线

Fig. 6 Loss function curves at different magnifications

为了进一步验证 WDRN 模型的有效性,将它与 Bicubic, ESPCN<sup>[8]</sup>, VDSR<sup>[9]</sup>, DRCN<sup>[11]</sup>, LapSRN<sup>[30]</sup>, WDSR\_Mini<sup>[23]</sup>, IMDN<sup>[14]</sup> 和 LatticeNet<sup>[15]</sup> 8 种主流算法,在 Set5, Set14, BSD100 以及 Urban100 测试集及不同放大倍数下( $\times 2$ ,  $\times 3$ ,  $\times 4$ )的超分辨率重建结果分别进行客观评价和主观评价,对比结果如表 4 所示。

从表 4 可以看出, WDRN 网络在 3 种放大倍数下的超分辨率重建结果均明显优于 Bicubic, ESPCN, VDSR, DRCN, LapSRN, WDSR\_Mini, IMDN 及 LatticeNet 网络。此外, WDRN 采用的是由 16 个残差块构成的密集残差网络模型,通过各个测试集上对不同放大倍数的重建结果进行测试和对比分析,实验结果表明,本文模型在 PSNR 和 SSIM 指标上均优于同为 16 个残差块

构成的 WDSR\_Mini 模型。与 LatticeNet 网络相比,虽然在放大倍数为 2 时本文模型的重建性能在客观指标 PSNR 上略低于该网络,在 SSIM 指标上却优于 LatticeNet,表明本文算法重建出的 2 倍图像在结构细节方面保持得较好,这主要得益于本文模型采用空间特征转换层来获取低分辨率图像空间结构信息的原因。此外, WDRN 网络在 3 倍和 4 倍时的超分辨率重建性能的提升则更加显著。尤其是在放大倍数为 4 时,本文的 WDRN 网络在 Set5, Set14, BSD100 数据集上的 PSNR 和 SSIM 与 IMDN 和 LatticeNet 相比,分别平均提高了 0.137 dB, 0.0215 和 0.075 dB, 0.021。值得注意的是,在主要由城市建筑物规则图案组成的纹理特征复杂的 Urban100 数据集上, WDRN 网络在 PSNR 指标上也相较 LatticeNet 提升了 0.14 dB。结果表明,本文所提算法在图像超分辨率重建效果方面均优于其他对比算法。

由于放大倍数越大,重建出的图像边缘就会越模糊,因此为了进一步表明本文算法重建结果的相对优越性,本文着重展示了不同重建方法在放大 4 倍时的视觉效果图,如图 7~图 10 所示。首先,从较小数据集方面分析重建结果,图 7 是不同方法对 Set5 数据集中的 baby 图像重建后的结果。从该结果可以观察到,当图像中有细小的条纹时, Bicubic, ESPCN, VDSR, WDSR\_Mini 等模型重建后的睫毛线条仍有少许的伪影,与 IMDN, LatticeNet 模型相比,本文的 WDRN 模型有效缓解了伪影问题。图 8 是不同方法对 Set14 数据集中 monarch 图像重建后的结果,从图中可以看出, Bicubic, ESPCN 重建后的图像较为模糊,而 WDSR\_Mini 重建后的图像缺失原图像边缘轮

表 4 放大倍数为 2,3,4 时在基准数据集上的重建结果比较

Tab. 4 Reconstruction comparison on baseline dataset at magnifications of 2, 3 and 4

(dB)

| 放大<br>倍数 | 模 型                        | Set5         |              | Set14        |              | BSD100       |              | Urban100     |              |
|----------|----------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|          |                            | PSNR         | SSIM         | PSNR         | SSIM         | PSNR         | SSIM         | PSNR         | SSIM         |
| × 2      | Bicubic <sup>[5]</sup>     | 36.66        | 0.929        | 30.24        | 0.868        | 29.56        | 0.843        | 26.88        | 0.840        |
|          | ESPCN <sup>[8]</sup>       | 37.00        | 0.956        | 32.75        | 0.910        | 31.51        | 0.894        | 29.87        | 0.907        |
|          | VDSR <sup>[9]</sup>        | 37.54        | 0.959        | 33.03        | 0.912        | 31.90        | 0.896        | 30.76        | 0.914        |
|          | DRCN <sup>[11]</sup>       | 37.63        | 0.959        | 33.06        | 0.912        | 31.85        | 0.884        | 30.76        | 0.913        |
|          | LapSRN <sup>[30]</sup>     | 37.52        | 0.959        | 32.99        | 0.912        | 31.80        | 0.895        | 30.45        | 0.913        |
|          | WDSR_Mini <sup>[23]</sup>  | 37.96        | 0.960        | 33.50        | 0.914        | 32.13        | 0.899        | 31.79        | 0.921        |
|          | IMDN <sup>[14]</sup>       | 38.00        | 0.961        | 33.63        | 0.912        | 32.19        | 0.900        | 32.17        | 0.928        |
|          | LatticeNet <sup>[15]</sup> | <b>38.15</b> | <u>0.961</u> | <b>33.78</b> | <u>0.919</u> | <b>32.25</b> | <b>0.901</b> | <b>32.43</b> | <u>0.930</u> |
| WDRN     | <u>38.08</u>               | <b>0.966</b> | <u>33.59</u> | <b>0.970</b> | <u>32.20</u> | <u>0.900</u> | <u>32.31</u> | <b>0.938</b> |              |
| × 3      | Bicubic                    | 30.39        | 0.868        | 27.55        | 0.774        | 27.21        | 0.738        | 24.46        | 0.735        |
|          | ESPCN                      | 33.02        | 0.914        | 29.49        | 0.827        | 28.50        | 0.794        | 26.41        | 0.816        |
|          | VDSR                       | 33.65        | 0.921        | 29.78        | 0.831        | 28.82        | 0.798        | 27.14        | 0.828        |
|          | DRCN                       | 33.82        | 0.923        | 29.77        | 0.831        | 28.80        | 0.796        | 27.15        | 0.823        |
|          | LapSRN                     | 33.82        | 0.923        | 29.87        | 0.832        | 28.83        | 0.798        | 27.08        | 0.828        |
|          | WDSR_Mini                  | 34.375       | 0.925        | 30.31        | 0.840        | 29.06        | 0.803        | 27.97        | 0.847        |
|          | IMDN                       | 34.36        | 0.927        | 30.32        | 0.842        | 29.09        | <u>0.805</u> | 28.17        | 0.852        |
|          | LatticeNet                 | <u>34.53</u> | <u>0.928</u> | <u>30.39</u> | <u>0.842</u> | <u>29.15</u> | <b>0.806</b> | <u>28.33</u> | <u>0.854</u> |
| WDRN     | <b>34.58</b>               | <b>0.938</b> | <b>30.47</b> | <b>0.910</b> | <b>29.17</b> | <b>0.806</b> | <b>28.45</b> | <b>0.857</b> |              |
| × 4      | Bicubic                    | 28.42        | 0.810        | 26.00        | 0.703        | 25.96        | 0.668        | 23.14        | 0.657        |
|          | ESPCN                      | 30.66        | 0.865        | 27.71        | 0.756        | 26.98        | 0.712        | 24.60        | 0.736        |
|          | VDSR                       | 31.35        | 0.884        | 28.01        | 0.767        | 27.30        | 0.725        | 25.18        | 0.751        |
|          | DRCN                       | 31.53        | 0.885        | 28.03        | 0.767        | 27.24        | 0.723        | 25.14        | 0.751        |
|          | LapSRN                     | 31.54        | 0.886        | 28.19        | 0.772        | 27.33        | 0.726        | 25.21        | 0.756        |
|          | WDSR_Mini                  | 32.17        | 0.893        | 28.59        | 0.783        | 27.56        | 0.731        | 25.95        | 0.778        |
|          | IMDN                       | 32.21        | 0.895        | 28.58        | <u>0.787</u> | 27.56        | 0.735        | 26.04        | 0.784        |
|          | LatticeNet                 | <u>32.30</u> | <u>0.896</u> | <u>28.68</u> | 0.783        | <u>27.62</u> | <u>0.737</u> | <u>26.25</u> | <b>0.787</b> |
| WDRN     | <b>32.33</b>               | <b>0.909</b> | <b>28.77</b> | <b>0.854</b> | <b>27.66</b> | <b>0.738</b> | <b>26.39</b> | <u>0.786</u> |              |

注:加粗字体表示每列最优值,下划线表示每列次优值。

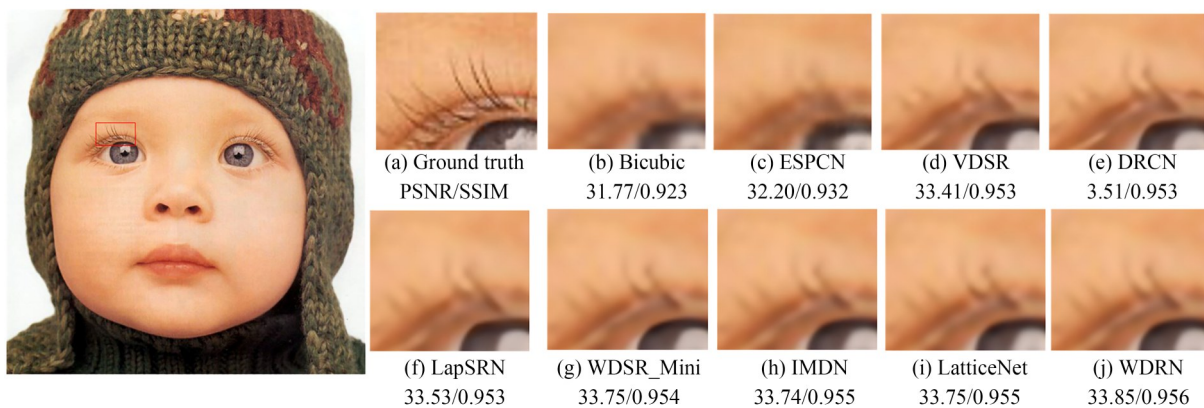


图 7 不同方法在 Set5\_baby 上的重建效果对比

Fig. 7 Comparison of reconstruction effects by different methods on Set5\_baby

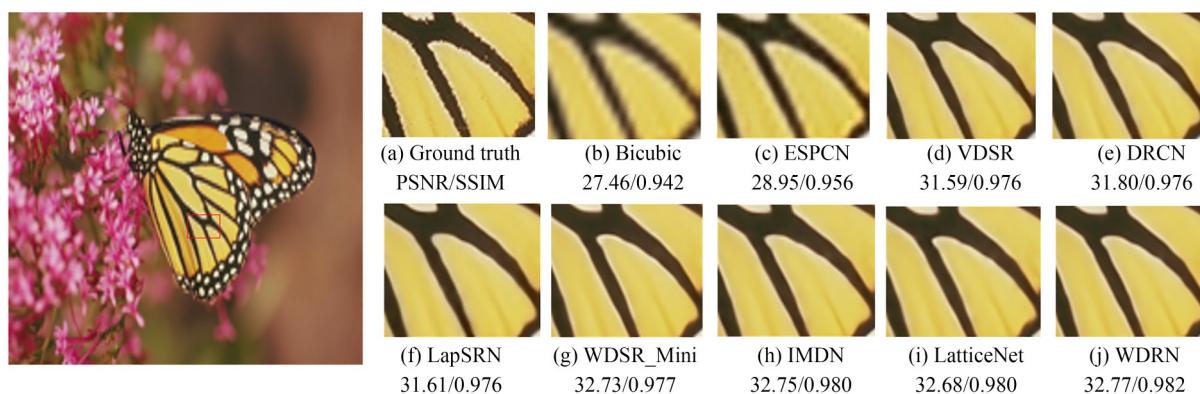


图 8 不同方法在 Set14\_monarch 上的重建效果对比

Fig. 8 Comparison of reconstruction effects by different methods on Set14\_monarch

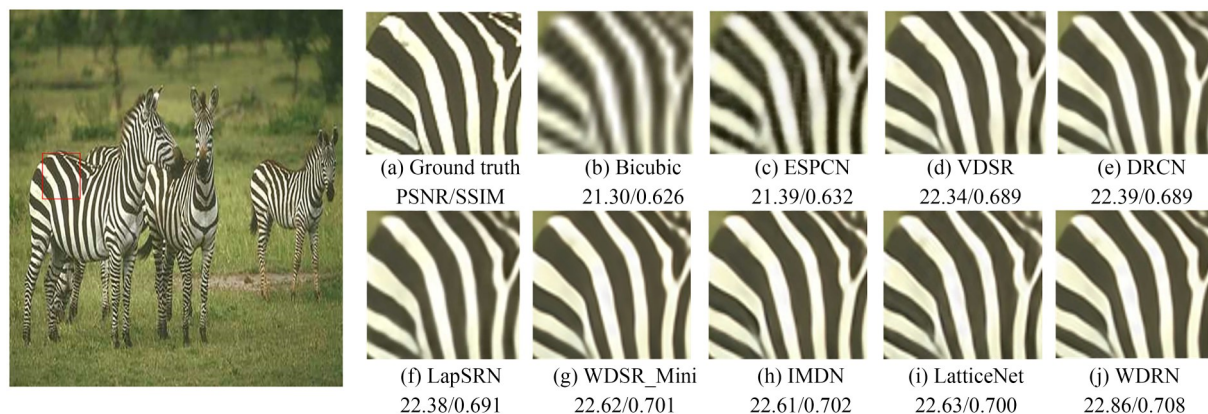


图 9 不同方法在 BSD100\_253027 上的重建效果对比

Fig. 9 Comparison of reconstruction effects by different methods on BSD100\_253027

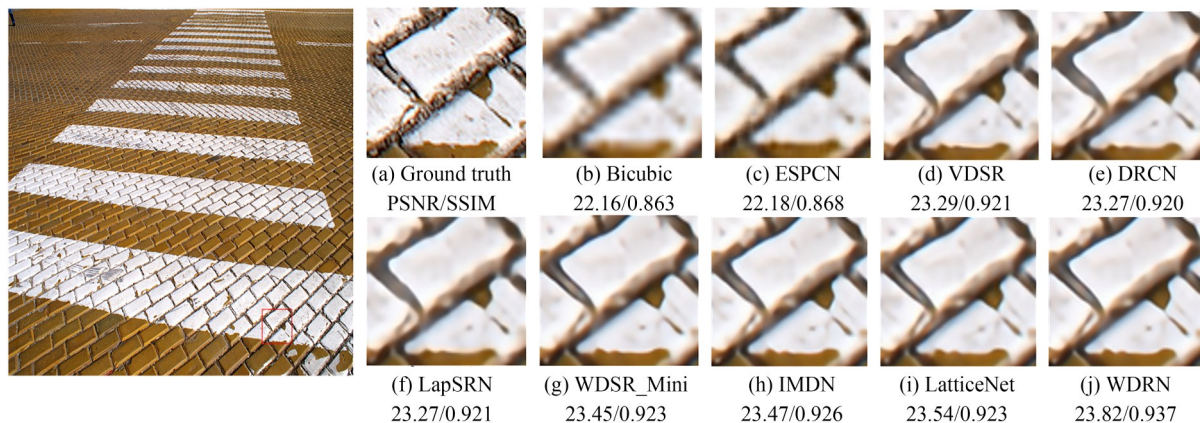


图 10 不同方法在 Urban100\_img091 上的重建效果对比

Fig. 10 Comparison of reconstruction effects by different methods on Urban100\_img091

廓微小的变化,IMDN 和 LatticeNet 得到了相对清晰的重建结果,但与原图和 WDRN 方法重建后的图像相比,上述方法表现得过于平滑,致使在纹理细节方面存在丢失的现象。

图 9 和图 10 是在较大数据集上各种方法的重建结果对比,其中图 9 是对 BSD100 数据集中编号为 253027 的斑马图像的重建结果。从图中可以看出,Bicubic,ESPCN 重建后的纹理信息非

常模糊,主观视觉上表现也最差,WDSR\_Mini模型重建后的图像整体相较原图像过于明亮并且边缘稍显模糊,IMDN重建后的图像在整体上有轻微的颜色失真,LatticeNet模型的重建结果则存在过曝光问题,而本文重建出的斑马图像身上的线条轮廓更为明显,整体也最为真实。图10中的原始图像选自图像细节丰富且重建难度较大的Urban100数据集,从编号为img091红框内地面砖块部分可以看出,本文算法明显改善了Bicubic,ESPCN两种方法在重建时出现锯齿状效应严重的问题。相比于VDSR,DRCN,LapSRN,WDSR\_Mini算法重建后均出现边缘细节模糊的现象,本文算法的重建效果也得到了明显的提升。此外,相比于IMDN,LatticeNet算法存在重建后图像过于平滑的现象,本文算法重建后的图像效果在主观上也更加接近原图。因此,WDRN网络模型无论是在主观评价上还是客观评价上均明显优于现有的主流图像超分辨率重建方法。

#### 参考文献:

- [1] 耿铭昆,吴凡路,王栋.轻量化火星遥感影像超分辨率重建网络[J].*光学精密工程*,2022,30(12):1487-1498.  
GENG M K, WU F L, WANG D. Lightweight Mars remote sensing image super-resolution reconstruction network[J]. *Opt. Precision Eng.*, 2022, 30(12):1487-1498. (in Chinese)
- [2] 李勇,王珂,张立保,等.多断层融合的肺CT肿瘤靶区超分辨率重建[J].*光学精密工程*,2010,18(5):1212-1218.  
LI Y, WANG K, ZHANG L B, *et al.* Super-resolution reconstruction of pulmonary nodules based on CT multi-section fusion[J]. *Opt. Precision Eng.*, 2010, 18(5):1212-1218. (in Chinese)
- [3] 王杰,徐国明,马健,等.轻量级注意力级联网络的偏振计算成像超分辨率重建[J].*光学精密工程*,2022,30(19):2404-2419.  
WANG J, XU G M, MA J, *et al.* Polarization computational imaging super-resolution reconstruction with lightweight attention cascading network [J]. *Opt. Precision Eng.*, 2022, 30(19):2404-2419. (in Chinese)
- [4] LI X, ORCHARD M T. New edge-directed interpolation[J]. *IEEE Transactions on Image Process-*

## 4 结 论

本文提出了一种基于注意力和宽激活密集残差网络的图像超分辨率重建模型,通过将图像的多尺度浅层局部特征和全局特征作为空间特征转换层的先验信息用以引导深层高频特征的提取,使得图像的空间信息一直作用于整个特征提取过程。然后,将轻量级注意力机制融入到宽激活残差块中,在不增加网络参数数量的情况下尽量扩大特征图数目,对高频信息自适应地分配更多的注意力,从而获得了更加丰富的高频细节特征信息。实验结果表明,本文所提方法在Urban100数据集上4倍超分辨率时,相比于LatticeNet模型PSNR指标提升了0.14 dB,并且模型参数量维持在较低水平,重建后的图像在主观视觉感受上边缘轮廓也更为清晰,纹理细节信息更加丰富。

*ing*, 2001, 10(10): 1521-1527.

- [5] LIU X M, ZHAO D B, ZHOU J T, *et al.* Image interpolation via graph-based Bayesian label propagation [J]. *IEEE Transactions on Image Processing*, 2014, 23(3): 1084-1096.
- [6] DAI S Y, HAN M, XU W, *et al.* SoftCuts: a soft edge smoothness prior for color image super-resolution [J]. *IEEE Transactions on Image Processing*, 2009, 18(5): 969-981.
- [7] DONG C, LOY C C, HE K M, *et al.* Image super-resolution using deep convolutional networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(2): 295-307.
- [8] 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]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 1874-1883.
- [9] 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)*. June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016:

- 1646-1654.
- [10] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 770-778.
- [11] KIM J, LEE J K, LEE K M. Deeply-recursive convolutional network for image super-resolution [C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 1637-1645.
- [12] TAI Y, YANG J, LIU X M. Image super-resolution via deep recursive residual network [C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 2790-2798.
- [13] 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)*. July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 1132-1140.
- [14] HUI Z, GAO X B, YANG Y C, *et al.* Lightweight image super-resolution with information multi-distillation network [C]. *Proceedings of the 27th ACM International Conference on Multimedia*. October 21 - 25, 2019, Nice, France. New York: ACM, 2019: 2024-2032.
- [15] LUO X T, XIE Y, ZHANG Y L, *et al.* *LatticeNet: Towards Lightweight Image Super-resolution with Lattice Block* [M]. Computer Vision - ECCV 2020. Cham: Springer International Publishing, 2020: 272-289.
- [16] TONG T, LI G, LIU X J, *et al.* Image super-resolution using dense skip connections [C]. 2017 *IEEE International Conference on Computer Vision (ICCV)*. October 22-29, 2017, Venice, Italy. IEEE, 2017: 4809-4817.
- [17] ZHANG Y L, TIAN Y P, KONG Y, *et al.* Residual dense network for image super-resolution [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 2472-2481.
- [18] 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)*. June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 1646-1654.
- [19] DAI T, CAI J R, ZHANG Y B, *et al.* Second-order attention network for single image super-resolution [C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 15-20, 2019, Long Beach, CA, USA. IEEE, 2020: 11057-11066.
- [20] WANG X L, GIRSHICK R, GUPTA A, *et al.* Non-local neural networks [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 7794-7803.
- [21] CHEN H, GU J, ZHANG Z. Attention in attention network for image super-resolution [EB/OL]. 2021: *arXiv*: 2104.09497. <https://arxiv.org/abs/2104.09497>.
- [22] LEDIG C, THEIS L, HUSZÁR F, *et al.* Photo-realistic single image super-resolution using a generative adversarial network [C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 105-114.
- [23] YU J, FAN Y, YANG J, *et al.* Wide activation for efficient and accurate image super-resolution [J]. *arXiv preprint arXiv*: 1808.08718, 2018.
- [24] SZEGEDY C, LIU W, JIA Y Q, *et al.* Going deeper with convolutions [C]. 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 7-12, 2015, Boston, MA, USA. IEEE, 2015: 1-9.
- [25] LI X, WANG W H, HU X L, *et al.* Selective kernel networks [C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 15-20, 2019, Long Beach, CA, USA. IEEE, 2020: 510-519.
- [26] WANG X T, YU K, DONG C, *et al.* Recovering realistic texture in image super-resolution by deep spatial feature transform [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 606-615.
- [27] ZEYDE R, ELAD M, PROTTER M. *On Single*

- Image Scale-up Using Sparse-representations* [M]. Curves and Surfaces. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012: 711-730.
- [28] 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. ICCV. July 7-14, 2001, Vancouver, BC, Canada.* IEEE, 2002: 416-423.
- [29] HUANG J B, SINGH A, AHUJA N. Single im-  
age super-resolution from transformed self-exem-  
plars [C]. *2015 IEEE Conference on Computer Vi-  
sion and Pattern Recognition (CVPR). June 7-  
12, 2015, Boston, MA, USA.* IEEE, 2015:  
5197-5206.
- [30] LAI W S, HUANG J B, AHUJA N, *et al.* Deep  
Laplacian pyramid networks for fast and accurate  
super-resolution [C]. *2017 IEEE Conference on  
Computer Vision and Pattern Recognition  
(CVPR). July 21-26, 2017, Honolulu, HI,  
USA.* IEEE, 2017: 5835-5843.

作者简介:



寇旗旗(1988—),男,河南许昌人,博士,讲师,主要从事视频/图像处理、智能检测与模式识别的研究。E-mail: kouqiqi@cumt.edu.cn



程德强(1979—),男,博士,教授,博士生导师,主要从事机器视觉与模式识别、图像智能检测与信息处理的研究。E-mail: chengdq@cumt.edu.cn