

基于残差的端对端图像超分辨率

华 臻^{1,3} 张海程^{2,3} 李晋江^{2,3}

(山东工商学院信息与电子工程学院 山东 烟台 264000)¹

(山东工商学院计算机科学与技术学院 山东 烟台 264000)²

(山东工商学院山东省高等学校协同创新中心未来智能计算 山东 烟台 264000)³

摘 要 深度卷积神经网络使图像超分辨率在准确性方面得到了很大改善。针对基于卷积神经网络的超分辨率重建方法网络结构简单、收敛速度慢、重建纹理模糊等问题,提出了一种基于残差学习的端对端深层卷积神经网络。该网络由局部残差网络和全局残差网络联合训练得到,增加了网络的宽度,能学习到不同的有效特征。局部残差网络包括特征提取、上采样和多尺度重建 3 个阶段,通过残差密集块密集连接卷积层提取有效的局部特征,采用多尺度卷积层获得丰富的上下文信息,利于高频信息的恢复;全局残差网络中采用渐进上采样的方式实现不同尺度的图像重建,通过残差学习提高收敛速度。在基准数据集 Set5, Set14, B100 和 Urban100 上进行放大 2 倍、3 倍和 4 倍的定量和定性评估。在这 4 种数据集下,所提算法在放大 3 倍时平均 PSNR/SSIM 指标分别为 34.70dB/0.9295, 30.54dB/0.8490, 29.27dB/0.8096 和 28.81dB/0.8653, 与其他方法相比有较大提升。在定性比较方面,所提方法重建出了更加清晰的图像,能更好地保留图像中的边缘细节。实验结果表明,所提方法在主观视觉和客观量化方面都有了较大改进,能有效提高图像重建的质量。

关键词 超分辨率, 残差学习, 卷积神经网络, 联合训练, 端对端

中图分类号 TP391

文献标识码 A

DOI 10.11896/j.issn.1002-137X.2019.06.037

End-to-end Image Super Resolution Based on Residuals

HUA Zhen^{1,3} ZHANG Hai-cheng^{2,3} LI Jin-jiang^{2,3}

(College of Information and Electronic Engineering, Shandong Technology and Business University, Yantai, Shandong 264000, China)¹

(College of Computer Science and Technology, Shandong Technology and Business University, Yantai, Shandong 264000, China)²

(Future Intelligent Computing of Collaborative Innovation Center of Shandong Higher Education Institutions, Shandong Technology and Business University, Yantai, Shandong 264000, China)³

Abstract Image super-resolution reconstruction technology is widely used in real life. An end-to-end deep convolutional neural network (CNN) based on residual learning was proposed to solve the problems of simple network structure, slow convergence rate and reconstructed texture blur in the network super-resolution CNN to further improve the quality of image reconstruction. The network is jointly trained by the local residual network and the global residual network, which increases the width of the network and learns different effective features. The local residual network includes three stages: feature extraction, upsampling and multi-scale reconstruction. The effective local features are extracted by densely concatenated blocks and the rich context information is obtained by multi-scale reconstruction, which is beneficial to the recovery of high-frequency information. In the global residual network, progressive upsampling is used to achieve multi-scale image reconstruction, and the convergence speed is improved by residual learning. Quantitative and qualitative evaluations are performed on the benchmark datasets Set5, Set14, B100, and Urban100 for scale factor of 2, 3, and 4. The proposed algorithm shows improved performances by 34.70dB/0.9295, 30.54dB/0.8490, 29.27dB/0.8096, and 28.81 dB/0.8653 on scale factor of 3. In terms of qualitative comparison, the proposed method reconstructs a clearer image, and preserves the edge details in the image better. The experimental results show that the proposed method has been greatly improved in subjective vision and objective quantization, which can improve the quality of image reconstruction effectively.

Keywords Super resolution, Residual learning, Convolutional neural network, Joint training, End-to-end

到稿日期:2018-10-12 返修日期:2019-01-23 本文受国家自然科学基金(61772319, 61472227, 61602277)资助。

华 臻(1966—),女,教授,主要研究方向为图像处理、机器学习, E-mail: huazhen66@foxmail.com(通信作者); 张海程(1994—),女,硕士生,主要研究方向为图像处理; 李晋江(1978—),男,博士,教授, CCF 会员,主要研究方向为图形图像处理、计算机视觉、机器学习。

1 引言

图像成像过程中,往往会受到降采样、噪声、模糊等因素的影响。图像超分辨率(Super-Resolution, SR)重建就是将质量较差的低分辨率(Low Resolution, LR)图像重建为接近真实图像的高分辨率(High Resolution, HR)图像。由于图像退化模型的不确定性,以及重建约束的不唯一性,SR重建本质上是一个病态问题^[1]。解决 SR 重建问题的方法主要分为 3 种:基于插值的方法^[2]、基于重建的方法和基于学习的方法。其中,基于插值的方法重建效果较差,基于重建的方法严重依赖于 HR 图像的先验信息,因此这两种方法应用较少;基于学习的方法是研究热点,其基本思想是通过训练样本得到 LR 图像到 HR 图像的映射关系,从而预测出 HR 图像。训练的样本库既可以来自 LR 图像内部^[3-5],也可以来自外部图像^[6-8]。马尔可夫网络^[6]、局部线性嵌入法^[7]、稀疏编码^[8-10]、锚定邻居^[11-12]等大量的学习算法被提出,但是这些方法在学习过程中的特征提取和表达能力有限。

最近,神经网络在图像 SR 重建领域表现出了巨大的优越性。Dong 等^[13]首先引入了一个三层结构的卷积神经网络(Convolutional Neural Network, CNN),实现了 SR 重建。与传统方法相比,该方法表现出了优越的性能,但是学习的特征较少。研究发现,深度较大的网络对图像恢复任务是非常有效的,可以根据更大的区域获得更多的信息,进而推断出更多目标区域的像素。但是,随着网络深度的增加,训练更加困难,容易出现梯度爆炸或梯度消失^[14]现象。残差学习框架^[15]可以解决这一问题,其跳跃连接可以使网络间的数据传输更为顺畅,也改善了因梯度消失而形成的欠拟合现象。VDSR^[16]和 LapSRN^[17]通过全局残差仅学习 HR 图像和 LR 图像之间高频部分的残差,大大减轻了训练的负担,提高了收敛速度和性能。

尽可能地重建出与 HR 图像高度相似的图像是 SR 任务的重点。网络模型中的每一层与 LR 图像和 HR 图像之间是高度相关的。如何将不同规模、不同级别的网络层的特征信息提取出来是非常关键的问题。已有方法大都采用单流和更深的网络获取更多的上下文信息。Lim 等利用残差块构造了一个非常大的残差缩放,建立了一个单尺度的 EDSR^[18]和一个多尺度的 MDSR^[18],获得了 NTIRE2017 超分辨率挑战赛的冠军,但是其每一层都需要大量的滤波器,且时间成本巨大^[19]。对于单流网络来说,在不同的感受野下融合互补的多尺度信息是比较困难的^[20]。文献^[21]证明了对于深层网络来说,增加宽度比增加深度更加有效,因此采用多个子网共同训练也是提高 SR 性能的有效解决方案^[22-23]。

本文提出了一种新的多子网联合训练的端对端网络,其提高了重建的性能,恢复出了更多的图像细节。将局部残差网络和全局残差网络联合训练,从而学习到不同的有效特征。两个网络的训练过程是独立的,可以避免参数的不适应问题,使得网络的宽度和参数增加。局部残差网络中引入了残差密集块(Residual Dense Block, RDB)^[24],充分利用了分层特征,并采用多尺度卷积层重建了丰富的上下文信息,利于高频信

息的恢复。全局残差网络中使用残差学习和渐进结构^[17],降低了网络训练的难度。

2 相关工作

2.1 超分辨率重建方法

目前,基于深度 CNN 的 SR 重建大致可以分为 4 种类型^[25],如图 1 所示。

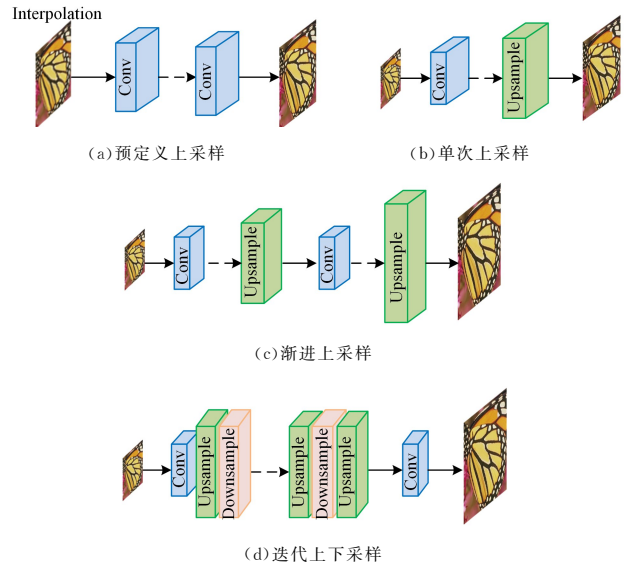


图 1 基于深度 CNN 的 SR 重建方法图

Fig. 1 SR-based reconstruction based on depth CNN

(1)预定义上采样利用传统的插值方法,如双三次插值方法,来获得目标空间的尺寸。以 SRCNN^[13]为代表,随后 VDSR^[16],DRRN^[25]使用残差网络,训练更深层次的网络架构,证明跳跃连接和递归卷积可以减轻超分辨率中携带身份信息负担。但是,这样的操作是不理想的,因为会增加额外的计算开销,还可能伤害原始 LR 图像中的关键信息。

(2)单次上采样采用亚像素卷积层(如 ESPCN^[26])或反卷积层(如 FSRCNN^[27])实现上采样操作。特征的减小,可以减少大量的计算,不会损失模型容量。但是,这些方法的网络结构相对简单,无法完全学习到 LR 空间到 HR 空间的复杂映射,也不能像 VDSR 那样处理单个框架中的多尺度问题。

(3)渐进上采样利用拉普拉斯金字塔网络逐渐预测 SR 图像^[17],是单个上采样的堆叠,有学习复杂映射的能力,能实现不同尺度因子的重建,对于大尺度因子更加有效。

(4)迭代上下采样由 DBPN^[19]提出,可以在不同阶段利用相互连接的上下采样模块在不同深度获得 HR 图像特征。

2.2 残差块

He 等^[15]首次提出残差学习框架 ResNet(见图 2(a))。残差结构是在标准的前馈卷积网络上添加一个跳跃以绕过一些层的连接,每绕过一层就产生一个残差块,每一层通过输入图像来学习残差函数而非原始函数。此后,这种基于残差学习的思想被大量应用到 SR 重建工作中。众所周知,LR 图像和 HR 图像有相似的低频信息,训练时加上这部分信息会多花费大量的时间,仅学习 LR 图像和 HR 图像之间的高频信息(即残差)就可以大大减轻训练的负担。VDSR 和 LapSRN

就采用了这种直接学习残差的方式,可以被认为是一种全局残差学习。VDSR 的残差学习结构由卷积层和整流线性单元 (Rectified Linear Units, ReLU)^[28] 组成,而 LapSRN 的结构由 Conv 和漏型整流线性单元 (Leaky Rectified Linear Units, LReLU) 组成。后来,SRResNet^[29] 在网络生成部分包含多个残差块,其去掉了 ResNet 后面的 ReLU 层,如图 2(b) 所示。EDSR 又删除了 SRResNet 中的批量归一化层 (见图 2(c)), 通过扩大网络尺寸来提升重建质量,取得了优越的性能。

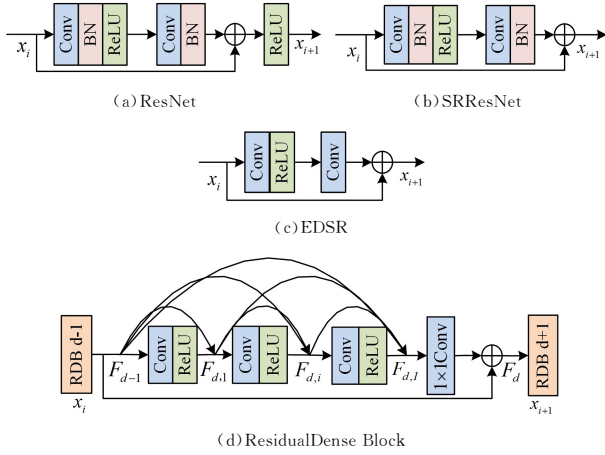


图 2 残差块的比较

Fig. 2 Comparison of residual blocks

以上方法都是以链式堆叠的方式构建模块,忽视了分层

特征。DenseNet^[30], SRDenseNet^[31] 允许密集块内的任意两层之间直接连接,取得了不错的效果,这说明不同深度特征之间包含的信息是互补的。Zhang 等^[24] 在密集网络的基础上提出了一个残差密集网络 (Residual Dense Network, RDN), 使用 RDB 充分利用来自 LR 图像的所有分层特征。RDB 是 RDN 的构建模块,包含密集连接层、局部特征融合和局部残差学习,残差结构见图 2(d)。RDB 中的每个卷积层都可以访问后续层,并传递所需的信息。

3 基于残差的图像 SR 重建网络

为了高效地学习到深度网络产生的分层特征,并能够进行多尺度重建,本文基于上述图像 SR 重建方法,采用单次上采样和渐进上采样相结合的网络结构,提出了全局残差网络和局部残差网络联合训练的网络模型。

图 3 显示了包含局部残差网络和全局残差网络的集合体系结构。两路网络有着相同的 LR 图像作为输入,通过联合训练得到最终的 HR 图像。模型中引入分离层^[32] 来构造一个两层的并列网络结构,增加了网络参数,并可以有效地防止过拟合。设计两路网络结构可以捕获不同的有效特征:局部残差网络有效地保留了网络的分层特征,有效重建了图像细节;全局残差网络的训练速度快,对于高尺度因子也有较好的恢复效果。两个网络可以更准确地恢复详细的 HR 图像,提高重建质量。

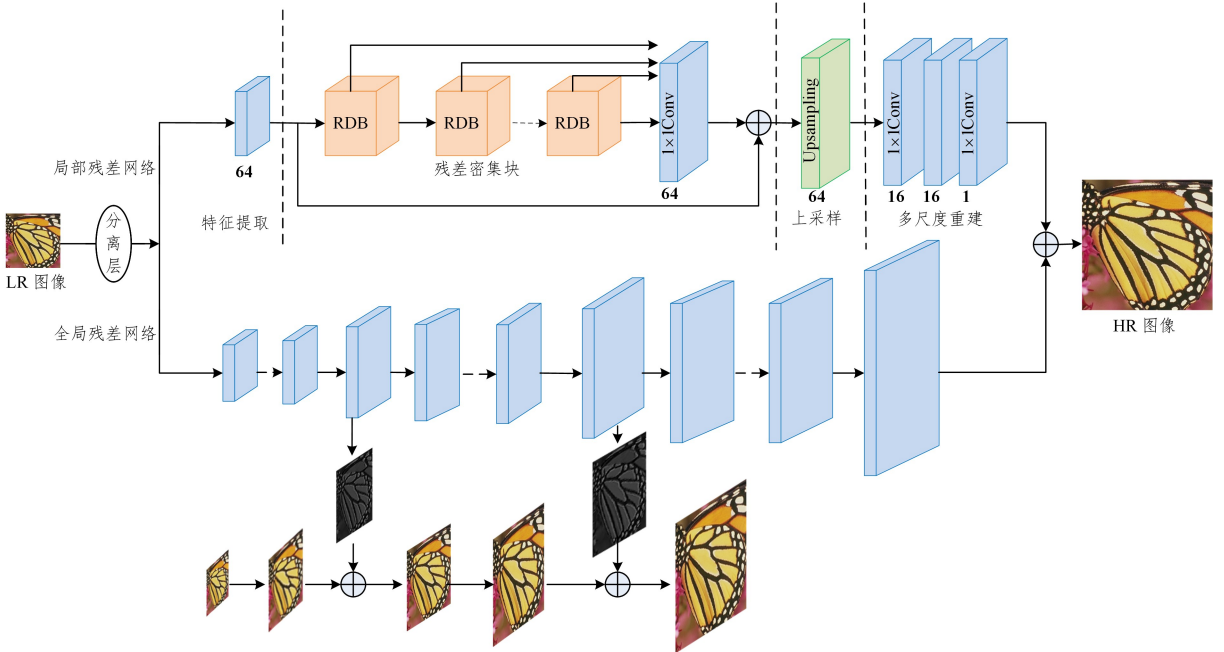


图 3 局部残差网络和全局残差网络联合训练的网络结构

Fig. 3 Network structure of joint training for local residual network and global residual network

3.1 局部残差网络

局部残差网络分为特征提取、残差密集块、上采样和多尺度重构 4 个模块。传统的特征提取方法大都用一阶、二阶梯度的方法对输入图像进行滤波;深度学习的方法则不需要手动设计滤波器,而是从训练数据中自动学习。

3.1.1 特征提取

此模块采用一个卷积层学习浅层特征,卷积核大小为 3×3 , 并加入 ReLU 进行非线性映射。该卷积层可以表示为

$$F_0 = \max(0, \omega_0 * I_L + b_0) \tag{1}$$

其中, ω_0 表示该卷积层的权重, b_0 表示该层的偏置, 运算符号

* 代表卷积操作, I_L 表示输入的 LR 图像, 最终通过 ReLU 激励函数得到该层的特征图 F_0 , 该卷积层共生成 64 个特征映射。

3.1.2 残差密集块

特征提取完成后, 将提取到的浅层特征 F_0 作为 RDB 的输入, 用 RDB 学习高层特征。假设有 D 个 RDB, 则第 d 个 RDB 可以通过式(2)获得:

$$F_d = O_{\text{RDB},d}(F_{d-1}) = O_{\text{RDB},d}(O_{\text{RDB},d-1}(\cdots(O_{\text{RDB},1}(F_0)))) \quad (2)$$

其中, $O_{\text{RDB},d}$ 是第 d 个 RDB 操作, 相当于式(1)中的卷积操作和 ReLU 操作; F_d 是第 d 个 RDB, 完全由块内的每个卷积层产生, 可被看作局部特征。这样, 便能够让前面的层对后续层进行访问。

图 2(d) 是第 d 个 RDB 的结构图, 通过将第 $d-1$ 个 RDB 的信息传递到第 d 个 RDB 的每一层来实现信息的连续传输^[24], 使得 RDB 中的每一层都与后续层直接相连。第 d 个 RDB 的第 i 个卷积层的输出可以表示为:

$$F_{d,i} = \max(0, \omega_{d,i} * [F_{d-1}, F_{d,1}, F_{d,2}, \cdots, F_{d,i-1}] + b_i) \quad (3)$$

其中, $\omega_{d,i}$ 是第 i 个卷积层的权重; $[F_{d-1}, F_{d,1}, F_{d,2}, \cdots, F_{d,i-1}]$ 表示由 $d-1$ 个 RDB 通过 $1, \cdots, i-1$ 个卷积层在第 d 个 RDB 中生成的特征映射, 每个卷积层后面都加入 ReLU 以增加非线性度。假设 F_{d-1} 有 64 个特征映射, 每个卷积层产生 g 个特征映射, 则最终共产生 $64 + (i-1)g$ 个特征映射。

随着每层特征映射数的增加, 训练也更加困难。为了减少特征映射的数量, 使用了 1×1 卷积层控制输出信息, 这个问题在文献^[33]中已经证明。最终将特征映射数减少到 64。在每个 RDB 中引入局部残差学习来提高网络的收敛速度, 增强了信息间的流动。

使用一组 RDB 学习多级局部密集特征后, 需要进行上采样操作。但是, 如果直接将 RDB 得到的所有特征串联起来后输入到反卷积层, 会产生巨大的计算开销, 因此同样添加一个核大小为 1×1 的卷积层来减少特征数量。这个卷积层把不同层次的特征融合起来生成 F_{LF} , 具体公式如下:

$$F_{\text{LF}} = O_{1 \times 1}([F_1, \cdots, F_D]) \quad (4)$$

其中, $[F_1, \cdots, F_D]$ 是由 RDB 产生的 D 个特征映射, $O_{1 \times 1}$ 是卷积核为 1×1 的卷积操作。为了将 LR 图像中的浅层特征与深层的密集特征连接起来, 加入跳跃连接。通过公式 $F_{\text{GF}} = F_0 + F_{\text{LF}}$ 将浅层特征图 F_0 与 F_{LF} 的输出相结合, 得到最终的密集特征图 F_{GF} , 生成 64 个特征映射。这样, 所有的特征图都被连接起来, RDB 使得来自 LR 图像的所有分层特征得到了充分利用, 有利于高频信息的恢复。

3.1.3 上采样

上采样模块连接特征提取和重构模块, 在提出的 SR 方法中起着关键作用。提取到 LR 空间的特征后, 根据单次上采样的思想, 使用反卷积层将 LR 图像上采样到目标 HR 图像的空间大小, 以更好地恢复图像细节。反卷积操作是卷积操作的逆操作, 以 h 倍的方式放大输入的特征图。适当增加卷积核的大小可以提高上采样质量, 但是如果太大会增加计算开销。因此, 如果上采样因子为 2, 3 和 4, 那么内核大小分

别设为 $14 \times 14, 15 \times 15$ 和 16×16 。

值得一提的是, 本文使用填充零的方式保证所有输出的特征映射的大小是相同的。事实证明, 这种方法的效果很好, 图像边界附近的像素也能被正确预测^[16]。在这里, 将输出特征映射的每一侧进行 6 个像素的零填充, 以保证空间大小相同。

3.1.4 多尺度重建

多尺度重建可以有效地聚合局部信息^[22,34], 从而进行更准确的重建。本文使用多尺度卷积来进行 HR 重建, 以恢复更多的上下文信息。首先, 使用 1×1 的降维层将上采样后的 64 个通道的特征图输出到 16 个通道。随后, 添加多尺度卷积层, 包括 4 个卷积运算; 卷积核分别为 $1 \times 1, 3 \times 3, 5 \times 5$ 和 7×7 。这 4 个卷积层同时作用于输入特征图, 产生 4 个 16 个通道的特征映射。然后, 将它们连接成 1 个 64 通道的特征映射, 使得不同尺度的上下文信息融合在一起。最后, 将级联的特征图再通过一个 1×1 卷积层输出到一个通道的卷积层, 以重建最终的 HR 图像。

3.2 全局残差网络

全局残差网络中, 本文借鉴了 LapSRN 中的渐进结构, 同样将网络结构进行分级, 通过全局残差学习逐步预测 $\log_2 S$ 级或 $\log_3 S$ 级的残差图像, 其中 S 为比例因子。这意味着: 如果要重建放大尺度为 8 的图像, 则网络由 3 级组成; 如果要重建放大尺度为 3 的图像, 则网络由 1 级组成。而 LapSRN 只能放大 2 倍、4 倍或 8 倍。只预测残差不仅可以加速网络的训练, 使其更容易收敛, 还可以达到比非残差网络更好的重建效果。网络模型分为特征提取分支和图像重建分支。

3.2.1 特征提取

第一级中, 将原始的 LR 图像作为输入(摒弃了内插放大的方式)。与 LapSRN 不同的是, 本文在进行特征提取的卷积层后面加入了 ReLU 层, 以进行非线性映射。每一级中, 都通过 $d-1$ 对卷积层和 ReLU 层进行特征提取, 后面加入一个反卷积层将规模放大 2 倍或 3 倍。反卷积层有两个输出: 一个输出继续进行下一层的特征提取; 另一个输出进行残差学习, 以预测出这一级的残差图像。与局部残差模型一样, 每个卷积层都采用 3×3 大小的卷积核, 生成 64 个特征映射。这样不仅可以学习到更多的网络参数, 还可以有效地避免过拟合现象, 而且在大型图像区域上可以有效利用更多的上下文信息。上采样层采用 5×5 大小的反卷积层。为了保证所有特征映射的大小相同, 在应用卷积之前都要在边界周围填零。

3.2.2 图像重建

使用双线性插值内核对上采样层进行初始化, 并且它与其他层共同优化。然后将上采样后的初始图像与预测出的残差图像进行结合, 生成 HR 图像。最后将本级别生成的 HR 图像送到下一级, 并按照同样的方法进行图像重建。

由于每一级的图像是在上一级的基础上得到的, 因此根据每一级的结构相似性, 各个级别的参数可以共享, 这样大大减少了参数的数量。此网络采用逐级预测残差的方式, 在进行高倍上采样时, 也能得到中间低倍上采样结果的输出。由于尺寸是逐步放大的, 不是所有的操作都在大尺寸特征上进行, 因此训练速度比较快。

3.3 训练

局部残差网络和全局残差网络不共享权重,两者都将相同的 LR 图像作为输入,独立地进行 SR 重建。训练时,可以将这两个网络看作网络的集合。最终的 HR 图像通过式(5)得到。

$$\hat{y} = H_L(x, \theta_L) + H_G(x, \theta_G) \quad (5)$$

其中, x 表示输入的 LR 图像, $H_L(\cdot, \theta_L)$ 和 $H_G(\cdot, \theta_G)$ 分别表示由 θ_L 和 θ_G 参数化的局部残差和全局残差网络的 HR 输出, \hat{y} 是联合训练出的最终 HR 图像。

给定 N 个训练图像对 $\{x^{(i)}, y^{(i)}\}_{i=1}^N$, 其中, x 表示输入的 LR 图像, y 表示对应的 HR 图像。需要使最终预测的 HR 图像 \hat{y} 与真实的 HR 图像 y 的均方误差 $\frac{1}{2} \|\hat{y} - y\|^2$ 最小, 以得到最终的参数优化模型。

训练时,使用基于反向传播的小批量梯度下降方法对目标进行优化,即:

$$\theta_L^*, \theta_G^* = \underset{\theta_L, \theta_G}{\operatorname{argmin}} \frac{1}{2N} \sum_{i=1}^N \|\hat{y}^{(i)} - y^{(i)}\|_2^2 + \lambda \Phi(\theta_L, \theta_G) \quad (6)$$

其中, $\Phi(\theta_L, \theta_G)$ 表示对网络参数施加的权重衰减, λ 表示折衷参数。

需要注意的是,本文使用了简单的残差学习来解决全局残差网络训练过程中的梯度爆炸/消失问题^[35]。由于输入图像与输出图像大部分是相似的,因此定义一个残差图像。级别为 S 的残差图像可以表示为 $r_S = y_S - x_S$, 其中 x_S 表示放大后的 LR 图像, y_S 表示级别 S 预测的 HR 图像。本文将每个级别的真实 HR 图像 y 通过双三次插值算法下采样到 y_S 。由式(5)可以知道, θ_G 是此路径中要优化的网络参数集,令 $\hat{y}_G = H_G(x, \theta_G)$ 。此处使用 LapSRN 提出的新的 loss 函数进行训练,公式如下:

$$\begin{aligned} \Psi(y_G, y; \theta_G) &= \frac{1}{N} \sum_{i=1}^N \sum_{s=1}^M \rho(\hat{y}_S^{(i)} - y_S^{(i)}) \\ &= \frac{1}{N} \sum_{i=1}^N \sum_{s=1}^M \rho((\hat{y}_S^{(i)} - x_S^{(i)}) - r_S^{(i)}) \end{aligned} \quad (7)$$

其中, $\rho(x) = \sqrt{x^2 + \epsilon^2}$ 是 Charbonnier 惩罚函数^[36], M 是金字塔中的级别数。此模型通过预测不同级别的残差图像,重建出多尺度图像。

本文通过式(5)~式(7)可以得到网络最终的优化参数模型,但是训练得到的深度模型可能无法在一定范围内收敛。VDSR 提出了一种提高学习率的方法——渐变裁剪^[37],其在抑制爆炸梯度的同时提供了一个可以调节的梯度幅度来最大限度地提高速度。在使用梯度下降进行优化时,若使用高学习率,则需要将梯度调整到很小以避免爆发梯度。若学习率很小,则有效梯度会接近于零。为了解决这个问题,将梯度调整到 $[-\eta, \eta]$ 来最大化地收敛速度,其中 $\eta = 0.4$ 。

4 实验分析

4.1 数据集

使用 Timofte 等发布的数据集 DIV2K^[38] 来训练网络模型。DIV2K 数据集由 800 个训练图像、100 个验证图像和 100 个测试图像组成。本文使用了 800 个训练图像,并在训练过程中选取了 6 幅验证图像进行验证。测试数据集则采用了

Set5^[39], Set14^[40], B100^[41] 和 Urban100^[42] 这 4 个基准数据集。

4.2 参数设置

实验环境包括硬件设备和软件配置,测试所用计算机的配置为 Intel Core i7-5820K CPU@3.30 GHz x12, GPU 为 NVIDIA GeForce TITANX, RAM 为 16 GB, 操作系统为 Win10, 编译软件为 MATLAB 2017a, 利用 Caffe^[43] 搭建卷积神经网络。

本文将两路网络联合训练,设置批量大小为 256, 动量为 0.9, 权重衰减为 0.009 (L_2 惩罚乘以 0.009)。对于权重的初始化,本文参考了 He 等^[44] 提出的方法。网络中卷积层的滤波器由零均值高斯分布随机初始化,标准差为 0.01。反卷积层中的滤波器由双线性插值内核初始化得到。所有层的初始学习率设置为 1×10^{-4} , 每 100 个 epochs 学习率减少一半。

对于每个放大倍数,从训练图像中将 HR 图像随机裁剪为 96×96 的图像块,然后使用双三次插值算法将 HR 图像生成对应的 LR 训练样本。为了避免过拟合并且进一步提高准确性,采用水平、垂直翻转和 90° 旋转等方式将数据集进行扩增。一个 epoch 由 1000 次反向传播迭代构成。由于人类对细节变化比颜色变化更敏感,与其他方法一样^[11-12], 本文也仅将本算法应用于 YCbCr 颜色空间中的亮度通道,将双三次插值方法应用于另外两个颜色通道。

由表 1 可知,网络越深越好。随着 RDB 中数量和卷积层数的增加,易获得更优的性能,而且高的增长率也有助于性能的提高^[15]。因此,将局部残差网络设计为更深、更广泛的网络,可以获得更多的分层特征。RDB 的个数 $D = 16$, 卷积层个数 $i = 8$, 每个卷积层产生 64 个特征映射。对于全局残差网络,每一级的网络同样是越深越好(见表 1),若放大尺度为 2 或者 4,则将网络深度设为 9,以在保证性能的基础上提高速度。若放大的尺度因子为 8,则将深度设为 5。

表 1 不同水平下的深度选择表

Table 1 Choice of depth at different levels

深度	B100		Urban100	
	PSNR/dB	时间/s	PSNR/dB	时间/s
3	31.46	0.053	31.48	0.058
5	31.52	0.061	31.58	0.063
10	31.69	0.069	31.70	0.074
15	31.74	0.089	31.82	0.090

4.3 算法比较

为了验证所提网络的有效性,将其与最新的方法进行定量和定性比较。参与比较的方法包括两种传统算法(双三次插值和锚定领域法(A+)^[11]),以及 4 种深度学习方法(SRCNN^[13], VDSR^[16], EEDS^[22] 和 EDSR^[31])。

4.3.1 定量比较

为了有效地评价重建效果的优劣,采用峰值信噪比(Peak Signal to Noise Ratio, PSNR)和结构相似度(Structural Similarity, SSIM)来衡量图像的质量。PSNR 是全参考评价方式,其值越大,表示图像的重建效果越好;SSIM 是指两幅图像之间结构信息的相似程度,也是一种全参考的图像质量评价指标,它分别从亮度、对比度、结构 3 个方面度量图像的相似性。SSIM 的取值范围为 $[0, 1]$, 其值越大,表示重建效果越好。

$$PSNR=10\lg\frac{PQ}{\|y-\hat{y}\|^2} \quad (8)$$

$$SSIM=\frac{(2\mu_y\mu_{\hat{y}}+C_1)(\sigma_{y\hat{y}}+C_2)}{(\mu_y^2+\mu_{\hat{y}}^2+C_1)(\sigma_y^2+\sigma_{\hat{y}}^2+C_2)} \quad (9)$$

其中, P,Q 分别为高、低分辨率图像的尺寸, y 为原始 HR 图像, \hat{y} 为重建后的 HR 图像, μ_y 和 σ_y 分别为原始 HR 图像的平均灰度值和方差, $\mu_{\hat{y}}$ 和 $\sigma_{\hat{y}}$ 分别为重建后图像的灰度平均值

和方差, $\sigma_{y\hat{y}}$ 为原始 HR 图像和重建 HR 图像的协方差, C_1 和 C_2 皆为常数。

表 2 展示了在 4 个公共数据集下分别进行 2 倍、3 倍和 4 倍放大后,SR 重建的定量比较结果,其中 PSNR 的单位为 dB。通过平均 PSNR 和 SSIM 总结了各方法对超分辨率的效果。在重建过程中,将任意大小的原始 LR 图像作为输入,直接重建相应的 HR 图像。

表 2 不同算法在不同噪声水平下的平均 PSNR/SSIM 的对比

Table 2 Average PSNR/SSIM comparison of different algorithms at different noise levels

数据集	尺度	Bicubic	A+	SRCNN	EEDS	VDSR	EDSR	本文算法
Set5	2	33.66/0.9299	36.54/0.9544	36.66/0.9542	37.29/0.9579	37.53/0.9587	38.11/0.9602	38.17/0.9605
	3	30.39/0.8682	32.58/0.9088	32.75/0.9090	33.47/0.9191	33.66/0.9213	34.66/0.9280	34.70/0.9295
	4	28.42/0.8104	30.28/0.8603	30.48/0.8628	31.14/0.8783	31.35/0.8838	32.46/0.8968	32.50/0.9013
Set14	2	31.24/0.8688	32.28/0.9056	32.42/0.9063	32.81/0.9105	33.03/0.9124	33.92/0.9195	33.99/0.9210
	3	27.55/0.7742	29.13/0.8188	29.28/0.8209	29.60/0.8284	29.77/0.8314	30.52/0.8462	30.54/0.8490
B100	2	29.56/0.8431	31.21/0.8863	31.36/0.8879	31.64/0.8928	31.90/0.8960	32.32/0.9013	32.31/0.9012
	3	27.21/0.7349	28.29/0.7835	28.41/0.7863	28.64/0.7925	28.82/0.7976	29.25/0.8093	29.27/0.8096
	4	25.96/0.6675	26.82/0.7087	26.90/0.7101	27.11/0.7200	27.29/0.7251	27.71/0.7420	27.73/0.7419
Urban100	2	26.88/0.8403	29.20/0.8938	29.50/0.8946	30.59/0.8710	30.76/0.9140	32.93/0.9351	32.94/0.9351
	3	24.46/0.7349	26.03/0.7973	26.24/0.7989	26.82/0.803	27.14/0.8279	28.80/0.8653	28.81/0.8653
	4	23.14/0.6577	24.32/0.7183	24.52/0.7221	24.70/0.7145	25.18/0.7524	26.64/0.8033	26.64/0.8031

从表 2 中可以看出,基于深度学习的算法明显优于 Bicubic 和 A+ 这类传统方法,而本文算法在大多数数据集上都优于其他算法。Bicubic 使用简单的插值方法,性能最差。A+ 算法采用了锚定领域回归的方法,重建效果得到了改进,PSNR 指数比 Bicubic 提高了 1~2 dB,SSIM 提高了大约 0.03~0.06。后来,深度学习的方法使重建精度得到了很大的提升。与 A+ 的 PSNR 指数相比, SRCNN 大约提升了 0.1~0.2 dB, EEDS 大约提升了 0.5~0.9 dB, VDSR 大约提高了 1~2 dB, EDSR 大约提高了 1.2~2.1 dB, 而本文方法大约提升了 1.1~2.2 dB。

对于 2 倍和 3 倍的放大尺度,本文方法更具优势。随着尺度因子的增加,所有方法的性能都会降低,但是本文方法始终略高于 EDSR 算法。这说明对于大尺度因子,本文方法仍能显示出较好的性能。大部分 Urban100 图像都包含自相似结构,使用较大的训练层和图像块都可以更好地利用感受野来获得更多的信息。EDSR 也使用了非常深的网络层,输入了更大的图像块,因此本文网络在此数据集上的 PSNR 和 SSIM 值与 EDSR 的值相差不大,但是与其他算法相比具有很大的优势。

4.3.2 定性比较

PSNR 和 SSIM 作为评价超分辨率质量的传统指标,用于评价像素间的差别,但无法完全代表人类的视觉效果。因此,本文在定量比较的前提下,结合视觉上的效果对图像质量进行定性分析。

图 4—图 9 展示了本文方法与最先进方法的定性比较结果。每张图的下方是方框位置放大后的局部效果,图 8 是“zebra”图像的局部放大图。Bicubic 是基于平滑假设的原理,出现了边缘模糊的情况,重建效果最差。A+ 算法的效果虽然得到了改善,但是出现了伪影。与传统方法相比, SRCNN

在视觉上有了较大提高,比如图 4 中的“bird”图像和图 7 中的“comic”图像都明显比传统算法清晰。但是, SRCNN 由于学习到的特征结构较为单一,重建细节的能力仍有不足,也会出现伪影和重建结果过度平滑的问题,甚至在深度不连续区域会出现边缘模糊的情况。比如,“bird”图像的眼珠部分、“comic”图像的项链和耳坠、“img_005”中的细线都没能得到较好的恢复。VDSR 采用了全局残差来提高收敛速度,重建效果比 SRCNN 好,但是仍与原图相差较大。SRCNN 和 VDSR 由于使用了双三次插值进行预处理,因此产生了明显的伪影。EEDS 虽然选择了单次上采样的方法学习原始像素的特征,但是没有采用更深层的网络结构,因此高频特征的重建效果不佳。

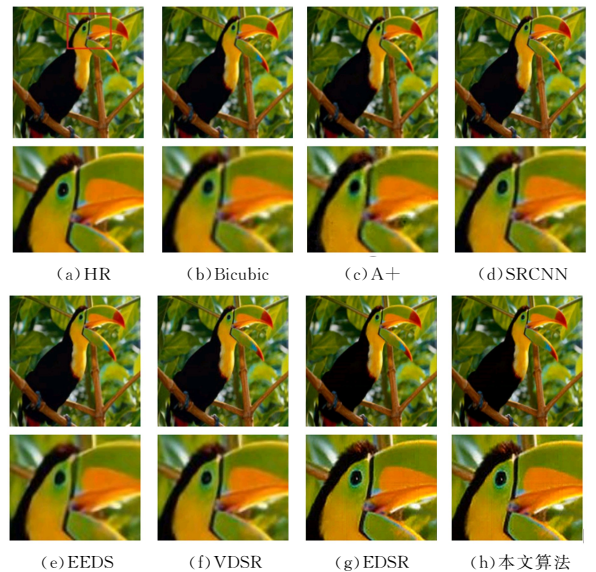


图 4 Set5(bird)图像在比例因子 $\times 2$ 下的超分辨率效果
Fig. 4 Super-resolution results for scale factor $\times 2$ on Set5 (bird image)

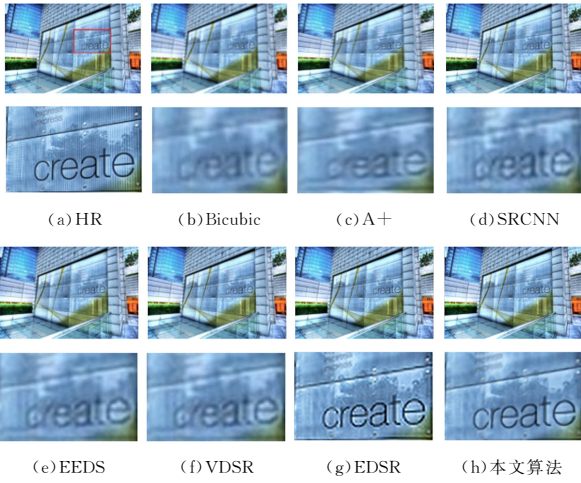


图5 Urban100(img_060)图像在比例因子 $\times 2$ 下的超分辨率效果
Fig. 5 Super-resolution results for scale factor $\times 2$ on Urban100 (img_060)

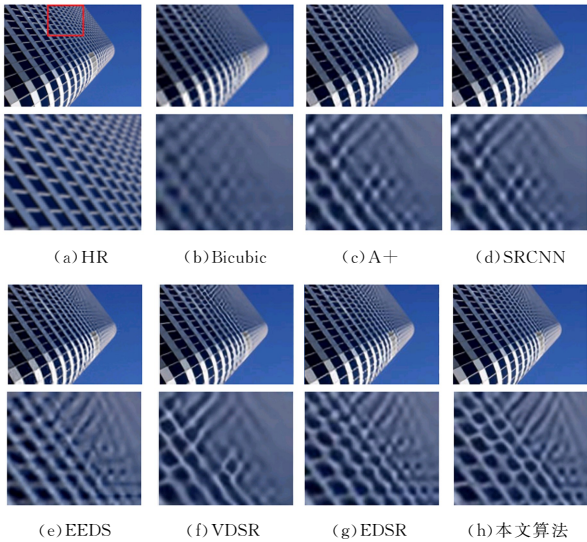


图6 Urban100 (img_005)图像在比例因子 $\times 4$ 下的超分辨率效果
Fig. 6 Super-resolution results for scale factor $\times 4$ on Urban100 (img_005)



图7 Set14 (comic) 图像在比例因子 $\times 3$ 下的超分辨率效果
Fig. 7 Super-resolution results for scale factor $\times 3$ on Set14 (comic image)

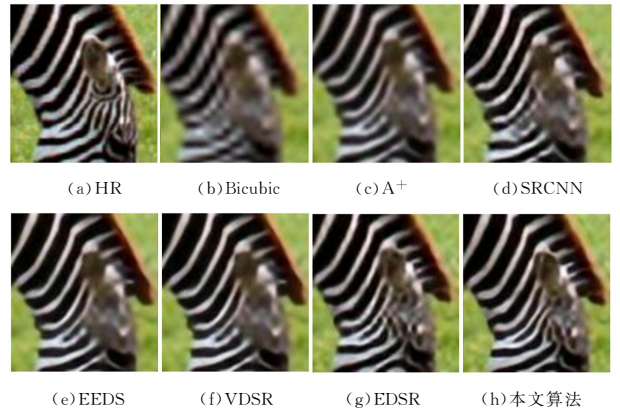


图8 Set14(zebra)图像在比例因子 $\times 4$ 下的超分辨率局部放大效果
Fig. 8 Super-resolution partial magnification results for scale factor $\times 4$ on Set14 (zebra image)

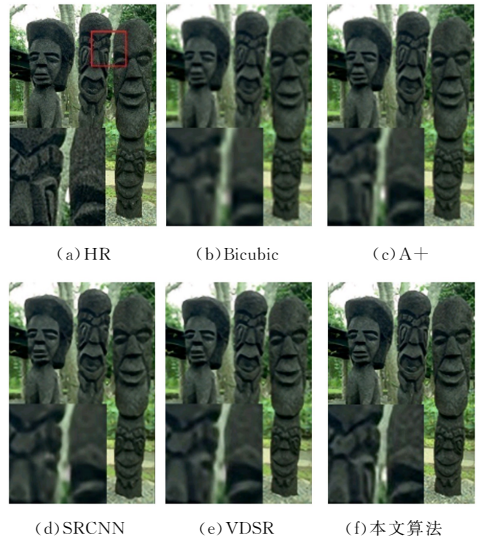


图9 B100(img_001)在比例因子 $\times 4$ 下的超分辨率效果
Fig. 9 Super-resolution results for scale factor $\times 4$ on B100(img_001)

EDSR 采用了更深的网络结构,能学习到更多的图像特征。从图像“img_060”中可以看出,EDSR 中的“create”单词和背景更加清晰。对于放大 3 倍的“comic”来说,EDSR 时颈部项链的重建效果都比其他算法的结果清晰,重建了更多的边缘。从图 6 可以更明显看出 EDSR 的重建优势。

本文方法与 EDSR 都重建了较多细节,但是本文方法略优于 EDSR。“zebra”图像眼部周围的毛发和“comic”中的眼睫毛等效果都得到了明显改善。而且,本文方法重建的斑马眼部旁的斑纹比 EDSR 的效果更加真实。这是因为本文方法通过更深的网络结构以及残差密集块,学习得到了更多的高频特征信息,使得重建的深度图像的视觉效果更好,接近真值。虽然图 6 中 EDSR 比本文重建了更多的纹理,但是本文算法重建的纹理更加清晰。随着放大尺度的增加,训练会更加困难,重建效果会越来越差,但本文采用的渐进式重建和残差学习不仅可以加快收敛速度,而且能够在避免边缘模糊的同时有效地解决边缘失真和伪影的问题,对大尺度因子也有较好的呈现效果。

4.4 消融研究

为了进一步了解所提网络的贡献,本文通过实验对模型中的不同组件进行评估。实验按照 4.2 节中的参数设置进行训练。由于不同放大尺度的结果相似,因此下面只给出了放大因子为 3 的结果,数据集为 Set5。

(1)联合训练:为了证明并行结构的有效性,将本文的网络(ESRR)拆分为局部残差网络(LRM)和全局残差网络(GRM)。图 10 是 Set5 数据集上 3 个网络的收敛图,可以看出,GRM 网络使用了较少的迭代次数进行收敛,但是 PSNR 性能相对较弱。LRM 网络和 ESRR 网络的 PSNR 性能明显高于 GRM,且 ESRR 网络的性能最优,大约达到了 34.7 dB。LRM 虽然性能比 GRM 好,但是训练比较困难,性能表现不稳定。LRM 的深层架构足够复杂,并且加入了残差密集块来学习更多的特征,因此可以恢复更精细的细节;但其可能会陷入局部极小值。

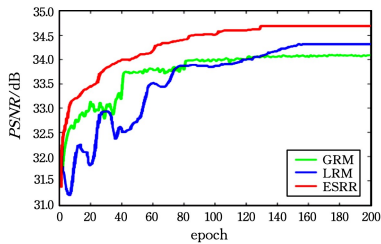


图 10 Set5 数据集上不同网络模型的收敛图

Fig. 10 Convergence diagram of different network models on Set5 data set

本文通过联合训练局部残差网络和全局残差网络来解决这个问题。在联合训练中,一开始,ESRR 网络依赖于 GRM 网络的收敛(见图 10)。当 GRM 网络学习到图像的主要特征时,深层次的 LRM 网络获得了更多的高频细节,学习并纠正了 GRM 网络的错误。从图 10 中还可以看出,联合训练的方法比 GRM 和 LRM 网络收敛得更快,而且获得了最好的 PSNR 性能。因此,联合训练可以达到最优的性能。

联合训练网络相当于加宽网络,增加网络的参数和特征数量,能有效提高重建的视觉效果;而且本文由两个并列的支路构成,捕捉了不同的有效图像特征,更多的有效特征也有利于提高重建质量。图 11 展示了两条支路的输出效果,每张图的下方是方框位置的局部放大图。



(a)HR (b)GRM 输出 (c)LRM 输出 (d)本文输出

图 11 所提网络和支路网络放大 3 倍的输出效果图

Fig. 11 Output of proposed network and subnetworks with scale factor $\times 3$

图 11(b)是全局残差网络的输出结果,其重建了图像的

光滑信息,但是含有伪像和模糊的内容。图 11(c)获得了更多的高频信息,弥补了一条支路的不足。通过结合两个网络的结果,所提网络获得了更准确的结果(见图 11(d))。

(2)残差学习:为了研究局部残差学习和全局残差学习的优势,将残差学习模块进行删除以观察效果。实验从局部残差网络中移除 RDB 开始和结束之间的跳跃连接(图 3 中带箭头的虚线)。因此,将所有 RDB 获得的特征融合后的输出 F_{LF} 直接馈送到上采样网络中,再通过重建网络获得最终的预测结果。全局残差学习的优势在文献[16-17]中已经得到了很好的证明。表 3 展示了不同组件在 Set5 数据集上放大 3 倍的消融研究结果,其中,局部残差学习表示为 LRL,全局残差学习表示为 GRL。可以看到,去掉 LRL 和 DRL 后,SR 重建的性能最低,去掉其中一个也会使性能降低。这是因为残差学习减少了网络的权重,改善了信息流和梯度,大大加快了收敛速度,并获得了很好的性能表现。

表 3 所提组件对重建质量的消融研究

Table 3 Ablation study of proposed component for reconstruction quality

LRL	GRL	RDB	MS1	MS3	MS5	MS7	PSNR
×	×	√	√	√	√	√	32.94
×	√	√	√	√	√	√	33.73
√	×	√	√	√	√	√	34.08
√	×	×	√	√	√	√	34.47
√	√	√	√	×	×	×	34.33
√	√	√	×	√	×	×	34.39
√	√	√	×	×	√	×	34.46
√	√	√	×	×	×	√	34.52
√	√	√	√	√	√	√	34.70

(3)残差密集块:为了展示 RDB 的优越性,本文使用了与 RDB 相同数量的卷积层替换模型中的单个 RDB 块。同样地,在第一个卷积层的开始和最后一个卷积层的结束加入跳跃连接,卷积核的大小为 3×3 ,并生成 64 个特征映射。从表 3 可以看出,加入 RDB 改进了重建性能,因为 RDB 能够充分利用并保留所有层的特征,有利于恢复更多的高频信息。

(4)多尺度重建:本文的 SR 重建模块由 4 个多尺度卷积层构成,卷积核大小分别为 1,3,5,7。为了验证此模块在 SR 重建中的影响,根据 EEDS 的思路,设置了 4 个单尺度的重建模块。将重建模块中的所有卷积层中的滤波器都设置为相同大小,分别为 $1 \times 1, 3 \times 3, 5 \times 5$ 和 7×7 。将这 4 个网络分别表示为 MS1,MS3,MS5,MS7。表 3 表明:多尺度的重建模块相比单尺度的重建模块,具有更好的性能表现。本文将 4 种尺度融合到一起重建出了更多的上下文信息,有效地改善了细节恢复问题。

结束语 为了有效提高深度图像的分辨率,同时获得更好的重建效果,本文提出了一种非常深的基于残差的图像超分辨率方法。不同于一般的深度学习方法在 HR 空间提取特征,本文直接将 LR 图像输入到网络中,在 LR 空间下提取特征,使用全局残差网络和局部残差网络联合学习不同的特征。引入残差密集块,使得每层之间密集连接,将每个 RDB 的输出结果融合到一起以训练更广泛的网络,使用局部残差使信息流和梯度得到进一步改善。采用反卷积层上采样到 HR 空间大小,并使用多尺度卷积层重构上下文信息。采用渐进上

采样学习残差的方式,大大提高了收敛速度,并重建了不同尺度的 HR 图像。实验结果表明,本文提出的方法在重建效果和客观指标两方面都优于其他算法,不仅能够重建更多的边缘细节,而且比 EDSR 的重建效果真实。在未来的工作中,将寻求更好的优化策略,进一步提高重建精度。

参 考 文 献

- [1] BAKER S, KANADE T. Limits on super-resolution and how to break them[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(9): 1167-1183.
- [2] XIAO J S, XU Z M, LI S, et al. Interpolation algorithm based on Improved adaptive shock filter image super-resolution[J]. *Chinese Journal of Computers*, 2015, 38(6): 1131-1139. (in Chinese)
肖进胜, 饶天宇, 贾茜, 等. 改进的自适应冲击滤波图像超分辨率插值算法[J]. *计算机学报*, 2015, 38(6): 1131-1139.
- [3] QIN X J, SHAN Y Y, XIAO J J, et al. Self-learning single image super-resolution reconstruction based on compressive sensing and SVR[J]. *Computer Science*, 2017, 44(S2): 169-174. (in Chinese)
秦绪佳, 单扬洋, 肖佳吉, 等. 基于压缩感知和 SVR 的自学习单幅图像超分辨率重建[J]. *计算机科学*, 2017, 44(S2): 169-174.
- [4] 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*. Los Alamitos: IEEE Computer Society Press, 2015: 5197-5206.
- [5] XIAO J S, GAO W, PENG H, et al. Detail enhancement image super-resolution algorithm based on SVD and local self-similarity [J]. *Chinese Journal of Computers*, 2016, 39(7): 1393-1406. (in Chinese)
肖进胜, 高威, 彭红, 等. 基于局部自相似性和奇异值分解的超采样图像细节增强[J]. *计算机学报*, 2016, 39(7): 1393-1406.
- [6] FREEMAN W T, PASZTOR E C, CARMICHAEL O T. Learning low-level vision[J]. *International Journal of Computer Vision*, 2000, 40(1): 25-47
- [7] CHANG H, XIONG Y M, YEUNG D Y. Super-resolution through neighbor embedding[C]// *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2004: 275-282
- [8] YANG J H, WRIGHT J, HUANG T S, et al. Image super-resolution via sparse representation[J]. *IEEE Transactions on Image Processing*, 2010, 19(11): 2861-2873
- [9] LI J H, WU Y R, LV J J. Online single image super-resolution algorithm based on group sparse representation[J]. *Computer Science*, 2018, 45(4): 312-318. (in Chinese)
李键红, 吴亚榕, 吕巨建. 基于组稀疏表示的在线单幅图像超分辨率算法[J]. *计算机科学*, 2018, 45(4): 312-318.
- [10] YANG C Y, YANG M H. Fast Direct Super-Resolution by Simple Functions[C]// *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, 2014: 561-568.
- [11] TIMOFTE R, DE V, VAN GOOL L. Anchored neighborhood regression for fast example-based super-resolution[C]// *Proceedings of the IEEE International Conference on Computer Vision*. Los Alamitos: IEEE Computer Society Press, 2013: 1920-1927
- [12] TIMOFTE R, SMET V D, GOOL L V. A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution[C]// *Proceedings of the Asian Conference on Computer Vision*. Cham: Springer, 2014: 111-126.
- [13] DONG C, CHEN C L, HE K, et al. Learning a Deep Convolutional Network for Image Super-Resolution[C]// *Proceedings of the European Conference on Computer Vision*. Cham: Springer, 2014: 184-199.
- [14] BENGIO Y, SIMARD P, FRASCONI P. Learning long-term dependencies with gradient descent is difficult[J]. *IEEE Transactions on Neural Networks*, 2002, 5(2): 157-166.
- [15] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society, 2015: 770-778.
- [16] KIM J, LEE J K, LEE K M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks[C]// *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2016: 1646-1654.
- [17] LAI W S, HUANG J B, 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*. Los Alamitos: IEEE Computer Society, 2017: 5835-5843.
- [18] LIM B, SON S, KIM H, et al. Enhanced Deep Residual Networks for Single Image Super-Resolution[C]// *Proceedings of the Computer Vision and Pattern Recognition Workshops*. IEEE, 2017: 1132-1140.
- [19] HARIS M, SHAKHAROVICH G, UKITA N. Deep Back-Projection Networks For Super-Resolution[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2018.
- [20] HU Y, GAO X, LI J, et al. Single Image Super-Resolution via Cascaded Multi-Scale Cross Network[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2018.
- [21] ZAGORUYKO S, KOMODAKIS N. DiracNets: Training Very Deep Neural Networks Without Skip-Connections[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2017.
- [22] WANG Y, WANG L, WANG H, et al. End-to-End Image Super-Resolution via Deep and Shallow Convolutional Networks[C]// *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*. IEEE, 2016.
- [23] REN H, EL-KHAMY M, LEE J. Image Super Resolution Based

- on Fusing Multiple Convolution Neural Networks[C] // Proceedings of the Computer Vision & Pattern Recognition. IEEE, 2017.
- [24] ZHANG Y, TIAN Y, KONG Y, et al. Residual Dense Network for Image Super-Resolution[C] // Proceedings of the Computer Vision and Pattern Recognition, 2018.
- [25] KIM J, LEE J K, LEE K M. Deeply-Recursive Convolutional Network for Image Super-Resolution[C] // Proceedings of the Computer Vision and Pattern Recognition. IEEE, 2016; 1637-1645.
- [26] 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] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016; 1874-1883.
- [27] DONG C, CHEN C L, TANG X. Accelerating the Super-Resolution Convolutional Neural Network[C] // Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2016; 391-407.
- [28] XU B, WANG N Y, CHEN T Q, et al. Empirical evaluation of rectified activations in convolutional network[C] // Proceedings of the 32th International Conference on Machine Learning; Deep Learning Workshop. Lille, France; ICML, 2015.
- [29] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2017; 105-114.
- [30] HUANG G, LIU Z, MAATEN L V D, et al. Densely Connected Convolutional Networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2017; 2261-2269.
- [31] TONG T, LI G, LIU X, et al. Image Super-Resolution Using Dense Skip Connections[C] // IEEE International Conference on Computer Vision. IEEE Computer Society, 2017; 4809-4817.
- [32] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines[C] // Proceedings of the International Conference on Machine Learning. Omnipress, 2010; 807-814.
- [33] SZEGEDY C, VANHOUCHE V, IOFFE S, et al. Rethinking the Inception Architecture for Computer Vision[C] // Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition. IEEE, 2016; 2818-2826.
- [34] SZEGEDY C, LIU W, JIA Y, et al. Going Deeper with Convolutions[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2015.
- [35] BENGIO Y, SIMARD P, FRASCONI P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE Transactions on Neural Networks, 2002, 5(2): 157-166.
- [36] BRUHN A, WEICKERT J, SCHN, et al. Lucas/Kanade meets Horn/Schunck; combining local and global optic flow methods [J]. International Journal of Computer Vision, 2005, 61(3): 211-231.
- [37] PASCANU R, MIKOLOV T, BENGIO Y. On the difficulty of training recurrent neural networks[C] // Proceedings of the International Conference on International Conference on Machine Learning. JMLR. org, 2013; 1310-1318.
- [38] TIMOFTE R, AGUSTSSON E, GOOL L V, et al. NTIRE 2017 Challenge on Single Image Super-Resolution; Methods and Results[C] // Proceedings of the Computer Vision and Pattern Recognition Workshops. IEEE, 2017; 1110-1121.
- [39] BEVILACQUA M, ROUMY A, GUILLEMOT C, et al. Low-Complexity Single Image Super-Resolution Based on Nonnegative Neighbor Embedding[C] // Proceedings of the British Machine Vision Conference. BMVA Press, 2012; 135. 1-135. 10.
- [40] ZEYDE R, ELAD M, PROTTER M. On single image scale-up using sparse-representations[C] // Proceedings of the International Conference on Curves and Surfaces. Berlin Heidelberg: Springer, 2010; 711-730.
- [41] MARTIN D R, FOWLKES C, TAL D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics[C] // Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2002; 1110-1121.
- [42] HUANG J B, SINGH A, AHUJA N. Single image super-resolution from transformed self-exemplars[C] // Proceedings of the Computer Vision and Pattern Recognition. IEEE, 2015; 5197-5206.
- [43] JIA Y Q, SHELHAMER, et al. Caffe: Convolutional Architecture for Fast Feature Embedding[C] // Proceedings of the ACM International Conference Multimedia. ACM, 2014; 675-678.
- [44] HE K, ZHANG X, REN S, et al. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification[C] // Proceedings of the IEEE International Conference on Computer Vision. IEEE Computer Society, 2015; 1026-1034.