

残差卷积注意网络的图像超分辨率重建

湛贵辉, 陈 伍, 李忠兵, 易 欣, 刘会康, 韩春阳

西南石油大学 电气信息学院, 成都 610500

摘 要:针对极深神经网络图像超分辨率重建过程中,存在图像特征提取少、信息利用率低,平等处理高、低频信息通道的问题,提出了残差卷积注意网络的图像超分辨率重建算法。构造多尺度残差注意块,最大限度地提高网络提取到多尺寸特征信息,引入通道注意力机制,增强高频信息通道的表征能力。引入卷积注意块的特征提取结构,减少高频图像细节信息的丢失。在网络的重建层,引入全局跳远连接结构,进一步丰富重建的高分辨率图像信息的流动。实验结果表明,所提算法在Set5等基准数据集上的PSNR、SSIM比其他基于深度卷积神经网络的方法均明显提升,验证了提出方法的有效性与先进性。

关键词:图像超分辨率重建;特征提取;多尺度残差注意块;卷积注意块

文献标志码:A **中图分类号:**TP391.4 **doi:**10.3778/j.issn.1002-8331.2003-0376

Image Super Resolution Reconstruction Based on Residual Convolution Attention Network

CHEN Guihui, CHEN Wu, LI Zhongbing, YI Xin, LIU Huikang, HAN Chunyang

School of Electrical Information, Southwest Petroleum University, Chengdu 610500, China

Abstract: In the process of image super-resolution reconstruction, there are some problems such as less image feature extraction, low information utilization, equal processing of high and low frequency information channels. The multi-scale residual attention block is constructed to maximize the network extraction of multi-dimensional feature information, and the channel attention mechanism is introduced to enhance the representation ability of high-frequency information channel. The convolution attention block feature extraction structure is introduced to reduce the loss of high-frequency image detail information. In the reconstruction layer of the network, the global long jump connection structure is introduced to further enrich the reconstruction high resolution image information flow. The experimental results show that the PSNR and SSIM of the proposed algorithm on Set5 and other benchmark datasets are significantly better than those of other deep convolutional neural networks, which verify the effectiveness and the advanced nature of the proposed method.

Key words: image super-resolution reconstruction; feature extraction; multi-scale residual attention block; convolution attention block

在人脸识别、医学成像、卫星图像和视频检测监控等应用领域中,为了取得图像更加丰富的高频信息细节,需要对含噪声模糊不清的低分辨率图像(Low-Resolution, LR)通过一定的处理技术进行处理,得到清晰可靠的高分辨率(High-Resolution, HR)图像^[1-3]。最常用的处理方法是采用图像超分辨率重建技术。超分辨率重建一直是计算机视觉领域的研究热点。

超分辨率重建技术大致分为三类:基于插值^[4]的方法、基于重建^[5]的方法和基于学习^[6]的方法。2014年, Dong 等人首次提出了基于卷积神经网络的图像超分辨率重建算法 SRCNN^[7](Super Resolution Convolutional

Neural Network)。开启了深度学习应用于图像超分辨率重建。2016年 Dong 等提出了沙漏型结构的基于快速的卷积神经网络超分辨率重建算法 FSRCNN^[8](Accelerating the Super-Resolution Convolutional Neural Network),解决了 SRCNN 算法中,将低分辨率图像进行上采样的预处理过程,会进一步引入噪声等问题。Kim 等人受残差网络结构的启发,通过增加网络层数来加深网络结构,提出了极深卷积神经网络的图像超分辨率算法 VDSR^[9](Accurate Image Super-Resolution Using Very Deep Convolutional Networks),取得了更好的图像重建效果。Shi 等人提出了基于亚像素卷积网络的图像超

基金项目:南充市机器人工程与智能制造重点实验室项目(19SXHZ0033)。

作者简介:湛贵辉(1971—),男,教授,研究领域为图像处理、机器视觉、计算机仿真及模拟技术;陈伍(1995—),男,硕士研究生,CCF 会员,研究领域为图像超分辨率重建、图像处理,E-mail:1136894382@qq.com。

收稿日期:2020-03-24 **修回日期:**2020-06-22 **文章编号:**1002-8331(2021)12-0193-08

分辨率重建算法 ESPCN^[10] (Efficient Sub-Pixel Convolutional Network), 通过直接在低分辨率图像结构上进行特征提取, 然后用亚像素卷积对提取到的特征图像进行上采样, 降低了特征提取过程的计算复杂度, 提高了重建效率。Ledig 等人提出了基于生成对抗网络的图像超分辨率重建算法 SRGAN^[11] (Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network), 通过将生成对抗网络思想应用到超分辨率重建中, 结合图像感知损失和对抗损失作为网络的损失函数, 使得重建效果更加接近真实高清图像。

VDSR、ESPCN 等方法表明: 网络深度的加深对超分辨率图像重建质量有至关重要的影响。但训练深度卷积神经网络难以收敛, 在训练过程会出现梯度消失和梯度爆炸等问题。同时, 未完全考虑到图像全局上下文的信息对提取区域的影响, 没有重点关注到图像边缘和纹理等图像细节。对提取的图像特征通道、特征位置和特征空间的信息平等对待, 进一步导致重建的图像缺乏全局特征信息。

本文为解决以上问题, 在 VDSR 网络结构的基础上, 提出了残差卷积注意网络的图像超分辨率算法。本文中提出的算法主要由低维特征提取结构、多尺度特征融合结构、特征注意提取结构和重建结构这四部分构成。运用多尺度卷积核提取不同尺度的特征信息, 并重点关注图像通道域信息进行特征信息的提取。为防止网络过程必要的低频信息的丢失和网络难以收敛等问题, 引入短跳跃连接和长跳远连接的残差结构, 构建多尺度残差注意特征提取模块, 提高图像特征信息的获取精度。同时, 本文算法关注图像特征的通道域注意、空间域注意这两种注意力提取的特征信息, 引入卷积块注意力模块 CBAM^[12] (Convolutional Block Attention Module), 通过显示建模特征通道间的信息和空间位置注意的相关性, 自适应地提取图像特征。网络最后, 通过反卷积跳远连接结构进行对重建图像进行恒等映射, 进一步提高重建图像质量。实验结果验证了本文算法的有效性, 与现有的代表性方法比较, 无论是主观视觉评价, 还是客观指标评价度量, 都表明本文算法的超分辨率重建性能优于其他比较有代表性的方法。

1 相关工作

1.1 多尺度的卷积核

不同尺度卷积核提取的特征信息不同, 本文采取 1×1 、 3×3 和 5×5 不同像素尺度的卷积核来对图像进行特征提取。通过多尺度的卷积核来对图像进行卷积处理, 可以从低分辨率图像中较好地得到更多高频图像细节信息。为节省计算力, 不采取尺寸 7×7 卷积核。受 inception 结构的影响, 将这三种卷积核加宽, 来代替传统加深的结构。多尺度卷积核基本结构如图 1 所示。

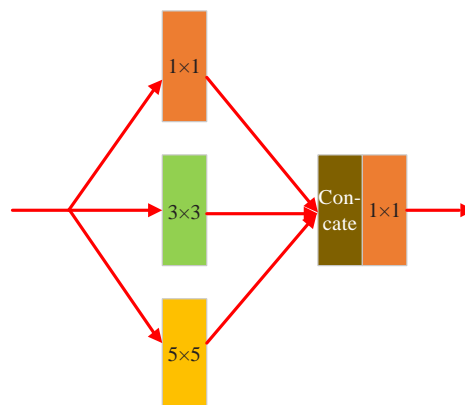


图1 多尺度卷积核基本结构

对多尺度卷积核基本结构的输入 I_L , 进行卷积核为 1×1 、 3×3 和 5×5 的卷积操作:

$$F_{1 \times 1} = H_{1 \times 1}(I_L) \quad (1)$$

$$F_{3 \times 3} = H_{3 \times 3}(I_L) \quad (2)$$

$$F_{5 \times 5} = H_{5 \times 5}(I_L) \quad (3)$$

$$F = H_{\text{concat}, 1 \times 1}(F_{1 \times 1}, F_{3 \times 3}, F_{5 \times 5}) \quad (4)$$

式中, I_L 表示多尺度卷积核输入的低分辨率图像, $H_{1 \times 1}(I_L)$ 、 $H_{3 \times 3}(I_L)$ 和 $H_{5 \times 5}(I_L)$ 分别表示使用 1×1 、 3×3 和 5×5 尺寸大小的卷积核对低分辨率图像 I_L 的卷积操作, 并经过 ReLU 激活函数。 $F_{1 \times 1}$ 、 $F_{3 \times 3}$ 和 $F_{5 \times 5}$ 分别表示经过对应卷积核操作和 ReLU 激活函数后提取到的图像特征图。 $H_{\text{concat}, 1 \times 1}(I_L)$ 表示使用 concat 函数, 将不同尺度卷积核下得到的图像特征图进行特征融合, 同时为保证输入和输出的图像特征图的维度保持不变, 使用 1×1 卷积核进行卷积操作, 降低特征图维度, 减少参数量。

1.2 通道注意力机制

在进行图像卷积的过程中, 对所有通道特征进行平等对待处理。为了解决这个问题, He 等人在文献[13]中提出的通道注意力 CA 机制, 重新校准先前获得的通道相互依赖的特征, 获得信息可以最大限度地利用通道, 通过挤压和激励两种操作显式建模通道依赖项的详细信息。

挤压操作中, 假设 $U = [u_1, u_2, \dots, u_k]$ 是 k 个空间维数为 $w \times h$ 的不同输入通道。特征压缩首先通过全局平均池化对空间维数进行压缩。空间信息放入通道描述符中, 通过使用全局平均池化生成通道统计量。即 2D 的每个特征通道都变成在某种程度上具有全局接受域的统计量。实际上, $Z \in \mathbb{R}^k$ 刻画了通道特征并允许输入附近的层来获取全局接受野。通道信息统计量 $Z \in \mathbb{R}^k$ 是通过其空间维数 $w \times h$ 收缩 U 得到。在挤压操作中, 输出维度匹配, 输入特征通道的数目。因此统计 Z 的第 c 个元素可以定义为:

$$z_c = G_{sq}(u_c) = \frac{1}{w \times h} \sum_{i=1}^h \sum_{j=1}^w u_c(i, j) \quad (5)$$

式中, $G_{sq}(\cdot)$ 表示全局平均池化操作, $u_c(i, j)$ 表示第 c 个

通道特征 u_c 在 (i,j) 处的值。转换 U 的输出可以理解为统计描述符的集合,通道信息统计量是整个图像信息的表现形式,采取的聚合方式为全局平均池化操作。

激励操作,利用挤压中聚集的图像信息统计量,接下来进行激励操作。这是一种类似于门结构的机制,在长期短期记忆(LSTM)中提出。这种门机制为每个特征生成通道权重。通道权重为学会显式地建模特征通道之间的相关性,采取sigmoid函数作为门控机制函数。具体定义如下:

$$s = G_{ex}(z, W) = f(W_2 \sigma(W_1 z)) = \sigma(W_2 \delta(W_1 z)) \quad (6)$$

式中, δ 表示ReLU激活函数, $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$, $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ 。为了限制模型的复杂性和辅助泛化,通过形成参数化浇注机制。即两层完全连(FC)的瓶颈非线性层,具有减速比 r (本文算法中 r 取16)、ReLU和维度增加层返回到转换的通道维度输出 U 。减速比 r 其目的是减少通道的数量然后减少计算力。实际上,整个过程分为四个部分:降维层、ReLU激活、增维层和sigmoid激活层。最后一个sigmoid函数功能是为了生成最终的特征图像信息 s ,这是该CA机制的核心。CA机制的输出特征通过 s 重新调整信道,如下所述:最终输出通过使用激活重新缩放 U :

$$\tilde{U}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c \quad (7)$$

式中, $\tilde{U} = [\tilde{u}_1, \tilde{u}_2, \dots, \tilde{u}_c]$ 和 $F_{scale}(u_c, s_c)$ 表示的是通道注意力校准的标量 s_c 与原始输入特征图 $u_c \in \mathbb{R}^{h \times w \times k}$ 之间逐通道相乘。而 s 表示最终的信道统计量。对于每个通道元素有 $\tilde{u}_c = s_c \cdot u_c$ 。通道注意力CA机制示意图如图2所示。

2 残差卷积注意网络

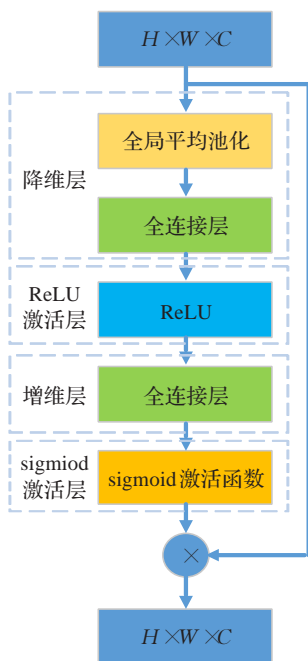


图2 通道注意力CA机制示意图

本文整体神经网络框架如图3所示,残差卷积注意力神经网络由低维特征提取结构、多尺度特征融合结构、特征注意提取结构和重建结构这四部分构成。

2.1 低维特征提取结构

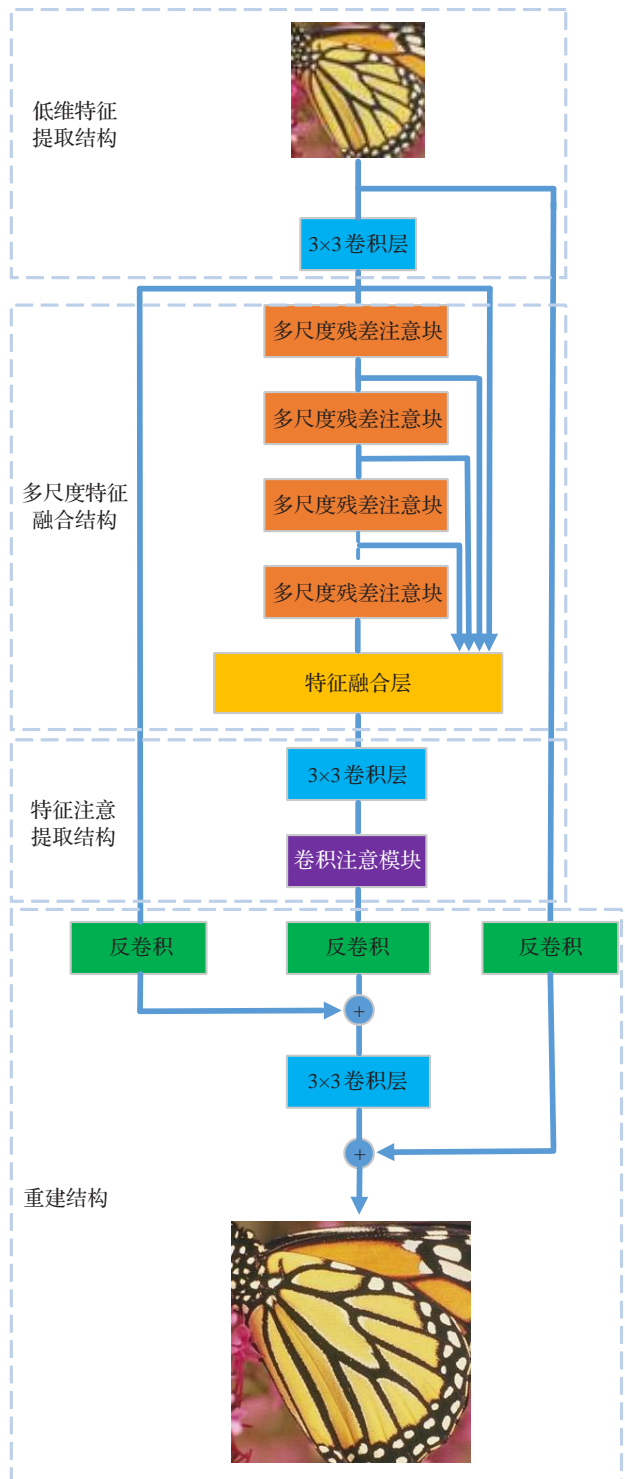


图3 残差卷积注意神经网络结构

在基于残差卷积注意网络的浅层结构中,采用尺寸大小为3×3的卷积核进行网络低维特征的提取,卷积层表示为:

$$I_0 = F_1(Y) = f(W_1 \times I_{LR} + b_1) \quad (8)$$

式中, I_{lr} 表示原始输入的低分辨率图像, W_i 表示 n 个大小为 $k \times k \times c$ 的滤波器, 偏置项为 b_i , f 表示激活函数, 本文采用 ReLU 激活函数。 $F_i(Y)$ 表示低维特征提取结构得到的 n 张特征图, I_o 表示输出特征图。

本文 k 取 3, n 取 64。不同于以往在亮度通道上进行处理, 本文的图像超分辨率重建算法直接在低分辨率的 RGB 图像上进行操作, 避免了忽略其他颜色通道的信息, 因此 c 取 3。在每次卷积操作前都对图像边界进行补零 padding 操作, 充分挖掘图像边缘特征信息。

2.2 多尺度特征融合结构

本文提出的多尺度特征融合结构由多尺度残差注意块和多级特征融合层结构组成, 可以更好地利用多尺度卷积核提取到不同的特征信息, 并经过融合层结构, 对提取到的多种图像特征信息进行自适应地融合。经过低维转换层的输出特征 I_o 作为多尺度残差注意块单元结构的输入特征图。经过实验, 采取 8 个多尺度残差注意块, I_o 输入到 8 个依次顺序连接的多尺度残差注意块。设计的多尺度残差注意块如图 4 所示。

多尺度残差注意块对 I_o 进行不同尺度的特征信息

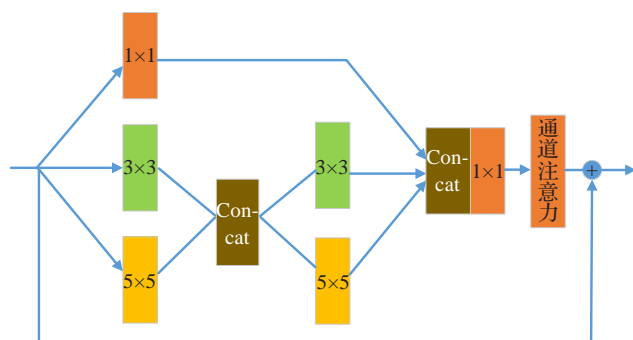


图4 多尺度残差注意力块

提取:

$$I_i = F_i(I_{i-1}, W_i + b_i) \quad (9)$$

式中, F_i 是第 i 个多尺度残差注意块的映射函数, W_i 是 F_i 的权重矩阵参数, b_i 为 F_i 的偏置项矩阵参数, 考虑到计算力本身, 常常忽略偏置项 b_i , I_i 是第 i 个多尺度残差注意块的输出。

多级特征融合层结构将每一级多尺度残差注意结构的输出特征 I_i 进行并联连接, 生成特征信息 I_M , 再通过 2 个卷积层, 卷积核大小为 1×1 , 进行进一步的特征信息融合和降维, 对提取到的特征信息最大利用化:

$$I_G = F_M(I_M, W_M + b_M) \quad (10)$$

式中, F_M 是多级特征融合层的映射函数, W_M 是 F_M 的权重矩阵参数, b_M 为 F_M 的偏置项矩阵参数。 I_M 是每一级多尺度残差注意块 I_i 并联连接的特征, I_G 是多级特征融合层的输出特征。

多尺度残差注意块和多级特征融合层的联合的特征信息提取结构, 提出的网络模型能够提取到不同深

度、不同卷积核尺度和不同通道之间的特征信息。同时, 在末端引入通道注意力层, 能够捕捉更好的建模全局依赖关系^[14]以及降低无关图像域特征信息^[15], 更好地学习到通道之间的相互联系。

2.3 特征注意提取结构

注意力机制能够将更多的注意力集中在感兴趣的信息上, 对应图像超分辨率重建中, 图像的边缘和纹理等高频信息是更关注的信息。在本文的特征注意提取结构中采用通道注意力和空间注意力模块的联合模块进行特征注意提取, 即引入卷积注意模块 CBAM^[12]来进行图像超分辨率注意力信息的提取。

CBAM^[12]的映射由一维通道注意映射 $M_c \in \mathbb{R}^{1 \times 1 \times C}$ 和二维空间注意映射 $M_s \in \mathbb{R}^{1 \times 1 \times C}$, 图 5 显示通道注意力的计算原理过程, 空间注意力的计算原理, 如图 6 所示。

整个 CBAM 注意力过程可以表示为如下:

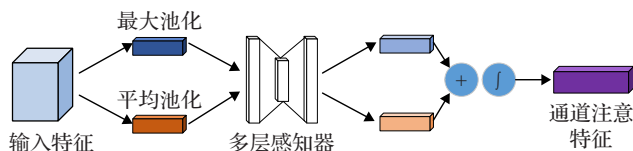


图5 通道注意力计算原理图

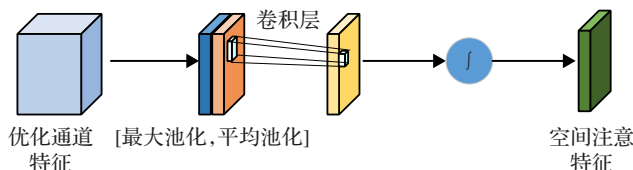


图6 空间注意力计算原理图

$$\begin{cases} F' = M_c(F) \otimes F \\ F'' = M_s(F') \otimes F' \end{cases} \quad (11)$$

式中, \otimes 表示逐元素相乘。在相乘期间, 注意力值按顺序进行处理: 通道注意力值是在空间维度上处理的, 同样地, 空间注意力值是在图像通道上处理的。 F'' 表示的是最后得到的图像注意力值。图 7 显示卷积注意模块结构图。

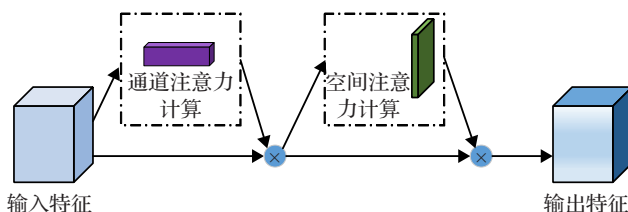


图7 卷积注意模块结构图

2.4 反卷积重建

在神经网络进行卷积操作提取特征信息时, 会导致输出的图像越来越小。本文算法在进行上采样的过程中, 考虑到这个因素, 采取反卷积来进行上采样, 能够放大多尺度特征映射图。

采取双三次插值对输入图像做预处理,会引入过多噪声,且会限制学习速度^[16]。反卷积进行上采样会减少人工因素的影响。在图像超分辨率重建领域,可以用于图像的放大过程,代替传统的双三次插值操作。本文算法,在进行图像上采样放大过程中,使用反卷积操作进行图像放大,得到想要的图像尺寸大小。

2.5 损失函数

在图像超分辨率重建领域使用最广泛的损失函数是平均绝对误差损失函数L1(MAE)和均方误差损失函数L2(MSE)。L1损失函数具有可以达到较好的主观视觉效果和客观评价指标,可以防止重建图像过于平滑,导致图像失真现象的优点。本文采用平均绝对误差损失函数L1(MAE)来估计网络参数 θ 。平均绝对误差损失函数表示为:

$$L(\theta) = \frac{1}{N} \sum_i \|F(Y_i; \theta) - X_i\|_1 \quad (12)$$

式中, $L(\theta)$ 表示平均绝对误差, Y_i 表示输入低分辨率图像(LR), X_i 和 $F(Y_i; \theta)$ 分别表示原始高分辨率图像(HR)和重建图像(SR), N 代表训练集样本的总数目,网络参数 θ 为各层权值 w_i 和偏置值 b_i 参数的集合,即 $\theta = (w_1, w_2, \dots, w_i, b_1, b_2, \dots, b_i)$ 。通过最小化重建图像与对应原始高分辨率图像的L1损失来估计网络参数 θ 。为比较L1与L2重建图像的优势,在3.3节对损失函数增加消融实验。

3 实验结果分析

3.1 实验配置

本文提出的算法实验所需要的实验环境为:Ubuntu16.04 LTS操作系统,采用深度学习框架Pytorch网络模型进行实验优化,搭建cuda10.0,cudnn7.6.2用于本文算法实验的加速,在进行测试时,使用Matlab R2014a。本文实验硬件设备配置为Intel® Core™ i5-9600K@3.7 GHz处理器,16 GB运行内存,NVIDIA Geforce GTX2060(6 GB)显卡。

3.1.1 实验数据集

整个实验过程中所使用的数据集,采用的是近年来新提出的一种高清(2K分辨率)的图像数据集DIV2K,其中包含有训练集图像800张,验证集图像100张,测试集图像100张。由于DIV2K仅包含800张训练集图像和100张验证集图像,其100张测试集图像尚未发布,本文采用图像超分辨率领域的基准数据集Set5、Set14和Urban100^[17]来进行测试。在进行图像预处理时,采用将DIV2K中的800张训练集图像,分别进行双三次插值下采样2倍、3倍和4倍的降质处理,作为网络的输入低分辨率图像。

3.1.2 实验细节

网络模型中超分辨率网络参数设置,如表1。表中

H 、 W 分别是输入图像的高度和宽度,ReLU是线性修正单元,Conv是卷积层,MaxPool是最大池化层,AvgPool是全局平均池化层,ConvT是反卷积层,FC层是全连接层。其中网络模型参数设置:每批16个训练图像对,全部训练图像对共6 720批,作为一个训练周期(epoch),共训练3 000个周期;训练优化算法Adam;初始学习率为 10^{-4} ;迭代200个周期,学习率减半;卷积滤波器初始化采用MSRA算法;反卷积滤波器采用均值为0、标准差为0.001的高斯分布进行初始化。卷积算子采用的是 1×1 、 3×3 和 5×5 的尺寸。网络层数大体分为14层,包含低维特征提取结构层、多尺度残差注意块层、特征融合层、卷积注意模块层和反卷积重建层等。

表1 超分辨率网络结构的参数设置

结构组件名称	滤波器	输入尺寸	输出尺寸
低维提取结构	Conv3×3	$H \times W \times 3$	$H \times W \times 64$
	ReLU	$H \times W \times 64$	$H \times W \times 64$
多尺度残差注意块×8	Conv1×1	$H \times W \times 64$	$H \times W \times 64$
	Conv3×3	$H \times W \times 64$	$H \times W \times 64$
	Conv5×5	$H \times W \times 64$	$H \times W \times 64$
	Conv1×1	$H \times W \times 128$	$H \times W \times 64$
	Conv1×1	$H \times W \times 192$	$H \times W \times 64$
特征融合层	Conv1×1	$H \times W \times 512$	$H \times W \times 64$
	Conv3×3	$H \times W \times 64$	$H \times W \times 64$
	AvgPool	$H \times W \times 64$	$1 \times 1 \times 64$
	FC层	$1 \times 1 \times 64$	$1 \times 1 \times 64$
	ReLU	$1 \times 1 \times 64$	$1 \times 1 \times 64$
特征注意提取结构	FC层	$1 \times 1 \times 64$	$H \times W \times 64$
	MaxPool7×7	$H \times W \times 64$	$H \times W \times 64$
	Conv3×3	$H \times W \times 64$	$H \times W \times 32$
	ReLU	$H \times W \times 32$	$H \times W \times 32$
	Conv3×3	$H \times W \times 32$	$H \times W \times 32$
反卷积重建	ReLU	$H \times W \times 32$	$H \times W \times 32$
	Conv3×3	$H \times W \times 32$	$H \times W \times 64$
	ConvT4×4	$H \times W \times 64$	$2H \times 2W \times 3$
	ConvT4×4	$H \times W \times 3$	$2H \times 2W \times 3$
LR反卷积重建	Conv3×3	$2H \times 2W \times 3$	$2H \times 2W \times 3$
	ConvT4×4	$H \times W \times 3$	$2H \times 2W \times 3$

3.2 实验客观评价指标

本文重建图像的重建效果的衡量标准,采用图像超分辨率中常用的客观评价指标:峰值信噪比(PSNR)和图像结构相似性(SSIM)。重建图像的峰值信噪比(PSNR)和结构相似性(SSIM)值在0到1的范围,PSNR是衡量对应像素点间的误差,即误差敏感图像质量评价。

一般情况下,越高的PSNR值表示越好的高分辨率图像,通过均方误差(Mean Square Error, MSE)对峰值信噪比(PSNR)进行定义。

均方误差可表示为:

$$MSE = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (I(i, j) - K(i, j))^2 \quad (13)$$

图像峰值信噪比(PSNR)的计算方法可表示为:

$$f_{\text{PSNR}} = 10 \lg \frac{(2^n - 1)^2}{MSE} \quad (14)$$

式中, I 表示 $H \times W$ 大小的重建图像, K 表示 $H \times W$ 大小的原始高分辨率图像, H 、 W 分别表示图像的高度和宽度, n 一般取 8, 表示每个像素的 bit 数值。峰值信噪比 PSNR 的单位是 dB, PSNR 数值越大, 重建图像失真越小, 重建效果越好。

图像结构相似性 (SSIM) 的计算表示为:

$$f_{\text{SSIM}} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_x\sigma_y + C_2)(\sigma_{xy} + C_3)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)(\sigma_{xy} + C_3)}$$

式中, x 表示重建图像, y 表示原始高分辨率图像, μ_x 表示 x 的平均值, μ_y 表示 y 的平均值, σ_x 表示 x 的方差, σ_y 表示 y 的方差, σ_{xy} 表示 x 和 y 的协方差。SSIM 取值在 $[0, 1]$, 越接近 1, 表明重建图像与原始高分辨率的相似度越高, 效果越好。

3.3 消融实验分析

为比较 L1 与 L2 重建图像的优势, 对多尺度残差注意块数量和损失函数的选择增加消融实验。

对 L1、L2 损失函数和基础块数量的选择进行算法验证。在 L1 和 L2 损失函数下对不同数量的基础块进行实验, 并在 Set5、Set14 和 urban100 测试集上进行测试, 得到了测试集上的平均峰值信噪比。

L1 和 L2 损失函数、多尺度残差注意块的 $\times 2$ 超分辨率重建效果在 Set5、Set14 和 Urban100 上的平均峰值信噪比, 如图 8、图 9 所示。由图知, 无论是采用 L1 损失函数还是 L2 损失函数, 当多尺度残差注意块数量为 8 时, 重建效果最好。

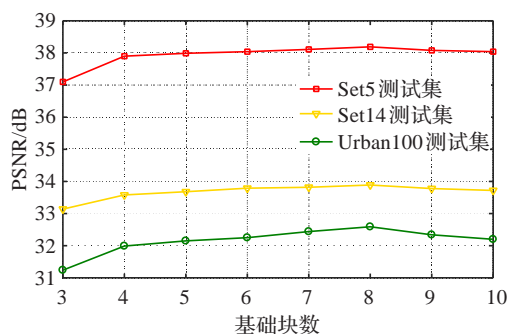


图8 L1 损失函数与多尺度残差注意块折线图

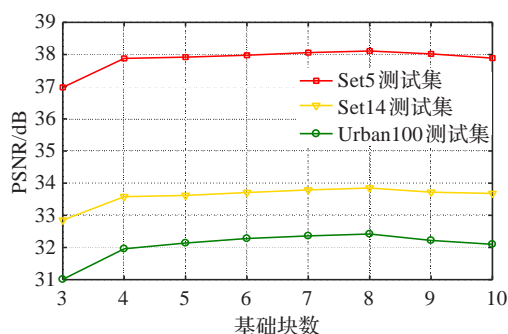


图9 L2 损失函数与多尺度残差注意块折线图

经过实验验证, 当基础块数量为 8, 选择 L1 损失函数时, 具有最好的重建效果。不同损失函数和基础块设置的网络模型对 $\times 2$ 的超分辨率性能的影响, 参见表 2。

表2 不同变量的网络模型 $\times 2$ 超分辨率

损失函数	基础块数量	PSNR/dB		
		Set5	Set14	Urban100
L1	3	37.10	33.14	31.24
	4	37.90	33.58	31.99
	5	37.99	33.68	32.15
	6	38.04	33.79	32.25
	7	38.11	33.82	32.44
	8	38.19	33.89	32.59
	9	38.08	33.78	32.34
	10	38.04	33.72	32.20
L2	3	36.98	32.85	31.01
	4	37.88	33.58	31.96
	5	37.92	33.62	32.14
	6	37.98	33.71	32.28
	7	38.06	33.79	32.36
	8	38.11	33.85	32.42
	9	38.02	33.72	32.22
	10	37.89	33.68	32.10

从表 2 中可看出, 当使用 L1 损失函数与多尺度残差注意块数量为 8 的时候, 重建指标最高, 而采用 L2 损失函数和多尺度残差注意块数量为 3 的基线模型重建指标最差, 在 Set5 测试集的 $\times 2$ 尺度中落后于最优配置 1.21 dB, 在 Set14 测试集的 $\times 2$ 尺度中落后 1.04 dB, 在 Urban100 测试集的 $\times 2$ 尺度中落后 1.58 dB, 证明了采用 L1 损失函数和多尺度残差注意块为 8 时, 提出的重建算法具有最好的重建效果。

采用 L1 损失函数的重建效果好于 L2 损失函数, L1 具有可以达到较好的重建图像主观视觉效果和客观评价指标, 这是 L2 损失函数所不具有的作用。当多尺度残差注意块数量增加时, 重建图像质量越来越高; 当注意块数量增加到 8 时, 图像的重建效果最好; 当继续增加注意块数量, 图像重建质量变低。可能是网络在提取高频图像信息的同时, 重建图像所需必要的低频信息被丢弃, 使得重建质量变低。这表明当注意块数量为 8 时, 网络具有能够最有效地提取到高频图像信息, 同时也保留了必要的重建图像低频信息的能力。

3.4 实验结果分析

为了证明提出方法的先进性与有效性, 验证所提出的算法与近年优秀的重建算法的对比效果。选取了单帧图像超分辨率领域先进的五种方法深度学习方法, VDSR^[8]、ESPCN^[9]、DRCN^[18] (Deeply-Recursive Convolutional Network for image super-resolution) 和 CARN^[19] (Accurate, and Light weight Super-Resolution with Cascading Residual Network) 和 IDN^[20] (Fast and accurate single image super-resolution via Information

表3 不同测试数据集上的PSNR(dB)、SSIM指标对比结果

Dataset	Scale	PSNR/SSIM						
		Bicubic	ESPCN	VDSR	DRCN	CARN	IDN	OUR
Set5	×2	33.66/0.930	36.86/0.956	37.53/0.959	37.63/0.960	37.76/0.959	37.83/0.960	38.19/0.961
	×3	30.39/0.869	32.95/0.915	33.67/0.921	33.82/0.923	34.29/0.926	34.11/0.925	34.61/0.929
	×4	28.43/0.810	30.29/0.867	31.35/0.883	31.53/0.882	32.13/0.894	31.82/0.890	32.35/0.897
Set14	×2	30.24/0.869	32.75/0.910	33.05/0.913	32.98/0.914	33.52/0.917	33.30/0.915	33.89/0.920
	×3	27.55/0.774	29.49/0.827	29.77/0.830	29.76/0.833	30.29/0.841	29.99/0.836	30.52/0.846
	×4	26.01/0.703	27.73/0.764	28.02/0.768	28.04/0.771	28.60/0.781	28.25/0.773	28.77/0.786
Urban100	×2	26.88/0.840	29.90/0.801	30.77/0.910	30.76/0.913	31.92/0.926	31.27/0.920	32.59/0.933
	×3	24.46/0.735	26.63/0.801	27.14/0.830	27.15/0.828	28.06/0.849	27.42/0.836	28.57/0.860
	×4	23.14/0.658	24.86/0.734	25.18/0.750	25.14/0.753	26.07/0.784	25.41/0.763	26.42/0.796

Distillation Network)算法与非深度学习图像重建领域传统的Bicubic算法,分别在Set5、Set14和Urban100标准测试集上进行重建图像2倍、3倍和4倍的性能验证实验。结果如表3所示。从表3中可以看出,无论是在2倍、3倍和4倍尺度下,在不同测试集下,提出的算法重建图像的PSNR与SSIM指标均高于这几种算法。上采样因子为4时,在Set5测试集上,本文算法的PSNR值比VDSR提高了1 dB,比CARN提高了0.22 dB。在Set14测试集上,本文算法比VDSR提高了0.75 dB,比CARN提高了0.17 dB。在Urban100测试集上,本文算法比VDSR提高了1.24 dB,比CARN提高了0.35 dB。

对不同先进算法的主观视觉效果进行了对比评价,所采用的对比方法的实现均来自作者公开的源代码,使用原论文中的参数和预训练模型得到。图10表示Set5测试集上本文超分辨率重建算法和其他算法的对比,本文算法中的蝴蝶图像放大后,可以更加清楚地观察到纹理细节。图11表示Set14测试集上和其他算法的对比,Set14中的猴子图像从放大眼睛及其边缘的细节可以看出,本文重建算法的猴子眼角纹理和眼睛瞳孔细节对比其他算法均更加清晰。图12表示Urban100测试集上和其他算法的对比,Urbn100中栏杆在放大栏杆的细节后可以看出,本文重建算法的高频细节与其他算法重建的图像对比更加清晰,恢复出了主栏杆的细节信息,更接近原始高清图像。因此,提出残差卷积注意网络的图像

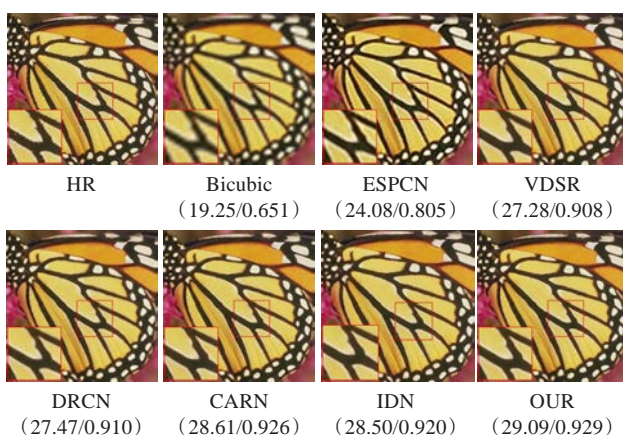


图10 Set5测试数据集上的对比结果

超分辨率重建效果从视觉效果来看有所提高。本文在极深卷积神经网络超分辨率重建算法的基础上,改进的残差卷积注意网络的图像超分辨率重建算法,在测试数据集上明显优于目前主流算法。甚至优于2018年在ECCV会议收录的CARN算法和2018年CVPR会议收录的IDN算法。

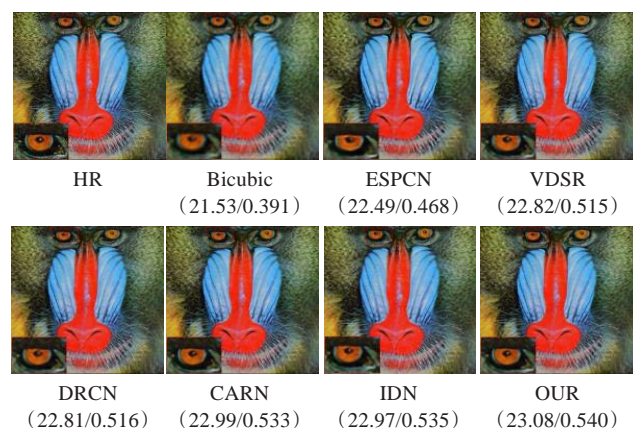


图11 Set14测试数据集上的对比结果

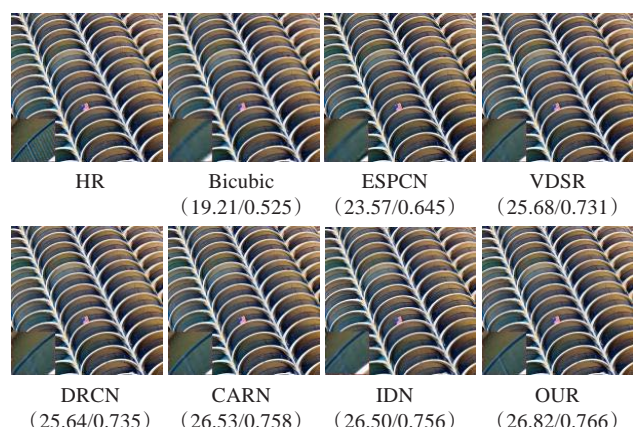


图12 Urban100测试数据集上的对比结果

4 结论

先前基于深度学习的图像超分辨率重建方法中,存在特征提取少、信息利用率低和平等处理图像各个通道的问题。本文受极深卷积神经网络超分辨率重建算法(VDSR)启发,结合卷积注意模块,提出了残差卷积注

意网络的超分辨率重建方法。在网络中采取残差注意块局部残差和网络全局残差相结合的方式降低网络训练难度,取消双三次插值,在最后增加反卷积层,在网络结构中加入concat层进行多尺度特征融合、通过 1×1 的卷积核进行降维和扩维。构造多尺度残差注意块来学习高分辨率和低分辨率图像之间的映射关系。本文提出的算法能够更好地恢复重建图像的高频纹理图像信息。无论从客观评价指标,还是主观视觉上,均优于目前的主流方法。

参考文献:

- [1] 苏衡,周杰,张志浩.超分辨率图像重建方法综述[J].自动化学报,2013,39(8):1202-1213.
- [2] CHEN H, HE X, TENG Q, et al. Single image super resolution using local smoothness and nonlocal self similarity priors[J]. Signal Processing Image Communication, 2016, 43: 68-81.
- [3] ZHANG H, ZHANG L, SHEN H. A blind super-resolution reconstruction method considering image registration errors[J]. International Journal of Fuzzy Systems, 2015, 17(2): 353-364.
- [4] BATZ M, EICHENSEER A, SEILER J, et al. Hybrid super-resolution combining example-based single-image and interpolation-based multi-image reconstruction approaches[C]//2015 IEEE International Conference on Image Processing (ICIP), 2015: 58-62.
- [5] KIM K I, KWON Y. Single-image super-resolution using sparse regression and natural image prior[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(6): 1127-1133.
- [6] LIAN Q S, ZHANG W. Super-resolution reconstruction algorithm based on classification of sparse representations of image blocks[J]. Journal of Electronics, 2012, 40(5): 920-925.
- [7] DONG C, CHEN C L, HE K, et al. Image super-resolution using deep convolutional networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 38(2): 295-307.
- [8] DONG C, CHEN C L, TANG X. Accelerating the super-resolution convolutional neural network[C]//European Conference on Computer Vision. Cham: Springer, 2016: 391-407.
- [9] 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.
- [10] 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]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1874-1883.
- [11] 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, Honolulu, USA, 2017: 4681-4690.
- [12] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.
- [13] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770-778.
- [14] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-attention generative adversarial networks[J]. arXiv: 1805.08318, 2018.
- [15] ZHANG Yulun, LI Kunpeng, LI Kai, 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.
- [16] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [17] HUANG J B, SINGH A, AHUJA N. Single image super-resolution from transformed self-exemplars[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 5197-5206.
- [18] KIM J, LEE J K, LEE K M. Deeply-recursive convolutional network for image super-resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1637-1645.
- [19] AHN N, KANG B, SOHN K A. Fast, accurate, and lightweight super-resolution with cascading residual network[C]//Proceedings of the European Conference on Computer Vision, 2018.
- [20] HUI Z, WANG X, GAO X. Fast and accurate single image super-resolution via information distillation network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 723-731.