

文章编号: 1002-1582(2022)06-0731-11

# 结合注意力和残差聚合的 图像超分辨率算法

姜继升, 徐开雄, 李华锋, 李凡\*

- (1. 昆明理工大学 信息工程与自动化学院, 云南 昆明 650500)
- (2. 云南省人工智能重点实验室, 云南 昆明 650500)

**摘要:** 为解决残差块间的层级特征利用不充分导致生成的图像结构扭曲、视觉模糊等问题, 提出结合注意力和残差聚合的图像超分辨率重建算法。网络通过浅层特征聚合模块获取多尺度的特征并输入到残差聚合网络, 采用渐进式融合策略从局部和全局两方面对各个残差块的特征进行聚合, 以达到充分利用残差块层级特征的目的。为进一步增强特征表示, 利用双重注意力机制分别从空间和通道来关注特征之间的相互依赖性。实验结果表明, 与 SRCNN、FSRCNN 等方法相比, 算法重建的图像结构清晰且细节信息丰富。

**关键词:** 超分辨率; 残差聚合; 注意力机制; 层级特征

中图分类号: TP751.1 文献标识码: A

DOI:10.13741/j.cnki.11-1879/o4.2022.06.015

## Image super-resolution method combining attention and residual aggregation

JIANG Jisheng, XU Kaixiong, LI Huafeng, LI Fan\*

- (1. School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China)
- (2. Yunnan Provincial Key Laboratory of Artificial Intelligence, Kunming 650500, China)

**Abstract:** In order to solve the problems of image structure distortion and visual blur caused by insufficient utilization of hierarchical features between residual blocks, an image super-resolution reconstruction algorithm combining attention and residual aggregation is proposed. The network obtains multi-scale features through the shallow feature aggregation module and inputs them to the residual aggregation network, and adopts the progressive fusion strategy to aggregate the features of each residual block from both local and global aspects, so as to make full use of the level features of residual block. To further enhance the feature representation, a dual attention mechanism is used to focus on the interdependencies between features from space and channel, respectively. The experimental results show that, compared with SRCNN, FSRCNN and other methods, the reconstructed images have clearer structure and richer details.

**Key words:** super-resolution; residual aggregation; attention mechanism; hierarchical features

收稿日期: 2022-05-29; 收到修改稿日期: 2022-06-27

基金项目: 国家自然科学基金项目(62161015); 云南省科技厅科技计划项目(基础研究专项)(202101AT070136)

作者简介: 姜继升(1996—), 男, 硕士研究生, 从事计算机视觉、数字图像处理、模式识别研究。2661951972@qq.com

李凡(1986—), 男, 副教授, 博士研究生, 从事计算机视觉、数字图像处理、模式识别研究。

\* 通讯作者: 478263823@qq.com

## 0 引言

图像超分辨率重建技术可将低分辨率图像进行处理,生成细节丰富且质量高清的图像,因此该技术广泛应用于视频监控、目标检测<sup>[15]</sup>和医学成像等领域,具有非常重要的研究价值。但是,对于任何一个 LR(low resolution, LR)输入,都有多种 HR(high resolution, HR)解决方案与其对应,导致了该技术的高度不确定性。为了解决这一问题,大量的方法被提出,例如基于插值的方法<sup>[29]</sup>和基于重建的<sup>[30]</sup>方法。

随着深度学习的迅速发展,基于该学习的卷积神经网络也被用于解决这一问题。最早是由 Dong 等设计了一个三层卷积神经网络 SRCNN<sup>[2]</sup>,该网络建立了插值法合成的 LR 与 HR 图像之间一一对应的映射关系,为后续的超分辨率重建方法奠定了坚实的基础。接着 Kim 等提出了 VDSR<sup>[10]</sup>和 DRCN<sup>[11]</sup>,通过增加网络深度和引入残差学习进一步提升网络性能,同时大大减轻训练难度。除此之外,还有很多方法<sup>[12,13,16,33,36,37]</sup>被用于单一图像的超分辨率重建。

尽管上述方法在该领域取得了很大成就,但它们都存在一些共有的缺陷。随着网络深度的增加,会出现梯度消失或者梯度爆炸,这必然会影响到网络训练。为了解决这些问题,一种残差框架<sup>[9]</sup>被提出,用于解决层数增加带来的梯度消失或者梯度爆炸。接着一种密集连接<sup>[8]</sup>的方式被提出,用于促进层间信息的流通。由于残差框架中的残差连接和残差块有助于模型重建出高清质量的图像,所以超分辨率模型常常会堆叠若干残差块来提取特征。在重建过程中,每个残差块会获取不同的特征,它们包含了许多输入图像的细节信息,但是上述方法都没有充分利用这些残差块与残差块间的层级特征,而这些特征对于图像的重建至关重要。

为此,本文提出一个双重调节残差聚合网络(Dual Adjustment Residual Aggregation Network, DARAN),充分利用残差块层级特征,提高图像重建质量。总体来讲,本文的贡献如下:

(1)本文对已有的浅层特征提取部分进行了优化,提出了一个浅层特征聚合模块(Shallow Feature Aggregation, SFA),该模块一方面利用由不同大小的卷积核构成的残差块来提取特征,另一方面将融合后的多尺度特征与最原始的浅层特征再次进行融合,其目的就是为了聚合更多输入图像的细节信息,实验结果进一步证明了该模块与已有模块的差异性

与优越性。

(2)本文提出了一个局部残差特征聚合网络(Local Residual Feature Aggregation, LRFA),与以往的串联式残差结构不同,该网络利用并联式残差结构来提取特征,以便利用不同大小的卷积核提取多个尺度的特征。采取的融合策略为渐进式融合,首次融合与二次融合均是从局部的角度出发,将残差块的特征两两融合,最后的融合从全局的角度出发,将所有的残差特征充分利用起来,生成更有利于重建的特征表示。为了进一步提高该网络的性能,本文使用了通道调节模块(Channel Adjustment Block, CAB)和空间调节模块(Spatial Adjustment Block, SAB),实验结果表明,二者的合理布局能够使残差聚合特征分别在通道和空间进行自适应调节,关注到细节信息更丰富的区域,使生成的特征具有更强的可鉴别性。

## 1 相关工作

### 1.1 基于深度学习的图像超分辨率重建

深度卷积神经网络得益于深度学习的进步,其性能已远远超越了基于传统方法的网络。最早是三层的 SRCNN<sup>[2]</sup>被提出,该方法利用插值法将低分辨率图像放大至目标图像尺寸,接着利用卷积神经网络进行非线性映射,得到超分辨率图像,这为后续出现的超分辨率模型奠定了基础。接着 Kim 等增加了网络的深度并引入了残差学习,设计了 VDSR<sup>[10]</sup>和 DRCN<sup>[11]</sup>。前者将插值后变成目标尺寸的低分辨率图像作为网络的输入,再将网络学到的残差与该图像相加得到最终的高分辨率图像;后者首次将递归神经网络应用在超分辨率网络中并利用残差思想提高了性能。然后, Tai 等提出了递归块和内存块,分别设计了 DRRN<sup>[24]</sup>和 MemNet<sup>[25]</sup>。相比 DRCN, DRRN 采用的是多路径模式的局部残差学习和全局残差学习以及多权重的递归学习,而 MemNet 是引入了一个包含递归单元和门控单元的内存块,以期通过自适应学习过程挖掘持续性记忆。以上这些方法都是以插值法预处理合成的 LR 图像作为输入,这种方式耗费大量的内存和计算时间。为了解决这个问题, ESPCN<sup>[21]</sup>提出了一个亚像素卷积层,该卷积层通过简单的数学运算,取代了普通的上采样,效率非常高。后续提出的很多深度卷积神经网络都依靠亚像素卷积层获得了更高的性能。接着,一种非常深的 EDSR<sup>[12]</sup>被提出,该网络利用去除 BN 层的残差块来提取特征,而 RDN<sup>[32]</sup>采用密集连接的方式来获取特征,进行超分辨率重建,性能获

得了进一步提升。后来,SPSR<sup>[3]</sup>被提出,该网络利用图像的梯度先验和结构信息,使得重建后的图像视觉质量更佳,接着 LESRCNN<sup>[42]</sup>被提出,该网络采用异构结构,结合低频和高频特征,以此提升网络性能。其中,异构结构能够在不降低 SR 性能的基础上减少网络参数同时提高 SR 网络效率,低频和高频特征的相互结合更能进一步提升网络的性能。此外,LESRCNN 在盲超分辨率邻域也取得了巨大进步。之后出现的 Cross-MPI<sup>[39]</sup>是一个端到端的 RefSR(Reference Image Super Resolution)网络,其中的核心模块平面感知注意力机制充分利用隐藏的场景结构进行高效的基于注意力的对应搜索,进一步结合温和的由粗到细引导上采样策略,实现稳健而准确的细节传输。最近出现的 SERAN<sup>[40]</sup>对医学图像进行超分重建,取得了重大突破,该网络先使用二阶注意力池化操作将全局空间信息压缩为空间描述,这使得模型更加专注于磁共振图像中更具信息性的结构和区域,然后建立基元之间的关系并引入图卷积来聚合信息,最终实现准确的细节和纹理重建。而 ENLCA<sup>[41]</sup>更是提出了一种高效的非局部对比注意力进行单一图像的超分辨率重建,以往的 NLA(Non-Local Attention, NLA)利用自然图像的内在相关性为图像的超分重建带来了显著的改进,但该注意力赋予噪声信息很大的权重,并且相对于输入大小消耗二次计算资源,限制了性能。ENLCA 中的 ENLA(Efficient Non-Local Attention)通过核函数逼近和矩阵乘法的结合律,将计算复杂度从二次显著降低到线性,此外,ENLA 还应用了对比学习,进一步加强了特征的相关性。

尽管以上方法在图像超分辨率领域取得了巨大成就,但这些方法都只是通过简单堆叠残差块和残差连接来提取特征,并没有充分利用残差块间的层级特征来恢复丢失的高频信息,因此本文提出残差聚合网络,获取更多包含细节信息的残差特征。

## 1.2 基于注意力机制的超分辨率重建

注意力机制可以倾向性选择聚焦区域,产生更具分辨性的特征表示,因此该机制广泛应用于计算机视觉图像处理领域中。非局部神经网络 Non-Local<sup>[26]</sup>被用于视频分类。具体来讲,该网络是从传统的 Non-local means 中吸取经验,接着在神经网络中应用了这个思想,直接融合了全局的信息,而不仅仅是通过堆叠多个卷积层获得较为全局的信息,这样可为后续卷积层带来更丰富的语义。SENet<sup>[7]</sup>网络利用核心 SE 模块对特征图进行处理变换。考虑到在大量的通道中,有些通道的特征图有用,有些通

道的特征图无效,SE 的想法就是通过学习的方式获得每个特征通道的重要程度,根据这个重要性来增强有用的特征,抑制无效的特征。接着出现了一些基于注意力机制的超分辨率模型,残差通道注意力网络 RCAN<sup>[31]</sup>借鉴了残差学习的思想,在残差中嵌入了残差,通过局部和全局残差的耦合,使信息流的传递更好,其目的也是为了加深网络的深度,在训练过程中更好收敛,同时该网络中加入了通道注意力机制,在注意力的引导下进行超分辨率重建。SAN<sup>[3]</sup>网络提出了一个二阶注意力机制进行相关性学习,以便获得更好的特征表达,同时该网络使用 non-locally 来捕获长距离的空间内容信息,使得重建后的图像质量更高。最新出现的 ENLCA<sup>[41]</sup>更是对传统的 NLA 提出了挑战,该网络提出了一种新颖高效的非局部对比注意力来执行远程视觉建模并利用了更多的非局部相关特征。具体来讲,ENLCA 中的 ENLA 模块采用核方法来降低计算复杂度,而 SA(Sparse Aggregation)模块用来专注于信息特征,最终网络的目标就是尽可能聚合所有目标的相关特征,保持非局部模块的稀疏性并大大降低其计算成本。

以上这些方法使用注意力机制或者进行图像分类,或者进行超分重建,取得了较好的视觉效果,因此,本文在深层聚合网络中也引入两个注意力模块,分别从空间维度和通道维度对网络进行引导,使残差聚合特征聚焦于信息更丰富的区域。

## 2 网络模型

如图 1 所示,DARAN 主要分为四部分:浅层特征聚合(Shallow Feature Aggregation, SFA)、局部残差特征聚合(Local Residual Feature Aggregation, LRFA)、全局残差特征聚合(Global Residual Feature Aggregation, GRFA)和重建部分(Reconstruction)。将  $I_{LR}$  和  $I_{SR}$  分别记作 DARAN 的输入和输出。本节将一一介绍各部分在网络中的作用以及具体细节。

### 2.1 浅层特征聚合模块

为了减轻训练难度,使输入到局部残差特征聚合网络(LRFA)的特征具有更强的可鉴别性,本文对已有的浅层特征提取模块进行了优化,提出了浅层特征聚合模块(Shallow Feature Aggregation, SFA)。如图 1 所示,与以往的特征提取不同,该模块在特征聚合阶段先将提取到的浅层特征  $X_0$  分别通过三个由  $3 \times 3$ 、 $5 \times 5$  和  $7 \times 7$  搭建的残差块,得到的特征又与原始的浅层特征  $X_0$  进行融合,这样就可

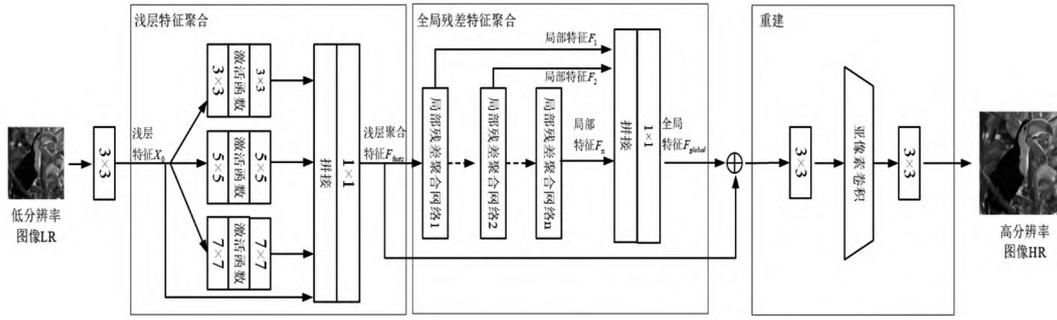


图 1 双重调节残差聚合网络

以聚合更多输入图像的特征,生成的浅层聚合特征  $F_{\text{feats}}$  表示为

$$X_0 = M_{\text{SFA1}}(I_{\text{LR}}) \quad (1)$$

$$F_{\text{feats}} = M_{\text{SFA2}}(X_0) \quad (2)$$

式中,  $M_{\text{SFA1}}(\cdot)$  表示  $3 \times 3$  卷积操作;  $M_{\text{SFA2}}(\cdot)$  表示浅层特征聚合。为了进一步验证该模块的性能,本文从已有的模型中抽取多个和 SFA 思想类似的模块并进行对比实验,如表 4 所示,数据表明,该优化确实有助于减轻网络训练难度,提升网络性能。

### 2.2 局部残差特征聚合

为了恢复丢失的高频信息,为最终的重建打下坚实的基础,本文提出了局部残差聚合模块(Local Residual Feature Aggregation, LRFA),如图 2 所示,吸取已有网络的经验后,该模块对残差块进行了重新分配。残差结构由以往的串联式改为现在的并联式,其目的就是为了利用不同大小的卷积核提取多个尺度的特征。网络采用渐进式融合策略,具体来讲,LRFA 分为上下两个分支,共有残差块 1、残差块 2、残差块 3 和残差块 4 四个残差块,其中残差块 1 和残差块 2 由大小为  $3 \times 3$  的卷积搭建而成,残差块 3 和残差块 4 由大小为  $5 \times 5$  的卷积搭建而成。首次融合体现在将第  $n$  个局部残差聚合网络的输入  $F_{n-1}$  分别输入到前两个残差块(Residual Block, RB)残差块 1 和残差块 3 中,得到的特征和  $F_{n-1}$  进

行拼接,为了加强特征表示,改变通道维度后将生成的特征  $F_{\text{concat}}^1$  输入到空间调节模块(SAB)中,从空间维度对特征进行关注;二次融合体现在将输出特征  $F_{n-1}^{\text{SAB}}$  分别输入到残差块 2 和残差块 4 中,同上,将得到的特征与  $F_{n-1}^{\text{SAB}}$  进行拼接,改变通道维度后生成特征  $F_{\text{concat}}^2$ ;末端融合体现在将  $F_{\text{concat}}^2$  与前四个残差块模块得到的特征再次进行拼接,生成特征  $F_{\text{concat}}^3$ ,使用  $1 \times 1$  卷积改变通道维度后将其输入到通道调节模块(CAB)中,从通道维度对特征进行关注,生成特征  $F_{n-1}^{\text{CAB}}$ 。最后,局部残差聚合网络生成的特征  $F_n$  的表示为

$$F_{\text{concat}}^1 = H_{1 \times 1}([F_{\text{RB}}^1, F_{\text{RB}}^3, F_{n-1}]) \quad (3)$$

$$F_{n-1}^{\text{SAB}} = H_{\text{SAB}}(F_{\text{concat}}^1) \quad (4)$$

$$F_{\text{concat}}^2 = H_{1 \times 1}([F_{n-1}^{\text{SAB}}, F_{\text{RB}}^2, F_{\text{RB}}^4]) \quad (5)$$

$$F_{\text{concat}}^3 = H_{1 \times 1}([F_{\text{concat}}^2, F_{\text{RB}}^1, F_{\text{RB}}^2, F_{\text{RB}}^3, F_{\text{RB}}^4]) \quad (6)$$

$$F_{n-1}^{\text{CAB}} = H_{\text{CAB}}(F_{\text{concat}}^3) \quad (7)$$

$$F_n = F_{n-1} + F_{n-1}^{\text{CAB}} \quad (8)$$

式中,  $H_{1 \times 1}(\cdot)$ 、 $H_{\text{SAB}}(\cdot)$  和  $H_{\text{CAB}}(\cdot)$  分别代表  $1 \times 1$  卷积、空间调节模块和通道调节模块;  $[\ ]$  表示拼接操作;  $F_{\text{RB}}^1$ 、 $F_{\text{RB}}^2$ 、 $F_{\text{RB}}^3$  和  $F_{\text{RB}}^4$  分别代表四个残差块生成的特征;  $F_{\text{concat}}^1$ 、 $F_{\text{concat}}^2$  和  $F_{\text{concat}}^3$  分别代表三次拼接生成的特征;  $F_{n-1}^{\text{CAB}}$  和  $F_{n-1}^{\text{SAB}}$  分别代表输入到通道调节模块和空间调节模块生成的特征;  $F_n$  表示第  $n$  个局部残差特征聚合网络生成的特征。

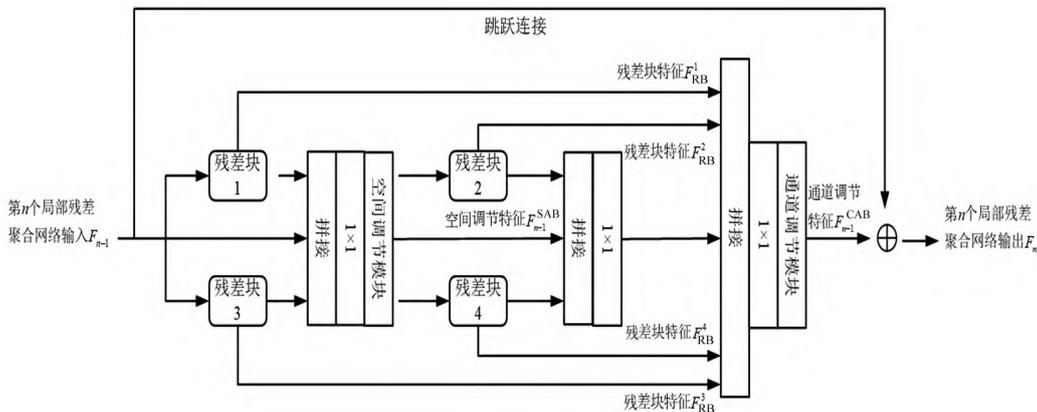


图 2 局部残差特征聚合网络

首次融合与二次融合从局部的角度出发,将残差块与相应的输入特征两两融合,而末端融合从全局的角度出发,充分利用所有残差块的特征,其目的就是利用包含更多输入图像细节信息的特征,恢复更多高频信息,进行最终的重建。总体来讲,本文的残差聚合网络能够更大程度利用到各个残差块生成的特征,这些特征可直接传播到 LRFA 末端,没有任何损失和干扰,从而使生成的特征包含更多的细节信息。

### 2.2.1 空间调节模块

为了加强聚合后的特征在空间中彼此的相互依赖性,本文引入了空间调节模块(Spatial Adjustment Block, SAB),该模块与 ESA<sup>[17]</sup>类似,区别在于 SAB 中加入了残差连接,进一步加强特征。如图 3 所示,作为一种简单而有效的注意力模块,在卷积神经网络中给定一个中间特征图,SAB 便可将注意力映射沿着特征图的空间维度进行输入,归一化后对输入特征进行加权,以达到让输入特征自适应细化的目的。具体来讲,先利用  $1 \times 1$  卷积降低通道维度,接着为了增大感受野,该模块使用了跨步卷积(conv\_stride)和最大池化层(max\_pool)。跨步卷积的卷积核大小为  $3 \times 3$ ,步长为 2,该卷积与普通卷积的区别就在于步长的变化,而最大池化层的卷积核大小为  $7 \times 7$ ,步长为 3。二者经常搭配使用来快速降低特征的空间维度,这种做法在图像分类中很常见。SAB 末端采用双线性插值法(bilinear)和  $1 \times 1$  卷积分别恢复特征的空间维度和通道维度,归一化后利用生成的权重 value( $value \in R^{N \times C \times H \times W}$ )对输入特征进行加权,最后输入特征与加权后的特征进行融合,生成特征  $F_{SAB}$ 。

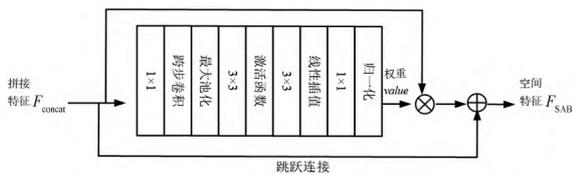


图 3 空间调节模块

### 2.2.2 通道调节模块

受到 CBAM<sup>[28]</sup>的启发,本文先从空间聚合残差特征进行关注,接着使用通道调节模块(Channel Adjustment Block, CAB)进一步加强特征在通道上的相互依赖性,该模块的细节如图 4 所示。假设输入特征  $F_{concat}$  大小为  $N \times C \times H \times W$ ,该模块利用自适应平均池化(avg\_pool)将特征维度降为  $N \times C \times 1 \times 1$ ,使用  $1 \times 1$  卷积先减少特征的通道维度,然

后将通道数复原,归一化后得到权重 value( $value \in R^{N \times C \times 1 \times 1}$ ),权重中的各个元素与特征的通道数一一对应,最后 CAB 采用逐元素相乘的方式对  $F_{concat}$  进行加权,生成特征  $F_{CAB}$ 。

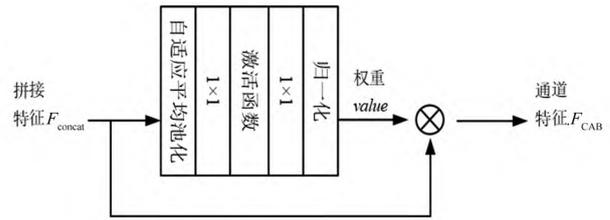


图 4 通道调节模块

### 2.3 全局残差特征聚合

为了充分利用各个 LRFA 生成的特征,与 RDN<sup>[32]</sup>的全局学习策略类似,本文提出了全局残差特征聚合(Global Residual Feature Aggregation, GRFA),该残差聚合的细节如图 1 所示。其目的就是将在  $n$  个 LRFA 生成的特征全部利用起来,生成较强的特征表示。最终,输入到重建部分的特征  $F_{global}$  表示为

$$F_n = M_{LRFA}^n(F_{n-1}) \tag{9}$$

$$M_{LRFA}^n(F_{n-1}) = M_{LRFA}^n(M_{LRFA}^{n-1}(\dots(M_{LRFA}^1(F_{feats}))\dots)) \tag{10}$$

$$F_{global} = H_{1 \times 1}([F_1, F_2, F_3, \dots, F_n]) \tag{11}$$

式中,  $M_{LRFA}^n(\cdot)$  表示第  $n$  个残差聚合网络函数;  $F_{n-1}$  代表第  $n$  个局部残差特征聚合网络的输入,  $F_n$  就是相应的输出。  $H_{1 \times 1}(\cdot)$  表示  $1 \times 1$  卷积操作,主要用于改变特征的通道维度,整合特征;  $[ ]$  表示拼接操作;  $F_1$  到  $F_n$  依次代表第 1 个 LRFA 到第  $n$  个 LRFA 生成的特征。

### 2.4 重建

重建部分(Reconstruction)结构比较简单,由于本文设计的网络较深,所以采用亚像素卷积<sup>[21]</sup>进行上采样,对图像的尺寸进行变换,大大提高了计算效率。同时,网络模型中加入跳跃连接来减轻训练难度。最后,重建图像  $I_{SR}$  表示为

$$I_{SR} = R_{recon}(F_{global} + F_{feats}) = \text{DARAN}(I_{LR}) \tag{12}$$

式中,  $R_{recon}(\cdot)$  表示重建操作;  $\text{DARAN}(\cdot)$  表示本文的模型;  $F_{feats}$  表示浅层聚合特征。

## 3 目标函数

在训练过程中,DARAN 使用重建损失  $L_{pixel}$ 、感知损失  $L_{per}$  和梯度损失  $L_{grad}$  分别从像素层面、特征层面和边缘信息层面对模型进行优化。取  $M$  对低分辨率图像  $LR$  和对应的高分辨率图像  $HR$ , 损

失函数表示为

$$L_{\text{pixel}} = \frac{1}{M} \sum_{i=1}^M \| \text{DARAN}(I_{\text{LR}}) - I_{\text{HR}} \|_1 \quad (13)$$

$$L_{\text{per}} = \frac{1}{M} \sum_{i=1}^M \| \Phi(\text{DARAN}(I_{\text{LR}})) - \Phi(I_{\text{HR}}) \|_2 \quad (14)$$

$$L_{\text{grad}} = \frac{1}{M} \sum_{i=1}^M \| G(\text{DARAN}(I_{\text{LR}})) - G(I_{\text{HR}}) \|_1 \quad (15)$$

$$L_{\text{total}} = L_{\text{pixel}} + \beta L_{\text{per}} + \gamma L_{\text{grad}} \quad (16)$$

式中,  $I_{\text{HR}}$  表示低分辨率图像对应的标签;  $\text{DARAN}(\cdot)$  表示本文的模型;  $\Phi(\cdot)$  表示 VGG16 模型函数, 主要用来提取  $I_{\text{HR}}$  和  $I_{\text{SR}}$  的特征;  $G(\cdot)$  表示梯度函数, 主要用来获取  $I_{\text{HR}}$  和  $I_{\text{SR}}$  的梯度图,  $\beta$  和  $\gamma$  分别表示感知损失和梯度损失的权重, 在实际训练过程中,  $\beta$  和  $\gamma$  分别设为 0.1 和 1。

## 4 实验分析

### 4.1 实验细节

本文使用浅层特征聚合(SFA)、局部残差特征聚合网络(LRFA)、通道调节模块(CAB), 和空间调节模块(SAB)搭建双重调节残差聚合网络(DARAN)。DARAN 使用 20 个局部残差特征聚合网络, 每个 LRFA 有 4 个 RB 模块。网络中使用的卷积有  $1 \times 1$ 、 $3 \times 3$ 、 $5 \times 5$  和  $7 \times 7$ 。 $1 \times 1$  卷积主要用于改变特征的通道维度,  $5 \times 5$  和  $7 \times 7$  卷积主要是为

了增大感受野, 从低分辨率图像中提取到更多细节信息。此外, 文中除 SAB 和 CAB 外, 其余卷积的滤波器数量均为 64。

### 4.2 实验设置

数据集选取: 与参考文献[17]和[35]一样, 本文以 DIV2K<sup>[23]</sup> 数据集作为训练集, 以 Set5<sup>[1]</sup>、Set14<sup>[34]</sup>、Urban100<sup>[18]</sup> 和 BSD100<sup>[6]</sup> 作为测试集。训练设置: 在每个训练批次中, 随机选取 16 张大小为  $64 \times 64$  的低分辨率 RGB 图片作为输入, 同时将图片随机水平翻转或旋转进行数据增强。训练过程中使用 Adam 优化器更新参数, 初始学习率为  $1 \times 10^{-4}$ , 总共训练 300 个 epoch, 每个 epoch 迭代 1000 次, 训练到 150 个 epoch 后学习率减少为原来的一半, 训练每个 epoch 时, 参数量约为 7.93M。训练结果使用 SSIM 和 PSNR 两个指标进行评估。网络训练所用的平台为 ubuntu, 编程框架为 Pytorch1.9, 处理器为 Intel(R) Core(TM) i9-11900K @ 3.50GHz  $\times$  16, 系统内存为 64G, 显卡为 GTX3090, 24G 显存容量, Cuda11.4 版本。

### 4.3 实验结果分析

#### 4.3.1 与其他方法的对比分析

将本文提出的 DARAN 网络与双三次插值法 Bicubic、SRCNN<sup>[2]</sup>、FSRCNN<sup>[4]</sup>、VDSR<sup>[10]</sup> 和 LESRCNN<sup>[42]</sup> 五种方法进行对比, 各个网络在不同的缩放因子下所取得的 PSNR 和 SSIM 如表 1 所示。

表 1 缩放因子在 2、3、4 时在标准数据集下的指标对比

数据集	缩放因子	PSNR/SSIM					
		Bicubic	SRCNN	FSRCNN	VDSR	LESRCNN	DARAN(本文)
Set5	$\times 2$	33.66/0.929	36.66/0.954	37.12/0.957	37.56/0.959	37.57/0.958	<b>37.73/0.960</b>
	$\times 3$	30.39/0.868	32.75/0.909	33.18/0.914	33.67/0.921	33.98/0.920	<b>34.00/0.922</b>
	$\times 4$	28.42/0.810	30.48/0.862	30.72/0.866	31.35/0.878	<b>31.88/0.880</b>	31.71/0.880
Set14	$\times 2$	30.25/0.868	32.45/0.906	32.59/0.909	33.01/0.912	<b>33.30/0.914</b>	33.15/0.916
	$\times 3$	27.52/0.774	29.30/0.821	29.37/0.824	29.78/0.832	<b>30.16/0.833</b>	29.59/0.835
	$\times 4$	26.00/0.702	27.50/0.751	27.61/0.755	27.97/0.768	<b>28.43/0.760</b>	28.00/0.768
BSD100	$\times 2$	29.56/0.843	31.36/0.887	31.53/0.892	31.90/0.896	31.95/0.897	<b>32.02/0.899</b>
	$\times 3$	27.21/0.738	28.41/0.786	28.53/0.791	28.83/0.796	<b>28.94/0.799</b>	28.73/0.799
	$\times 4$	25.96/0.667	26.90/0.710	26.98/0.715	27.29/0.726	<b>27.47/0.732</b>	27.08/0.715
Urban100	$\times 2$	26.88/0.840	29.50/0.894	29.88/0.902	30.77/0.914	31.39/0.920	<b>31.42/0.932</b>
	$\times 3$	24.46/0.734	26.24/0.798	26.43/0.808	27.14/0.829	27.70/0.840	<b>27.71/0.845</b>
	$\times 4$	23.14/0.657	24.52/0.722	24.62/0.728	25.18/0.754	25.74/0.759	<b>25.76/0.769</b>
均值		27.79/0.786	29.67/0.833	29.88/0.838	30.37/0.849	<b>30.71/0.851</b>	30.57/0.853

注: 黑色加粗为最优。

通过表 1 中数据可知,DARAN 网络的性能明显高于其他网络。当缩放因子为 2 时,与其他五种方法相比,该网络在 Set5 上的 PSNR 分别超出了 4.07dB、1.07dB、0.61dB、0.17dB 和 0.16dB,SSIM 分别超出了 0.031、0.006、0.003、0.001 和 0.002;当缩放因子为 3 时,该网络在 Set5 上的 PSNR 分别超出了 3.61dB、1.25dB、0.82dB、0.33dB 和 0.02dB,SSIM 分别超出了 0.054、0.013、0.008、0.001和 0.002;缩放因子为 4 时,该网络在 Urban100 上的 PSNR 分别超出了 2.62dB、1.24dB、1.14dB、0.58dB 和 0.02dB,SSIM 分别超出了 0.112、0.047、0.041、0.015 和 0.01。

为了进一步验证本文模型相对于 Bicubic、SRCNN、FSRCNN 和 VDSR 这几种对超分重建贡献巨大的算法性能的优越性,将 Set5 和 Set14 中的 19 张图片进行缩放因子为 2 的超分辨率重建,分别计算每张图片在不同算法下的 PSNR 和 SSIM,计算结果如表 2 和表 3 所示。由表中数据可知,与其

他四种算法相比,DARAN 在 Set5 的 Bird 上的 PSNR 分别超出了 6.09dB、1.99dB、1.17dB 和 0.41dB,SSIM 分别超出了 0.019、0.004、0.002 和 0.001;在 Butterfly 上的 PSNR 分别超出了 7.67dB、2.33dB、1.54dB 和 0.58dB,SSIM 分别超出了 0.062、0.011、0.004 和 0.001;在 Woman 上的 PSNR 分别超出了 4.15dB、0.93dB、0.53dB 和 0.23dB,SSIM 分别超出了 0.03、0.006、0.004 和 0.002。DARAN 在 Set14 上的 Monarch 上的 PSNR 分别超出了 6.91dB、2.11dB、1.28dB 和 0.42dB,SSIM 分别超出了 0.02、0.004、0.001 和 0.001;在 Ppt3 上的 PSNR 分别超出了 7.45dB、2.8dB、3.01dB 和 1.97dB,SSIM 分别超出了 0.046、0.011、0.009 和 0.006;在 Zebra 上的 PSNR 分别超出了 3.83dB、0.97dB、0.37dB 和 0.27dB,SSIM 分别超出了 0.04、0.005、0.003 和 0.002。整体分析表 2 和表 3 数据,DARAN 的实验结果几乎都处于最优和次优,进一步验证了该网络性能的优越性。

表 2 缩放因子为 2 时在 Set5 上的实验结果对比

图像	PSNR/SSIM				
	Bicubic	SRCNN	FSRCNN	VDSR	DARAN (本文)
Baby	37.09/0.953	38.54/0.966	38.69/0.967	<b>38.78/0.967</b>	38.77/0.966
Bird	36.81/0.971	40.91/0.986	41.73/0.988	42.49/0.989	<b>42.90/0.990</b>
Butterfly	27.41/0.913	32.75/0.964	33.54/0.971	34.50/0.974	<b>35.08/0.975</b>
Head	34.86/0.861	35.72/0.886	35.88/0.888	<b>35.96/0.890</b>	35.62/0.883
Woman	32.14/0.945	35.36/0.969	35.76/0.971	36.06/0.973	<b>36.29/0.975</b>
均值	33.66/0.929	36.66/0.954	37.12/0.957	37.56/0.959	<b>37.73/0.960</b>

注:黑色加粗为最优。

表 3 缩放因子为 2 在 Set14 上的实验结果对比

图像	PSNR/SSIM				
	Bicubic	SRCNN	FSRCNN	VDSR	DARAN(本文)
Baboon	24.86/0.695	25.74/0.767	25.66/0.772	25.94/0.778	<b>26.10/0.800</b>
Barbara	28.00/0.843	28.64/0.877	28.27/0.877	28.42/0.879	<b>28.66/0.890</b>
Bridge	26.58/0.791	27.83/0.850	29.15/0.861	28.04/0.858	<b>29.32/0.870</b>
Coastguard	29.12/0.793	30.83/0.851	30.84/0.851	<b>30.98/0.851</b>	30.88/0.851
Comic	26.02/0.853	28.52/0.917	29.05/0.925	29.40/0.933	<b>29.66/0.942</b>
Face	34.83/0.862	35.70/0.884	35.86/0.888	<b>35.93/0.890</b>	35.62/0.884
Flowers	30.37/0.900	33.32/0.937	33.83/0.942	34.34/0.946	<b>34.59/0.949</b>
Foreman	34.14/0.948	36.47/0.966	35.11/0.970	<b>37.40/0.972</b>	36.20/0.972
Lenna	34.70/0.912	36.64/0.930	36.87/0.931	<b>37.06/0.933</b>	36.45/0.918
Man	29.52/0.846	31.04/0.889	31.17/0.891	31.40/0.896	<b>31.40/0.897</b>
Monarch	32.94/0.961	37.74/0.977	38.57/0.980	39.43/0.980	<b>39.85/0.981</b>
Pepper	34.95/0.899	36.87/0.919	36.50/0.917	<b>37.38/0.925</b>	36.61/0.927
Ppt3	26.87/0.945	31.52/0.980	31.31/0.982	32.35/0.985	<b>34.32/0.991</b>
Zebra	30.63/0.906	33.49/0.942	34.09/0.944	34.19/0.945	<b>34.46/0.947</b>
均值	30.25/0.868	32.45/0.906	32.59/0.909	33.01/0.912	<b>33.15/0.916</b>

注:黑色加粗为最优。

为进一步验证 DARAN 充分利用残差块间的层级特征重建高清图像性能的提升,从 Set5、Set14、BSD100 和 Urban100 中抽取一些图片进行对比,各个算法重建后的结果如图 5~8 所示。

如图 9 和表 4 所示,本文针对 SFA 的优化,进行了对比实验说明,在实验过程中,各个模块统一运行 5 个 epoch,每个 epoch 迭代 1000 次,最终的结果以 5 次实验数据的平均值为主,确保实验的公平性。数据表明,此次优化更有助于提高模型 DARAN 的

性能。

表 4 不同特征提取模块对 DARAN 的影响

设置	PSNR/SSIM
RFAWA+a	35.69/0.947
RFAWA+b	35.72/0.947
RFAWA+c(SFA)	<b>35.82/0.948</b>
RFAWA+d	35.71/0.947

注:黑色加粗表示最优

#### 4.3.2 消融实验

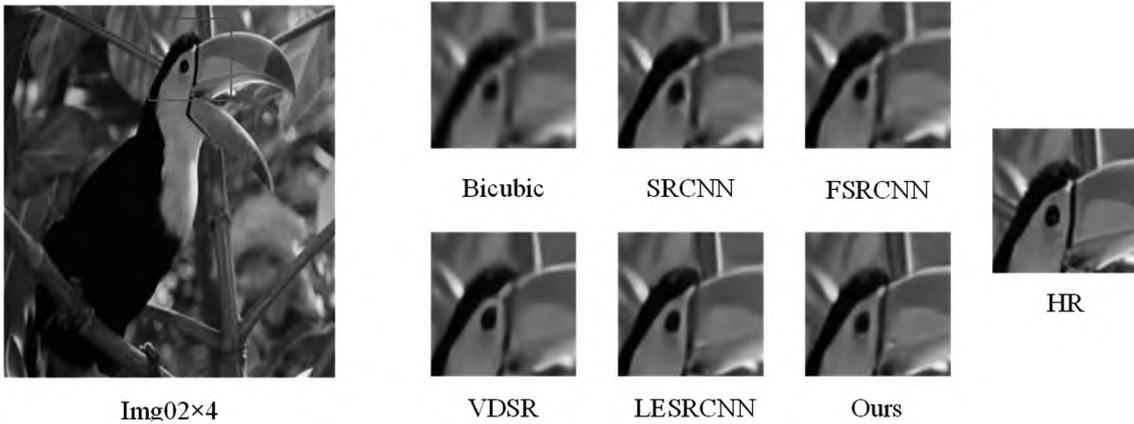


图 5 缩放因子为 4 时各个模型在 Set5 上的重建结果

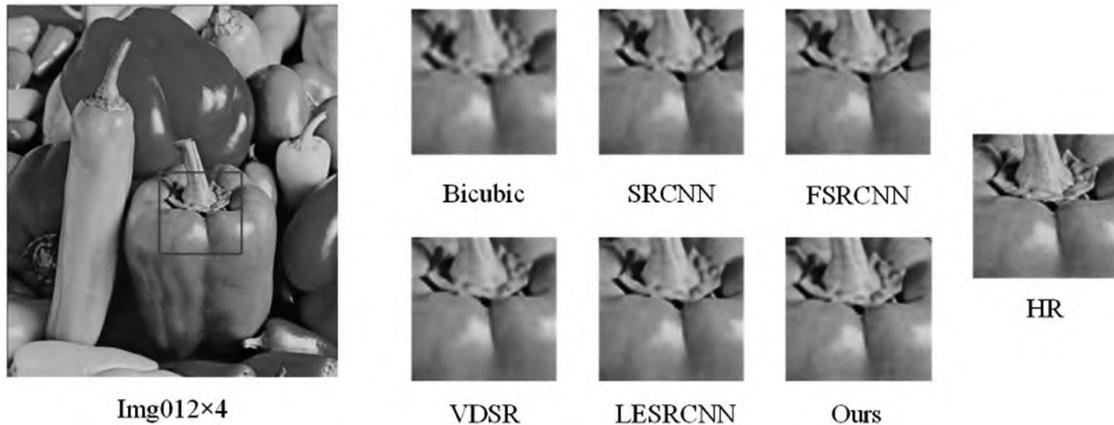


图 6 缩放因子为 4 时各个模型在 Set14 上的重建结果

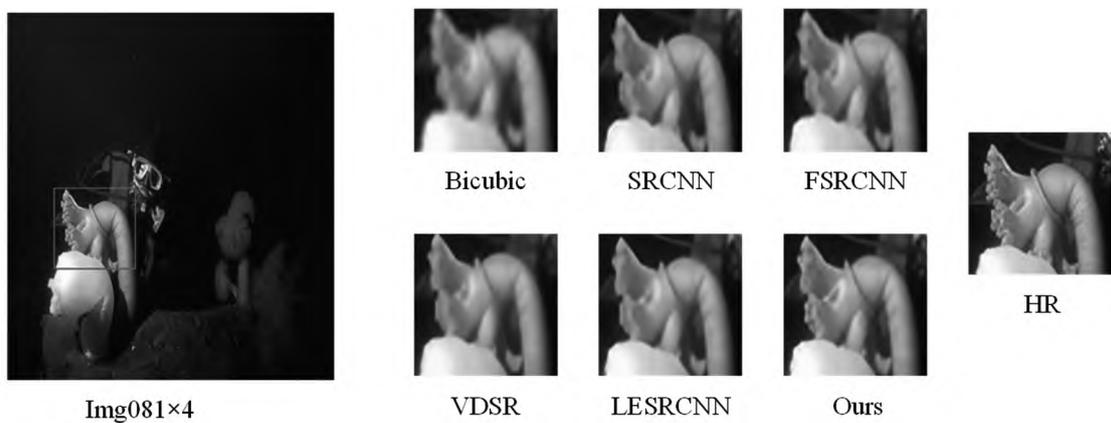


图 7 缩放因子为 4 时各个模型在 BSD100 上的重建结果

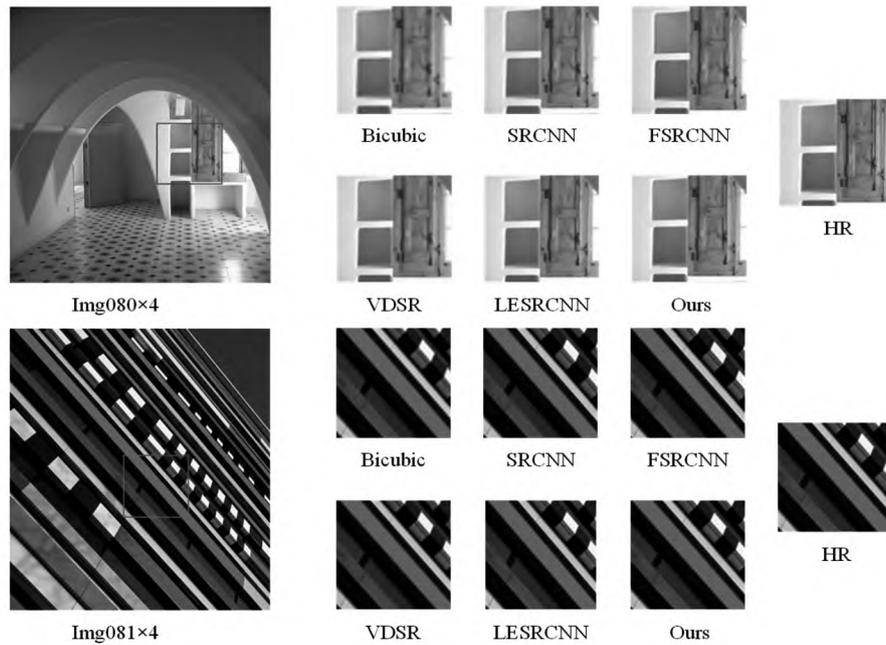


图 8 缩放因子为 4 时各个模型在 Urban100 上的重建

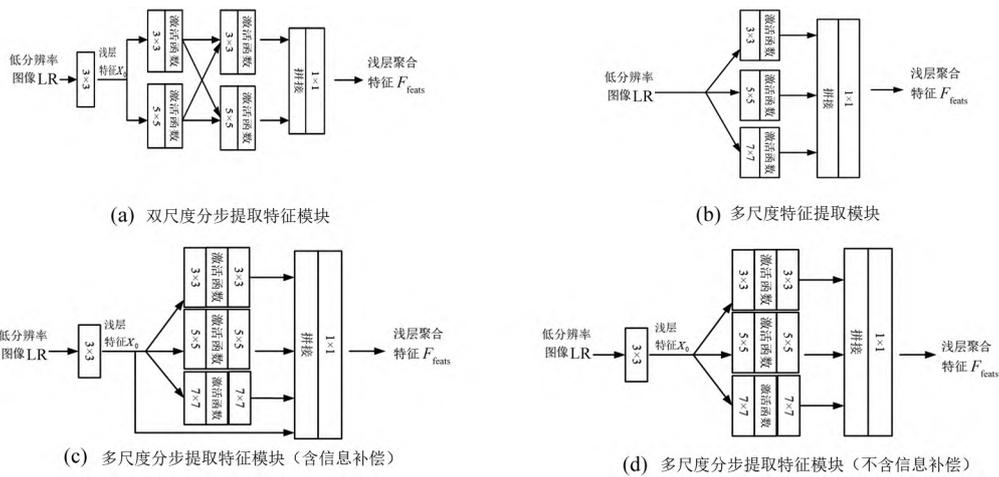


图 9 不同模型的多尺度特征提取模块

为了验证各个模块在网络中的有效性,本文依次对 RFAWA (Residual Feature Aggregation Without Attention), SFA 和 SACA (Spatial And Channel Adjustment Block) 模块进行了验证。在验证过程中,各个模块都在缩放因子为 2 的 DIV2K 数据集上进行训练,在 Set5 数据集进行测试,确保实验的公正性与合理性。每次验证运行 5 个 epoch,每个 epoch 迭代 1000 次,最终的结果以 5 次实验数据的平均值为主。

实验结果如表 5 所示。由表中数据可知,加入 SFA 后比仅用 RFAWA 的 PSNR 高 0.22dB, SSIM 高 0.001;加入 SACA 要比仅用 RFAWA 的 PSNR 高 0.1dB, SSIM 高 0.001;加入 SFA 要比加入 SACA 的 PSNR 高 0.12dB;同时加入 SFA 和 SACA 要比仅用 RFAWA 的 PSNR 高 0.3dB, SSIM 高 0.001。

表 5 不同模块对 DARAN 的影响

设置	PSNR/SSIM
RFAWA	35.60/0.947
RFAWA+SFA	35.82/0.948
RFAWA+SACA	35.70/0.948
RFAWA+SFA+SACA	<b>35.90/0.948</b>

注:黑色加粗表示最优

## 5 结论

本文提出了一个结合注意力和残差聚合的双重调节残差合网络(DARAN)。该网络打破了以往的串联式残差结构,对残差块重新分配,形成了并联式局部残差特征聚合网络(LRFA)。每个 LRFA 利用不同大小的卷积核提取多尺度特征并且从局部和全局充分利用了各个残差块以及聚合网络间的层级特

征,导致输入到重建部分的特征具有更强的可鉴别性。为了进一步增强特征表示,文中利用浅层特征聚合模块(SFA)、通道调节模块(CAB)和空间调节模块(SAB),分别从输入、通道和空间三大方面来对特征进行引导处理。

实验结果表明,与其他五种方法相比,DARAN在残差块层级特征的利用程度、图像细节重建能力、重建后的视觉效果等方面都有所提升。尽管某些图片重建后的指标不是最优,但也是次优水平,从某种程度上证明了DARAN的有效性。不足之处在于本文使用了较为传统的注意力机制来提升网络性能,缺乏一定的创新性。下一步将认真思考超分辨率重建体系现有的缺陷并构建新的模型,同时加入一些全新的模块来解决问题。预期该想法的实施将会获得更好的视觉效果。

#### 参考文献:

- [1] Bevilacqua M, Roumy A, Guillemot C, et al. Low-complexity single-image super-resolution based on nonnegative neighbor embedding[C]// British Machine Vision Conference (BMVC). Newcastle, British: BMVC, 2012.
- [2] Dong C, Loy C, He K, et al. Learning a deep convolutional network for image super-resolution[C]// European Conference on Computer Vision. Zurich, Switzerland: IEEE, 2014: 184—199.
- [3] Dai T, Cai J, Zhang Y, et al. Second-order attention network for single image super-resolution [C] // Proceedings of The IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019: 11065—11074.
- [4] Dong C, Loy C, Tang X. Accelerating the super-resolution convolutional neural network[C]// European Conference on Computer Vision. Springer: Cham, 2016: 391—407.
- [5] Hu X, Mu H, Zhang X, et al. Meta-SR: A magnification-arbitrary network for super-resolution[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019: 1575—1584.
- [6] Huang J, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars[C]// Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015: 5197—5206.
- [7] Hu J, Shen L, Sun G. Squeeze and excitation networks[C]// Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 7132—7141.
- [8] Huang G, Liu Z, Van D, et al. Densely connected convolutional networks[C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017: 2261—2269.
- [9] He K, Zhang X, Ren S et al. Deep residual learning for image recognition[C]// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 770—778.
- [10] Kim J, Lee J, Lee K. Accurate image super-resolution using very deep convolutional networks [C] // Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 1646—1654.
- [11] Kim J, Lee J, Lee K. Deeply recursive convolutional network for image super-resolution[C]// Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 1637—1645.
- [12] 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. Hawaii, USA: IEEE, 2017: 136—144.
- [13] Lai W, Huang J, 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. Honolulu, USA: IEEE, 2017: 624—632.
- [14] Li X, Orchard M. New edge-directed interpolation[J]. IEEE Transactions on Image Processing, 2001, 10(10): 1521—1527.
- [15] 李红艳, 李春庚, 安居白, 等. 注意力机制改进卷积神经网络的遥感图像目标检测[J]. 中国图象图形学报, 2019, 24(8): 1400—1408.
- Li Hongyan, Li Chungeng, An Jubai, et al. Attention mechanism improves CNN remote sensing image object detection[J]. Journal of Image and Graphics, 2019, 24(8): 1400—1408.
- [16] Li Z, Yang J, Liu Z, et al. Feedback network for image super resolution[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019: 3867—3876.
- [17] 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. Seattle, USA: IEEE, 2020: 2359—2368.
- [18] 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. Vancouver, Canada: IEEE, 2001, 2: 416—423.
- [19] Radu T, Vincent D, Luc V. Anchored neighborhood regression for fast example-based super-resolution[C]// Proceedings of The IEEE International Conference on Computer Vision. Sydney, Australia: IEEE, 2013: 1920—1927.
- [20] Radu T, Vincent D, Luc V. A+: Adjusted anchored neighborhood regression for fast super-resolution[C]// Asian Conference on Computer Vision. Springer: Cham, 2014: 111—126.
- [21] Shi W, Caballero J, Huszar 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. Las Vegas, USA: IEEE, 2016: 1874—1883.
- [22] Tomer P, Michael E. A statistical prediction model based on sparse representations for single image super-resolution [J]. IEEE Trans. Image Processing, 2014, 23(6): 2569—2582.
- [23] Timofte R, Agustsson E, Van G, et al. Ntire2017 challenge on single image super resolution: Methods and results[C]// Proceed-

- ings of The IEEE Conference on Computer Vision and Pattern Recognition Workshops. Hawaii, USA; IEEE, 2017: 1110—1112.
- [24] Tai Y, Yang J, Liu X. Image super resolution via deep recursive residual network[C] // Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA; IEEE, 2017: 3147—3155.
- [25] Tai Y, Yang J, Liu X, et al. MemNet: A persistent memory network for image restoration[C] // Proceedings of The IEEE International Conference on Computer Vision. Venice, Italy; IEEE, 2017: 4539—4547.
- [26] Wang X, Girshick R, Gupta A, et al. Non-local neural networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE, 2018: 7794—7803.
- [27] Wang Z, Bovik A, Sheikh H, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600—612.
- [28] Woo S, Park J, Lee J, et al. CBAM: Convolutional block attention module[C] // Proceedings of The European Conference on Computer Vision Munich, Germany; ECCV, 2018: 3—19.
- [29] 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.
- [30] Zhang K, Gao X, Tao D, et al. Single image super-resolution with non-local means and steering kernel regression[J]. IEEE Trans. Image Processing, 2012, 21(11): 4544—4556.
- [31] Zhang Y, Li K, Wang L, et al. Image super resolution using very deep residual channel attention networks[C] // Proceedings of The European Conference on Computer Vision. Munich, Germany; ECCV, 2018: 286—301.
- [32] 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. Salt Lake City, USA; IEEE, 2018: 2472—2481.
- [33] Zhang Y, Li K, Zhong B, et al. Residual non-local attention networks for image restoration[C] // International Conference on Learning Representations. Louisiana, USA; ICLR, 2019.
- [34] Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations[C] // International Conference on Curves and Surfaces. Springer, Berlin, Heidelberg; ICCS, 2010: 711—730.
- [35] Zhao B, Hu R, Jia X, et al. Multi-scale residual fusion network for super-resolution reconstruction of single image[J]. IEEE Access, 2020, 8: 155285—155295.
- [36] 李晓萌, 陈兆学. 基于 GAN 的人脸超分辨率重建算法研究[J]. 光学技术期刊, 2021, 47(01): 101—106.  
Li Xiaomeng, Chen Zhaoxue. Research on face super-resolution reconstruction algorithm based on GAN[J]. Journal of Optical Technology, 2021, 47(01): 101—106.
- [37] 王溢琴, 董云云, 刘慧玲. 基于 GoogLeNet 和空间谱变换的高光谱图像超分辨率方法[J]. 光学技术期刊, 2022, 48(1): 93.  
Wang Yiqin, Dong Yunyun, Liu Huiling. Super-resolution method of hyperspectral image based on GoogLeNet and spatial spectrum transformation[J]. Journal of Optical Technology, 2022, 48(1): 93.
- [38] Ma C, Rao Y M, Cheng Y A, et al. Structure-preserving super resolution with gradient guidance[C] // Proceedings of The IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA; IEEE, 2020: 7769—7778.
- [39] Zhou Y, Wu G, Fu Y, et al. Cross-MPI: Cross-scale stereo for image super-resolution using multiplane images[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Online; IEEE, 2021: 14842—14851.
- [40] Zhang Y, Li K, Li K, et al. Mr image super-resolution with squeeze and excitation reasoning attention network[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Online; IEEE, 2021: 13425—13434.
- [41] Xia B, Hang Y, Tian Y, et al. Efficient Non-local contrastive attention for image Super-resolution[J]. arXiv preprint arXiv: 2201.03794, 2022.
- [42] Tian C W, Zhuge R B, Wu Z H, et al. Lightweight image Super-resolution with enhanced CNN[J]. Knowledge-Based Systems, 2020, 205: 106235.