

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2019)03-0410-10

论文引用格式: Ying Z L, Long X. Single-image super-resolution construction based on multi-scale dense residual network [J]. Journal of Image and Graphics, 2019, 24(03): 0410-0419. [应自炉, 龙祥. 多尺度密集残差网络的单幅图像超分辨率重建[J]. 中国图象图形学报, 2019, 24(03): 0410-0419. ] [DOI: 10. 11834/jig. 180431]

# 多尺度密集残差网络的单幅图像超分辨率重建

应自炉, 龙祥

五邑大学信息工程学院, 江门 529020

**摘要:** 目的 近几年应用在单幅图像超分辨率重建上的深度学习算法都是使用单尺度卷积核提取低分辨率图像的特征信息, 这样很容易造成细节信息的遗漏。另外, 为了获得更好的图像超分辨率重建效果, 网络模型也不断被加深, 伴随而来的梯度消失问题会使得训练时间延长, 难度加大。针对当前存在的超分辨率重建中的问题, 本文结合 GoogleNet 思想、残差网络思想和密集型卷积网络思想, 提出一种多尺度密集残差网络模型。方法 本文使用 3 种不同尺度卷积核对输入的低分辨率图像进行卷积处理, 采集不同卷积核下的底层特征, 这样可以较多地提取低分辨率图像中的细节信息, 有利于图像恢复。再将采集的特征信息输入残差块中, 每个残差块都包含了多个由卷积层和激活层构成的特征提取单元。另外, 每个特征提取单元的输出都会通过短路径连接到下一个特征提取单元。短路径连接可以有效地缓解梯度消失现象, 加强特征传播, 促进特征再利用。接下来, 融合 3 种卷积核提取的特征信息, 经过降维处理后与  $3 \times 3$  像素的卷积核提取的特征信息相加形成全局残差学习。最后经过重建层, 得到清晰的高分辨率图像。整个训练过程中, 一幅输入的低分辨率图像对应着一幅高分辨率图像标签, 这种端到端的学习方法使得训练更加迅速。结果 本文使用两个客观评价标准 PSNR (peak signal-to-noise ratio) 和 SSIM (structural similarity index) 对实验的效果图进行测试, 并与其他主流的方法进行对比。最终的结果显示, 本文算法在 Set5 等多个测试数据集上的表现相比于插值法和 SRCNN 算法, 在放大 3 倍时效果提升约 3.4 dB 和 1.1 dB, 在放大 4 倍时提升约 3.5 dB 和 1.4 dB。结论 实验数据以及效果图证明本文算法能够较好地恢复低分辨率图像的边缘和纹理信息。关键词: 单幅图像超分辨率; 多尺度卷积核; 残差网络; 密集型卷积网络; 特征提取单元

## Single-image super-resolution construction based on multi-scale dense residual network

Ying Zilu, Long Xiang

School of Information Engineering, Wuyi University, Jiangmen 529020, China

**Abstract:** **Objective** Single-image super-resolution aims to generate a visually pleasing high-resolution image from its degraded low-resolution measurement. Single-image super-resolution is used in various computer vision tasks, such as security and surveillance imaging, medical imaging, and image generation. However, image super-resolution is an ill-posed inverse problem because of a multitude of solutions for any low-resolution input. In recent years, a series of convolution neural networks model has been proposed for single-image super-resolution. The deep learning algorithms applied to single-image

收稿日期: 2018-07-05; 修回日期: 2018-09-04; 预印本日期: 2018-09-11

基金项目: 国家自然科学基金项目 (61771347, 61072127, 61372193); 广东省自然科学基金项目 (S2013010013311, 10152902001000002, S2011010001085, S2011040004211)

**Supported by:** National Natural Science Foundation of China (61771347, 61072127, 61372193); Natural Science Foundation of Guangdong Province, China (S2013010013311, 10152902001000002, S2011010001085, S2011040004211)

super-resolution reconstructions have used single-scale convolutional kernels to extract feature information of low-resolution images, thereby causing omissions of detailed information easily. Moreover, to obtain better image super-resolution reconstruction effects, the network model is constantly deepened, and the accompanying problems of gradients vanished, thereby resulting in longer training time and difficulty. A multi-scale dense residual network model based on GoogleNet, residual network, and intensive convolution network ideas is proposed to address these existing super-resolution reconstruction problems. **Method** Different from the traditional single-scale feature extraction convolutional kernel, this study uses three different scales of convolution kernels  $3 \times 3$ ,  $5 \times 5$ , and  $7 \times 7$  to perform convolution processing on the input low-resolution images and collects the underlying features of different convolution kernels. Therefore, more detailed information on low-resolution images, which are beneficial to image restoration, are extracted. Then, the collected feature information is inputted into the residual block. Each residual block contains a number of feature extraction units consisting of convolutional and active layers. In addition, the output of each feature extraction unit is connected to the next feature extraction unit through a short path. Short-path connections can effectively alleviate the disappearance of gradients, enhance the propagation of features, and promote the reuse of features. Then, the feature information extracted by the three convolution kernels is merged, and the feature information extracted by the  $3 \times 3$  pixels convolution kernel is added after dimensionality reduction processing to form a global residual learning. After a final reconstruction, a clear, high-resolution image is obtained. Throughout the training process, an input low-resolution image corresponds to a high-resolution image tag. This end-to-end learning method results in faster training. We use the mean squared error as the loss function. The loss is minimized using stochastic gradient descent with the standard backpropagation. We use a training data of 1 000 images from DIV2k, and the flipped and rotated versions of the training images are considered. We rotate the original images by  $90^\circ$  and  $270^\circ$ . **Results** This study uses two objective evaluation criteria, namely, peak signal-to-noise ratio and structural similarity index to test the effect map of the experiment and compare it with other mainstream methods. The final results show that compared with interpolation method and SRCNN algorithm on Set5 dataset, the proposed algorithm improves approximately 3.4 dB and 1.1 dB at three times of magnification and 3.5 dB and 1.4 dB at four times of magnification, respectively. **Conclusion** We propose a multi-scale dense residual network for single-image super-resolution. The experimental data and the effect graph confirm that the proposed algorithm can better recover the edge and texture information of low-resolution images. However, our networks have a large number of parameters because our algorithm uses three channels to recover image details. Therefore, our algorithm requires more convergence time. We will also reduce the number of weight parameters by decomposing the convolution kernel.

**Key words:** single image super-resolution; multi-scale convolution kernel; residual network; dense convolutional network; feature extraction unit

## 0 引言

单幅图像超分辨率(SISR)技术可以恢复低分辨率(LR)图像中的细节信息,得到高分辨率(HR)图像。该技术可以应用在城市安防、卫星遥感、医学图像等多个领域,具有重要的研究价值,所以自图像超分辨率(SR)这个概念被提出后就一直倍受学术界关注,各种方法层出不穷,包括基于插值法<sup>[1]</sup>、基于重建法<sup>[2]</sup>、基于学习法<sup>[3-5]</sup>等。

深度学习这个概念是由著名计算机科学家Hinton等人<sup>[6]</sup>于2006年在《科学》杂志上发表的论文中所提出的。2014年,Dong等人<sup>[7]</sup>提出基于卷积神经网络的图像超分辨率(SRCNN),首次将深度

学习方法应用在SR。SRCNN是一个3层的卷积神经网络,该网络建立了经过插值处理后的LR图像和HR图像之间端到端的映射关系。SRCNN算法的提出为之后各种基于深度学习算法的出现奠定了基础。Shi等人<sup>[8]</sup>提出一种利用亚像素卷积层的超分辨率重建(ESPCN)方法,该方法在最后一层重新对特征图进行排列,得到高分辨率图像。在随后的发展过程中,研究人员发现深度网络可以得到更好的效果。但是,一味地增加网络深度会导致梯度弥散或者梯度爆炸,使得网络难以收敛。Kim等人<sup>[9]</sup>引入He等人<sup>[10]</sup>提出的残差网络(ResNet)思想,提出基于深度卷积神经网络的图像超分辨率(VDSR)。VDSR网络采用跳跃连接(skip connection)方式,通过特征图累加的方式解决因为网络层

数增加带来的梯度弥散问题。VDSR 不仅有较快的收敛速度,而且适用于多个尺度图像超分辨率网络。增加网络深度不仅会带来梯度弥散问题,还会增加大量的网络参数,这就在很大程度上延长了网络的收敛时间。Kim 等人<sup>[11]</sup>提出的基于深度循环神经网络的图像超分辨率重建(DRCN)和 Tai 等人<sup>[12]</sup>提出的基于深度循环残差网络的图像超分辨率重建(DRRN)均在网络中加入循环神经网络,大幅度减少网络参数。以上网络都需要在图像的预处理部分将 LR 图像放大到期望的大小。这个预处理过程不仅增加了网络的计算复杂度,而且得到的 HR 图像过于平滑,不符合视觉的真实感受。

2016 年, Dong 等人<sup>[13]</sup>在 SRCNN 的基础上进行改进,用反卷积层放大图像尺寸,就可以直接将原始 LR 图像输入至网络中,降低网络的计算复杂度。另外,使用了更小的卷积核,加快了网络的收敛速度。Tong 等人<sup>[14]</sup>使用密集跳跃连接实现图像超分辨率重建(SRDenseNet),直接引用 He 等人<sup>[10]</sup>提出的残差网络思想,在训练时间和参数存储上都取得了较好的表现。但是,残差网络提出的最初目的是用于解决计算机视觉方面的问题,比如检测、分类等等。如果不进行改进直接应用在图像超分辨率重建技术上,可能泛化性较差。Lim 等人<sup>[15]</sup>提出的增强深度残差网络(EDSR)修改了残差块的设计,去除了其中的 BN(batch normalization)层。实验证明,去除 BN 层后可以节省约 40% 的 GPU 内存空间。

以上所有图像超分辨率重建方法,都是用单个尺度的卷积核对 LR 图像进行特征提取,没有充分地提取和利用原始 LR 图像的信息。针对上述问题,本文提出一种基于多尺度残差网络的单幅图像超分辨率重建方法。使用多个不同尺度的卷积核对原始图像进行卷积操作,提取多种底层特征。然后分别将提取的特征图作为残差块的输入,通过这样的方式学习到图像的多种高频信息。再将学习到的高频信息进行串接后降维,与原始低分辨率图像重建出高分辨率图像。

## 1 理论基础

### 1.1 卷积神经网络

近年来,卷积神经网络(CNN)以其独到的优势成为计算机视觉领域科研人员关注的热点。CNN 对图像的几何变换、形变、光照等具有一定程度的不变性,以较小的计算量就能完成整幅图像的特征扫描。端到端的深度学习算法,直接通过学习 CNN 提取的特征,基于原始数据的特征学习,逐层进行特征提取和特征融合。随着网络的不断加深,越深的 CNN 层提取的特征越抽象,越抽象的特征越可以较大幅度地提升目标检测和识别精度。一般 CNN 由以下几个核心部分组成:输入层、隐藏层、输出层。其结构如图 1 所示。

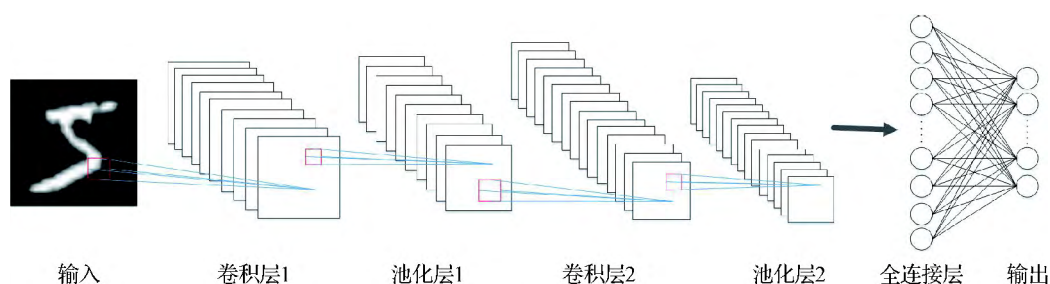


图 1 卷积神经网络结构图

Fig. 1 Architecture of convolutional neural network

输入层是将经过裁减、转换颜色空间等预处理步骤的图像输入网络。隐藏层包括卷积层、池化层和全连接层。其中卷积层是通过卷积核对输入图像进行卷积运算得到多通道的特征图像,可以有效地增强输入图像的特征,降低噪声干扰。池化层可以减少网络参数,降低计算量,同时保持图像的旋转不

变性。全连接层采用 softmax 全连接,得到的激活值即网络提取的图像特征。输出层得到 CNN 输出的分类结果。CNN 通过 3 个方法,即局域感受野、权值共享和次抽样实现识别图像的位移、缩放和扭曲不变性。局域感受野指的是每一层网络的神经元只与上一层的一个小邻域内的神经单元连接,通过局

域感受野,每个神经元可以提取初级的视觉特征;权值共享使得CNN具有更少的参数,需要相对较少的训练数据;次抽样可以减少特征的分辨率,实现对位移、缩放和其他形式扭曲的不变性。卷积层之后通常用次抽样层来减少计算时间、建立空间和结构的不变性。在图像超分辨率重建中,卷积神经网络通过提取LR图像的特征信息重建出HR图像,计算输入图像与目标图像之间的误差,不断调整参数使得HR图像更接近目标图像。

## 1.2 残差网络

随着计算机硬件计算能力的突飞猛进,神经网络的层数也在不断提升。在解决复杂问题上,深度神经网络比浅层网络具有更好的效果。目前优化神经网络的方法都是基于反向传播的思想,即根据损失函数计算的误差通过梯度反向传播的方式,指导深度网络权值的更新优化。但是网络的加深使得梯度弥散问题容易出现。一般解决梯度弥散的方法为:1)选择合适的激活函数,如Relu;2)批规范化

(BN)。通过规范化操作将输出信号规范化到均值为0、方差为1,从而保证网络的稳定性。Relu激活函数在 $x > 0$ 时可以有效地缓解梯度弥散问题。但是随着训练的进行,当 $x < 0$ 时对应的权值参数无法更新,影响网络的收敛。BN法的使用伴随的是网络参数的大量增加,同样会使得网络收敛时间延长。相比以上两种方法,He等人<sup>[10]</sup>提出的残差网络(ResNet)方法效果更佳。通过计算残差网络的偏导数(式(1))可以看出,ResNet不仅使得网络的层数大幅度加深,还很好地缓解了网络梯度弥散问题

$$\frac{\partial X_2}{\partial X_1} = \frac{\partial X_1 + \partial F(X_1, W_1, b_1)}{\partial X_1} = 1 + \frac{\partial F(X_1, W_1, b_1)}{\partial X_1} \quad (1)$$

式中 $X_1$ 为残差单元输入参数, $X_2$ 为残差单元输出参数, $W_1$ 为权值参数, $b_1$ 为偏置参数, $F(X_1, W_1, b_1)$ 表示残差单元学习到的映射。残差网络的结构如图2所示。

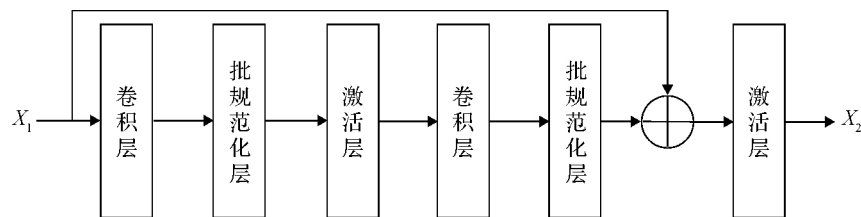


图2 残差网络结构图

Fig. 2 Architecture of residual neural network

## 2 本文方法

图像超分辨率重建是通过提取低分辨率图像的细节信息、重建等步骤得到其相应的高分辨率图像。这种端到端的非线性映射关系可以使得网络能够学习到低分辨率图像的高频信息。前人应用在图像超分辨率上的深度学习方法,全部都是使用一种尺度的卷积核提取底层特征信息,这样就会在低分辨率图像上遗漏较多的细节信息。本文提出的网络由底层特征提取模块、密集残差块模块和重建模块组成。其中底层特征模块采用多个不同尺度的卷积核对输入图像进行特征提取,本网络中使用尺度分别为 $3 \times 3$ 像素、 $5 \times 5$ 像素和 $7 \times 7$ 像素的卷积核提取底层特征。在对提取的底层特征图的后续处理上采用密集残差模型,这样可以有效避免因为网络的深度

导致的梯度弥散问题。在重建部分使用反卷积,不需要在输入前对低分辨率图像进行插值处理,不仅减少大量的参数,而且加快网络的收敛速度。多尺度残差网络如图3所示。

由图3可知,多尺度残差网络结构主要由输入、底层特征提取(SFE)、残差密集块(RDB)、局部特征融合(LFF)、全局特征融合(GFF)、全局残差学习(GRL)、重建和输出组成。

SFE模块包括3种不同的卷积核,尺度分别为 $3 \times 3$ 像素、 $5 \times 5$ 像素和 $7 \times 7$ 像素。该模块引入Inception Net<sup>[16]</sup>思想,使用不同尺度的卷积核对输入图像进行卷积处理,相比前人使用单个卷积核的图像超分辨率深度学习算法,本文算法可以从低分辨率图像中得到更多的细节信息,有利于低分辨率图像的细节恢复

$$F = H_{\text{SFE}, 3 \times 3}(I_{\text{LR}}) \quad (2)$$

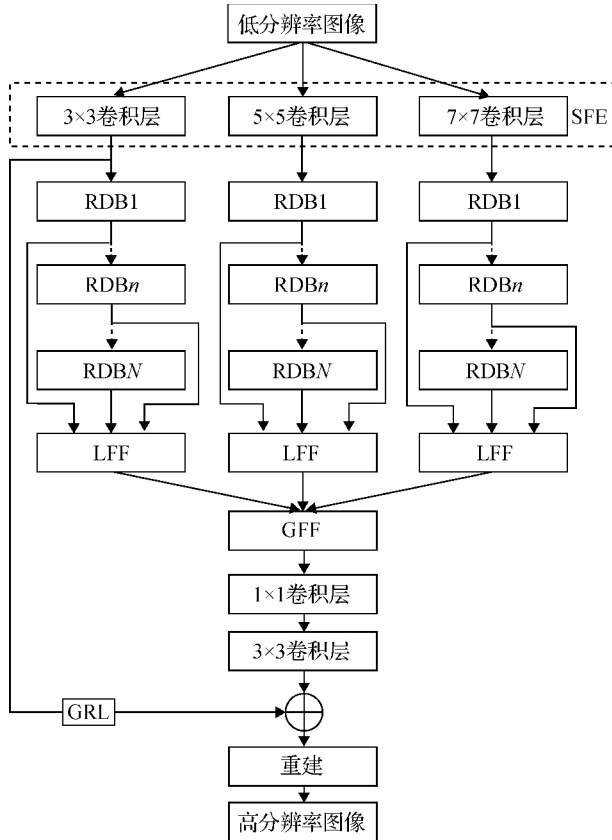


图3 多尺度残差网络结构图

Fig. 3 Architecture of multi-scale dense residual network

式中  $H_{SFE\ 3\times 3}(\cdot)$  表示用  $3\times 3$  像素的卷积核对低

分辨率图像进行卷积处理。同样,用  $5\times 5$  像素和  $7\times 7$  像素的卷积核对低分辨率图像进行卷积处理得到不同的特征图。

RDB 模块由卷积层和激活层组成的特征提取单元(FEU)多次重复串接组成,这样可以获得较深的网络层数。该模块引入 DCCN(densely connected convolutional network)<sup>[17]</sup> 思想,在保证深层结构的同时,为了确保网络中各层之间的最大信息流,网络中采用了残差连接的短路径方式。通过该方式将所有具有相同特征映射大小的特征图连接起来,每一层的输入都附加了前面所有层的输出,这种从前面层到后续层的短路径不仅可以缓解梯度消失问题,而且增强了特征传播,促进了特征再利用,保留了前向传播的特性。这种卷积层和激活层的输出附加到后面卷积层的输入的方法使得网络具有连续记忆的功能,每一层的状态得以连续传递。为了防止残差块中特征融合(FF)带来的参数增长,残差块中加入  $1\times 1$  像素的卷积层降低特征图的维度,减少参数,缓解训练难度。本文没有使用 He 等人<sup>[10]</sup> 提出的残差网络中批规范化层(BN),因为使用 BN 来归一化特征会使得网络参数大量增加,占用较多内存,不利于收敛,所以不使用 BN 层。残差块结构如图 4 所示。

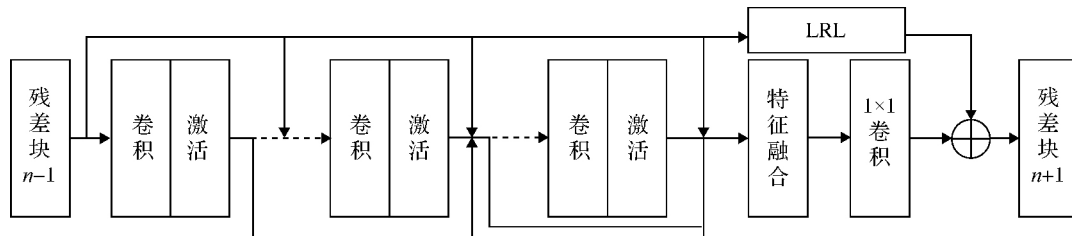


图4 残差块结构图

Fig. 4 Architecture of residual dense block

上一残差块的输出经过特征提取单元中的卷积层和激活层为

$$F_{FEU\ j} = \max(c(RDB_{n-1})\ \rho) \quad (3)$$

特征图经过  $d$  个特征提取单元的结果为

$$F_{FEU\ d} = F_{FEU\ d-1}(\cdots(F_{FEU\ j})\cdots) \quad (4)$$

为了充分利用特征信息和保持梯度存在的状态,在特征图经过残差块的同时,残差块的输出层加入了上一残差块的输出,以形成局部残差学习(LRL)

$$f_{RDB_{n+1}} = f_{RDB_{n-1}} + F_{1\times 1}(P_{FF}[F_{FEU\ j}, \cdots, F_{FEU\ m}]) \quad (5)$$

式中  $c$  表示卷积处理,  $m$  表示特征提取单元的数目,  $n$  表示残差块的数目。  $P_{FF}$  表示特征融合,  $F_{1\times 1}$  表示用  $1\times 1$  像素的卷积核卷积,对融合后的特征图进行降维处理。

LFF 模块是将每个残差块输出的特征图进行连接,形成密集特征融合。局部残差学习的使用能提高网络的表达能力,网络效果更佳

$$I_{LFF} = P_{FF}[RDB_1, \cdots, RDB_n] \quad (6)$$

GFF 模块在提取多层局部密集特征后,进一步进行全局特征融合。以全局方式自适应地保留分层特征。这里同样使用  $1\times 1$  像素的卷积层对融合后

的特征图进行降维处理,以减少网络参数。

GRL模块是在特征图进行重建之前,将输入图像经过底层卷积操作后得到的特征图再次融合到经过残差块提取的特征图中,形成全局残差学习。这样做有利于网络训练和梯度的保持。在本文中只将 $3 \times 3$ 像素卷积核输出的特征图附加到全局特征融合后的特征图,既可以补充细节信息,又可以提高训练效率

$$F_{\text{GRL}} = F_{3 \times 3}(F_{1 \times 1}(g_{\text{GFF}})) + F_{3 \times 3, \text{SFE}} \quad (7)$$

式中 $F_{3 \times 3, \text{SFE}}$ 表示对输入图像进行 $3 \times 3$ 像素卷积处理后提取的特征图, $g_{\text{GFF}}$ 表示全局特征融合后的输出特征图。

重建模块使用反卷积层对特征图进行扩充处理。反卷积有以下优势:1)加速超分辨率图像的重建过程。在网络的末端加入反卷积层后,使得整个卷积过程在LR空间完成,减少计算复杂度。2)大量的LR图像的纹理信息可以被用于推断出HR图像的细节信息。

### 3 实验结果及分析

#### 3.1 实验参数设置

实验所用计算机的CPU为InterCore i3-7350k, GPU为GeForce GTX1080, 16 GB内存, 操作系统为Ubuntu16.04, 深度学习框架为tensorflow, 加速库为CUDA Toolkit 8.0, OpenCV3.4。在学习率的设置上, 本文采用学习率衰减的方法防止训练时间过长。在本实验中学习率初始化为0.001, 每训练30个epoch, 学习率衰减一半。采用随机梯度下降法(SGD)训练网络。

#### 3.2 训练和测试数据集

本文采用的训练集为DIV2K数据集。该数据集由1000幅高清图像组成, 每幅图像的像素在200万左右。图像的细节部分清晰, 非常适合用来作为超分辨率训练。通过旋转、缩放等方式对数据进行增强, 使得数据扩充数倍, 可以有效避免过拟合。测试集使用Set5、Set14、BSD100、Urban100等标准图像超分辨率重建评价数据集。

#### 3.3 训练过程

首先对训练数据集中每幅图像按照步长为16进行截取, 得到若干个分辨率为 $128 \times 128$ 像素的原始HR图像块, 然后随机取64个图像块作为一个

batch。对每一个batch中的图像块做4倍的下采样, 得到 $32 \times 32$ 像素的LR图像。LR图像分别经过3个不同尺度的卷积核提取出图像的底层特征信息。在本文算法模型中, 残差块的数目设置为8个, 每个残差块中的特征提取单元的数目设置为5个。因为本模型的残差块中含有较多的卷积层, 所以卷积核的大小设置为 $3 \times 3$ 像素, 64通道, 这样可以有效避免大卷积核带来的模型参数增加、难以训练问题。为了缓解因为特征图融合带来的参数增加问题, 本文采用 $1 \times 1$ 像素, 64通道的卷积核对融合后的特征图进行降维处理。在训练过程中每一幅输入的LR图像都有其对应的HR图像来计算它们之间的损失值, 神经网络模型通过LR到HR这种端到端的映射关系学习到网络中的权值参数。本文算法使用的是L1损失函数, 而不是使用可以得到更大峰值信噪比值的L2损失函数。根据VDSR<sup>[9]</sup>实验效果, 使用L2损失函数使得最终的效果图过于平滑, 失真较严重, 与人的直观视觉不符。所以本文选用L1函数作为网络的损失函数

$$L(\theta) = \frac{1}{n} \sum_{i=1}^n \|F(Y_i; \theta) - X_i\|^2 \quad (8)$$

式中 $Y_i$ 表示LR,  $X_i$ 表示HR,  $\theta$ 是权值参数和偏置的集合,  $i$ 表示遍历整个训练集的图像的序号。

#### 3.4 实验结果

##### 3.4.1 实验评价标准

为了检测本文算法的实际效果, 分别对3个测试集进行实验, 并且与当前主流的图像超分辨率重建算法进行对比。实验结果采用通用的图像质量客观评价标准: 峰值信噪比(PSNR)和结构相似性(SSIM)。PSNR是基于对应像素点间的误差, 即误差敏感图像质量评价。PSNR的值越高表示图像质量越好。SSIM是衡量两幅图像相似度的指标, SSIM的值越接近1表示重建的图像越接近原分辨率图像。其中, PSNR计算方法为

$$f_{\text{PSNR}} = 10 \times \lg\left(\frac{(2^n - 1)^2}{f_{\text{MSE}}}\right) \quad (9)$$

式中 $f_{\text{MSE}}$ 为网络的损失函数,  $n$ 一般取8。

SSIM计算方法为

$$f_{\text{SSIM}}(X, Y) = l(X, Y) \cdot c(X, Y) \cdot s(X, Y) \quad (10)$$

$$l(X, Y) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1} \quad (11)$$

$$c(X,Y) = \frac{2\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2} \tag{12}$$

$$s(X,Y) = \frac{2\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3} \tag{13}$$

式中  $X$  为重建后的 HR 图像  $Y$  为原始的 HR 图像 ,  
 $\mu$  表示图像的均值 , $\sigma$  表示图像的方差。  $C_1$ 、 $C_2$ 、 $C_3$   
均为常数 ,用于避免分母为 0。通常

$$C_1 = (K_1 \cdot L)^2 \tag{14}$$

$$C_2 = (K_2 \cdot L)^2 \tag{15}$$

$$C_3 = \frac{C_2}{2} \tag{16}$$

一般地 ,取  $K_1 = 0.01$ 、 $K_2 = 0.03$ 、 $L = 255$ 。

3.4.2 实验结果分析与对比

实验部分首先将本文算法与插值法、SRC-  
NN<sup>[7]</sup>、DRCN<sup>[11]</sup>和 LapSRN<sup>[18]</sup>算法在 Set5 数据集的

单幅图像上进行测试 ,各效果图的 PSNR 值如表 1  
所示。表 2 测试的是包括本文算法在内的 5 种超分  
辨率算法在 Set5、Set14、Urban100 和 BSD100 4 个测  
试数据集上的 3 种放大倍数的平均 PSNR 和 SSIM  
值。综合表 1 和表 2 的实验数据 ,本文方法相比于  
当前主流图像超分辨率重建算法有更好的实验效  
果。本文算法使用的模型包括多个不同尺度的卷积  
核对低分辨率图像进行特征提取 这就表示有更多的  
信息被利用 ,而且本文引入密集型卷积网络和残差网  
络思想 密集型残差块中的短路径连接使提取到的特  
征信息不断被再利用 ,以保证图像的细节信息不被遗  
漏。残差网络的引入使得本文模型在加深网络的同  
时又保证梯度能够持续存在。综合以上 3 点优势 ,本  
文算法恢复的高分辨率图像效果要优于其他算法。

表 1 本文算法与其他算法在 Set5 数据集的 PSNR 值对比

Table 1 Comparison of PSNR values between our algorithm and other algorithms on Set5 dataset

						/dB
Set5	尺度	Bicubic <sup>[1]</sup>	SRCNN <sup>[7]</sup>	DRCN <sup>[11]</sup>	LapSRN <sup>[18]</sup>	本文算法
Baby	×2	36.91	39.30	40.41	41.43	41.58
Bird	×2	36.58	40.01	40.63	41.02	41.31
Butterfly	×2	29.37	34.21	35.74	33.60	34.32
Head	×2	33.88	35.54	35.49	35.61	35.90
Woman	×2	31.56	34.24	35.88	35.94	36.49

表 2 超分辨率算法在 4 个测试集上的 PSNR( /dB) /SSIM 平均值

Table 2 Average PSNR( /dB) and SSIM comparison of super-resolution algorithm on four datasets

数据集	尺度	Bicubic <sup>[1]</sup>	SRCNN <sup>[7]</sup>	DRCN <sup>[11]</sup>	LapSRN <sup>[18]</sup>	本文算法
Set5	×2	33.66/0.929 9	36.66/0.954 2	37.63/0.958 8	37.52/0.959 1	37.92/0.959 1
	×3	30.39/0.868 2	32.75/0.909 0	33.81/0.922 6	33.82/0.922 7	33.84/0.923 0
	×4	28.42/0.810 4	30.48/0.862 8	31.53/0.885 4	31.54/0.885 5	31.91/0.890 1
Set14	×2	30.24/0.868 8	32.45/0.906 7	33.04/0.911 8	33.08/0.913 0	32.98/0.916 5
	×3	27.55/0.774 2	29.30/0.821 5	29.76/0.831 1	29.79/0.832 0	29.90/0.834 7
	×4	26.00/0.702 7	27.50/0.751 3	28.02/0.767 0	28.19/0.772 0	28.69/0.775 3
Urban100	×2	26.88/0.840 3	29.50/0.894 6	30.75/0.913 3	30.41/0.910 1	31.57/0.923 4
	×3	24.46/0.734 9	26.24/0.798 9	27.15/0.827 6	27.07/0.827 2	27.83/0.835 6
	×4	23.14/0.657 7	24.52/0.722 1	25.14/0.751 0	25.21/0.755 3	26.39/0.795 0
BSD100	×2	29.56/0.843 1	31.36/0.887 9	31.85/0.894 2	31.80/0.895 0	32.26/0.901 1
	×3	27.21/0.738 5	28.41/0.786 3	28.80/0.796 3	28.82/0.797 3	29.05/0.800 3
	×4	25.96/0.667 5	26.90/0.710 1	27.23/0.723 3	27.32/0.728 0	27.58/0.733 8



在本文的网络模型中是将  $3 \times 3$  像素卷积核提取的底层特征图用于 GRL。在表 3 中分别将 3 种尺度底层特征提取卷积核提取的特征图附加在重建模块前,形成 3 种情况下的 GRL。将 Set5 数据集中的 5 幅图像分别进行 4 倍下采样,得到 5 幅低分辨率图像。再将 5 幅低分辨率图像分别输入多尺度密集残差网络,得到效果图,分别测得它们在不同尺度卷积核的 PSNR 值如表 3 所示。从表 3 中可以看出  $3 \times 3$  像素的底层特征提取卷积核提取的特征图附加在重建模块前得到的 PSNR 值最高,相比于其他两种卷积核  $3 \times 3$  像素的卷积核能够提取较多的细节信息,所以得到的实验效果更好。

为了体现多尺度密集残差网络算法的有效性,将本文算法与插值法、SRCNN<sup>[7]</sup>、DRCN<sup>[11]</sup> 和 LapSRN<sup>[18]</sup> 图像超分辨率算法进行对比。在实验部分,分别在 Set5、Set14、BSD100 和 Urban100 测试数

据集中各随机抽取一幅图像,将它们都进行 3 倍下采样处理,再输入到各个网络模型中。实验效果对比如图 5 所示。通过观察图 5,本文算法相比于插值法和 SRCNN 算法,恢复的图像中边缘和纹理细节更加清晰,和原图相比,本文算法的效果图在锐度上稍强一点。

表 3 不同卷积核 Set5 数据集的 PSNR 值  
Tab. 3 PSNR comparison of different convolutional kernel scale on Set5 dataset

Set5				/dB
	$3 \times 3$	$5 \times 5$	$7 \times 7$	
Baby	31.35	30.61	29.68	
Bird	32.68	31.75	30.53	
Butterfly	34.21	32.86	30.87	
Face	30.43	29.73	28.49	
Women	30.89	30.25	28.97	

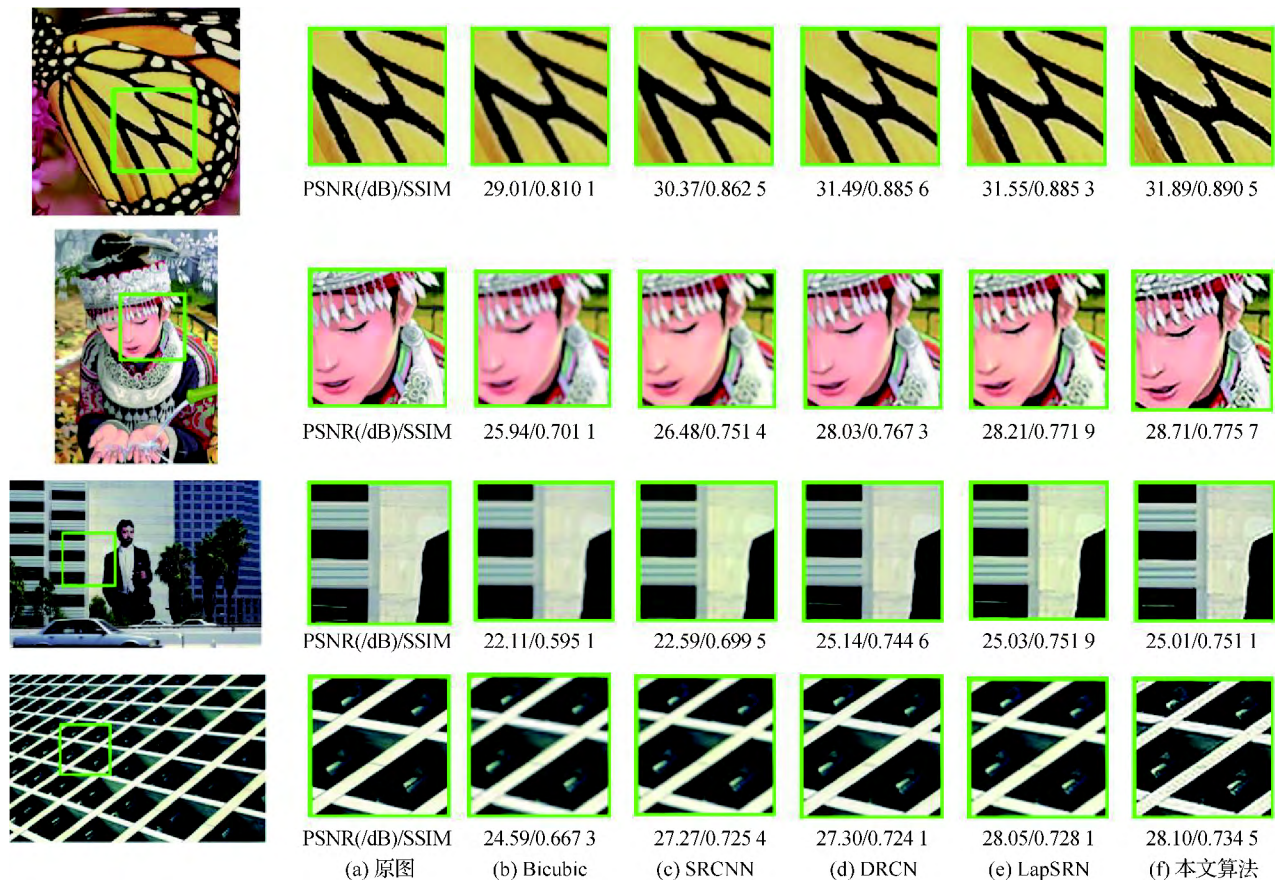


图 5 本文算法与其他算法重建效果对比

Fig. 5 Qualitative comparison of our method with other methods

(( a ) original images; ( b ) Bicubic; ( c ) SRCNN; ( d ) DRCN; ( e ) LapSRN; ( f ) our method)



## 4 结 论

提出一种多尺度密集型残差网络的图像超分辨率重建算法以解决当前主流算法中对于低分辨率图像中的细节信息利用不够充分的问题。本文模型使用多种尺度的卷积核提取 LR 图像的底层特征信息。另外,引入密集型卷积网络思想,每个残差块之间使用短路径相连,这样不仅可以保证每一层的特征信息不丢失,而且还可以使得网络梯度一直保持存在的状态,不会发生梯度消失现象。

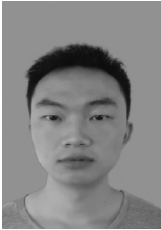
通过最终的实验结果和数据对比,可以看出本文算法可以从低分辨率图像中恢复出较多细节信息。在以后的优化实验中可以尝试增加残差块的数目来改进实验效果。在网络的参数方面,可以通过分离卷积核的方式来降低权值参数的数目以减少训练时间。以上两点将是接下来的改进方向。

## 参考文献(References)

- [ 1 ] Zhang L , Wu X L. An edge-guided image interpolation algorithm via directional filtering and data fusion[J]. IEEE Transactions on Image Processing , 2006 , 15( 8 ) : 2226-2238. [DOI: 10. 1109/ TIP. 2006. 877407]
- [ 2 ] Zhang K B , Gao X B , Tao D C , et al. Single image super-resolution with non-local means and steering kernel regression [J]. IEEE Transactions on Image Processing , 2012 , 21( 11 ) : 4544-4556. [DOI: 10. 1109/TIP. 2012. 2208977]
- [ 3 ] Timofte R , De V , Van Gool L. Anchored neighborhood regression for fast example-based super-resolution [C]//Proceedings of 2013 IEEE International Conference on Computer Vision. Sydney , NSW , Australia: IEEE , 2013: 1920-1927. [DOI: 10. 1109/ICCV. 2013. 241]
- [ 4 ] Timofte R , De Smet V , Van Gool L. A + : adjusted anchored neighborhood regression for fast super-resolution [C]//Proceedings of 12th Asian Conference on Computer Vision. Singapore: Springer , 2014: 111-126. [DOI: 10. 1007/978-3-319-16817-3\_8]
- [ 5 ] Peleg T , Elad M. A statistical prediction model based on sparse representations for single image super-resolution [J]. IEEE Transactions on Image Processing , 2014 , 23( 6 ) : 2569-2582. [DOI: 10. 1109/TIP. 2014. 2305844]
- [ 6 ] Hinton G E , Salakhutdinov R R. Reducing the Dimensionality of Data with Neural Networks [J]. ScienceNew Series , 2006 , 313( 5786 ) : 504-507. [DOI: 10. 1126/science. 1127647]
- [ 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. [DOI: 10. 1109/TPAMI. 2015. 2439281]
- [ 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]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas , NV , USA: IEEE , 2016: 1874-1883. [DOI: 10. 1109/CVPR. 2016. 207]
- [ 9 ] Kim J , Lee J K , Lee K M. Accurate image super-resolution using very deep convolutional networks [C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas , NV , USA: IEEE , 2016: 1646-1654. [DOI: 10. 1109/CVPR. 2016. 182]
- [ 10 ] He K M , Zhang X Y , Ren S Q , et al. Deep residual learning for image recognition [C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas , NV , USA: IEEE , 2016: 770-778. [DOI: 10. 1109/CVPR. 2016. 90]
- [ 11 ] Kim J , Lee J K , Lee K M. Deeply-recursive convolutional network for image super-resolution [C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas , NV , USA: IEEE , 2016: 1637-1645. [DOI: 10. 1109/CVPR. 2016. 181]
- [ 12 ] Tai Y , Yang J , Liu X M. Image super-resolution via deep recursive residual network [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu , Hawaii , USA: IEEE Computer Society , 2017: 2790-2798. [DOI: 10. 1109/CVPR. 2017. 298]
- [ 13 ] Dong C , Loy C C , Tang X O. Accelerating the super-resolution convolutional neural network [C]//Proceedings of 14th European Conference on Computer Vision. Amsterdam , The Netherlands: Springer , 2016: 391-407. [DOI: 10. 1007/978-3-319-46475-6\_25]
- [ 14 ] Tong T , Li G , Liu X J , et al. Image super-resolution using dense skip connections [C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice , Italy: IEEE Computer Society , 2018: 4809-4817. [DOI: 10. 1109/ICCV. 2017. 514]
- [ 15 ] Lim B , Son S , Kim H , et al. Enhanced deep residual networks for single image super-resolution [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu , HI , USA: IEEE , 2017: 1132-1140.
- [ 16 ] Szegedy C , Liu W , Jia Y Q , et al. Going deeper with convolutions [C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston , MA , USA: IEEE , 2015: 1-9.
- [ 17 ] Huang G , Liu Z , Van Der Maaten L , et al. Densely connected convolutional networks [C]//Proceedings of 2017 IEEE Confer-

ence on Computer Vision and Pattern Recognition. Honolulu , Hawaii , USA: IEEE Computer Society , 2017: 2261-2269. [DOI: 10.1109/CVPR.2017.243]

- [18] Lai W S , Huang J B , Ahuja N , et al. Deep laplacian pyramid networks for fast and accurate super-resolution [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu , Hawaii , USA: IEEE Computer Society , 2017: 5835-5843. [DOI: 10.1109/CVPR.2017.618]



龙祥 通信作者 ,男 ,硕士研究生 ,主要研究方向为生物特征识别和图像超分辨率重建。  
E-mail: xianglong946@163. com

# 作者简介



应自炉 ,1966 年生 ,男 ,教授 ,主要研究方向为生物特征识别、人机自然交互。  
E-mail: ziluy@163. com