

基于深度特征学习的图像超分辨率重建

胡长胜¹ 詹曙¹ 吴从中¹

摘要 基于学习的图像超分辨率 (Super-resolution, SR) 算法利用样本先验知识来重建图像, 相较于其他重建方法拥有明显的优势, 也是近年来研究的热点. 论文首先分析了影响图像重建质量的因素, 然后对基于卷积神经网络的图像超分辨率重建算法 (Super-resolution convolutional neural network, SRCNN) 提出了两点改进: 我们用随机线性纠正单元 (Randomized rectified linear unit, RReLU) 去避免原有网络学习中对图像某些重要的信息过压缩, 同时我们用 NAG (Nesterov's accelerated gradient) 方法去加速网络的收敛并且避免了网络在梯度更新的时候产生较大的震荡. 最后通过实验验证了我们改进网络可以获得更好的主观视觉评价和客观量化评价.

关键词 超分辨率, 图像复原, 深度学习, 卷积神经网络, 特征映射

引用格式 胡长胜, 詹曙, 吴从中. 基于深度特征学习的图像超分辨率重建. 自动化学报, 2017, 43(5): 814–821

DOI 10.16383/j.aas.2017.c150634

Image Super-resolution Based on Deep Learning Features

HU Chang-Sheng¹ ZHAN Shu¹ WU Cong-Zhong¹

Abstract Learning-based image super-resolution method is a research hotspot in recent years which uses prior knowledge of sample to reconstruct the image and has obvious advantages over other reconstruction methods. In this paper, we first analyze the factors of reconstructed image quality. Then we use randomized rectified linear unit (RReLU) to solve the problem of over compression in the original network. Besides, Nesterov's accelerated gradient (NAG) is invoked to accelerate convergence and avoid large oscillations. Finally, we conduct a quantitative experiments to prove the validity of the proposed algorithm.

Key words Super-resolution (SR), image restoration, deep learning, convolution neural network (CNN), feature map

Citation Hu Chang-Sheng, Zhan Shu, Wu Cong-Zhong. Image super-resolution based on deep learning features. *Acta Automatica Sinica*, 2017, 43(5): 814–821

图像超分辨率 (Super-resolution, SR) 重建是图像复原的一个重要分支, 首先由 Tsia 等于 80 年代首次提出^[1]. 在随后的 30 多年时间, 由于广泛的实用价值与理论价值, 超分辨率图像重建技术成为计算机视觉与图像处理领域的一个研究热点, 引起了研究者的广泛关注. 实际上由于图像退化模型的不定性, 以及重建约束的不唯一性, 超分重建本质上是一个病态的问题^[2]. 所以如何由一幅或者多幅低分辨率图像 (Low resolution, LR) 产生一个具有良好视觉效果和更加接近真实图像的高分辨率图像

(High resolution, HR) 成为图像超分辨率重建的主要目的. 一般来说, 由一个低分辨率图像获得一个高分辨率图像有许多可行的方法, 比如使用某种形式的正则化 (高分辨率的先验知识) 来保证对图像扩展的唯一性和稳定性. 因此目前主流的超分辨率重建方法^[3] 大致分为 3 类: 基于插值、基于重建和基于学习的方法.

基于插值^[4] 的方法是最早提出的也是最基本的图像重建方法, 经典的插值方法有最近邻插值和三次样条插值方法, 其通过直接利用自然图像的先验知识, 产生了很好的效果. 但是简单的基于插值的方法在增大图像幅面的同时也会产生具有震荡和锯齿状的过平滑图像. 基于重建^[5] 的方法强制约束平滑和下采样, 使高分辨率图像与低分辨率图像保持一致, 但是这些基于重建的方法的性能表现严重依赖施加在高分辨率图像的先验知识, 并且当放大因子较大或者可用的输入图像较小时, 重建质量会迅速下降. 在这种情况下, 结果同样会过平滑, 缺失重要的高频细节. 基于学习的方法^[6] 是图像超分辨率重建目前研究的热点方向之一, 其可以使用机器学习的理论来避免上述的问题. 通过对低分辨率

收稿日期 2015-10-13 录用日期 2016-06-17
Manuscript received October 13, 2015; accepted June 17, 2016
国家自然科学基金 (61371156), 安徽省科技攻关项目基金 (1401B042019), 中科院自动化所复杂系统管理与控制国家重点实验室开放课题 (20130107) 资助
Supported by National Natural Science Foundation of China (61371156), Anhui Province Science and Technology Research Programs (1401B042019), and State Key Laboratory of Management and Control for Complex System of Institute of Automation Chinese Academy of Sciences Open Project (20130107)
本文责任编辑 贾珈
Recommended by Associate Editor JIA Jia
1. 合肥工业大学计算机与信息学院 合肥 230009
1. School of Computer and Information, Hefei University of Technology, Hefei 230009

图像和对应的高分辨率图像的字典学习^[7]等, 学习它们之间的统计学联系, 然后利用学习到的联系来重建低分辨率图像以获得对应的高分辨率图像. 例如, Freeman 等^[8]首先提出利用马尔科夫网络来建立图像局部区域和场景之间的关系模型来重建图像, Polatkan 等^[9]则提出了一种新的贝叶斯非参数模型, 使用 β -伯努利过程从一组称之为字典元素的数据中来学习一组循环的视觉形态. 近些年来, 随着稀疏表达被引入图像超分辨率重建领域, 其立刻成为了基于学习方法中最具代表性的一个方法. Yang 等^[10]提出了一个基本的稀疏表示模型 (Sparse coding for super-resolution, ScSR) 来重建高分辨率图像, 通过对低分辨率图像块字典和高分辨率图像块字典的联合训练, 强化低分辨率和高分辨率图像块对应真实字典稀疏表示的相似性, 从而使低分辨率图像块的稀疏表示和高分辨率超完备字典联合作用重建出高分辨率图像块, 最终由高分辨率图像块连接得到最终完整的高分辨率图像. 这种方法虽然很高效, 但是在实际操作中需要预先指定字典元素的数量以及噪声方差模型, 因此对于现实生活中的自然图像很难获得最优的参数.

最近, 深度学习^[11-13]在计算机视觉领域表现出巨大的潜力, 它可以通过建立强有力的模型和设计高效的学习策略来克服过度拟合, 并且神经网络可以灵活地通过增加新的非线性激活函数或特定功能的层来更好地拟合训练数据. 因此在图像超分辨率重建领域, 越来越多的研究者开始探索卷积神经网络 (Convolutional neural network, CNN) 和图像超分辨率之间的联系, 并且获得了令人惊喜的结果^[14-15].

本文, 在基于原有的超分重建的卷积神经网络 (Super-resolution convolutional neural network, SRCNN)^[14]的算法思想上, 提出了对网络的改进, 在主观视觉和客观数据上都获得了比原算法更好的实验结果. 我们采用了一个三层的卷积神经网络去学习高分辨率图像和低分辨率图像之间端到端的映射, 首先我们采用随机纠正线性单元 (Randomized rectified linear unit, RReLU)^[16]去克服原有网络中存在的信息过压缩现象, 然后使用了 NAG (Nesterovs accelerated gradient)^[17-18]方法去加速损失函数的收敛和避免采用随机梯度下降方法 (Stochastic gradient descent, SGD) 时产生的较大的震荡.

1 方法与理论

1.1 预处理

为了得到满足要求的低分辨率图像, 在网络训

练之前, 我们首先进行一定的预处理过程, 即用双三次插值扩展图像去获得满足要求的低分辨率图像, 相应的在测试阶段也进行了相同的预处理. 事实上, 可以把双三次插值操作看做一个卷积操作, 在网络中用一个卷积层实现. 但在实际应用中由于这一层的输出尺寸大于输入尺寸, 会存在分式的步幅, 因此为了更好地在深度学习框架 Caffe^[19]进行训练, 我们未将这个预处理层包含在训练网络中.

1.2 特征提取

这个操作是从低分辨率图像中提取图像块并把每个块表示成一个高维向量. 这些向量组成一组特征图, 其数量和向量的维度相等.

传统的神经元模型使用的激活函数是 sigmoid, 是从神经科学上边仿生过来的, 用它来模拟神经元从受到刺激, 接收到的电信号超过一定的阈值就产生兴奋这个特性, 它有一个严重的问题就是其容易产生饱和效应, 也称梯度弥散效应, 即在 sigmoid 函数两侧梯度非常小 (接近 0). 因此在后来的发展中, 诸如 $abs(x)$, $\max(x, 0)$, $\min(x, 0)$ 等函数开始进入研究者的视线, 对于计算机来说, 它们可以通过调整符号位很容易实现. 同时因为它们的梯度很简单, 即 1, -1 或者 0, 因此无论在网络中传播多少层, 其梯度和都会维持在一个相对稳定的数量级上.

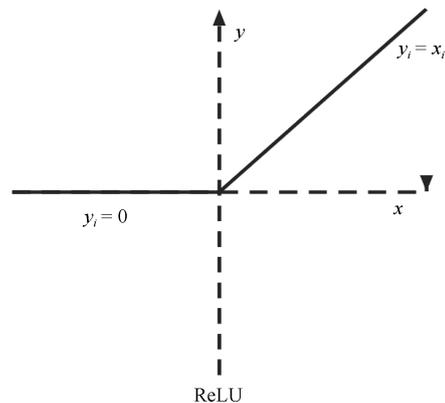


图 1 ReLU 函数的示意图

Fig. 1 An illustration of ReLU

因此在网络中, 将第一层的操作表示为以下形式:

$$F_1(X) = \max(0, W_1 * X + B_1) + \frac{(\min(0, W_1 * X + B_1))}{a} \quad (1)$$

其中, W_1 为滤波器, B_1 为偏差, “*” 表示卷积操作, W_1 的大小为 $c \times f_1 \times f_1$, c 是输入图像中通道的数量, f_1 是滤波器的空间大小, n_1 为滤波器的数量, a 服从一个均匀分布: $a \sim U(l, u)$. 显然, 在这里我们使用了随机线性纠正单元 (RReLU^[16]) 来代替线性

纠正单元 (ReLU^[20]). 由图 1 可以看到, 当我们使用 ReLU 的时候, 仅在 $x > 0$ 的象限中的信息被保留下来, 而其他区域中的信息则在一定程度上被压缩 (忽略), 因此在重建的时候这些因过压缩而丢失的信息将无法恢复, 这在实际应用中是不合理的. 而在图 2 中, 可以看到, 当 $x > 0$ 的时候, RReLU 可以通过设置和 ReLU 保持完全一致, 这样不仅继承了 ReLU 函数的优良特性, 同时在 $x < 0$ 的时候通过设置合理的均匀分布来保留一些有用的信息而忽略冗余信息来进一步提高重建质量.

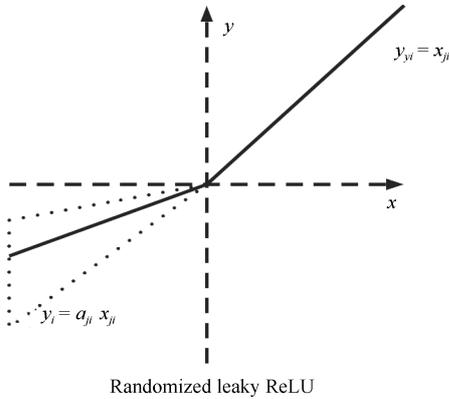


图 2 RReLU 函数的示意图 (其中 a_{ji} 为在抽样给定范围类的一个随机变量, 同时为了方便, 在测试阶段, 我们通常根据实际情况取一个固定值来进行测试)

Fig. 2 An illustration of RReLU (a_{ji} is a random variable of in the given sampling scope. And in the testing phase, we usually take a fixed value to test according to actual condition.)

1.3 非线性映射

在第一层中, 通过卷积操作来对每个图像块提取 n_1 特征, 而在第二层中, 我们通过卷积层把每一个 n_1 维向量映射到一个 n_2 维向量中. 第二层的操作可表示为

$$F_2(X) = \max(0, W_2 * F_1(X) + B_2) +$$

$$\frac{(\min(0, W_2 * F_1(X) + B_2))}{a} \quad (2)$$

同样 W_2 对应 n_2 个大小为 $n_1 \times f_2 \times f_2$ 大小的滤波器, B_2 为 n_2 的偏差. 每一个 n_2 维向量在概念上表示可用于重建的高分辨率图像块.

1.4 图像重建

在文献 [9] 中, 预测的重叠的高分辨率图像块通过平均来产生最终的完整图. 平均步骤可以看作是一个在一组特征图上一个预先定义的滤波器, 受这个启发, 我们定义一个卷积层去产生最终的高分辨率图像:

$$F(X) = W_3 * F_2(X) + B_3 \quad (3)$$

其中, W_3 为一个大小为 $n_2 \times f_3 \times f_3 \times c$ 的线性滤波器, B_3 为 c 维的偏差. 整个网络的概述性描述见图 3.

1.5 训练与优化

为了学习端到端的映射函数 F , 本文使用均方差 (Mean square error, MSE) 作为损失函数来估计网络参数 $\Theta = \{W_1, W_2, W_3, B_1, B_2, B_3\}$, 在实验中我们也尝试使用其他更接近人类视觉感知的损失函数来替代 MSE, 但实验结果显示 MSE 可以通过相对简单的优化来使图片质量在主观上获得显著提升. MSE 形式如下:

$$L(\Theta) = \frac{1}{n} \sum_{i=1}^n \|F(Y_i, \Theta) - X_i\|^2 \quad (4)$$

其中, n 为训练样本的数量.

而对于改进使用包含斜率为负部分的激活函数是否使目标函数在梯度更新时仍然可微^[21-22], 可以参考 PReLU (Parametric rectified linear unit)^[23] 函数可微性证明做出类似推论. 对于 RReLU, 为了方便书写, 我们重写表达式如下:

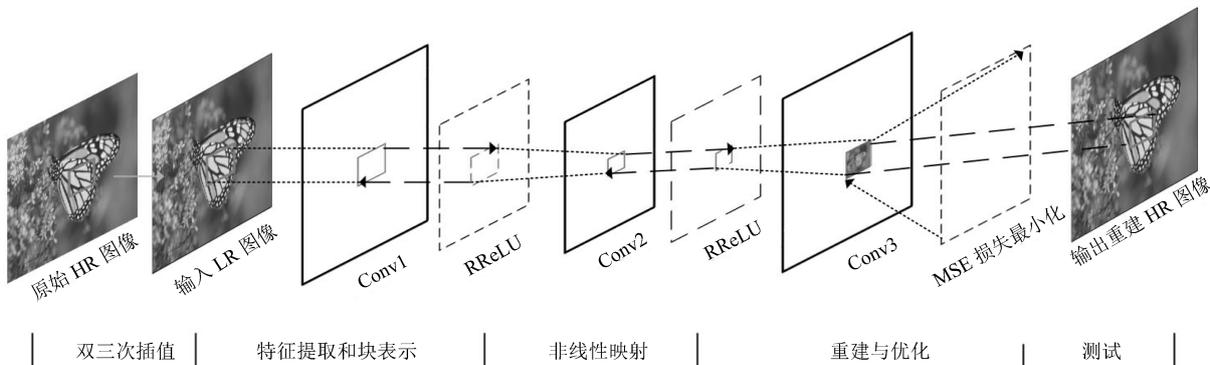


图 3 超分重建卷积神经网络结构示意图

Fig. 3 The structure chart of CNN for super-resolution

$$f(x) = \begin{cases} x, & \text{若 } x > 0 \\ cx, & \text{若 } x < 0 \end{cases} \quad (5)$$

其中, $c = 1/a$, 很显然 c 为负值部分的斜率. 我们可以使用反向投影进行训练同时将它与其他层一起优化, 其 c 的更新方式我们可以根据链式法则来进行简单的推导. 对于某一层 c 的梯度如下:

$$\frac{\partial \varepsilon}{\partial c} = \sum_x \frac{\partial \varepsilon}{\partial f(x)} \frac{\partial f(x)}{\partial c} \quad (6)$$

其中, ε 表示目标函数, $\partial \varepsilon / \partial f(x)$ 项是来自更深层的梯度传播, 则对于整个激活函数, 其梯度如下:

$$\frac{\partial f(x)}{\partial c} = \begin{cases} 0, & \text{若 } x > 0 \\ x, & \text{若 } x < 0 \end{cases} \quad (7)$$

随机梯度下降方法通过一个负梯度 $\nabla L(W)$ 线性组合的经典势能 (Classical momentum, CM) 方式来更新权重 W . 其形式如下:

$$\begin{cases} \Delta_{i+1} = \mu \Delta_i + \eta \nabla L(W_i^l) \\ W_{i+1}^l = W_i^l + \Delta_{i+1} \end{cases} \quad (8)$$

其中, $l \in [1, 2, 3]$, 表示卷积层号, i 为迭代次数, $\nabla L(W) = \partial L / (\partial W_i^l)$ 为导数, η 是学习率 (即负梯度的权重, 其意义为当前批处理的梯度多大程度上影响最终更新方向), μ 为前一次迭代权重且 $\mu \in [0, 1]$.

但是在本文中, 我们采用 NAG 的方法来更新权重, 其权重更新的方式与上面的随机梯度下降方法相似:

$$\begin{cases} \Delta_{i+1} = \mu \Delta_i + \eta \nabla L(W_i^l + \mu \Delta_i) \\ W_{i+1}^l = W_i^l + \Delta_{i+1} \end{cases} \quad (9)$$

NAG^[23-24] 首先对 W_i^l , 执行局部更新, 计算 $W_i^l + \mu \Delta_i$, 这与 W_{i+1}^l 相似 (见图 4), 但忽略了未知的修正. 同时考虑到加上 $\mu \Delta_i$ 项立即会导致目标函数异常增加的情况, 所以在 $W_i^l + \mu \Delta_i$ 的位置对向量 Δ_i 进行梯度矫正. 这样做对于出现 $\mu \Delta_i$ 更新缓慢的情况, $\nabla L(W_i^l + \mu \Delta_i)$ 相较于 $\nabla L(W_i^l)$ 对 W_i^l 的回溯性更强, 从而相对于 CM 方法给 Δ_i 提供了一个更大且更高效的矫正. 而这种差异也促使 NAG 可以使 Δ 更新的更快, 获得更好的响应. NAG 方法在很多情况下它比 CM 方法更加稳定 (如图 5), 尤其是在 μ 比较大的情况下, 因为他可以避免 CM 方法在优化路径经过高曲率垂直方向上出现的大的震荡.

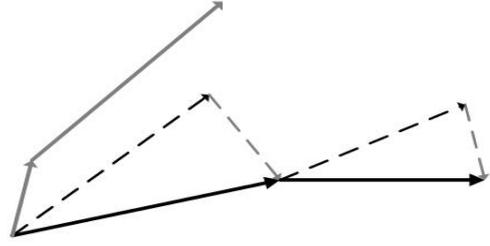


图 4 NAG 方法更新方法示意 (首先按照原有路径方向更新一个步长 (黑色虚线向量), 计算该位置的梯度值 (灰色虚线向量), 然后用这个梯度值进行修正, 得到最终的更新方向 (黑色实线向量). 图中描述了 NAG 更新两步的示意图, 其中灰色实线向量表示 CM 方法更新路径)

Fig. 4 An illustration of NAG method (which updates a step (the black dotted line vector in the figure) according to the original path direction, firstly. Then calculating the gradient value of the current position and correcting the update path (the gray dotted line vector in the figure). The black line vector is the final path of NAG and the gray line vector is the update path of CM.)

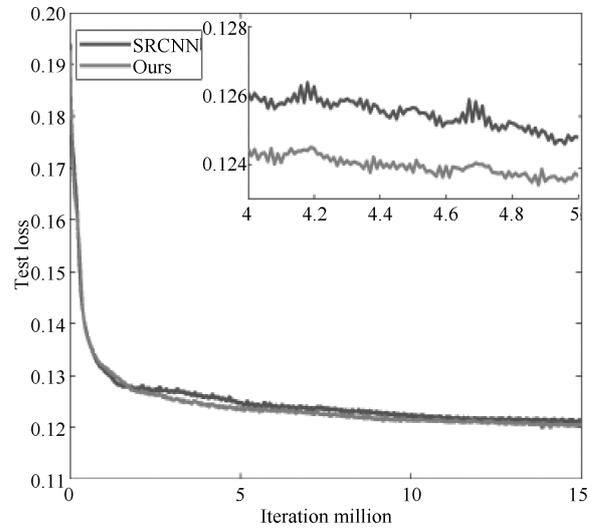


图 5 在 Set 5 测试集上, 随着迭代系数的增加, 不同方法的 Test Loss 曲线图

Fig. 5 The curve of Test Loss in Set 5 for different methods with the number of iterations increasing

2 实验结果与分析

2.1 网络初始化

在这部分工作中我们进行了大量的实验研究来验证我们改进的卷积神经网络的性能优势. 在实验中, 基本的参数设置如下: 由文献 [14] 的结论, 在速度和性能之间的权衡下, 我们在本文的实验中选择了最简单三层卷积网络 (9-1-5) 来进行实验, 即卷积层中对应的滤波器核大小分别为 $f_1 = 9, f_2 = 1, f_3 = 5$. 实验所使用的训练集均为

标准的 91 幅图像训练集^[10], 放大因子为 3, 同时在网络中设置 $n_1 = 64, n_2 = 32, \mu = 0.9$ ^[18], 网络中每一层的权重初始化为均值为 0, 标准差为 0.001 的高斯分布, 每一层的偏差我们初始化设置为 0, 在网络训练中学习率始终是固定的, 即每层学习率为 10^{-4} , 实验对图像的 R, G, B 三个通道分别进行重建, 实验平台为: 开源深度学习平台 CAFFE^[19] 及其 Matlab 接口。

2.2 综合量化评价

峰值信噪比 (Peak signal to noise ratio, PSNR) 和结构相似度 (Structural similarity, SSIM) 是广泛用来衡量图像质量的指标, 因此本文采用这两个指标 (式 (10) 和 (11)) 来客观评价各种方法的重建质量。

$$\text{PSNR} = 10 \lg \frac{MN}{\|f - \hat{f}\|^2} \quad (10)$$

$$\text{SSIM} = \frac{(2\mu_f\mu_{\hat{f}} + C_1)(\sigma_{f\hat{f}} + C_2)}{(\mu_f^2 + \mu_{\hat{f}}^2 + C_1)(\sigma_f^2 + \sigma_{\hat{f}}^2 + C_2)} \quad (11)$$

式中, M, N 为图像的尺寸, f 为真实高分辨率图像, \hat{f} 为重建后的超分辨率图像, μ_f 和 σ_f 分别为原始高分辨率图像的平均灰度值和方差, $\mu_{\hat{f}}$ 和 $\sigma_{\hat{f}}$ 分别为重建后图像的平均灰度值和方差, $\sigma_{f\hat{f}}$ 为原始高分辨率图像和重建图像的协方差, C_1, C_2 为常数。

我们首先分别定量比较了线性纠正单元 RReLU 与 ReLU 以及 NAG 与 SGD(CM) 方法在 Set 5 测试集上重建的高分辨率图像的平均 PSNR 值, 获得了令人满意的结果, 见图 6。

为了更加客观地比较算法的优劣, 我们依次用双三次插值法、ScSR^[10]、SRCNN^[14] 来重建相同的低分辨率图像。我们首先在 Set 5^[25] 测试集上进行了对比实验, 从表 1 可以看出, 我们改进后的网络相比于原网络平均 PSNR 和 SSIM 都有了明显的提升。

在 Set 14^[26] 测试集上, 我们进行了同样的实验, 得到了类似的结果, 见表 2。

图 7~10 分别展示了 Set 5 和 Set 14 部分图像的视觉效果, 通过对图像细节的放大, 可以看出我们的图像相对于双三次插值和 ScSR 重建方法在视觉效果上有较大的提升, 相比 SRCNN 方法, 我们方法重建的图像在细节上要更丰富一些。

为了探究改进后的网络相较于 SRCNN 网络学习了哪些更好的有用信息, 我们做了一些简单对比的实验。首先分别将重建的高分辨率图像与输入的低分辨率图像 (双三次插值图像) 进行比较, 获取两个网络学习到的有用信息。例如对于 Baby 图像 (部分), 图 11 (a) 和 (b) 分别显示了原有的 SRCNN 和

改进后的网络重建的 Baby 图像相对于双三次插值后的 Baby 图像, 学习后恢复的高频信息 (为了更加直观地显示这些信息, 我们对它们做了相同的增强对比度处理, 下同)。但是在这一步处理之后仍然无法直观地判断他们的优劣, 因此随后我们直接求解了两个网络重建的图像的差值, 如图 11 (c) 所示, 通过图 11 (c) 可以很直观地看到, 改进后的网络相较于原有的 SRCNN 网络学习到了更加细微的高频信息, 从而提高了图像的重建质量。同时综合比较图 11 的各个子图, 可以看出网络学习到的有用信息集中在图像纹理信息丰富的区域 (对于 Baby 图像来说, 下同, 集中于眼睛、鼻孔、帽子等), 而对于平滑区域 (脸庞等), 两个网络几乎都没有学到有用的高频信息。因此这也从另一方面给我们提供了训练集数据选择的技巧, 比如可以使用更好的特征提取方法来减少冗余信息或者剔除某些过平滑的图像块来减少计算复杂度等。

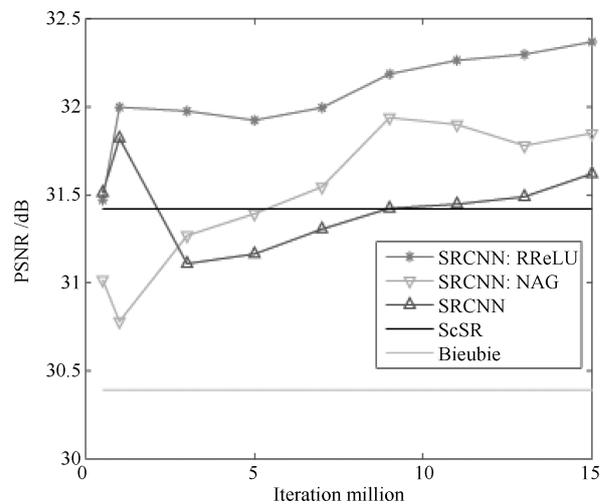


图 6 在 Set 5 测试集上, 随着迭代系数的增加, 不同方法的平均 PSNR (dB) 值的走势

Fig. 6 The average value of PSNR (dB) for different methods with the number of iterations increasing

表 1 在 Set 5 测试集上的 PSNR (dB), SSIM

Table 1 PSNR (dB) and SSIM for Set 5

图片	双三次插值	ScSR ^[10]	SRCNN ^[14]	本文方法
Baby	33.91	34.29	34.42	34.85
Bird	32.58	34.11	33.35	35.02
Butterfly	24.04	25.58	27.89	27.73
Head	32.87	33.17	31.79	33.44
Woman	28.56	29.94	30.67	30.80
Average	30.96	31.42	31.62	32.37
SSIM	0.8687	0.8821	0.8890	0.9039

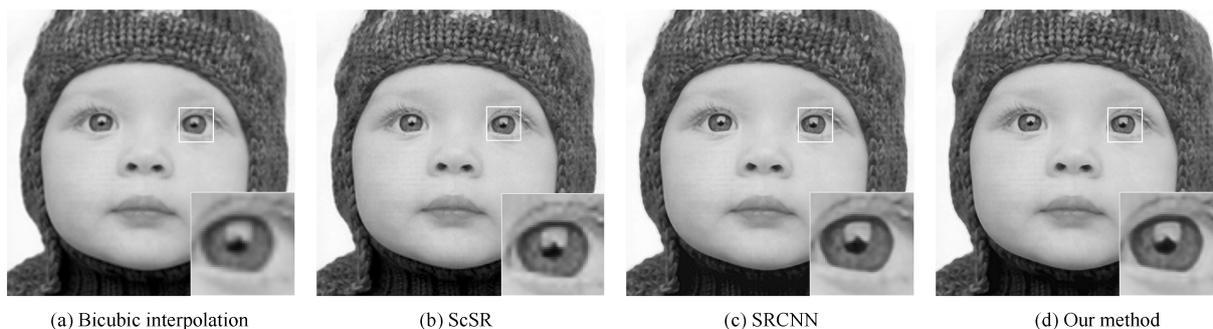


图 7 Set 5 中的 Baby_GT 重建对比图

Fig. 7 The quality of reconstruction comparison for image Baby_GT in Set 5

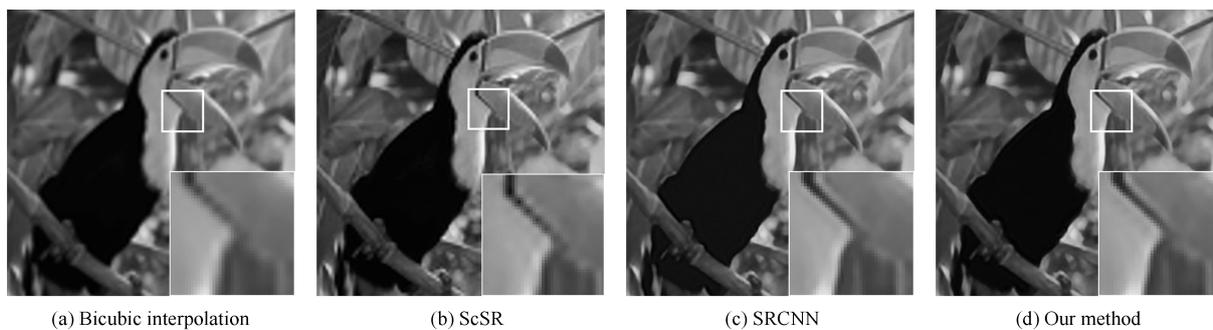


图 8 Set 5 中的 Bird_GT 重建对比图

Fig. 8 The quality of reconstruction comparison for image Bird_GT in Set 5

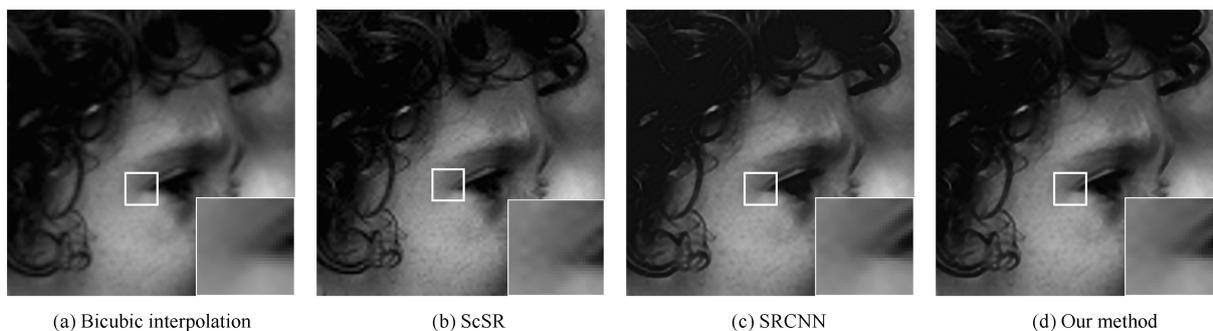


图 9 Set 14 中的 Face 重建对比图

Fig. 9 The quality of reconstruction comparison for image Face in Set 14



图 10 Set 14 中的 Pepper 重建对比图

Fig. 10 The quality of reconstruction comparison for image Pepper in Set 14

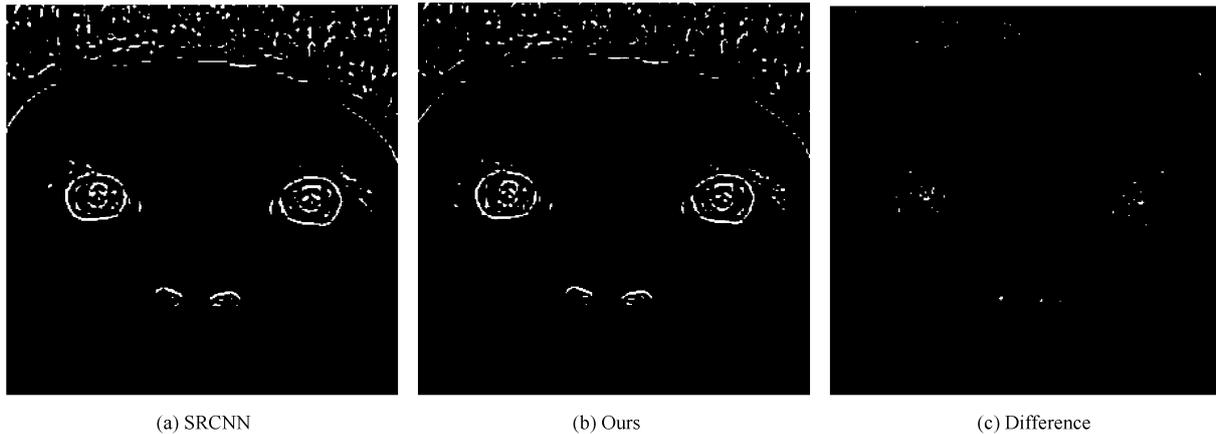


图 11 网络对 Baby 学习到的信息

Fig. 11 The information learned by network for Baby

表 2 在 Set 14 测试集上的平均 PSNR (dB), SSIM

Table 2 The average PSNR (dB) and SSIM for Set 14

图片	双三次插值	ScSR ^[9]	SRCNN ^[11]	本文方法
PSNR	27.47	28.19	28.84	28.92
SSIM	0.7722	0.7977	0.8137	0.8178

3 总结与展望

本文提出了一个改进的三层端到端的卷积神经网络去重建低分辨率图像获得对应的高分辨率图像, 我们分析了原有深度网络中使用线性纠正单元 (ReLU) 和随机梯度下降 (SGD) 来重建图像的局限性, 并提出了相应的改进. 在标准测试集 Set 5 和 Set 14 上的实验结果进一步证明了我们的改进算法相较于其他三种方法在主观视觉评价和客观评价量化上拥有更好的表现.

通过分析可知, 由于网络结构采用了通用的卷积神经网络, 所学到的“知识”全部来自训练图像, 而忽略了自然图像先验知识和退化模型等 SR 领域内的专业知识 (Domain expertise, DE), 这也是目前基于学习的方法共同面对的问题, 因此如何使用 DE 来更好地设计深度神经网络模型, 重建出更接近真实图像的高分辨率图像是下一步研究工作的方向之一.

References

- 1 Tsai R Y, Huang T S. Multiple frame image restoration and registration. *Advances in Computer Vision and Image Processing*. Greenwich: JAI, 1984. 317–339
- 2 Baker S, Kanade T. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(9): 1167–1183
- 3 Su Heng, Zhou Jie, Zhang Zhi-Hao. Survey of super-resolution image reconstruction methods. *Acta Automatica Sinica*, 2013, **39**(8): 1202–1213 (苏衡, 周杰, 张志浩. 超分辨率图像重建方法综述. *自动化学报*, 2013, **39**(8): 1202–1213)
- 4 Zhou F, Yang W M, Liao Q M. Interpolation-based image super-resolution using multisurface fitting. *IEEE Transactions on Image Processing*, 2012, **21**(7): 3312–3318
- 5 Lin Z C, Shum H Y. Fundamental limits of reconstruction-based super-resolution algorithms under local translation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(1): 83–97
- 6 Pan Zong-Xu, Yu Jing, Hu Shao-Xing, Sun Wei-Dong. Single image super resolution based on multi-scale structural self-similarity. *Acta Automatica Sinica*, 2014, **40**(4): 594–603 (潘宗序, 禹晶, 胡少兴, 孙卫东. 基于多尺度结构自相似性的单幅图像超分辨率算法. *自动化学报*, 2014, **40**(4): 594–603)
- 7 Lian Qiu-Sheng, Shi Bao-Shun, Chen Shu-Zhen. Research advances on dictionary learning models, algorithms and applications. *Acta Automatica Sinica*, 2015, **41**(2): 240–260 (练秋生, 石保顺, 陈书贞. 字典学习模型、算法及其应用研究进展. *自动化学报*, 2015, **41**(2): 240–260)
- 8 Freeman W T, Jones T R, Pasztor E C. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 2002, **22**(2): 56–65
- 9 Polatkan G, Zhou M Y, Carin L, Blei D, Daubechies I. A Bayesian nonparametric approach to image super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(2): 346–358
- 10 Yang J C, Wright J, Huang T S, Ma Y. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 2010, **19**(11): 2861–2873
- 11 Yu D, Deng L. Deep learning and its applications to signal and information processing. *IEEE Signal Processing Magazine*, 2011, **28**(1): 145–154
- 12 Yu D, Deng L, Seide F. The deep tensor neural network with applications to large vocabulary speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 2013, **21**(2): 388–396
- 13 Hutchinson B, Deng L, Yu D. Tensor deep stacking networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, **35**(8): 1944–1957

- 14 Dong C, Loy C C, He K M, Tang X O. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **38**(2): 295–307
- 15 Cui Z, Chang H, Shan S G, Zhong B E, Chen X L. Deep network cascade for image super-resolution. In: Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer, 2014. 49–64
- 16 Xu B, Wang N Y, Chen T Q, Li M. Empirical evaluation of rectified activations in convolutional network. In: Proceedings of the 32th International Conference on Machine Learning: Deep Learning Workshop. Lille, France: ICML, 2015.
- 17 Nesterov Y. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Soviet Mathematics Doklady*, 1983, **27**(2): 372–376
- 18 Sutskever I, Martens J, Dahl G, Hinton G. On the importance of initialization and momentum in deep learning. In: Proceedings of the 30th International Conference on Machine Learning. Atlanta, Georgia, USA: JMLR, 2013. 1139–1147
- 19 Jia Y Q, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T. Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia. Orlando, Florida, USA: ACM, 2014. 675–678
- 20 Nair V, Hinton G F. Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning. Haifa, Israel: ICML, 2010. 807–814
- 21 Nesterov Y. *Introductory Lectures on Convex Optimization: A Basic Course*. US: Springer, 2004. 63–66
- 22 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. In: Proceedings of the 26th Annual Conference on Neural Information Processing Systems. Lake Tahoe, USA: Curran Associates, Inc., 2012. **25**(2): 1097–1105
- 23 He K M, Zhang X Y, Ren S Q, Sun J. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015. 1026–1034
- 24 Lan G H. An optimal method for stochastic composite optimization. *Mathematical Programming*, 2012, **133**(1–2): 365–397
- 25 Bevilacqua M, Roumy A, Guillemot C, Morel M L A. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: Proceedings of the 2012 British Machine Vision Conference. Guildford, UK: University of Surrey, 2012.
- 26 Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations. In: Proceedings of the 7th International Conference on Curves and Surfaces. Avignon, France: Springer, 2010. 711–730



胡长胜 合肥工业大学硕士研究生. 2014 年获得安徽师范大学物理与电子信息学院通信工程系学士学位. 主要研究方向为图像超分辨率重建.

E-mail: hucley@mail.hfut.edu.cn

(**HU Chang-Sheng** Master student at the Hefei University of Technology. He received his bachelor degree from Anhui Normal University in 2014. His main research interest is image super-resolution.)



詹曙 合肥工业大学计算机与信息学院教授. 分别于 1990 年和 1993 年获得合肥工业大学电子工程系学士学位和硕士学位. 2000 年获得中国科学技术大学博士学位. 2002~2005 年日本东京大学, 进行博士后研究. 主要研究方向为模式识别, 计算机视觉和医学图像处理. 本文通信作者.

E-mail: shu.zhan@hfut.edu.cn

(**ZHAN Shu** Professor at the School of Computer and Information, Hefei University of Technology, China. He received his bachelor and master degrees in electronic engineering from the Hefei University of Technology in 1990 and 1993, China and the Ph.D. degree in electronic engineering from University of Science and Technology of China in 2000. He was a postdoctor at the University of Tokyo from 2002~2005, Japan. His research interest covers pattern recognition, computer vision and medical imaging. Corresponding author of this paper.)



吴丛中 合肥工业大学计算机与信息学院副教授. 主要研究方向为信号处理.

E-mail: zhanshu_hfut@163.com

(**WU Cong-Zhong** Associate professor at the School of Computer and information, Hefei University of Technology. His main research interest is signal processing.)