

利用多尺度卷积神经网络的图像超分辨率算法

陈书贞 解小会 杨郁池 练秋生
(燕山大学信息科学与工程学院, 河北秦皇岛 066004)

摘 要: 单幅图像超分辨率问题是典型的图像反问题。近年来深度学习广泛应用于图像超分辨率重建。为提高超分辨率算法的性能, 本文利用多尺度和残差训练的思想, 提出一种利用多尺度卷积神经网络的图像超分辨率算法。该算法采用多尺度的卷积核及收缩-扩展的网络结构来提取图像多尺度的信息, 并在网络结构中使用跳跃连接, 以便更好的传递信息并弥补由于使用下采样和上采样而造成的图像细节信息的损失, 来提高图像的重建质量。通过与其他算法的对比实验表明了本文算法不仅可以取得更好的性能, 并且训练的收敛速度较快。

关键词: 超分辨率; 深度学习; 多尺度卷积核; 残差训练; 跳跃连接

中图分类号: TP391.41 **文献标识码:** A **DOI:** 10.16798/j.issn.1003-0530.2018.09.003

Image Super-Resolution Algorithm Based on Multi-Scale Convolution Neural Network

CHEN Shu-zhen XIE Xiao-hui YANG Yu-chi LIAN Qiu-sheng
(School of Information Science and Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China)

Abstract: The single image super-resolution problem is a typical image inverse problem. In recent years, deep learning is widely used in image super-resolution. In order to improve the performance of super-resolution algorithms, in this paper, the ideas of multi-scale and residual training are utilized. An image super-resolution algorithm that exploits the multi-scale convolution neural network is proposed. This algorithm exploits the multi-scale convolution kernels and the shrinkage-extension structure to extract image multi-scale information. Skip connection is used in the network structure to improve the quality of image reconstruction, which can transmit information effectively. Moreover, it can compensate for the loss of image details resulting from the use of down-sampling and up-sampling. Compared with other algorithms, the experiments shows that our algorithm can not only achieve better performance, but also has the faster convergence speed.

Key words: super-resolution; deep learning; multi-scale convolution kernel; residual training; skip connection

1 引言

图像超分辨率算法的目标是通过低分辨率图像来重建高分辨率图像。图像超分辨率在图像处理和计算机视觉中有着很重要且广泛的应用, 在医学诊断^[1]、卫星遥感^[2]等领域都有着非常重要的应用。图像为人们提供了可视化的直观信息, 其质量的好坏直接影响识别、检测及分类等操作^[3]。图像的分辨率是评价图像质量的重要指标, 主要反映了

图像所含信息量的多少^[4]。图像的分辨率越高, 所含信息量就越多, 视觉效果也越好, 它对于视频监控、医学诊断等有着极其重要的作用。

图像超分辨率重建由 Harris 等人在 1964 年首次提出^[5], 之后有许多学者对其进行了深入研究。由于图像超分辨率重建是由低分辨率图像来重建高分辨率图像, 所以超分辨率重建是图像处理中典型的病态反问题^[6]。单幅图像超分辨率算法大体上可以分为两大类^[7]: 一类是基于插值的方法, 一

收稿日期: 2018-03-07; 修回日期: 2018-05-31

基金项目: 国家自然科学基金(61471313); 河北省自然科学基金(F2014203076)

类是基于学习的方法。基于插值的超分辨率算法^[8-9]比较简单,容易实现,但是对于重建质量较高的图像纹理和细节信息存在一定的困难。目前主流的基于学习的方法大致有三大类:基于稀疏表示的学习算法^[10-11]、基于近邻嵌入的学习算法^[12-14]和基于深度学习的算法^[15-20]。基于稀疏表示的方法是通过学习高、低分辨率图像块的完备字典来重建高分辨率图像。基于近邻嵌入的学习是利用图像特征空间上的结构相似性,对低分辨率图像块及其在特征空间上最近邻的图像块进行学习,以此来重建对应的高分辨率图像。近年来基于深度学习的图像超分辨率问题受到大家越来越多的关注和研究,尤其是卷积神经网络应用于图像超分辨率重建取得了很大的成功。卷积神经网络通过卷积运算能更好的利用一个像素点的周围像素来获取此像素点的信息,而且卷积神经网络可以处理比较复杂的映射关系并对数据进行并行处理。卷积神经网络的学习其本质是寻找输入与输出之间复杂的非线性映射关系;其应用于图像超分辨率重建,则是通过学习低分辨率图像块与高分辨率图像块之间端到端的映射关系来重建高分辨率的图像。

基于卷积神经网络的图像超分辨率算法发展比较迅速。最早用卷积神经网络实现图像超分辨率的是 Dong Chao 等人,其提出了 SRCNN^[15]的方法。该方法采用三层卷积网络来学习输入低分辨率图像块和输出高分辨率图像块之间的映射关系,三层卷积分别用于特征提取、非线性映射和重建。SRCNN 的优点是结构简单,相比较于插值、稀疏表示以及近邻嵌入方法可以取得更好的性能,缺点是由于训练的困难即使加深网络也无法提高图像的重建质量^[16]。随后,为了提高训练速度,Dong Chao 等人又提出了 FSRCNN^[17]的方法。FSRCNN 重新设计了 SRCNN 的结构,该网络直接输入原始的低分辨率图像,并在网络最后使用转置卷积层把图像放大到需要的大小。FSRCNN 算法无需图像的插值预处理,大大减少了计算的复杂度;而且在网络的映射部分选用尺寸更小的滤波器以及更多的映射层,为了降低网络的计算量在映射前缩小输入特征维数,映射后再扩大回来,以免影响图像的重建质量。经过这样的修改不仅加快了训练速度而且提高了

性能。由于深层网络模型在图像分类问题上取得很大突破,受此启发,Kim 等人提出了一个深层的卷积神经网络模型 VDSR^[18]。VDSR 采用 20 层卷积网络,利用残差训练,不再简单的学习低分辨率图像与高分辨率图像之间的映射,而是通过网络来学习高分辨率图像和低分辨率图像的残差图像。VDSR 方法提出的残差训练和梯度裁剪方法不仅解决了深层网络容易出现的梯度消失和爆炸的问题,而且 Kim 的实验^[18-19]也表明利用残差训练可以加快网络的收敛速度并提高图像的重建质量。虽然 VDSR 算法取得了较好的性能,但是其网络结构是卷积层的堆叠,网络结构越深,梯度消失现象就会越明显。

在 CVPR2017 中,Lai 等人提出 LapSRN 算法来快速精确实现超分辨率重建^[20],该算法采用深度拉普拉斯金字塔网络来逐级重建高分辨率图像。Tai 等人提出的 DRRN 算法^[21],用一个深度回归残差网络来重建高分辨率图像,可以有效提高算法的重建性能。U-net 是一种很好地用于图像语义分割的网络结构,其结构包括一个收缩支路来提取信息和一个对称的扩展支路来重建信息^[22]。受此启发,为提高超分辨率算法的性能,本文设计了一种利用多尺度卷积核的收缩-扩展残差网络。本文提出的算法有三大优点:本文算法利用多尺度卷积核的滤波器并且采用收缩-扩展的网络结构来提取不同尺度的图像信息,以此来提高重建图像的质量;利用网络结构的跳跃连接更有利地传播信息以及残差训练的方法来提高性能;损失函数不再使用 l_2 范数损失函数,而是采用 l_1 范数损失函数,使得性能有所提高。实验结果表明本文算法不仅可以较好的提高图像的重建质量,而且训练速度也相对较快。

2 网络结构

本文的网络结构如图 1 所示,该网络在整体上可以看做是一个具有收缩-扩展结构的残差网络。前一部分(左侧)是网络的收缩部分,采用两次下采样把训练图像缩放两次,即缩减为输入图像的四分之一;网络的后一部分(右侧)是与收缩部分完全对称的扩展部分,对应的使用两次上采样把缩小的图像扩大回原来的大小。

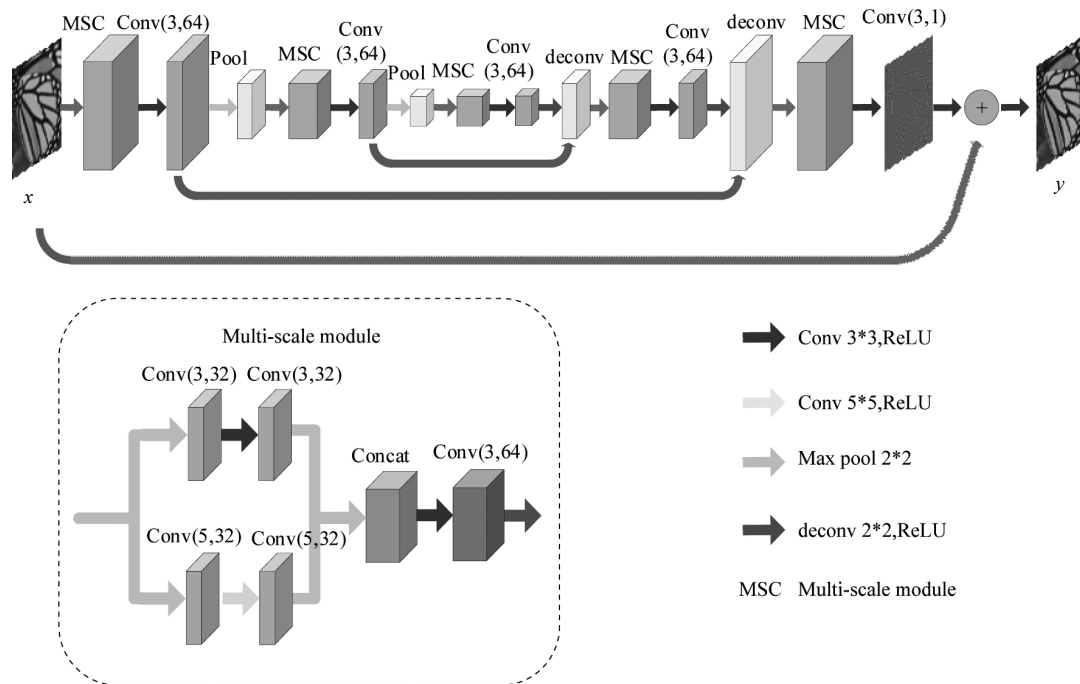


图1 多尺度卷积神经网络的网络结构图

Fig.1 Network structure of multi-scale convolution neural network

本文的网络结构包含 20 层卷积层、2 层下采样和 2 层上采样。 x 和 y 分别是网络的输入和输出, $\text{conv}(k, n)$ 中 k 为卷积核大小, n 为卷积核的个数。输入层首先连接的是一个多尺度模块, 该网络结构中共包含 5 个多尺度模块:

$$y_m = \Phi(w_m^{(3)} F(x_m; \theta) + b_m^{(3)}) \quad m = 1, \dots, 5 \quad (1)$$

其中 x_m 和 y_m 分别表示第 m 个多尺度模块的输入和输出, 且 $x_1 = x$, $m = 1$ 。每个多尺度模块由三个卷积层组成, 每个卷积层卷积核的个数为 64。 Φ 表示的是激活函数, 本网络选用的激活函数为 ReLU^[23]。 $w_m^{(3)}$ 和 $b_m^{(3)}$ 分别是第 m 个多尺度模块中第三层卷积的权重和偏置。 $F(x_m; \theta)$ 为多尺度模块中前两层多尺度卷积层对应的非线性映射, 是两个尺度非线性映射的拼接 (concat):

$$F(x_m; \theta) = \text{Concat}(f_1(x_m; \theta_1), f_2(x_m; \theta_2)) \quad (2)$$

其中 $\theta = \{\theta_1, \theta_2\}$ 表示两层多尺度卷积层的权重和偏置。 $f_i(x_m; \theta_i)$, $i = 1, 2$ 为每个多尺度卷积层对应的非线性映射的输出, $\theta_i = \{w_m^{(1i)}, w_m^{(2i)}, b_m^{(1i)}, b_m^{(2i)}\}$ 为每个尺度卷积层对应的滤波器的权重和偏置。多尺度卷积层采用两个尺度的卷积核, 即 64 个卷积核中有 32 个卷积核大小为 3×3 , 32 个卷积核为 5×5 , 每个尺度的非线性映射都有两层卷积组成:

$$f_i(x_m; \theta_i) = \Phi(w_m^{(2i)} \Phi(w_m^{(1i)} x_m + b_m^{(1i)}) + b_m^{(2i)}),$$

$$i = 1, 2 \quad (3)$$

其中 $w_m^{(ji)}$ 和 $b_m^{(ji)}$, $i = 1, 2$; $j = 1, 2$ 表示第 m 个多尺度模块中第 i 个尺度非线性映射中第 j 层卷积的权重和偏置。

在网络的收缩部分, 每个多尺度模块之后都连接一个卷积层和一个下采样层。卷积层采用的卷积核是 3×3 , 卷积核的个数为 64。下采样层采用的是最大池化, 池化操作的区域为 2×2 , 步幅为 2, 最大池化即找出池化区域中的最大值。

在网络的扩展部分, 则是在转置卷积层之后连接一个多尺度模块和一个卷积核为 3×3 的卷积层, 这两个多尺度模块之后连接的两个卷积层的卷积核个数分别是 64 和 1。转置卷积层选用的卷积核为 2×2 , 卷积核个数为 64, 该层的作用即为上采样。

除最后一层卷积层和下采样层之外, 每一层后都使用一个 ReLU 层。此外, 为了使网络能更高效的传递信息, 在每个下采样层之前的卷积层和上采样层之间都使用跳跃连接, 跳跃连接指的是这两层作和, 而卷积核个数不变。本文网络结构的优点是采用多尺度的卷积核和收缩-扩展的网络结构来提取图像多尺度的信息, 且采用跳跃连接和残差训练来提高图像的重建质量。

本文在网络的多尺度模块采用多尺度卷积核

的滤波器,即在同一个卷积层上应用不同尺寸的滤波器。采用多尺度的卷积核有两大优点,首先,多尺度的卷积核最大的优点是不同尺寸的卷积核可以提取图像不同尺度的特征,以便滤波器提取和学习更丰富的图像信息;其次,卷积神经网络训练模型时,是通过学习滤波器的参数(权重和偏置)来实现的,即不断地学习滤波器的参数,使其达到一个最贴近标签的最优值;本文采用多尺度的卷积核,目的是让某一卷积层具有多样性的滤波器,从而使权重和偏置的学习更加多样性,进而可以充分有效的提取并学习图像的有用信息。

本文采用的是完全对称的收缩-扩展网络结构,即先把图像缩小,共缩小两次,图像大小变为输入图像的四分之一,为了便于传播信息和重建图像再对称的扩大回原来的大小。收缩-扩展的网络结构中,网络的收缩部分用于学习和提取同一图像不同尺度的特征,可学习同一图像不同分辨率下的信息;扩展部分则用来合并和重建图像信息。利用多尺度的图像信息,使得训练图像的信息更加丰富,进而促进网络更高效的学习。不仅如此,该网络结构因为有对图像的收缩处理,会降低计算的复杂度,提高训练和测试的速度。

文中下采样层和上采样层分别用于缩小和扩大图像,下采样层可采用 2×2 的最大池化,或是步幅为 2 的卷积。实验结果表明采用步幅为 2 的卷积层与池化效果相差不多,但由于池化运算参数较少,训练速度更快,所以本文采用最大池化来进行下采样。上采样采用转置卷积层实现,转置卷积层是通过一系列转置卷积滤波器来上采样图像到相应的大小。

网络的跳跃连接体现在收缩部分每个下采样层之前的卷积层都会与扩展部分相应的转置卷积层相连。采用跳跃连接有利于图像信息在网络结构中的传播,在深层网络结构中,由于卷积层数增多、使用池化层或转置卷积层,会损失或退化图像的细节信息^[24]。本文采用对称的收缩-扩展结构,在图像的收缩(池化)和扩展(转置卷积)过程中均会损失一些细节信息。在这种情况下使用跳跃连接,连接前面的卷积层和后面对应的转置卷积层,可使更多的图像信息直接传播到后面的网络结构中,以此来弥补这些信息的损失。跳跃连接还有残差学习的作用,可以直接把前面的特征映射到后面,增加网络后层的特征输入,以此便于网络高效的学习和训练。此外,深层网络结

构在实际训练过程中常常遇到的一个困难就是梯度消失的问题,而跳跃连接在网络的反向传播过程中可以使梯度更好的向底层传播,可以有效地解决这一问题,使网络训练更容易。

多数用于超分辨率重建的网络都是直接由低分辨率图像块训练高分辨率图像块,而残差训练则是训练高分辨率图像与低分辨率图像的差 y , $y = Y - X$ 。其中 Y 和 X 分别是高分辨率图像和低分辨率图像。如图 1 所示,本文网络的输出为:

$$y = x + f_{\text{res}}(x) \quad (4)$$

其中 $f_{\text{res}}(x)$ 为网络的残差输出。使用残差训练不仅可以很好的避免深层网络易出现的梯度消失的问题,而且还可以加快训练的收敛速度并提高重建图像的质量。

在数多的网络结构中,训练的损失函数为 l_2 范数损失函数:

$$\text{Loss}(\Theta) = \frac{1}{2n} \sum_{i=1}^n \|y_i - f(x_i; \Theta)\|_2^2 \quad (5)$$

其中 n 为样本数, y_i 是高分辨率图像, $f(x_i; \Theta)$ 为网络的预测输出, $\Theta = \{w_1, w_2, \dots; b_1, b_2, \dots\}$ 。本文算法训练的损失函数选用 l_1 范数损失函数:

$$\text{Loss}(\Theta) = \frac{1}{2n} \sum_{i=1}^n \|y_i - f(x_i; \Theta)\|_1 \quad (6)$$

3 训练和测试

本文的训练实验是通过 Caffe^[25] 来实现,图形处理器(GPU)使用的是 GTX 1080。训练图像共 291 张,这 291 张图像与 VDSR 算法所用的训练图像完全一致,其中 91 张图像源自论文 Image super-resolution via sparse representation^[15],其余 200 张源自 Berkeley Segmentation Dataset^[26]。实际的训练图像还进行了数据增强处理,即把这 291 张图像进行旋转和缩放,经过这一数据增强的处理,原来的 291 张图像就变为了 5820 张图像。测试所用数据集为 'Set5' 和 'Set14',是两个用于测试的基准数据集。

使用网络结构进行训练前,要先对输入图像做预处理,选用的处理方法是与 SRCNN 方法相同的双三次插值法。使用卷积神经网络训练模型时,为了能高效快速的进行训练,一般对图像进行分块处理,因为批量处理更方便计算而且分块处理可得到更多数量的样本。本文实验把图像分成 40×40 的

图像块作为训练样本,取块步幅为 30,训练阶段批量处理的大小为 64,一共训练 30 轮,用 GTX 1080 训练需要 270 分钟。训练所采用的优化算法是 Adam^[27],Adam 是一种自适应时刻估计,相比较于 SGD,Adam 优化方法更灵活,Adam 利用梯度的一阶矩估计和二阶矩估计来动态的调整每个参数的学习率,每一次迭代更新后都会把学习率控制在一定范围内,使得参数的学习更加稳定。Adam 有两个动量参数,本文实验采用的两个参数分别是 0.9 和 0.999,学习率设置为 0.0001, l_2 范数正则化的权重衰减率设为 0.0001,采用 l_2 范数正则化的作用主要是为了防止网络的过拟合。

4 实验结果

4.1 与其他算法比较

为评价本文算法的性能,本文算法分别与 SRCNN^[15]、FSRCNN^[17]、VDSR^[18]、LapSRN^[20] 以及 DRRN^[21] 五种算法进行了对比实验。主要用于评价实验结果的指标为峰值信噪比(PSNR)和结构相似度(SSIM)。

实验结果如表 1、表 2 所示。分别给出了 SRCNN、FSRCNN、VDSR、LapSRN 算法与本文算法在 PSNR、SSIM 上的实验结果对比。本文所有对比实验的数据均来自各个论文中作者给出的代码和训练模型。如表 1 所示,通过比较五种算法在两个不同测试集(Set5 和 Set14)的平均 PSNR 可以看出,在不同放

大因子($\times 2$ 、 $\times 3$ 、 $\times 4$)下,本文所提出的算法相比较于 SRCNN 算法和 FSRCNN 算法 PSNR 有明显的提高。尤其与性能相对较好的 VDSR 算法相比,在 Set5 数据集上,不同放大因子($\times 2$ 、 $\times 3$ 、 $\times 4$)下本文算法的平均 PSNR 分别高出 VDSR 算法 0.27 dB、0.27 dB 和 0.28 dB,而在含有较多细节信息图像的 Set14 数据集上优势相对较小,不同放大因子($\times 2$ 、 $\times 3$ 、 $\times 4$)下分别比 VDSR 算法高 0.15 dB、0.02 dB 和 0.10 dB。由表 1 也可以看出在放大因子为 2 和 4 时,本文算法相比较于 LapSRN 算法在 Set5 数据集上的平均 PSNR 分别提高 0.34 dB 和 0.10 dB,在 Set14 中分别提高 0.25 dB 和 0.04 dB。并且,从表 2 可以看出,本文提出的算法在结构相似度(SSIM)上也有一定程度的提高。实验结果表明,本文所提出的算法在图像超分辨率重建问题中相比较于 SRCNN 算法、FSRCNN 算法、VDSR 算法以及 LapSRN 算法在性能上取得了较大的提高。

如表 3 所示,本文算法与 VDSR 算法在测试时间上进行了对比,本文采用与 VDSR 完全相同的训练集和测试集,测试平台为 GPU GTX1080 Matlab 2015b。VDSR 算法采用的是 20 层卷积网络的残差训练,而本文的网络结构一共 25 层,虽然本文的网络结构层数比 VDSR 多,由于本文采用的是收缩—扩展的网络结构,即在测试时先对图像进行收缩,然后再扩展,在图像与卷积核进行卷积运算时映射变小,有效减小了计算量,所以在运行时间上与 VDSR

表 1 五种超分辨率算法在 Set5 和 Set14 数据集上的 PSNR(dB) 对比

Tab.1 The five super-resolution algorithms are compared with PSNR(dB) on Set5 and Set14 data sets

Dataset	Scale	SRCNN	FSRCNN	VDSR	LapSRN	OURS
Set5	$\times 2$	36.66	37.00	37.53	37.46	37.80
	$\times 3$	32.74	33.15	33.64	—	33.91
	$\times 4$	30.48	30.70	31.34	31.52	31.62
Set14	$\times 2$	32.45	32.63	33.06	32.96	33.21
	$\times 3$	29.28	29.42	29.80	—	29.82
	$\times 4$	27.50	27.59	28.02	28.08	28.12

表 2 五种超分辨率算法在 Set5 和 Set14 数据集上的 SSIM 对比

Tab.2 The five super-resolution algorithms are compared with SSIM on Set5 and Set14 data sets

Dataset	Scale	SRCNN	FSRCNN	VDSR	LapSRN	OURS
Set5	$\times 2$	0.9541	0.9557	0.9588	0.9584	0.9597
	$\times 3$	0.9087	0.9139	0.9214	—	0.9235
	$\times 4$	0.8618	0.8646	0.8823	0.8843	0.8862
Set14	$\times 2$	0.9066	0.9086	0.9129	0.9119	0.9137
	$\times 3$	0.8218	0.8245	0.8326	—	0.8328
	$\times 4$	0.7518	0.7539	0.7678	0.7694	0.7702

算法相当。如表 3 所示,在 Set5 数据集上放大因子为 2 和 3 时以及 Set14 数据集上放大因子为 4 时,本文算法比 VDSR 算法快 0.01 s。

本文实验在 Intel Core i5-4460 CPU 和 GTX 1080 的 GPU 操作平台上进行,除与上述几种算法对比外,本文利用 Set5 数据集中的四张测试图像与 DRRN 算法对比了图像的重建结果。四张图像如图 2 所示,记为 Set4 数据集。本文算法与 DRRN 算法的对比结果如表 4 所示。对比实验中选用了 DRRN 算法在 1 个回归模块、9 个残差单元共包含 20 层卷积情况下的实验结果。由表 4 可知,在放大因子为 2 和 4 时,本文算法的性能要高于 DRRN 算法,但在 3 倍放大因子下本文算法的性能略低于 DRRN 算法。

DRRN 算法中每一个卷积层有 128 个卷积核,而本文算法的卷积层采用 64 个卷积核,并且本文在网络中采用收缩-扩展的网络结构,网络参数大大减少,所以从运行时间上来看,本文算法速度较快。由表 4 可知,本文算法的运行时间是 DRRN 算法的四分之一。

除了 PSNR、SSIM 以及测试时间等评价指标外,本文还选取了 Set5 和 Set14 两个数据集中比较有代表性的三张图像,将这几张图像显示出来进行对比,结果如图 3、图 4 和图 5 所示,三张图像的大小分别为:336×220,380×576,648×520。从图中可以看出,在放大因子为 4 时,本文算法的重建图像细节信息更丰富,相比较于其他三种算法图像明显更清晰。

表 3 两种超分辨率算法在 Set5 和 Set14 数据集上的测试时间对比(s)

Tab.3 Comparison of test time between the two super-resolution algorithms on Set5 and Set14 data sets (s)

Dataset	Scale	VDSR	OURS
Set5	×2	0.04	0.03
	×3	0.03	0.03
	×4	0.04	0.03
Set14	×2	0.06	0.06
	×3	0.07	0.06
	×4	0.06	0.06

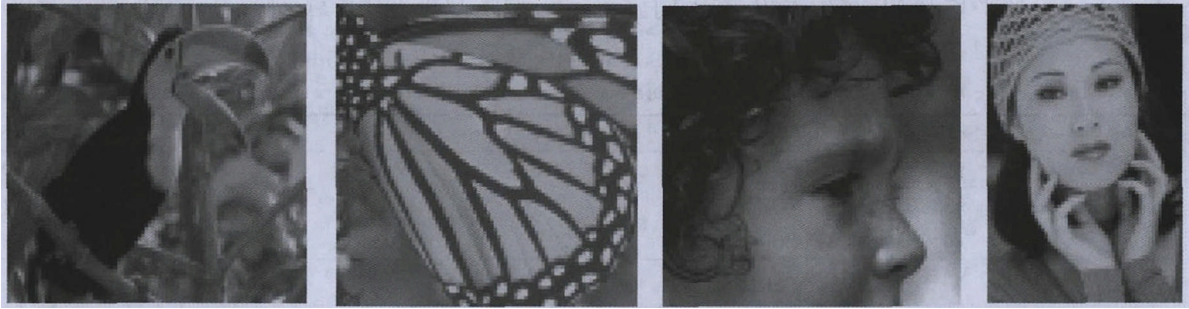


图 2 Set4 数据集的四张图像

Fig.2 Four images in Set4 dataset

表 4 两种超分辨率算法在 Set4 数据集上的 PSNR(dB) /SSIM/运行时间(s) 的对比

Tab.4 The two super-resolution algorithms are compared with PSNR(dB) /SSIM/test time(s) on Set4 data sets

Dataset	Scale	DRRN(B1U9)	OURS
Set4	×2	37.39/0.9570/0.12	37.53/0.9577/0.03
	×3	33.56/0.9227/0.12	33.52/0.9224/ 0.03
	×4	31.06/0.8835/0.12	31.14/0.8850/0.03



图 3 Set5 中 woman_GT 图像在放大因子为 4 时的视觉对比

Fig.3 Visual comparison of the woman_GT image in Set5 with the scale of 4

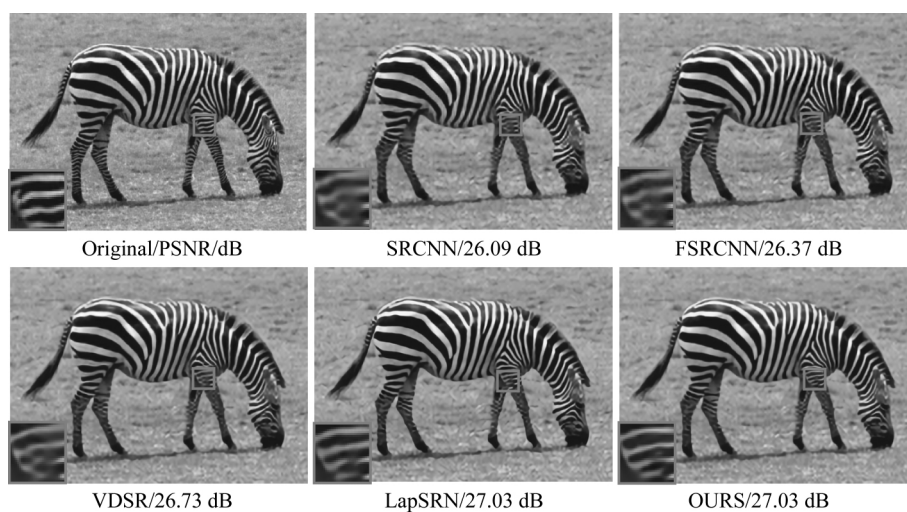


图 4 Set14 中 zebra 图像在放大因子为 4 时的视觉对比

Fig.4 Visual comparison of the zebra image in Set14 with the scale of 4

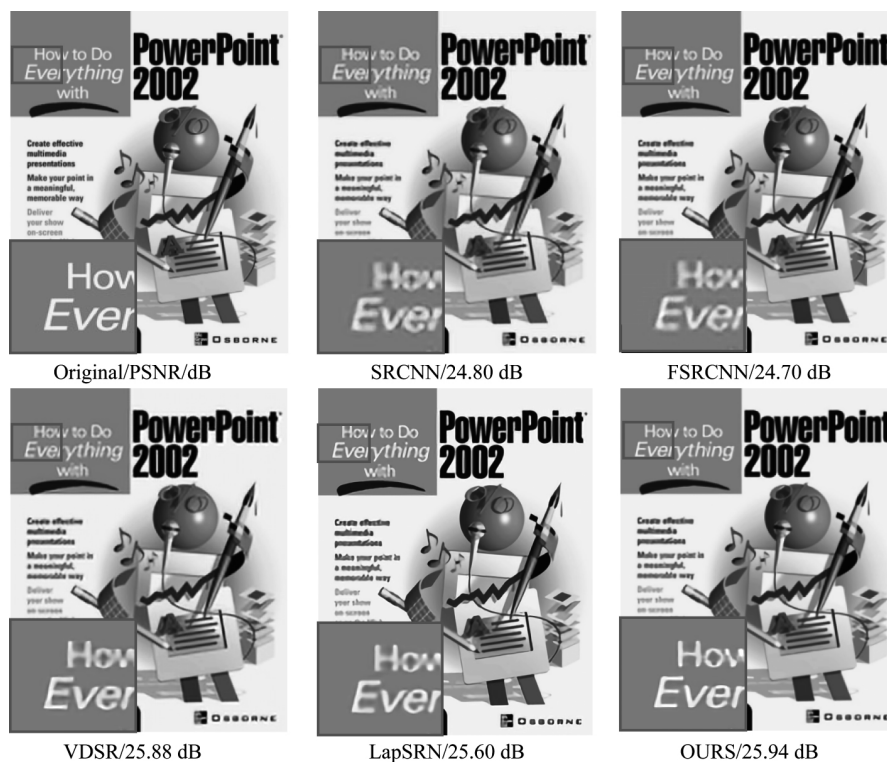


图5 Set14 中 ppt3 图像在放大因子为 4 时的视觉对比

Fig.5 Visual comparison of the ppt3 image in Set14 with the scale of 4

除了上述标准测试图像外,本文还增加了医学图像和遥感图像进行测试,分别选取了一张医学图像和一张遥感图像进行了视觉对比,如图6、图7所示。图像大小分别是 248×248 、 248×248 。其中,图7的遥感图像选自 NWPU-RESISC45 数据库。从图6、

图7可以看出本文算法相比较于其他三种算法重构图像的细节信息更清晰,并且相比较于 SRCNN、FSRCNN 及 VDSR 算法图6的 PSNR 分别提高 0.98 dB、0.67 dB、0.59 dB。图7的 PSNR 分别提高了 1.49 dB、0.94 dB 和 0.86 dB。

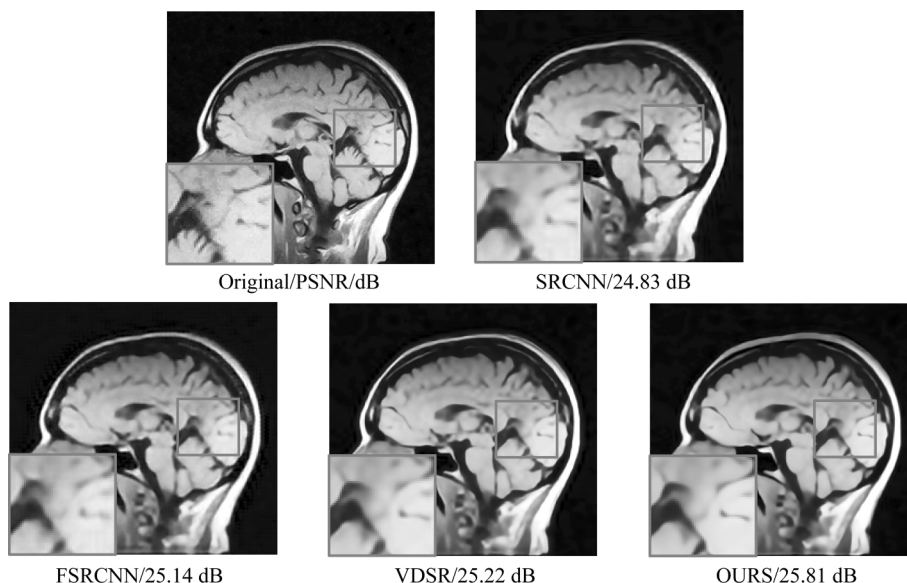


图6 Brain 图像在放大因子为 4 时的视觉对比

Fig.6 Visual comparison of the Brain image with the scale of 4

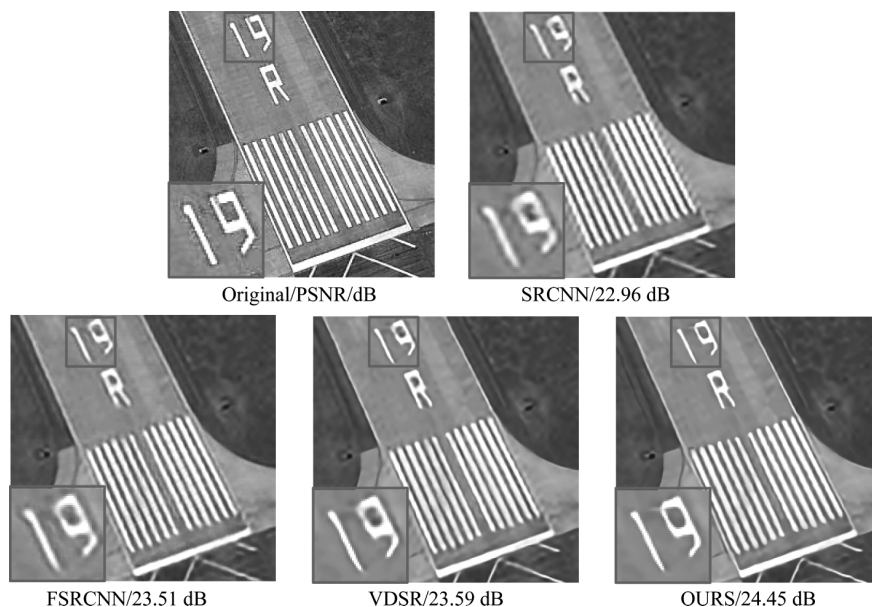


图 7 runway 图像在放大因子为 4 时的视觉对比

Fig.7 Visual comparison of the runway image with the scale of 4

4.2 不同损失函数下性能比较

l_2 范数损失函数广泛应用于图像处理优化问题中,在使用时一般假设噪声与图像的局部特性是相互独立的,但是人类视觉系统对噪声的灵敏度是依赖于图像的局部特性的^[28]。所以在使用 l_2 范数时具有一定的局限性。而 l_1 范数损失函数所得到的预测图像与清晰图像的误差是稀疏的,能更直观地反映实验结果。文献[29]指出从 PSNR 和 SSIM 的角度来说 l_2 范数损失函数并不能保证总能优于其他类型的损失函数,文献[29]提出的 EDSR 算法也使用了 l_1 范数损失函数来代替 l_2 范数损失函数,并取得了相对较好的性能。本文通过实验对比了两种损失函数的性能,实验结果如表 5 所示。

在 Set5 数据集上,不同放大因子($\times 2$ 、 $\times 3$ 、 $\times 4$)下,采用 l_1 范数损失函数的平均 PSNR 分别比 l_2 范数损失函数的平均 PSNR 高 0.1 dB、0.13 dB 和 0.13 dB,而在 Set14 数据集上取得提高比较小,不同放大因子($\times 2$ 、 $\times 3$)下分别提高 0.06 dB、0.01 dB,在放大因子为 4 时低了 0.01 dB。由表 5 也可以看出使用 l_1 范数损失函数的 SSIM 也相对较高一些。实验结果表明在网络训练时选用 l_1 范数损失函数比 l_2 范数损失函数所得到的重建图像性能更高。

4.3 两种下采样方法的对比实验

本文网络结构的下采样部分可以采用最大池化和步幅为 2 的卷积两种方法,如表 6 所示,两种方

法在 PSNR 和 SSIM 上都相差不多。在 Set5 数据集上,放大因子为 2 和 3 时池化方法的 PSNR 稍稍高了 0.03 dB,SSIM 相差不多;但是在 Set14 数据集上当放大因子为 3 和 4 时最大池化方法比步幅为 2 的卷积方法的 SSIM 分别低了 0.0005 和 0.0009。虽然在性能上两种方法相差不多,但是相比较于卷积,池化所需参数更少,训练速度较快。本文通过实验记录了两种方法在同样的数据集和训练参数下共训练 30 轮所需时间,在 GPU GTX1080 上池化方法所需训练时间比卷积方法少半个小时。所以本文实验选用池化方法进行下采样。

4.4 不同网络层数的对比实验

本文通过实验对比了不同网络层数对实验性能的影响。本文采用的网络结构共 25 层,在此网络基础上,在网络的每一个多尺度模块后分别减少和增加一个卷积层来进行对比实验。每个多尺度模块后减少一个卷积层后网络共 20 层,每个多尺度模块后增加一个卷积层后网络共 30 层。本次对比实验在放大因子为 3 的情况下进行,实验结果如表 7 所示。当网络层数从 25 层降为 20 层时,Set5 和 Set14 数据集的平均 PSNR 分别低了 0.09 dB 和 0.01 dB;当网络层数由 25 层增加为 30 层时,Set14 数据集的平均 PSNR 增加了 0.03 dB,但 Set5 数据集的平均 PSNR 反而下降了 0.07 dB。由表 7 实验结果可知,当网络层数达到一定数量后,即使增加网络层数,实验性能也不会有明显改善。

表 5 在 Set5 和 Set14 数据集上两种损失函数 PSNR(dB) /SSIM 对比
Tab.5 PSNR (dB) / SSIM comparison of two loss functions on Set5 and Set14 data sets

Dataset	Scale	l_2 loss	l_1 loss
Set5	×2	37.70/0.9594	37.80/0.9597
	×3	33.78/0.9224	33.91/0.9235
	×4	31.49/0.8842	31.62/0.8862
Set14	×2	33.15/0.9137	33.21/0.9137
	×3	29.81/0.8328	29.82/0.8328
	×4	28.13/0.7708	28.12/0.7702

表 6 在 Set5 和 Set14 数据集上两种下采样方法 PSNR(dB) /SSIM 对比
Tab.6 Comparison of PSNR (dB) / SSIM for two downsampling methods on Set5 and Set14 data sets

Dataset	Scale	Max-Pool	Conv(s=2)
Set5	×2	37.80/0.9597	37.77/0.9597
	×3	33.91/0.9235	33.88/0.9234
	×4	31.62/0.8862	31.62/0.8861
Set14	×2	33.21/0.9137	33.20/0.9136
	×3	29.82/0.8328	29.87/0.8333
	×4	28.12/0.7702	28.12/0.7711

表 7 Set5 和 Set14 数据集不同网络层数时 PSNR(dB) /SSIM 对比
Tab.7 Comparison of PSNR(dB) /SSIM for different network layers of Set5 and Set14 data sets

Dataset	25 Layers	20 Layers	30 Layers
Set5	33.91/0.9235	33.82/0.9229	33.84/0.9231
Set14	29.82/0.8328	29.81/0.8328	29.85/0.8332

表 8 Set5 和 Set14 数据集不同卷积核个数时 PSNR(dB) /SSIM 对比
Tab.8 Comparison of PSNR(dB) /SSIM for different number of convolution kernel on Set5 and Set14 data sets

Dataset	48	64	128
Set5	33.82/0.9230	33.91/0.9235	33.92/0.9240
Set14	29.81/0.8328	29.82/0.8328	29.81/0.8330

4.5 不同卷积核个数的对比实验

本文在放大因子为 3 的情况下对比了不同卷积核个数对网络性能的影响,如表 8 所示,分别对比了卷积核个数为 48 ,64 ,128 三种情况下的实验结果。从表 8 可以看出,卷积核个数为 48 时,与本文所选用的 64 个卷积核相比,Set5 和 Set14 数据集的平均 PSNR 分别低了 0.09 dB 和 0.01 dB;卷积核的个数为 128 时,相比较于 64 个卷积核,Set5 数据集的平均 PSNR 增加了 0.01 dB,Set14 数据集的平均 PSNR 下

降了 0.01 dB。由表 8 实验结果可知,本文算法选取卷积核个数为 64 是比较合适的。

4.6 不同训练样本数量的对比实验

本文在 291 张图像的基础上增加了 100 张图像作为训练集,经过数据增强处理后的训练图像由 5820 张增加到 7820 张,实验结果如表 9 所示。增加训练样本后,在 Set5 数据集上性能并无明显改善,在 Set14 数据集的平均 PSNR 增加了 0.05 dB。由此可知,当训练样本达到一定规模后,再增加样本,对实验性能也不会有显著提高。

表 9 Set5 和 Set14 数据集不同训练样本
数量时 PSNR(dB) /SSIM 对比

Tab.9 Comparison of PSNR(dB) /SSIM for different number
of train sample on Set5 and Set14 data sets

Dataset	5820	7820
Set5	33.91/0.9235	33.91/0.9238
Set14	29.82/0.8328	29.87/0.8346

5 结论

本文采用了一个具有多尺度卷积核的收缩—扩展残差网络,多尺度的卷积核使得训练时不仅可以提取不同尺度的特征信息还可以增加权重和偏置的多样性,以此高效的学习图像的有用信息。采用收缩—扩展的网络结构来提取多尺度的图像信息,使得训练更加高效,然而对网络进行收缩和扩展就要使用下采样和上采样层,这两层的使用会丢失图像的一些细节信息,所以本文使用跳跃连接来弥补这些损失的信息,使得信息更好的传播;并且采用残差训练的方式来提高重建图像质量。本文的实验结果表明,在相同训练集和测试集下,本文提出的网络结构用于图像的超分辨率重建,可以在不以时间为代价的前提下取得更高的性能。

参考文献

- [1] Shi W Z, Caballero J, Ledig C, et al. Cardiac image super-resolution with global correspondence using multi-atlas patch match [C] // Medical Image Computing and Computer-Assisted Intervention, Nagoya, Japan, 2013: 9-16.
- [2] Thornton M W, Atkinson P M, Holland D A. Sub-pixel mapping of rural land cover objects from fine spatial resolution satellite sensor imagery using super-resolution pixel-swapping[J]. International Journal of Remote Sensing, 2006, 27(3): 473-491.
- [3] 王民, 刘可心, 刘利, 等. 基于优化卷积神经网络的图像超分辨率重建[J]. 激光与光电子学进展, 2017, 54(11): 111005.
Wang M, Liu K X, Liu L, et al. Super-resolution reconstruction of image based on optimized convolution neural network[J]. Laser & Optoelectronics Progress, 2017, 54(11): 111005. (in Chinese)
- [4] 肖进胜, 刘恩雨, 朱力 等. 改进的基于卷积神经网络的图像超分辨率算法[J]. 光学学报, 2017, 37(3): 0318011.
- Xiao J S, Liu E Y, Zhu L, et al. Improved image super-resolution algorithm based on convolutional neural network [J]. Acta Optica Sinica, 2017, 37(3): 0318011. (in Chinese)
- [5] Harris J L. Diffraction and resolving power [J]. Journal of the Optical Society of America, 1964, 54(7): 931-936.
- [6] Zhu Y, Zhang Y, Yuille A L. Single image super-resolution using deformable patches [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014: 2917-2924.
- [7] Qu Y, Shi C, Liu J, et al. Single Image super-resolution via convolutional neural network and total variation regularization [C] // International Conference on Multimedia Modeling, MIAMI, USA, 2016: 28-38.
- [8] Hou H, Andrews H. Cubic splines for image interpolation and digital filtering [J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1978, 26(6): 508-517.
- [9] Keys R. Cubic convolution interpolation for digital image processing [J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1981, 29(6): 1153-1160.
- [10] Yang J, Wright J, Huang T S, et al. Image super-resolution via sparse representation [J]. IEEE Transactions on Image Processing, 2010, 19(11): 2861-2873.
- [11] Yang J, Wright J, Huang T, et al. Image super-resolution as sparse representation of raw image patches [C] // Computer Vision and Pattern Recognition, Anchorage, AK, USA, 2008: 1-8.
- [12] 曹明明, 干宗良, 陈杰, 等. 自适应邻域选取的邻域嵌入超分辨率重建算法[J]. 信号处理, 2015, 31(1): 8-16.
Cao M M, Gan Z L, Chen J, et al. Neighborhood embedding super-resolution reconstruction algorithm based on adaptive neighborhood selection [J]. Journal of Signal Processing, 2015, 31(1): 8-16. (in Chinese)
- [13] Yang J, Lin Z, Cohen S. Fast image super-resolution based on in-place example regression [C] // IEEE Conference on Computer Vision and Pattern Recognition, Portland, USA, 2013: 1059-1066.
- [14] Chang H, Yeung D Y, Xiong Y M. Super-resolution through neighbor embedding [C] // Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 2004, 1: 275-282.
- [15] Dong C, Loy C C, He K, et al. Image super-resolution using deep convolutional networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(2): 295-304.

- [16] Liang Y , Timofte R , Wang J , et al. Single image super resolution-when model adaptation matters [OL]. <http://arxiv.org/abs/1703.10889> , 2017.
- [17] Dong C , Loy C C , Tang X. Accelerating the super-resolution convolutional neural network [C] // European Conference on Computer Vision , Amsterdam , Netherlands , 2016: 391-407.
- [18] Kim J , Kwon Lee J , Mu 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 2016: 1646-1654.
- [19] Kim J , Kwon Lee J , Mu 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 , 2016: 1637-1645.
- [20] Lai W S , Huang J B , Ahuja N , et al. Deep laplacian pyramid networks for fast and accurate super-resolution [C] // The IEEE Conference on Computer Vision and Pattern Recognition. Hawaii , USA , 2017: 624-632.
- [21] Tai Y , Yang J , Liu X. Image super-resolution via deep recursive residual network [C] // The IEEE Conference on Computer Vision and Pattern Recognition. Hawaii , USA , 2017: 3147-3155.
- [22] Ronneberger O , Fischer P , Brox T. U-net: Convolutional networks for biomedical image segmentation [C] // International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich Germany 2015: 234-241.
- [23] Krizhevsky A , Sutskever I , Hinton G E. Imagenet classification with deep convolutional neural networks [C] // Advances in Neural Information Processing Systems , USA , 2012: 1097-1105.
- [24] Mao X J , Shen C , Yang Y B. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections [C] // Advances in Neural Information Processing Systems , Barcelona , Spain , 2016: 2802-2810.
- [25] Jia Y , Shelhamer E , Donahue J , et al. Caffe: Convolutional architecture for fast feature embedding [C] // Proceedings of the 22nd ACM International Conference on Multimedia , Orlando , Florida , USA , 2014: 675-678.
- [26] 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] // IEEE International Conference on Computer Vision , Vancouver , BC , Canada , 2001 , 2: 416-423.
- [27] Kingma D , Ba J. Adam: A method for stochastic optimization [C] // The International Conference on Learning Representations , San Diego , USA , 2015.
- [28] Zhao H , Gallo O , Frosio I , et al. Loss functions for image restoration with neural networks [J]. IEEE Transactions on Computational Imaging , 2017 , 3(1) : 47-57.
- [29] 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. Honolulu , HI , USA , 2017: 136-144.

作者简介



陈书贞 女, 1968 年生, 河北定州人。燕山大学信息科学与工程学院副教授, 主要研究方向为图像处理、压缩感知、相位恢复、深度学习。
E-mail: chen_sz818@163.com



解小会 女, 1992 年生, 河北秦皇岛人。燕山大学信息科学与工程学院硕士研究生, 主要研究方向为深度学习、图像超分辨率。
E-mail: 2415203368@qq.com



杨郁池 女, 1979 年生, 黑龙江泰来人。燕山大学信息科学与工程学院博士研究生, 河北经贸大学讲师, 主要研究方向为全息超分辨率和压缩感知。
E-mail: 43537195@qq.com



练秋生 男, 1969 年生, 江西遂川人。燕山大学信息科学与工程学院教授, 博士生导师, 主要研究方向为图像处理、稀疏表示、压缩感知、相位恢复及深度学习。
E-mail: lianqs@ysu.edu.cn