

## 自适应残差注意力的轻量级图像超分辨率重建模型

江明, 肖庆生, 易见兵\*, 曹锋

江西理工大学信息工程学院, 江西 赣州 341000

**摘要** 单幅图像超分辨率重建在众多领域应用广泛,然而现有算法大多通过扩展卷积神经网络深度和宽度来提取更多特征细节,导致算法计算复杂度的提高和模型参数数量的增大。为解决上述问题,提出一种自适应残差注意力的轻量级超分辨率网络算法。首先通过改进坐标注意力网络,生成了关注高频位置信息的注意力特征图;接着将改进的自适应残差注意力信息提取模块和坐标注意力模块双支路并行连接,使输出的特征信息包含更多图像细节;同时采用多尺度上采样方法,使训练得到的网络模型(单次训练)一次输出多个不同尺度的超分辨率图片。通过和经典的轻量级超分辨率算法比较,所提算法得到的4倍重建图片在4个公共数据集上的平均峰值信噪比(PSNR)提高了1.06 dB,模型参数数量减少了59%,图像观感更为清晰,包含更多高频细节。

**关键词** 图像超分辨率; 轻量级; 注意力机制; 卷积神经网络; 残差学习

中图分类号 TP394

文献标志码 A

DOI: 10.3788/LOP202259.1610007

## Lightweight Super-Resolution Image-Reconstruction Model with Adaptive Residual Attention

Jiang Ming, Xiao Qingsheng, Yi Jianbing\*, Cao Feng

School of Information Engineering, Jiangxi University of Science and Technology,  
Ganzhou 341000, Jiangxi, China

**Abstract** Super-resolution single-image reconstruction is widely used in many fields; however, most existing algorithms extract more feature details by extending the depth and width of the convolutional neural network, which increases the computational complexity as well as number of model parameters. To solve these problems, a lightweight super-resolution network algorithm with adaptive residual attention is proposed herein. This algorithm first generates an attention feature map, which focuses on high-frequency location information by improving the coordinate attention network, and then connecting the improved adaptive residual attention information-extraction module and coordinate attention module in parallel to obtain more image details in the output feature information. The multiscale upsampling method was used to enable the trained network model (single training) to output multiple super-resolution images with different scales at once. Compared with the classical lightweight super-resolution algorithm, the fourfold reconstructed image obtained using the proposed algorithm has an average peak signal-to-noise ratio (PSNR) improvement of 1.06 dB on four common datasets and the number of model parameters is reduced by 59%. Further, the image obtained using the proposed algorithm is clearer and contains more high-frequency details.

**Key words** image super-resolution; lightweight; attention mechanism; convolutional neural network; residual learning

## 1 引言

单幅图像超分辨率(SISR)是一种从低分辨率(LR)图像生成高分辨率(HR)图像的方法,该方法能够使生成图像具有较高的视觉质量和较好的纹理细

节。由于SISR的复杂性和在改善图像细节和纹理方面的实用价值,在当前该领域被广泛研究。超分辨率(SR)对于许多设备,如高清电视、电脑显示器、照相机、智能手机、平板电脑,都是至关重要的。此外,它还被应用于各种计算机视觉领域,如目标检测<sup>[1]</sup>、医学成

收稿日期: 2021-06-01; 修回日期: 2021-06-16; 录用日期: 2021-06-27

基金项目: 国家自然科学基金(61862031, 61665003)、江西省自然科学基金(20181BAB202004, 20192BAB207036)、江西省教育厅科技项目(GJJ190458, GJJ200818)、赣州市科技计划(GZKJ20206030)

通信作者: \*yjianbing8@163.com

像<sup>[2-3]</sup>、安全和监视成像<sup>[4]</sup>、人脸识别、遥感图像等。

自从 Dong 等<sup>[5]</sup>提出 SRCNN 算法以来,基于深度学习的 SR 算法开始逐渐出现。SRCNN 将一个经上采样后的低分辨率图像作为输入,经过三层卷积得到重建后的高分辨率图像,这种把上采样后的图像作为输入的方法需要额外的计算开销。为了解决上述问题,FSRCNN<sup>[6]</sup>和 ESPCN<sup>[7]</sup>通过对网络最后输出的特征图进行上采样,减少了大量计算和运行时间。与 SRCNN 算法的先上采样方法相比,后上采样策略能显著降低内存,但是由于没有足够的深层网络,没有提取到更深的特征信息,导致网络性能不足。Kim 等<sup>[8]</sup>提出的 VDSR 算法通过添加 20 层卷积来增加网络深度,并使用残差学习来缓解训练压力。为了减少模型参数, Kim 等<sup>[9]</sup>随后又提出了深度递归卷积网络,避免在增加网络深度的同时引入额外的参数,为了解决训练问题,作者采用了递归监督策略和跳跃连接。同样 Tai 等<sup>[10]</sup>提出深度递归残差网络,使用参数共享策略减少了深层网络参数,随后又在 MemNet<sup>[11]</sup>中提出了一种用于更深层网络的记忆模块,解决了 84 层卷积的长期依赖问题。Lim 等<sup>[12]</sup>通过扩展网络大小和去除残差块中的批量归一化层来增强残差块,从而得到增强型深度超分辨率网络(EDSR)。Zhang 等<sup>[13]</sup>提出了带有残差结构和密集跳跃连接的密集残差网络(RDN),该网络能够充分利用层次结构特征。

上述方法通过增加网络深度实现了更高的峰值信噪比(PSNR),但是参数量和计算量都有较大增加,由

此可能导致过拟合和对实际应用的限制。为解决上述问题,本文提出了一种基于自适应残差注意力的轻量级网络结构的超分辨率算法(ARASR),主要工作如下。1)在数据预处理中,将大图片裁剪成小尺寸图片作为输入,起到扩增数据集的作用,以减小过拟合的风险,且小尺寸图片能减少模型计算量。2)在网络结构的设计中,首先改进坐标注意力网络,使该网络不仅可以学习各个通道间的关系,还考虑了空间位置信息,生成关注高频位置信息的空间选择性注意力特征图;接着将改进的自适应残差注意力信息提取模块和坐标注意力模块双支路并行连接,使特征信息有效融合,达到输出的特征信息包含更多图像细节的目的。3)上采样模块通过亚像素级卷积和残差学习技术结合全局和局部特征,将低频特征转换为高频特征,同时多尺度上采样方法可以使训练得到的网络模型(单次训练)一次输出多个不同尺度的图片,由此可以节省大量的训练时间。

## 2 所提方法内容

### 2.1 网络结构概述

所提网络结构如图 1 所示,主要通过级联非线性特征映射模块和上采样重建模块来实现算法。 $I_{LR}$ 和  $I_{SR}$  分别代表模型的输入和最后的输出,使用一个卷积核大小为  $3 \times 3$  的卷积层来提取输入 LR 图像的浅层特征  $F_0$ :

$$F_0 = \text{Conv}_{3 \times 3}(I_{LR}). \quad (1)$$

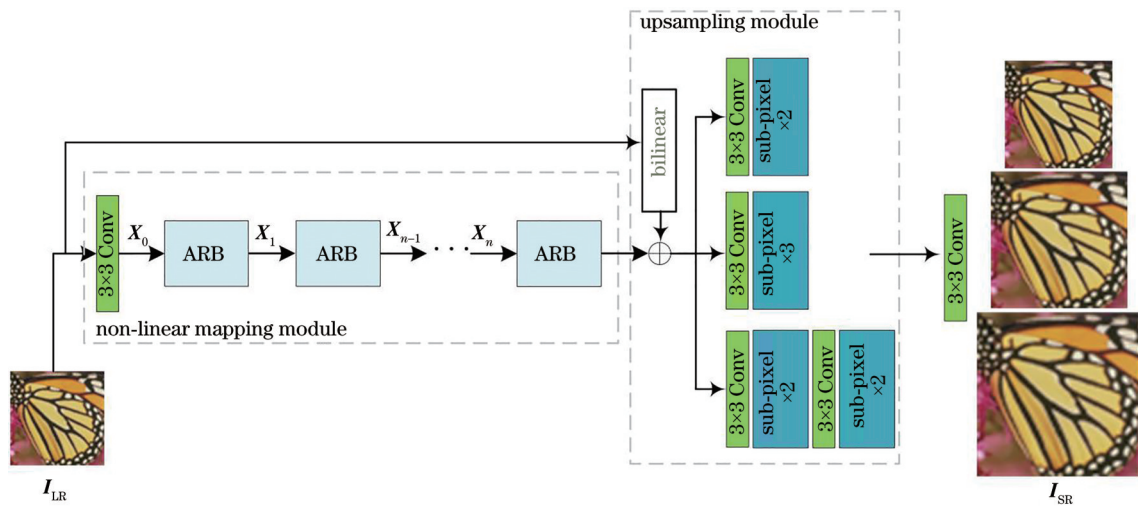


图 1 自适应残差注意力超分辨率网络结构

Fig. 1 Adaptive residual attention super-resolution network structure

采用  $3 \times 3$  卷积的目的是使用更少的参数来构建轻量级网络,同时在超分辨率任务中,特别是在第一层不宜采用感受野较大的核。由于下采样图像中的每个像素对应于原始图像中的一个小区域,因此在训练过程中大的感受野可能会引入不相关的信息。浅层特征  $F_0$  经过非线性特征映射模块完成信息提取,

该模块由 20 个自适应残差注意力信息提取模块(ARB)构成,在经过第  $n$  个自适应残差注意力信息提取模块时,特征  $X_n$  是从先前块输出特征  $X_{n-1}$  中提取到的,公式为

$$X_n = g_{ARB}^n(X_{n-1}), \quad (2)$$

式中:  $g_{ARB}^n$  表示第  $n$  个自适应残差注意力信息提取模

块。从最后一个残差注意力信息提取模块中获得的特征  $\mathbf{X}_L$  会经多尺度上采样模块,

$$\mathbf{X}_{\text{tail}} = F_{\text{tail}}(\mathbf{X}_L), \quad (3)$$

式中:  $F_{\text{tail}}$  表示多尺度上采样模块, 用于将  $\mathbf{X}_L$  上采样到目标特征  $\mathbf{X}_{\text{tail}}$ 。该模块集成非线性映射模块产生的多频信息, 并使用亚像素卷积对图像进行上采样。亚像素卷积(sub-pixel)出现在 ESPCN<sup>[7]</sup>中, 作为上采样的一种方式, 目的是使用亚像素卷积对一张低分辨率图像的特征图各通道间的像素进行重排列, 得到对应的高分辨率图像。为了使映射学习更容易, 引入原始低频信息, 保持低频精度, 采用与  $F_{\text{tail}}$  相同分量的双线性插值模块  $F_{\text{skip}}$ , 获得全局信息  $\mathbf{X}_{\text{skip}}$ :

$$\mathbf{X}_{\text{skip}} = F_{\text{skip}}(\mathbf{I}_{\text{LR}}). \quad (4)$$

最后, 目标图像  $\mathbf{I}_{\text{SR}}$  由  $\mathbf{X}_{\text{skip}}$  和  $\mathbf{X}_{\text{tail}}$  相加获得:

$$\mathbf{I}_{\text{SR}} = \mathbf{X}_{\text{tail}} + \mathbf{X}_{\text{skip}}. \quad (5)$$

表 1 展示了所提网络结构中卷积核的参数设置, 其中“ARB”表示自适应残差注意力信息提取模块, 数量为 20 个, 分为两个分支, 其中一个为“RB”残差分支, 另一个为“CoordA”坐标注意力分支; “kernel\_size”表示卷积核大小; “input\_channel”和“output\_channel”分别表示每个卷积核的输入通道数和输出通道数。另外在所提算法中上采样使用“pixelshuffle”亚像素卷积方法对输入图像进行放大, 损失函数使用 L1 正则化方法对算法进行约束。

表 1 所提网络结构的参数设置

Layer_name	Kernel_size	Input_channel	Output_channel
Conv1	(3, 3)	3	64
[ARB] × 20 (RB)	(1, 1)	3	32
	(3, 3)	32	32
[ARB] × 20 (CoordA)	(3, 3)	32	32
	(1, 1)	64	32
	(1, 1)	32	16
	(1, 1)	16	32
	(1, 1)	16	32
	(1, 1)	64	64
Upconv1	(3, 3)	64	256
Upconv2	(3, 3)	64	256
Conv_out	(3, 3)	64	3

## 2.2 坐标注意力

坐标注意力通过对特征图的水平方向和垂直方向进行一维全局平均池化, 在两个方向上都生成了具有空间方向的特征图, 然后经过不同的卷积运算, 生成关注位置信息的特征图。目前存在的基于注意力机制的超分辨率网络模型通常使用的是通道注意力网络和空间注意力网络, 这两种网络虽然能显著提高模型性能, 但位置信息却没有被考虑进去, 而位置信息对生成空

间选择性注意力特征图非常重要。受文献[14]启发, 对其中的坐标注意力网络进行改进: 由于图像的色彩分布信息经批量归一化层后被归一化, 破坏了原本对比度信息, 因此在该网络中去除其中的批量归一化层。

如图 2 所示,  $\mathbf{X}$  表示经  $1 \times 1$  卷积后的形状大小为  $(c, 64, 64)$  的输入特征,  $c$  表示通道数, 64 表示图像高和宽。分别对输入特征图的水平方向和垂直方向进行二元自适应平均池化(AdaptiveAvgPool2d); 通过从特征空间中自适应地提取特征的平均值来平滑和消除 LR 图像中的噪声, 降低维数, 并保留重要信息; 使用两个空间上的池化核  $(H, 1)$  和  $(1, W)$  分别沿横坐标和纵坐标对每个通道进行编码。输入  $\mathbf{X}$  的高度为  $h$ , 第  $c$  个通道的输出  $\mathbf{Y}$  表示为

$$\mathbf{Y}_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} \mathbf{X}_c(h, i). \quad (6)$$

同理, 输入  $\mathbf{X}$  的宽度为  $w$ , 第  $c$  个通道的输出  $\mathbf{Y}$  表示为

$$\mathbf{Y}_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} \mathbf{X}_c(j, w). \quad (7)$$

通过自适应池化层后, 得到一对对位置信息敏感的特征图  $\mathbf{Y}^h$  和  $\mathbf{Y}^w$ , 两者具有与输入  $\mathbf{X}$  相同的通道数, 形状大小分别为  $(c, 64, 1)$  和  $(c, 1, 64)$ , 然后对  $\mathbf{Y}^h$  和  $\mathbf{Y}^w$  进行合并操作, 并通过一个  $1 \times 1$  的卷积操作得到  $\mathbf{f}$ :

$$\mathbf{f} = \delta \left\{ \text{Conv}_{1 \times 1} \left[ \text{Cat}(\mathbf{Y}^h, \mathbf{Y}^w) \right] \right\}, \quad (8)$$

式中:  $\text{Cat}(\cdot)$  表示合并操作;  $\delta$  表示 H\_Sigmoid 非线性激活函数。然后输入  $\mathbf{f}$  沿空间维度分成两个独立张量  $\mathbf{f}^h$  和  $\mathbf{f}^w$ , 分别经两个  $1 \times 1$  卷积层, 转换成与输入  $\mathbf{X}$  具

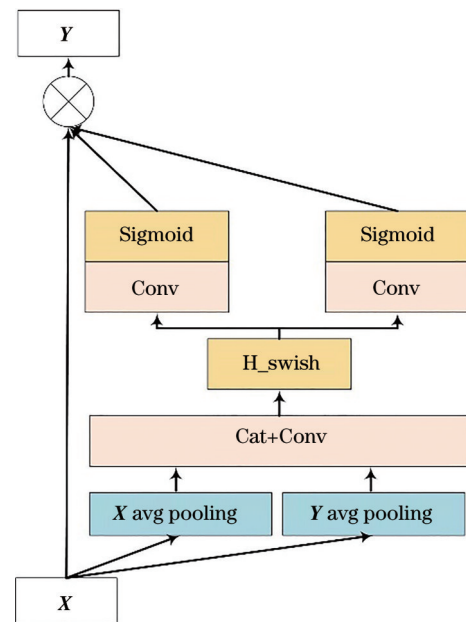


图 2 坐标注意力

Fig. 2 Coordinate attention

有相同通道数的张量,得到

$$g^h = \sigma[\text{Conv}_{1 \times 1}(f^h)], \quad (9)$$

$$g^w = \sigma[\text{Conv}_{1 \times 1}(f^w)], \quad (10)$$

式中: $\sigma$ 表示 Sigmoid 函数。输出的  $g^h$  和  $g^w$  作为注意力权重,最后从坐标注意力模块输出的  $Y$  表示为

$$Y_c(i, j) = X_c(i, j) * g_c^h(i) * g_c^w(j). \quad (11)$$

通过 Set14 数据集上的“baboon”图像在不同网络层下的注意力图的热力图来表示坐标注意力对图像高

频细节部分的影响,如图 3 所示,block1, block9, block15, block20 分别表示第 1, 9, 15, 20 个 ARB 的输出图像,在输出特征时采用 Sigmoid 函数使特征值约束到  $[0, 1]$ , 图像越亮的部分表示注意力更加关注的地方。从图 3 可以看到,在低层次网络中注意力更加关注图像的整体轮廓,在网络深处,注意力更加倾向于图像的高频细节,由此说明该坐标注意力对高频细节位置信息有更强的关注度。

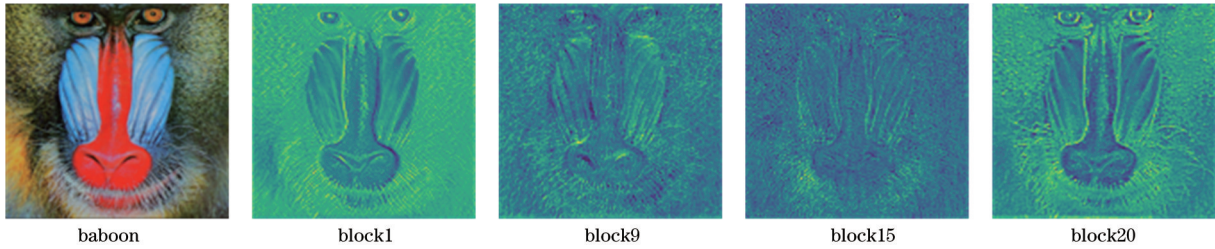


图 3 不同网络层下的注意力图热力图

Fig. 3 Heat maps of attention map under different network layers

### 2.3 自适应残差注意力信息提取模块

文献[15]提出的自矫正卷积通过拆分特征通道,使用两个分支对输入特征进行不同功能的处理。自矫正分支通过注意力模块得到注意力特征图,卷积分支则进行常规特征提取,将两者的特征融合作为最后输出,自矫正分支可以仅关注注意力图的位置信息,减少额外的学习参数。由于低分辨率图像包含大量的低频信息和少量的高频信息,低频信息往往更容易学习,同时为了更好地使模型学习到高频信息,采用残差块中的跳跃连接将低频信息与输出特征相加,使模型直接学习更复杂的高频信息,同时缓解训练压力。因此受自矫正卷积[15]启发,提出了自适应残差注意力信息提取模块,它是组成非线性映射模块的核心,与单一信息流的设计不同,自适应残差注意力信息提取模块主要采用残差模块与坐标注意力模块并行的模式,残差模块负责提取 LR 图像的特征信息,坐标注意力分支能自适应生成对位置信息敏感的特征图,然后对特征信息进行融合获得更好的结果。如图 4 所示,图中“RB”表示残差块,“CoordAttention”表示坐标注意力模块,输入特征  $X_{n-1}$  在经过两条支路前,会经过  $1 \times 1$  卷积层  $f^1$  和  $f^2$ , 它们的主要作用是分离通道,实现降维并减少运算量,表达式为

$$H_1 = f^1(X_{n-1}), \quad (12)$$

$$H_2 = f^2(X_{n-1}), \quad (13)$$

式中: $H_1, H_2$  分别表示两条支路经过  $1 \times 1$  卷积的输出。 $H_1, H_2$  分别经过残差模块和坐标注意力后,对输出特征进行融合,最后融合结果经过一个  $1 \times 1$  卷积层  $f^3$ , 采用跳跃连接的方式缓解梯度消失的问题,生成最后的  $X_n$ :

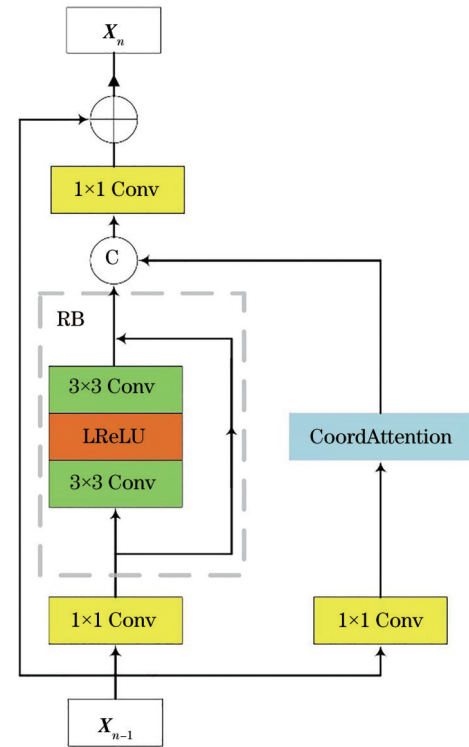


图 4 自适应残差注意力信息提取模块

Fig. 4 Adaptive residual attention information extraction module

$$X_n = f^3[\text{RB}(H_1), \text{CA}(H_2)] + X_{n-1}, \quad (14)$$

式中: $\text{RB}(\cdot)$ 表示残差模块的输出; $\text{CA}(\cdot)$ 表示坐标注意力的输出; $[\cdot, \cdot]$ 表示将两个输出特征拼接。

### 2.4 损失函数

使用 L1 损失函数优化网络,相比均方误差(MSE)损失,L1 损失对相对误差的惩罚力度更小,对图像纹理细节重建效果会更好。给定训练集  $\{I_{\text{LR}}^m, I_{\text{HR}}^m\}_{m=1}^M$ , 其中  $I_{\text{LR}}^m, I_{\text{HR}}^m$

分别表示低分辨率图像和高分辨率图像,  $M$  表示批量训练集大小, 参数集为  $\theta$ , 所提算法的损失函数表达式为

$$L(\theta) = \frac{1}{M} \sum_{m=1}^M |f_{\text{ARASR}}(I_{\text{LR}}^m) - I_{\text{HR}}^m|, \quad (15)$$

式中:  $f_{\text{ARASR}}$  表示所提网络模型。

### 3 实验

#### 3.1 实验环境

实验的硬件环境为: CPU 主频 3.6 GHz, 内存 32 GB, 显卡为 NVIDIA RTX 2080Ti。软件环境为: 操作系统为 Ubuntu18.04, 深度学习框架为 Pytorch, 编程语言为 Python3.7。

#### 3.2 数据集及参数设置

使用公共数据集 DIV2K<sup>[16]</sup> 中的 800 张高质量 RGB 训练图片作为训练集, Set5<sup>[17]</sup>、Set14<sup>[18]</sup>、BSD100<sup>[19]</sup>、Urban100<sup>[20]</sup> 四个数据集作为标准测试集, 其中, Set5、Set14 和 BSD100 包含自然场景图片, Urban100 包含具有挑战性的城市市场景图像, 细节分布在不同频段。通过 90°, 180°, 270° 旋转和水平翻转进行数据增强。为了评估模型性能, 在 YCbCr 通道的 Y 通道 (即亮度) 上计算峰值信噪比 (PSNR) 和结构相似性 (SSIM)。

在训练阶段, 采用 Adam 算法对模型收敛速度进行优化; 参数设置为  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ,  $\epsilon = 10^{-8}$ ; 初始学习率设置为  $1 \times 10^{-4}$ , 每 200 个 epoch 学习率减半, 模型总共训练 600 个 epoch; batchsize 设置为 32, 放大因子为  $2 \times$ 、 $3 \times$ 、 $4 \times$ , 分别使用大小为  $128 \times 128$ 、 $192 \times 192$ 、 $256 \times 256$  的 HR 子图作为标签,  $64 \times 64$  大小的 LR 子图作为输入。

#### 3.3 实验结果与分析

##### 3.3.1 客观评估结果与分析

###### 1) 算法中改进模块的性能评估

为了验证所提算法中的自适应残差注意力信息提取模块的有效性和改进坐标注意力网络的可行性, 为了保证实验的公平性, 以 EDSR 算法<sup>[12]</sup> 作为基准模型, 设置模块数都为 16。“Baseline”表示 EDSR 基准模型, 由 16 个残差模块构成; “Baseline+CA”表示残差模块与通道注意力模块级联; “Baseline+CoordA”表示残差模块与坐标注意力模块级联; “ARASR”表示残差模块与坐标注意力模块并行连接, 也是本文采用的网络结构。为了训练公平性, 实验都训练了  $5 \times 10^5$  迭代轮数。表 2 显示了不同模块组合在 4 倍放大时在 Set5 数据集上的 PSNR 值, 可以发现, 所提自适应残差注意力信息提取模块 ARASR 对比其他算法在参数量上有明显减少, PSNR 值也有一定的提高, 由此证明了自适应残差注意力信息提取模块的有效性。

###### 2) 算法中 ARB 模块数量对模型性能的影响

为了探究模型网络深度对模型性能的影响, 分别选取 16, 18, 20, 22 个 ARB 模块数量用于实验, 图 5 表

表 2 放大倍数为 4 时, 不同模块组合的参数数量及在 Set5 数据集下的 PSNR 值

Table 2 Number of parameters of different module combinations and PSNR values under Set5 dataset at a magnification of 4

Module	Number of parameters / $10^3$	PSNR / dB
Baseline	1517	25.98
Baseline+CA	1526	26.02
Baseline+CoordA	1544	26.03
ARASR(ours)	922	26.09

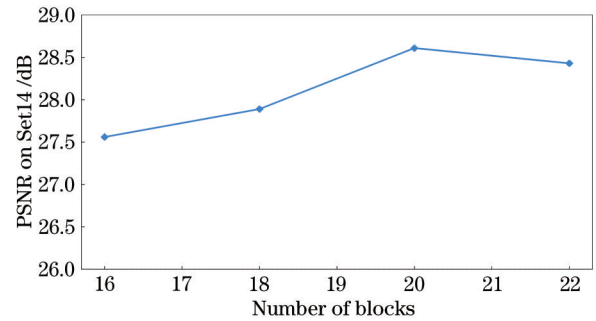


图 5 ARB 模块数量对模型性能的影响

Fig. 5 Effect of number of ARB modules on model performance

示了含有不同 ARB 模块数量的模型在 Set14 数据集上的 PSNR 值, 可以看出在网络不够深的情况下 PSNR 值偏低, 原因可能是网络没有提取足够深层的特征, 导致模型无法学习到图像的高频细节; 在网络加深到一定层后, 继续加深不会提高 PSNR 值; 在 ARB 模块数量为 20 时, PSNR 值最高, 模型提取特征能力趋于饱和, 因此最终选取 ARB 模块数量为 20。

###### 3) 算法整体性能评估

为评估所提算法的性能, 采用所提算法得到的结果与 SRCNN<sup>[5]</sup>、FSRCNN<sup>[6]</sup>、VDSR<sup>[8]</sup>、DRCN<sup>[9]</sup>、LapSRN<sup>[21]</sup>、DRRN<sup>[10]</sup>、CARN<sup>[22]</sup>、MemNet<sup>[11]</sup>、SRMDNF<sup>[23]</sup>、EDSR-baseline<sup>[19]</sup> 算法的实验结果进行对比, 分别比较了  $2 \times$ 、 $3 \times$ 、 $4 \times$  三种放大倍数下的 PSNR、SSIM 值及参数量大小。

表 3 分别展示了不同放大倍数下的模型参数量和客观质量评价指标值。可以看出: 在 3 种放大倍数下, 所提 ARASR 相比其他算法的 PSNR 和 SSIM 值都有一定的提高, 特别是在放大倍数为 2 时, 对比基准模型 EDSR, 在四个数据集上的 PSNR 值分别提高了 0.03 dB, 0.14 dB, 0.03 dB, 0.26 dB, 可以看到本文模型在复杂场景 Urban100 数据集中表现更为出色; 同时在参数量上, 比 CARN 和基准模型 EDSR 各减少了  $818 \times 10^3$ ,  $596 \times 10^3$ 。为了探究网络的宽度对模型性能的影响, 设置了两组实验, 参数条件为: ARASR-s 算法的 ARB 网络的模块数量为 20, 输入通道数为 40; 所提算法

表 3 四种数据集上,不同 SR 算法在  $2\times$ 、 $3\times$  和  $4\times$  上的平均 PSNR 和 SSIMTable 3 Average PSNR and SSIM of different SR algorithms under  $2\times$ ,  $3\times$ , and  $4\times$  magnification on the four datasets

Method	Scale	Number of parameters / $10^3$	Set5		Set14		BSD100		Urban100	
			PSNR /dB	SSIM	PSNR /dB	SSIM	PSNR /dB	SSIM	PSNR /dB	SSIM
Bicubic			33.66	0.9299	30.24	0.8688	29.56	0.8431	26.88	0.8403
SRCNN		57	36.66	0.9542	32.45	0.9067	31.36	0.8879	29.50	0.8946
FSRCNN		13	37.00	0.9558	32.63	0.9088	31.53	0.8920	29.88	0.9020
VDSR		666	37.53	0.9587	33.03	0.9124	31.90	0.8942	30.75	0.9133
DRCN		1774	37.63	0.9588	33.04	0.9118	31.85	0.8942	30.75	0.9133
LapSRN		251	37.52	0.9591	32.99	0.9124	31.80	0.8952	30.41	0.9103
DRRN	$2\times$	298	37.74	0.9591	33.23	0.9136	32.05	0.8973	31.23	0.9188
MemNet		678	37.78	0.9597	33.28	0.9142	32.08	0.8978	31.31	0.9195
EDSR-baseline		1370	37.99	0.9604	33.57	0.9175	32.16	0.8994	31.98	0.9272
SRMDNF		1511	37.79	0.9601	33.32	0.9159	32.05	0.8985	31.33	0.8204
CARN		1592	37.76	0.9590	33.52	0.9166	32.09	0.8978	31.92	0.9256
ARASR-s		307	37.88	0.9602	33.49	0.9171	32.12	0.8991	31.92	0.9266
ARASR		774	38.02	0.9606	33.71	0.9189	32.19	0.9000	32.24	0.9297
Bicubic			30.39	0.8682	27.55	0.7742	27.21	0.7385	24.46	0.7349
SRCNN		57	32.75	0.9090	29.30	0.8215	28.41	0.7863	26.24	0.7989
FSRCNN		13	33.18	0.9140	29.37	0.8240	28.53	0.7910	26.43	0.8080
VDSR		666	33.66	0.9213	29.77	0.8314	28.82	0.7976	27.14	0.8279
DRCN		1774	33.82	0.9226	29.76	0.8311	28.80	0.7963	27.15	0.8276
LapSRN		502	33.81	0.9220	29.79	0.8325	28.82	0.7980	27.07	0.8275
DRRN	$3\times$	298	34.03	0.9244	29.96	0.8349	28.96	0.8004	27.53	0.8378
MemNet		678	34.09	0.9248	30.00	0.8350	28.96	0.8001	27.56	0.8376
EDSR-baseline		1555	34.37	0.9270	30.28	0.8417	29.09	0.8052	28.15	0.8527
SRMDNF		1528	34.12	0.9254	30.04	0.8382	28.97	0.8025	27.57	0.8398
CARN		1592	34.29	0.9255	30.29	0.8407	29.06	0.8034	28.06	0.8493
ARASR-s		379	34.24	0.9262	30.27	0.8411	29.03	0.8038	27.98	0.8488
ARASR		959	34.37	0.9270	30.39	0.8433	29.11	0.8056	28.29	0.8552
Bicubic			28.42	0.8104	26.00	0.7027	25.96	0.6675	23.14	0.6577
SRCNN		57	30.48	0.8628	27.50	0.7513	26.90	0.7101	24.52	0.7221
FSRCNN		13	30.72	0.8660	27.61	0.7550	26.98	0.7150	24.62	0.7280
VDSR		665	31.35	0.8838	28.01	0.7674	27.29	0.7251	25.18	0.7524
DRCN		1774	31.53	0.8854	28.02	0.7670	27.23	0.7233	25.14	0.7510
LapSRN		813	31.54	0.8852	28.09	0.7700	27.32	0.7275	25.21	0.7562
DRRN	$4\times$	297	31.68	0.8888	28.21	0.7720	27.38	0.7284	25.44	0.7638
MemNet		678	31.74	0.8893	28.26	0.7723	27.40	0.7281	25.50	0.7630
EDSR-baseline		1518	32.09	0.8938	28.58	0.7813	27.57	0.7357	26.04	0.7849
SRMDNF		1552	31.96	0.8925	28.35	0.7787	27.49	0.7337	25.68	0.7731
CARN		1592	32.13	0.8937	28.60	0.7806	27.58	0.7349	26.07	0.7837
ARASR-s		365	32.06	0.8932	28.50	0.7790	27.51	0.7342	25.93	0.7807
ARASR		922	32.20	0.8951	28.61	0.7821	27.58	0.7367	26.15	0.7876

(ARASR)的 ARB 网络的模块数量为 20,输入通道数为 64。从表 3 可以看出,网络的宽度对模型参数数量的影

响巨大,在放大倍数为 4 时,对比 ARASR, ARASR-s 参数量减少了 60%,但模型性能有所下降。

图 6 表示在 2 倍放大时 Set14 数据集上各算法的模型参数量和 PSNR 值的对应关系,从图中可以看出所提模型在性能和参数量上有较好的权衡。

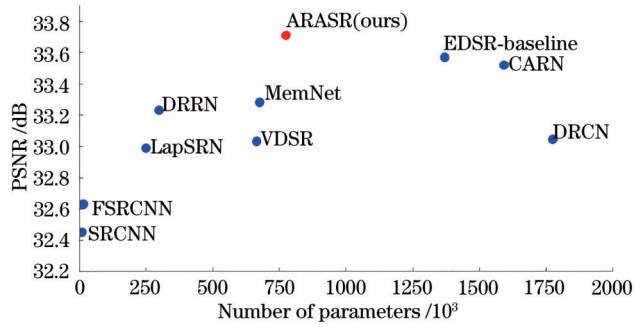


图 6 各算法的模型参数量和 PSNR 值的对应关系  
Fig. 6 Correspondence between number of model parameters and PSNR values for each algorithm

### 3.3.2 主观效果展示与分析

图 7 展示了在放大倍数为 4 时的模型生成效果图对比。可以看到:由双三次插值得到的斑马条纹图片整体模糊;从 SRCNN 等算法重建的效果图中能明显看到局部模糊的现象;所提算法对这种斑马条纹不规则图像最大程度还原了高频细节并极大减少了振铃效应。

图 8、图 9 展示了在放大倍数为 4 时 Urban100 数据集中的两张图片“img\_012”和“img\_046”的重建效果。从“img\_046”可以看出,所提算法能更好地恢复图像边缘和纹理细节,相比其他算法,更接近真实图像。从“img\_012”可以看出,其他算法在一定程度上都会对图像中的纹理方向产生错误的重建,同时会造成棋盘效应,所提算法能很好地重建出正确的纹理细节,重建图像更加接近真实图像。

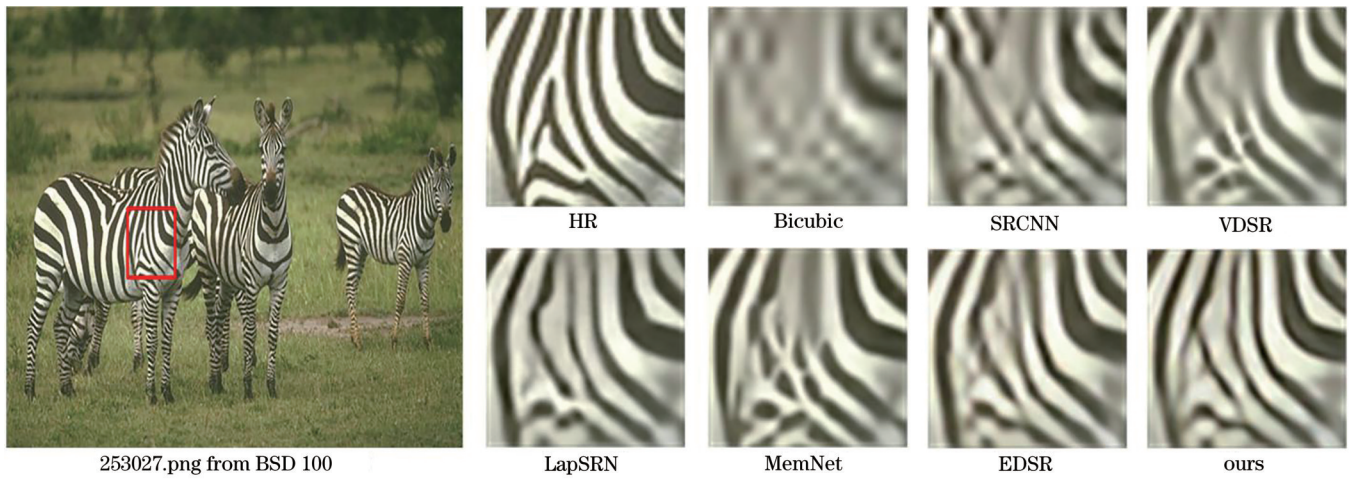


图 7 不同算法对 BSD100 数据集中的“253027.png”图像放大 4× 的效果  
Fig. 7 Results of different algorithms for the “253027.png” in BSD100 dataset under 4× magnification

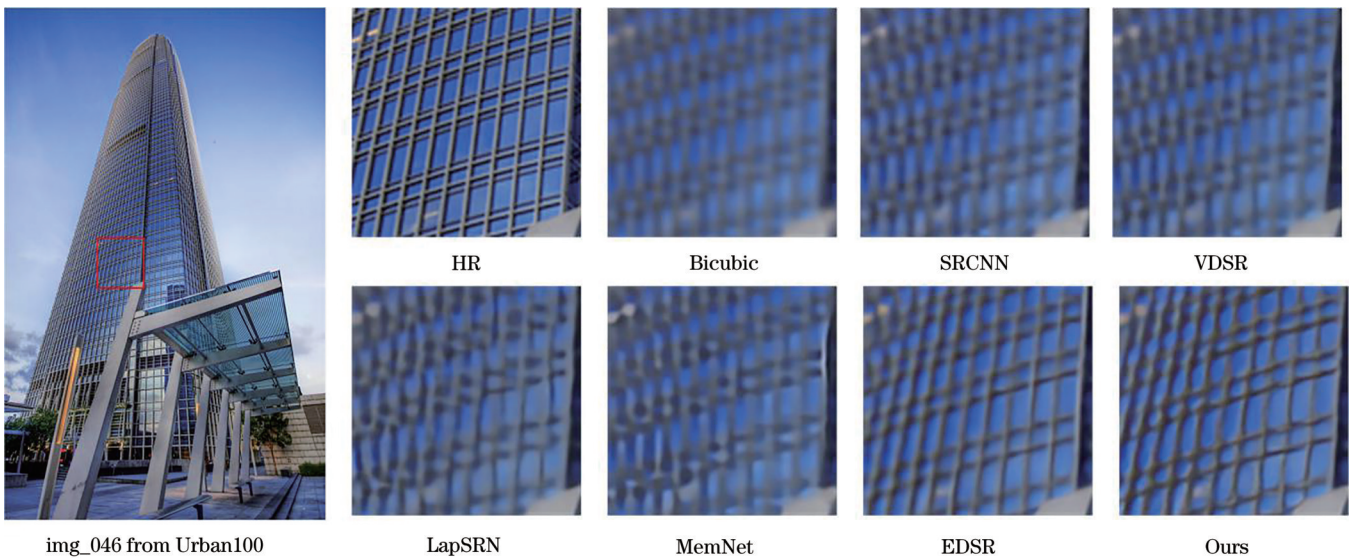


图 8 不同算法对 Urban100 数据集中的“img\_046”图像放大 4× 的效果  
Fig. 8 Results of different algorithms for the “img\_046” in Urban100 dataset under 4× magnification

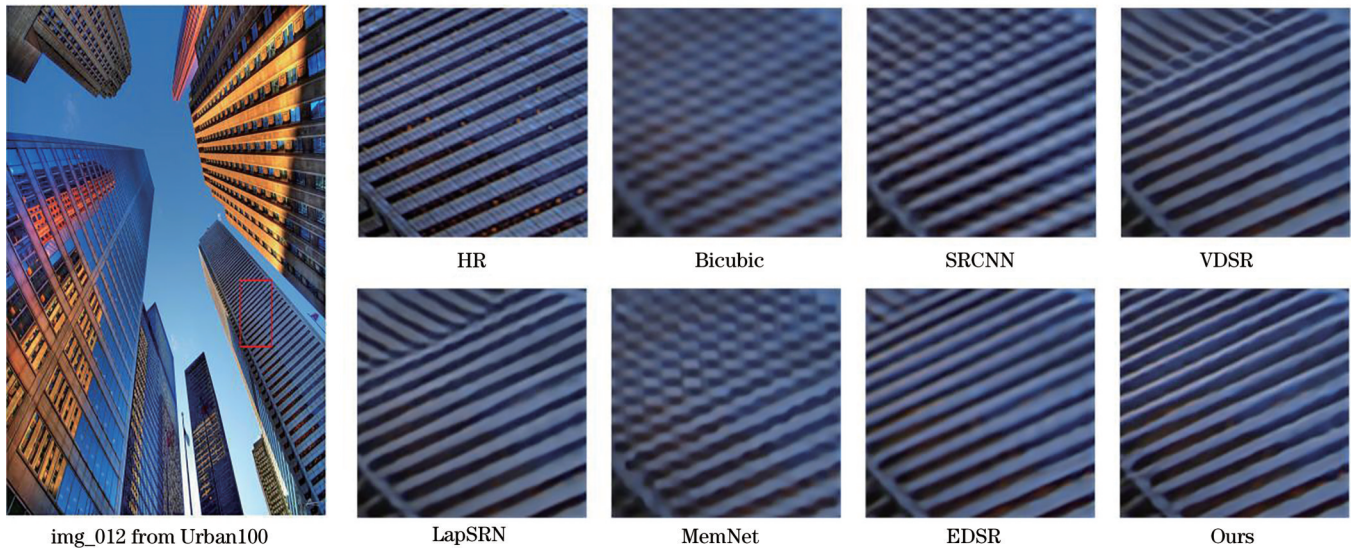


图9 不同算法对Urban100数据集中的“img\_012”图像放大4×的效果

Fig. 9 Results of different algorithms for the “img\_012” in Urban100 dataset under 4× magnification

## 4 结 论

在单幅图像超分辨率的任务上,同时减少模型参数量并提高模型性能一直是技术难点,提出了一种自适应残差注意力网络,可以实现轻量级、准确的单图像超分辨率。所提方法主要涉及级联自适应残差注意力信息提取模块。该模块由残差结构和坐标注意力并行方式构成,残差结构用于提取特征,缓解深层训练难度,坐标注意力模块用于自适应生成对高频细节位置信息敏感的特征图。实验结果表明:所提方法相比性能更好的算法在参数量上能有所减少,在客观评价指标上也能达到最优;在视觉质量上,所提方法能很好地处理模糊伪影,并对纹理细节进行更好的重建。同时可以看到网络宽度对参数量和模型性能影响较大,在后续工作中会考虑到如何有效减小网络宽度,并减小参数量和提高模型性能。

## 参 考 文 献

- [1] Girshick R, Donahue J, Darrell T, et al. Region-based convolutional networks for accurate object detection and segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(1): 142-158.
- [2] 胡芬, 林洋, 侯梦迪, 等. 基于深度学习的细胞骨架图像超分辨重建[J]. *光学学报*, 2020, 40(24): 2410001.  
Hu F, Lin Y, Hou M D, et al. Super-resolution reconstruction of cytoskeleton image based on deep learning[J]. *Acta Optica Sinica*, 2020, 40(24): 2410001.
- [3] 刘智, 罗泽伟, 王正印, 等. 基于结构照明的超分辨荧光显微成像重建算法[J]. *中国激光*, 2021, 48(3): 0307001.  
Liu Z, Luo Z W, Wang Z Y, et al. Super-resolution fluorescence microscopy image reconstruction algorithm based on structured illumination[J]. *Chinese Journal of Lasers*, 2021, 48(3): 0307001.
- [4] 张智敏, 黄宇然, 刘少聪, 等. 共路并行荧光辐射差分超分辨显微成像[J]. *中国激光*, 2021, 48(16): 1607002.  
Zhang Z M, Huang Y R, Liu S C, et al. Common-path parallel fluorescence emission difference super-resolution microscopy[J]. *Chinese Journal of Lasers*, 2021, 48(16): 1607002.
- [5] Dong C, Loy C C, He K M, et al. Learning a deep convolutional network for image super-resolution[M]// Fleet D, Pajdla T, Schiele B, et al. *Computer vision-ECCV 2014. Lecture notes in computer science*. Cham: Springer, 2014, 8692: 184-199.
- [6] Dong C, Loy C C, Tang X O. Accelerating the super-resolution convolutional neural network[M]// Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9906: 391-407.
- [7] 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, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1874-1883.
- [8] 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, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1646-1654.
- [9] 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, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1637-1645.
- [10] 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, July 21-



- 26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2790-2798.
- [11] Tai Y, Yang J, Liu X M, et al. MemNet: a persistent memory network for image restoration[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 4549-4557.
- [12] 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, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1132-1140.
- [13] 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. New York: IEEE Press, 2018: 2472-2481.
- [14] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 13708-13717.
- [15] Liu J J, Hou Q B, Cheng M M, et al. Improving convolutional networks with self-calibrated convolutions [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 10093-10102.
- [16] 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, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1122-1131.
- [17] Bevilacqua M, Roumy A, Guillemot C, et al. Low-complexity single-image super-resolution based on nonnegative neighbor embedding[C]//Proceedings of the British Machine Vision Conference 2012, September 19, 2012, Surrey. London: British Machine Vision Association, 2012: 1-10.
- [18] Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations[M]//Boissonnat J D, Chenin P, Cohen A, et al. Curves and surfaces. Lecture notes in computer science. Heidelberg: Springer, 2012, 6920: 711-730.
- [19] 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 of Eighth IEEE International Conference on Computer Vision, July 7-14, 2001, Vancouver, BC, Canada. New York: IEEE Press, 2001: 416-423.
- [20] Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 5197-5206.
- [21] 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, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5835-5843.
- [22] Ahn N, Kang B, Sohn K A. Fast, accurate, and lightweight super-resolution with cascading residual network[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11214: 256-272.
- [23] Zhang K, Zuo W M, Zhang L. Learning a single convolutional super-resolution network for multiple degradations[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 3262-3271.