

改进超分辨率卷积神经网络和字典学习的图像超分辨率重构算法

张海涛, 赵 焱

(辽宁工程技术大学 软件学院 辽宁 葫芦岛 125105)
E-mail: zhaoyi199413@gmail.com

摘 要: 为了提高图像超分辨率重构效果,在保留结构化信息的同时弥补高、低分辨率图像之间的高频信息损失,本文结合深度学习方法,提出了一种基于改进超分辨率卷积神经网络和字典学习的图像超分辨率重构方法。首先使用卷积神经网络所学习到的结构化的图像信息训练一个端到端的图像超分辨率重构模型,再采用字典学习模型对图像残差部分的高频信息进行补偿,从而获得具有更好视觉效果和峰值信噪比的高分辨率图像。实验结果表明,本文算法在主观评价上比其他代表性算法更加清晰并且较好的恢复了高频信息,提高了峰值信噪比值。

关键词: 图像超分辨率重构;深度学习;卷积神经网络;字典学习

中图分类号: TP391

文献标识码: A

文章编号: 1000-1220(2018)09-2090-08

Super-resolution Reconstruction Algorithm Based on Improved Super-resolution Convolutional Neural Network and Dictionary Learning

ZHANG Hai-tao, ZHAO Yi

(College of Software Liaoning Technical University, Huludao 125105, China)

Abstract: In order to improve the image super-resolution reconstruction effect, while preserving the structured information while making up the high-frequency information loss between high and low resolution images, this paper proposes an image super-resolution reconstruction method based on improved SRCNN and dictionary learning. First, an end-to-end image super-resolution reconstruction model is trained using the structured image information learned by the convolution neural network. And then use the dictionary learning model to compensate for the high frequency information of the image residual part, obtain a high resolution image with better visual effect and peak signal to noise ratio (PSNR). The experimental results show that the proposed algorithm is more clear and better in the subjective evaluation than other representative algorithms, and the PSNR value is higher.

Key words: image super-resolution reconstruction; deep learning; convolution neural network; dictionary learning

1 引言

图像超分辨率(Super-Resolution, SR)重构的目标是根据输入的一幅或者多幅低分辨率(Low-Resolution, LR)图像来估计高分辨率(High-Resolution, HR)图像。从本质上说,由于在图像降质过程中的信息丢失,这些降质过程基本上都是不可逆的,因此图像超分辨率重构是一个不适定问题。

目前,单帧图像超分辨率重构(Single Image Super-Resolution, SISR)可分为三大类:基于插值的方法^[1-4]、基于重构的方法^[5-8]和基于学习的方法^[9-16]。在早期研究中最常使用图像插值法。根据低分辨率图像的局部统计信息和空间关系来估计高分辨率图像。典型的图像插值方法包括二次线性插值法(Bilinear)、二次立方插值法(Bicubic)和边缘定向插值法(Edge Directed)。但是,基于插值的图像超分辨率算法更倾向于将图片处理得平滑。基于重构的方法是利用信号处理技术来求解成像系统的逆过程,进而得到高分辨率图像,但是其运

算较为复杂。基于学习的图像超分辨率重构算法由于具有较好的超分辨率重构质量成为近几年的研究热点,其主要思想是通过一个包含高、低分辨率图像的外部数据集学习得到先验知识,这种先验知识通常是低分辨率图像与其对应的高分辨率图像之间的映射关系,从而在给定低分辨率图像的情况下,通过优化方法获得相应的高分辨率图像。Chang等人提出基于局部线性嵌入的邻域嵌入方法^[9](Neighbor Embedding with Locally Linear Embedding, NE + LLE),该方法假设低分辨率图像块与高分辨图像块组成的两个流形空间具有相似的局部几何结构。根据这种假设,将低分辨率输入图像块与训练集中的K近邻之间的重构关系,映射到对应的高分辨率图像块上,合成需要的高分辨率图像。Yang等人提出基于稀疏表示的图像超分辨率重构算法^[10](Image super-resolution via sparse representation, ScSR),该方法假设高、低分辨率图像信号之间的线性关系可以通过低维投影精确表示出来。根据这个假设,先通过训练集学习到高、低分辨率之间的先验知识,称为字典,对于输入的低分辨率图像根据低分辨率图像字典求取其稀疏性

收稿日期: 2017-09-07 收修改稿日期: 2018-02-26 基金项目: 国家自然科学基金项目(61172144)资助。 作者简介: 张海涛,男,1974年生,博士,教授,CCF会员,研究方向为图形图像处理、高光谱图像压缩;赵 焱,男,1994年生,硕士研究生,研究方向为图形图像处理。

系数矩阵, 再通过稀疏性系数矩阵和高分辨率图像字典重构高分辨率图像. Timofte 等人基于 NE + LLE 和 ScSR 提出了锚定邻域回归^[11] (Anchored Neighborhood Regression, ANR), 改善了算法时间效率的问题. 首先, 使用与字典原子的相关性而不是欧式距离来计算最近的邻域; 第二, 使用全局协同编码带来可观的计算提速, 将超分辨率映射简化到预计算的投影矩阵; 第三, 提出了锚定邻域回归, 将低分辨率图像块的邻域嵌入到字典中最近的原子, 并预先计算相应的嵌入矩阵.

近年来, 深度学习在图像处理以及计算机视觉领域中异常火热, 出现大量基于深度学习图像超分辨率算法. 其中 Gao 等人提出了基于深度受限玻尔兹曼机^[12] (Restricted Boltzmann machine, RBM) 的图像超分辨率重构算法, 使用深度受限玻尔兹曼机学习高、低分辨率图像之间共享的稀疏表示系数. Cui 等人提出深度网络级联^[13] (Deep Network Cascade, DNC) 的思想, DNC 是级联的多个协同局部自编码器. 在级联的每个层中, 首先进行非局部自相似搜索, 以增强输入图像块的高频纹理细节, 然后将增强的图像块输入到协同局部自编码器中, 生成高分辨率图像块. Dong 等人提出了基于卷积神经网络的图像超分辨率重构算法^[14] (Super-Resolution Convolutional Neural Network, SRCNN), 首先利用 3 层卷积神经网络学习得到有效的高、低分辨率图像块之非线性映射关系, 然后根据这种映射关系将输入的低分辨率重构为高分辨率图像. Ledig 等人提出基于对抗生成网络的图像超分辨率重构算法^[15] (Super-Resolution Using a Generative Adversarial Network, SRGAN), 该算法取得了较好的视觉效果, 但是客观评价指标却低于其他算法, 其细节部分虽然在视觉上更清晰, 但是与原始高分辨率图像有较大差距.

虽然基于学习的图像超分辨率重构算法可以利用学习到的高、低分辨率图像之间的先验知识将低分辨率图像重构为高分辨率图像, 但这种方法在图像细节的恢复上还有待加强. 本文在 SRCNN 的基础上在网络层数、卷积核大小、卷积核数量以及激活函数的选取上进行改进并结合传统学习方法中的字典学习方法对原始高分辨率图像与改进 SRCNN 网络重构图像之间的残差进行学习, 并对 SRCNN 重构图像中缺少的残差部分的高频信息进行补偿, 以取得更好的图像超分辨率重构效果.

2 基于超分辨率卷积神经网络的图像超分辨率重构

2.1 卷积神经网络简介

卷积神经网络 (Convolutional Neural Network, CNN) 是最近几年发展起来并受到高度重视的一种前馈神经网络. CNN 的二维拓扑结构专门针对具有类似网络结构的数据设计的, 例如时间序列数据或者图像数据等. 因此, 多维图像可以直接作为网络的输入, 避免传统算法中对图像复杂的预处理、特征提取等过程. CNN 的非全连接、权值共享和局部感受野的特性在图像处理方面有其独特的优越性, 减少了网络训练所需的参数, 降低了网络复杂性, 并且这种特性对于位移、比例缩放和旋转等形式的变形具有高度不变性. 所以 CNN 在图像处理中得到广泛应用.

CNN 是一个多层神经网络, 其每层都由一组二维平面构成, 且每个平面包含多个独立神经元. CNN 的层级结构一般包含卷积层、激励层、池化层和全连接层. 卷积层的作用是对输入数据进行特征提取. 激励层通过激励函数添加非线性因素, 将线性映射转化成非线性映射. 因为通常每个卷积层后都连接激励层, 所以常把连接的卷积层和激励层视为一层卷积层对待. 池化层对卷积层得到特征进行降采样提取局部敏感特征, 一方面将特征压缩降低网络计算复杂性, 另一方面进行特征压缩提取主要特征. 全连接层将前面获取到的局部特征汇总为全局特征, 并且最后的全连接层作为输出层执行分类或回归等任务. CNN 通常在一个或几个卷积层后会连接一个池化层, 进而级联成深度神经网络, 最后由全连接层输出. 经典的卷积神经网络 LeNet-5 结构如图 1 所示.

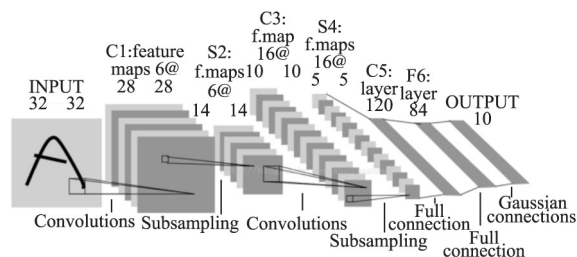


图 1 LeNet-5 网络结构

Fig. 1 LeNet-5 network structure

2.2 基于超分辨率卷积神经网络的图像超分辨率重构模型

作为一种基于学习的单帧图像超分辨率算法, 超分辨率卷积神经网络 (Super-Resolution Convolutional Neural Network, SRCNN) 旨在学习一个以低分辨率图像为输入, 高分辨率图像为输出的端到端的非线性映射. 这里将原始高分辨率图像设为 X , 将 X 降采样后用 Bicubic 插值恢复到与 X 相同大小的低分辨率图像设为 Y , 通过 SRCNN 学习到的非线性映射关系设为 F , 使得 $F(X)$ 与原始超分辨率图像 X 尽可能相似. SRCNN 网络结构包含 3 个卷积层, 相当于一个全卷积神经网络. 由于超分辨率重构需要更好的保存图像特征信息并且其输出也是不需要降维的完整图像, 因此不需要池化层和全连接层. 前两层使用激励函数为 ReLu (Rectified Linear Unit), 最后一层不使用激励函数. ReLu 函数公式如下:

$$y = \max(0, x) \quad (1)$$

SRCNN 这 3 个卷积层, 每个层负责特定的任务, 分别于传统的 SISR 算法中三个过程相对应:

图像块提取和表示: 该操作从低分辨率图像 Y 中提取多个特征块, 并将每个特征块表示为高维度向量, 这些向量构成第一层的特征图.

非线性映射: 这个操作是非线性地将来自第一层的每个高维向量映射到另一个高维向量上. 每个映射向量都是一个高分辨率图像块的表示. 这些映射向量构成了第二层的特征图.

重构: 这个操作集合了上一层的高分辨率图像块, 来生成最终的高分辨率图像. 这幅图像与原始高分辨率图像 X 近似.

上述操作在网络中的位置以及 SRCNN 整体网络结构如图 2 所示. 接下来详细介绍每层的具体操作.

特征块提取和表示阶段: 在传统算法中这一步骤通常使用 PCA、DCT 等操作来表示它们, 这相当于通过一组卷积核对图像进行卷积, 而这些卷积核是由人们设定好的模板. 在 SRCNN 中将模板加入网络优化中, 由网络训练得到, 设本层操作为 F_1 :

$$F_1(Y) = \max(0, W_1 \times Y + B_1) \quad (2)$$

其中, W_1 表示大小为 $c \times f_1 \times f_1 \times n_1$ 的卷积核, c 表示输入的低分辨率图像的通道数, f_1 表示卷积核的尺寸, n_1 表示卷积核的个数, B_1 表示维度 n_1 为偏置向量.

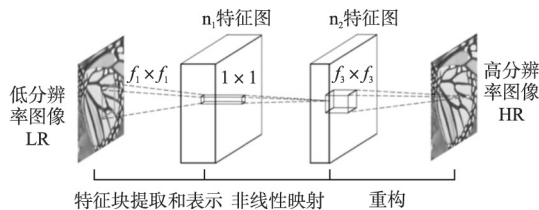


图2 SRCNN 网络结构

Fig.2 SRCNN network structure

非线性映射阶段: 在上一层对每个图像块提取到一个 n_1 维的特征图. 在本层中将这 n_1 维特征图非线性映射为 n_2 维的特征图. 设本层操作为 F_2 :

$$F_2(Y) = \max(0, W_2 \times F_1(Y) + B_2) \quad (3)$$

其中, W_2 表示大小为 $n_1 \times 1 \times 1 \times n_2$ 的卷积核, n_1 表示第一层特征图维度, n_2 表示第二层特征图维度, B_2 表示维度 n_2 为偏置向量.

重构: 通过求取重叠图像块的平均值来生成最终的高分辨率图像. 求取平均值也可视为使用卷积核对图像进行卷积. 因此, 设本层操作为 F_3 :

$$F_3(Y) = W_3 \times F_2(Y) + B_3 \quad (4)$$

其中, W_3 表示大小为 $n_2 \times f_3 \times f_3 \times c$ 的卷积核, n_2 表示第二层特征图维度, f_3 表示卷积核尺寸, c 表示输出的高分辨率图像的通道数, B_3 表示维度 c 为偏置向量. 由于求取平均值的操作是一种线性操作, 因此本层并未使用激活函数增加其非线性因素.

2.3 对超分辨率卷积神经网络的改进

SRCNN 网络的三层结构, 第一层选择卷积核尺寸 f_1 为 9, 卷积核个数 n_1 为 64; 第二层选择卷积核尺寸为 1, 卷积核个数 n_2 为 32; 第三层选择卷积核尺寸 f_3 为 5, 卷积核个数等于输出图像通道数, 即 9-1-5 的网络结构. 这种网络结构在实际使用中, 在网络训练阶段存在损失函数收敛速度差的情况. 为使 SRCNN 网络达到更好的重构效果, 本文从网络层数、卷积核尺寸与数量以及激活函数的选择上对网络进行改进, 优化网络的超分辨率重构效果.

深度学习方法一般情况下优于传统机器学习方法, 因为深度学习通常具有更深的网络层次结构. 对于深度学习来说, 这些更深网络层次结构可以提供更多可供训练学习的参数, 因此得到的效果往往更好. 对于 CNN 在 SISR 上的应用, 三层卷积的神经网络也许不是最优方案. 参考 AlexNet^[17]、VGG-Net^[18] 和 GoogleNet^[19] 等网络结构, 它们层数已经达到十几层甚至几十层之多. 因此, 我们也可以通过增加 CNN 的层次来提高图像超分辨率重构效果. 但通常上述网络针对的是分

类任务, 在网络中存在多个池化层, 通过对特征压缩从而降低网络计算复杂度. 对于 SRCNN 使用的全卷积神经网络, 过高的层数意味着有更多参数需要训练学习, 无疑增大了训练时间. 通过实验, 增加卷积层数到 4 层的网络模型是可行的, 训练时间可以接受. 如果继续增加层数, 训练网络所花费的时间代价较高. 因此, 我们选择 4 层的卷积神经网络来进行图像超分辨率.

加深卷积神经网络的方法除了增加层数, 也可以通过增加卷积层卷积核的个数、卷积核的尺寸等方法. 由于 CNN 是深度学习领域中一个著名的“黑盒”模型, 各参数均由网络训练迭代优化而来. 因此, SRCNN 的 3 层与传统 SISR 的 3 个步骤并非完全对应. 由于 SRCNN 的第一层和第二层都使用了激活函数增加非线性因素, 非线性映射并不仅仅针对第二层, 而是通过每层的激活函数逐步进行的, 因此 SRCNN 在第二层只使用较少参数的 1×1 卷积核仅仅是为了非线性映射而降低了对于特征进一步提取的能力. 我们可以增大第二层的卷积核尺寸. 在我们的 4 层卷积神经网络中, 每层卷积核尺寸分别为 9×9 、 7×7 、 7×7 和 5×5 , 即 9-7-7-5 的网络结构. 在卷积核数量上, 为了使提取到的特征逐步整合为高分辨率图像, 参考 SRCNN 逐步递减的方式, 每层卷积核数量分别为 64、32、16 和 c , c 表示输出高分辨率图像的通道数.

在激活函数的选择上, 我们使用比 ReLU 更好的 PReLU^[20] (Parametric Rectified Linear Unit) 作为激活函数, PReLU 即带参数的 ReLU, 函数公式如下:

$$y = \max(0, x) - \alpha \times \min(0, x) \quad (5)$$

其中, α 参数可由网络训练获得. 使用 PReLU 作为激活函数的优点是, 仅仅增加了少量的参数, 便可达到更好的效果, 同时网络的计算复杂性和网络过拟合的危险性仅仅小幅度增加.

修改后的网络各层操作分别为 F_1 、 F_2 、 F_3 和 F_4 , 其公式如下:

$$F_1(Y) = \max(0, W_1 \times Y + B_1) + \alpha_1 \times \min(0, W_1 \times Y + B_1) \quad (6)$$

$$F_2(Y) = \max(0, W_2 \times F_1(Y) + B_2) + \alpha_2 \times \min(0, W_2 \times F_1(Y) + B_2) \quad (7)$$

$$F_3(Y) = \max(0, W_3 \times F_2(Y) + B_3) + \alpha_3 \times \min(0, W_3 \times F_2(Y) + B_3) \quad (8)$$

$$F_4(Y) = W_4 \times F_3(Y) + B_4 \quad (9)$$

其中, W_1 、 W_2 、 W_3 和 W_4 分别表示大小为 $c \times f_1 \times f_1 \times n_1$ 、 $n_1 \times f_2 \times f_2 \times n_2$ 、 $n_2 \times f_3 \times f_3 \times n_3$ 和 $n_3 \times f_4 \times f_4 \times c$ 的卷积核, c 表示输入的低分辨率图像的通道数, f_1 、 f_2 、 f_3 和 f_4 表示各层卷积核的尺寸, n_1 、 n_2 和 n_3 表示各前三层卷积核的个数, B_1 、 B_2 、 B_3 和 B_4 表示各层偏置向量, α_1 、 α_2 和 α_3 表示前三层中非线性映射参数.

根据上文网络结构, 学习端到端的映射函数 F 所需要的网络参数为 $\theta = \{W_1, W_2, W_3, W_4, B_1, B_2, B_3, B_4, \alpha_1, \alpha_2, \alpha_3\}$, 通过随机梯度下降法与反向传播法^[21] (Back-Propagation, BP) 优化参数, 使重构图像 $F(Y; \theta)$ 和其对应的原始高分辨率图像之间的损失函数达到最小值, 此时的参数即为所需的参数. 给定一组高分辨率图像集 $\{X_i\}$ 和一组与之对应的降质后的低分辨率图像集 $\{Y_i\}$, 损失函数选择均方误差 (MSE), 其公如下:

$$L(\theta) = \frac{1}{n} \sum_{i=1}^n \|F(Y_i; \theta) - x_i\|^2 \quad (10)$$

其中 n 为训练样本个数。

3 基于字典学习的残差补偿

3.1 字典学习

字典学习^[22] (Dictionary Learning) 也可称为稀疏表示 (Sparse Representation)。在稀疏表示方面的成就表明, 高维信号之间的线性关系可以从其低维度投影中准确地恢复。其基本原理为, 自然图像可以视为稀疏信号, 当用一组完备字典线性表达出输入信号时, 在满足一定稀疏度的情况下, 展开系数可以获得对原始输入信号的良好近似。该方法在图像复原与超分辨率重构以及图像降噪中取得良好表现。通过矩阵分解的角度看待字典学习的过程: 给定样本数据集 $X = \{x_1, x_2, \dots, x_n\}$, 字典学习的目标是把 X 矩阵分解为 D, Z 矩阵:

$$X \approx D \cdot Z \quad (11)$$

其中 $D = \{d_1, d_2, \dots, d_k\}$ 称之为字典, D 的每一列 d_k 称之为原子, 同时 D 的每一个原子都是一个归一化向量。 $Z = \{z_1, z_2, \dots, z_n\}$ 称对应稀疏表示的系数矩阵, 同时 X 要求尽可能稀疏。字典学习优化问题的目标函数可以表示为:

$$\min_{D, Z} \{ \|X - D \cdot Z\|_2^2 + \lambda \cdot \|Z\|_0 \} \quad (12)$$

其中 X 表示训练数据集, D 表示学习的字典, Z 为所对应稀疏表示系数矩阵。但是由于这种表示中 l_0 正则项难以求解, 所以很多时候使用 l_1 正则项近似替代。因为目标函数中存在两个未知变量 D, Z , 构造字典的算法通常分为稀疏表示和字典更新两个步骤。

稀疏表示: 首先设定一个初始化的字典, 用该字典对给定数据进行稀疏表示, 得到系数矩阵 Z 。 d_i 表示字典 D 的列向量, z_i 表示稀疏表示系数矩阵 Z 的行向量。即:

$$D \cdot Z = \sum_{i=1}^K d_i z_i \quad (13)$$

字典更新: 初始字典通常不是最优的, 满足稀疏性的系数矩阵表示的数据和原数据会有较大误差, 我们需要在满足稀疏度的条件下逐行逐列对字典进行更新优化, 减小整体误差, 逼近可使用的字典。忽略字典中第 k 项 d_k 的影响, 计算当前稀疏表示误差矩阵:

$$E = X - \sum_{i \neq k} d_i z_i \quad (14)$$

MOD^[23] 和 K-SVD^[24] 等是字典学习的经典算法。这里使用 K-SVD 算法, K-SVD 的求解是一个迭代过程。首先, 假设字典 D 是固定的, 用匹配追踪 (Matching Pursuit, MP) 或者正交匹配追踪^[25] (Orthogonal Matching Pursuit, OMP) 等贪婪算法, 可以得到在字典 D 上 X 的稀疏表示的系数矩阵 Z , 然后让 Z 固定, 根据 Z 更新字典 D , 循环此操作直到误差收敛为止。与 MOD 不同, K-SVD 字典学习中字典 D 的更新不是整体进行而是逐列进行的。首先假设系数矩阵 Z 和字典 D 都是固定的, 将要更新的是字典的第 k 列 d_k , 系数矩阵 Z 中 d_k 对应第 k 行为 z_T^k , 则:

$$\begin{aligned} \|X - D \cdot Z\|_2^2 &= \|X - \sum_{j=1}^n d_j z_T^j\|_2^2 = \|(X - \sum_{j \neq k} d_j z_T^j) - d_k z_T^k\|_2^2 \\ &= \|E_k - d_k z_T^k\|_2^2 \end{aligned} \quad (15)$$

得到当前误差矩阵 E_k 后, 我们只需要调整 d_k 和 z_T^k , 使其乘

积与 E_k 的误差尽可能的小。

3.2 基于字典学习对改进超分辨率卷积神经网络模型的残差补偿

基于改进 SRCNN 的图像超分辨率重构算法可以有效恢复图像的大部分信息, 但在恢复图像细节信息 (如边缘细节等锐变部分) 方面还有不足。将训练集中原始高分辨率图像 X 与通过基于改进 SRCNN 图像超分辨率算法重构的图像 \tilde{X} 相减, $\bar{X} = X - \tilde{X}$ 所得图像 \bar{X} 称为残差图像, 残差图像包含了重构图像 \tilde{X} 中没有得到恢复的高频信息。为了补偿这些在基于改进 SRCNN 的图像超分辨率重构中没有得到恢复的高频信息, 提高重构图像质量, 最初我们依旧使用 CNN 针对残差图像进行训练, 但是训练中出现网络不收敛现象, 实际效果并不理想。因此, 受文献 [26, 27] 启发, 本文将字典学习引入其中, 即结合 SRCNN 与字典学习进行图像超分辨率重构。

引入字典学习模型, 针对已经通过基于改进 SRCNN 的图像超分辨率算法获得的重构图像 \tilde{X} 与残差图像 \bar{X} , 探究它们之间的相关性联系, 再通过这种联系重构图像中通过 SRCNN 重构阶段没有恢复出来的高频信息。

在本文中, 针对残差图像的字典学习优化模型如下:

$$\min_{\tilde{D}, Z} \sum_{i=1}^i (\|\tilde{x}_i - \tilde{D} \cdot Z_i\|_2^2 + \|\bar{X}_i - \bar{D} \cdot Z_i\|_2^2) + \lambda \cdot \|Z_i\|_1 \quad (16)$$

其中 \tilde{X} 为通过基于 SRCNN 的图像超分辨率算法获得的重构图像, \tilde{D} 为该重构结果图像所对应的字典; \bar{X} 原始高分辨率图像与改进 SRCNN 重构图像的残差, \bar{D} 为残差图像所对应的字典; Z 为对应稀疏表示的系数矩阵; λ 为用调节模型误差与稀疏度的参数。

通过上文所述的 K-SVD 方法训练得到重构图像对应的字典 \tilde{D} 和残差图像对应的字典 \bar{D} 。对于给定基于改进 SRCNN 图像超分辨率算法重构后的图像, 可通过匹配追踪和正交匹配追踪等贪婪算法估计残差图像 \bar{X} , 即基于改进 SRCNN 图像超分辨率算法的重构图像所缺失的高频信息部分, 得到近似残差图像 \bar{X} 。

由于该字典学习方法有益于高频信息的恢复, 本文将卷积神经网络与字典学习相结合, 应用于图像超分辨率重构中, 本文方法总体框架如图 3 所示。

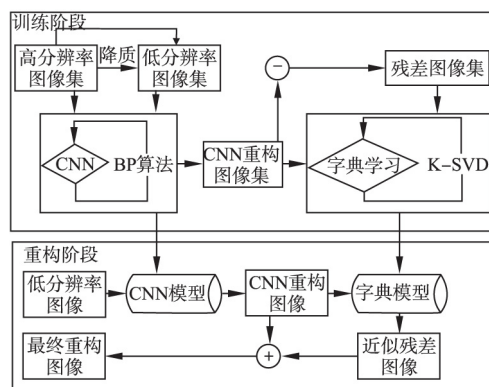


图3 本文方法总体框架图

Fig. 3 Overall framework of proposed method

在训练阶段, 将高、低分辨率图像集 $\{X_i\}, \{Y_i\}$ 作为训练集, 使用 CNN 训练得到 CNN 模型, 包括 CNN 中每层参数, 再

将原始高分辨率图像集与 CNN 重构图像集 $\{\tilde{X}_i\}$ 做差,得到残差图像集 $\{\tilde{X}_i'\}$. 再将 CNN 重构图像集和残差图像集通过字典学习得到字典模型,包括重构图像对应的字典 \tilde{D} 、残差图像对应的字典 \tilde{D}' 以及稀疏表示系数矩阵 Z .

在图像重构阶段,对于低分辨率测试图像 Y ,首先使用 CNN 对其进行重构,得到 CNN 重构图像 \tilde{X} ,再将该结果图像 \tilde{X} 通过字典模型估计近似残差图像 \tilde{X}' ,用来对 CNN 重构图像中缺失的高频信息进行补偿.最后,该方法的最终重构图像 $X' = \tilde{X} + \tilde{X}'$,即 CNN 重构图像 \tilde{X} 与近似残差图像 \tilde{X}' 的融合图像.

4 实验及分析

4.1 实验设置

本文的实验环境: CPU 为 Intel(R) Core(TM) i5-7500 CPU @ 3.40GHz; GPU 为 NVIDIA GeForce GTX 1060 6GB; 内存为 8GB; 操作系统为 Ubuntu 16.04 LTS. 其中卷积神经网络的训练在深度学习框架 Caffe 上实现,其他部分均在 Matlab 2017a 平台上完成.

在数据集的选取上,训练集选择与 SRCNN 相同的 Timofte 数据集(包含 91 幅自然图像),测试集选择在图像超分辨率重构领域中常用的两个数据集 Set5 与 Set14,分别包含 5 张图像与 14 张图像.在数据集处理上,本文使用与通常的 SISR 研究中相同的方法获取低分辨率图像,即对原始高分辨率图像进行降采样至其 $1/k$ 之后,再使用 Bicubic 插值方法放大 k 倍,得到与原始高分辨率图像相同尺寸的低分辨率图像.在使用卷积神经网络和字典学习训练图像超分辨率重构模型时,输入数据为高、低分辨率图像块.在卷积神经网络中以步长为 10 像素的滑动窗口在图像上截取大小 20×20 像素的具有重叠区域的图像块.由于卷积神经网络中的卷积核大小影响,导致输出图像小于输入图像,因此要对低分辨率图像块进行填充,填充后大小为 38×38 像素.在字典学习中,直接将 CNN 重构图像与残差图像分解成 9×9 的图像块.

本文使用 YCbCr 颜色空间,由于大多数图像超分辨率算法都是对 YCbCr 颜色空间中的亮度分量(Y)敏感,通常是将图像转换到 YCbCr 颜色空间,对 Y 分量进行图像超分辨率重构后在与原 Cb、Cr 分量合并得到重构的高分辨率图像.因此,本文在实验过程中仅考虑 Y 分量的单通道,但本文也适用多通道.

考虑到单通道,本文卷积神经网络模型中各项参数为: $c=1$, $\mu_1=64$, $f_1=9$, $\mu_2=32$, $f_2=7$, $\mu_3=16$, $f_3=7$, $f_4=5$, 基础学习率为 0.0001.在字典学习中,原子个数为 500.

4.2 实验结果与分析

为验证本文算法效果,在对比实验中,选取如下算法作为对比算法:图像超分辨率重构中的经典算法以及多数图像超分辨重建算法的预处理算法 Bicubic 插值算法;基于稀疏表示的图像超分辨率重建算法中的经典算法 SeSR^[10];结合邻域嵌入与局部线性嵌入的基于外部样例学习的图像超分辨率重建算法 NE+LLE^[9];结合了稀疏表示与邻域嵌入的基于外部样例学习的图像超分辨率重建算法 ANR^[11];以及本文的第一阶段原型算法,使用深度学习模型中卷积神经网络的 SRC-

NN^[14].

客观评价效果:按照图像超分辨率重构领域中常用实验方法,将上述算法同本文算法分别在放大系数 k 为 2、3、4 倍进行对比.主要客观评价指标为峰值信噪比(Peak Signal to Noise Ratio, PSNR),次要客观评价指标为结构相似性(Structural Similarity, SSIM).由于篇幅所限,仅在 Set5 数据集上放大系数为 3 倍的图像超分辨率重构中将每张图像的 PSNR 评价指标列出,其他不同倍数、不同评价指标的对比将按照数据集,以该数据集图像的平均值做对比.实验结果如表 1、表 2 和表 3 所示.

表 1 Set5 上 PSNR 对比结果($k=3$)

Table 1 PSNR contrast results on Set5 ($k=3$)

Image	Bicubic	SeSR	NE+LLE	ANR	SRCNN	本文
Baby	33.91	34.4	35.06	35.13	35.01	35.06
Bird	32.58	34.1	34.56	34.60	34.91	35.50
Butterfly	24.04	25.6	25.75	25.90	27.58	28.28
Head	32.88	33.2	33.60	33.63	33.55	33.64
Woman	28.56	29.9	30.22	30.33	30.92	31.26

对比上述客观评价实验数据,本文算法取得了较高的 PSNR 与 SSIM 的平均值,对比 SRCNN 大部分图像均有所提

表 2 Set5 上不同放大系数的 PSNR 与 SSIM 均值对比结果

Table 2 Comparison of PSNR and SSIM mean of different amplification coefficients on Set5

	k	Bicubic	SeSR	NE+LLE	ANR	SRCNN	本文
PSNR	2	33.66	-	35.77	35.83	36.34	36.53
	3	30.39	31.44	31.84	31.92	32.39	32.75
	4	28.42	-	29.61	29.69	30.08	30.36
SSIM	2	0.929	-	0.949	0.949	0.952	0.953
	3	0.868	0.882	0.895	0.896	0.900	0.909
	4	0.801	-	0.840	0.841	0.852	0.860

升,但是在部分图像上提升并不明显,如 Set5 上的 baby 图像.这是因为一些图像纹理精细、细节微小,在降采样之后损失大量有效信息,以至于重构效果提升不明显,下文主观对比

表 3 Set14 上不同放大系数的 PSNR 与 SSIM 均值对比结果

Table 3 Comparison of PSNR and SSIM mean of different amplification coefficients on Set14

	k	Bicubic	SeSR	NE+LLE	ANR	SRCNN	本文
PSNR	2	30.23	-	31.76	31.80	32.18	32.31
	3	27.54	28.31	28.60	28.65	29.00	29.24
	4	26.00	-	26.81	26.85	27.20	27.42
SSIM	2	0.868	-	0.899	0.900	0.903	0.904
	3	0.773	0.795	0.807	0.809	0.815	0.820
	4	0.701	-	0.733	0.735	0.741	0.749

中 bird 图像,鸟眼内部瞳孔的还原没有得到较好改进也是这个原因.而对比不同放大系数 k 的实验数据可以发现,在放大系数 k 为较大倍数(2、3)时,PSNR 值和 SSIM 值提升比放大系数 k 为较小倍数时明显,这是因为将原始图像降采样到其 $1/2$ 时,其图像信息保存相对较完整,大部分图像超分辨率算

法均可以较好的对其进行重构, 残差相对较小, 而本文中的字典学习方法是针对残差图像进行学习的, 在使用字典学习方法进一步对残差中的高频分量进行补偿时, 可能会引入一些不必要的冗余信息, 因此, 限制了其分辨率重构效果。

主观评价效果: 将本文算法得到的超分辨率重构图像同上述其他算法得到超分辨率重构图像进行对比, 选取放大系数 $k=3$ 上的三张图像进行主观对比, 结果如图 4、图 5 以及图 6 所示。

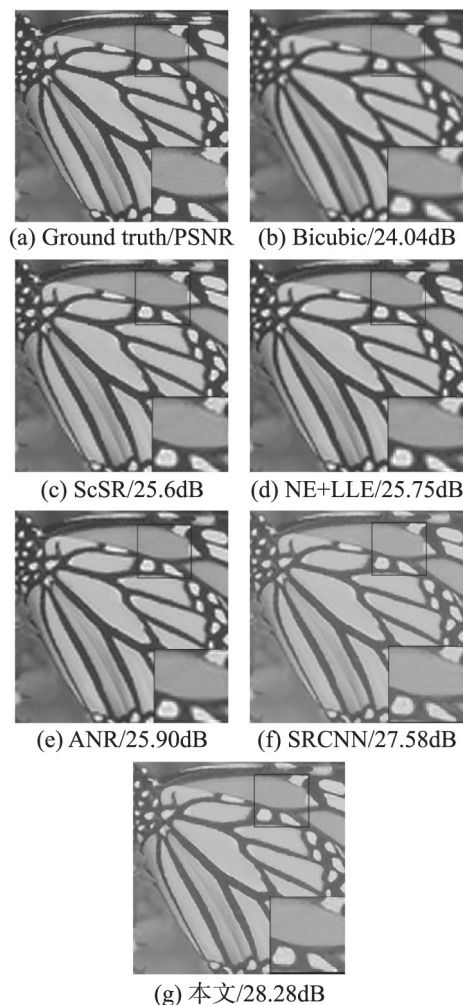


图 4 butterfly 图像主观对比结果

Fig. 4 Subjective contrast results of butterfly images

图 4 展现了不同算法在 butterfly 图像上的放大 3 倍的超分辨率重构与结果, 从放大区域可以看出, 本文算法对该图像超分辨率重构效果较好, 在边缘等细节上的重构效果要好于其他算法, 更加接近原始高分辨率图像, 同时图像中黑色部分噪声较少, 更为纯净, 而传统算法的结果图中可以看出边缘较为模糊。

图 5 展现了不同算法在 bird 图像上放大 3 倍的超分辨率重构与结果, 从放大区域可以看出, 虽然本文算法的超分辨率重构效果及边缘上的表现都好于其他算法, 如鸟嘴边缘的黑色羽毛处的重构, 但是对比本文算法结果图与原始高分辨率图像, 在鸟眼内部的细节还原上还欠佳, 同其他算法一样鸟眼内部的瞳孔并未得到还原。

图 6 展现了不同算法在 lena 图像上放大 3 倍的超分辨率重构与结果, 通过放大图像中的帽檐边缘可以看出, 本文算法得到的结果图像边缘更加锐利, 更接近原始高分辨率图像。

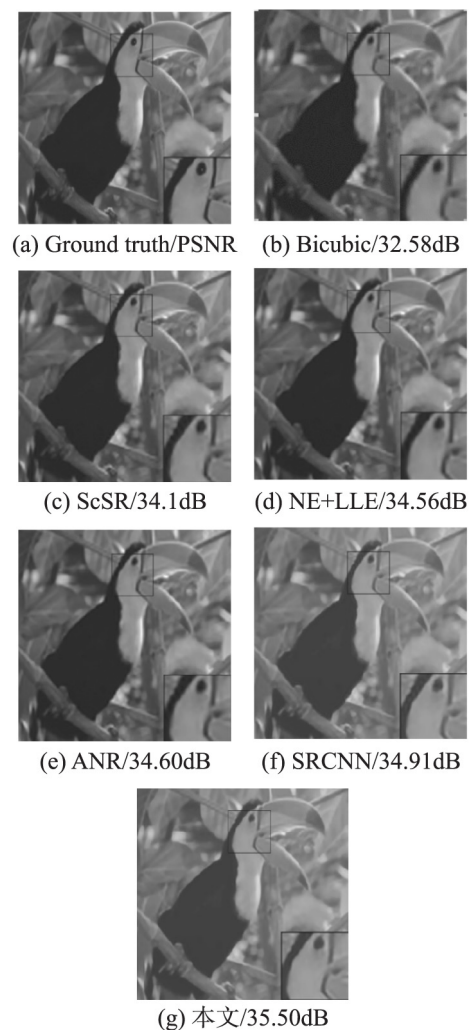


图 5 bird 图像主观对比结果

Fig. 5 Subjective contrast results of bird images

在时间效率方面, 对比算法中只有 SRCNN 与本文算法是深度算法, 与其他非深度算法对比训练时间效率意义不大, 因为在使用训练好的模型时只需执行一次, 深度算法需要花费大量时间训练, SRCNN 与本文算法通常训练需要迭代 100 万次能得到理想结果, 在输入图像块大小均为 38×38 , 使用相同的 GPU 环境下训练, 每进行 100 次迭代, SRCNN 花费时间平均为 1.8s, 本文算法花费时间平均为 3.2s, 因此每次迭代相差 0.014s, 以迭代 100 万次为训练完成计算, 本文算法在训练阶段比 SRCNN 需多花约 4 小时, 在使用模型进行重建时, 将本文算法与其他对比算法在 Set5 数据集上相同 CPU 环境下运行时间进行对比, 如表 4 所示, 本文算法要快于 ScSR 略慢与 SRCNN, 其中 Bicubic 为其他算法基础操作且运行速度约为 0.01s。

通过上文主观以及客观的评价效果, 与传统算法相比本文算法效果提升明显, 表明卷积神经网络可以学习到有效的高、低分辨率图像之间的映射关系, 同时与 SRCNN 相比, 本

文也取得了效果上的提升,表明基于字典学习的残差补偿方法能补偿 SRCNN 中未能恢复的高频信息,但时间成本也会因此增加.本文所提出的基于改进 SRCNN 和字典学习的算法能恢复更为丰富的细节信息和分明的图像边缘,得到视觉效果更好的超分辨率重构图像.



图6 lena 图像主观对比结果

Fig. 6 Subjective contrast results of lena images

表4 Set5 上时间效率对比结果($k=3$)

Table 4 Results of time and efficiency comparison on Set5 ($k=3$)

Image	Bicubic	ScSR	NE + LLE	ANR	SRCNN	本文
Baby	0.0s	24.6s	7.4s	1.0s	13.6s	15.7s
Bird	0.0s	7.3s	2.4s	0.3s	3.3s	4.1s
Butterfly	0.0s	5.7s	1.9s	0.3s	2.3s	4.6s
Head	0.0s	8.6s	2.2s	0.3s	2.4s	4.4s
Woman	0.0s	7.8s	2.1s	0.3s	2.4s	4.2s

5 结束语

本文提出了一种基于改进 SRCNN 和字典学习的图像超分辨率重构算法,该算法在超分辨率卷积神经网络的基础上,

对其网络结构进行改进,并针对原始高分辨率图像与 SRCNN 输出的重构图像的残差采用传统的字典学习方法进行学习,实现了高频信息补偿,具有较优的视觉效果及客观评价指标.在图像超分辨率的图像质量上,本文算法取得了一定研究成果,对比 SRCNN 在 Set5 和 Set14 测试集上平均 PSNR 分别提升了 0.36dB 和 0.24dB,但在网络训练时会消费较多时间,因此本文下一步将在保证图像质量的前提下研究如何提升算法的时间效率.

References:

- [1] Tsai R, Huang T. Multiple frame image restoration and registration [C]. Advances in Computer Vision and Image Processing, Greenwich, CT: JAI Press Inc, 1984: 317-339.
- [2] Komatsu T, Aizawa K, Igarashi T, et al. Signal-processing based method for acquiring very high resolution images with multiple cameras and its theoretical analysis [J]. Communications Speech & Vision Iee Proceedings I, 1993, 140(1): 19-24.
- [3] Ur H, Gross D. Improved resolution from subpixel shifted pictures [J]. Cvgip Graphical Models & Image Processing, 1992, 54(2): 181-186.
- [4] Komatsu T, Igarashi T, Aizawa K, et al. Very high resolution imaging scheme with multiple different-aperture cameras [J]. Signal Processing Image Communication, 1993, 5(5-6): 511-526.
- [5] Tappen M F, Russell B C, Freeman W T. Exploiting the sparse derivative prior for super-resolution and image demosaicing [C]. In IEEE Workshop on Statistical and Computational Theories of Vision 2003.
- [6] Kim K I, Kwon Y. Single-image super-resolution using sparse regression and natural image prior [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2010, 32(6): 1127-1133.
- [7] Dai S, Han M, Xu W, et al. Soft edge smoothness prior for alpha channel super resolution [C]. Computer Vision and Pattern Recognition, CVPR'07. IEEE Conference on. IEEE, 2007: 1-8.
- [8] Dai S, Han M, Xu W, et al. Softcuts: a soft edge smoothness prior for color image super-resolution [J]. IEEE Transactions on Image Processing, 2009, 18(5): 969-981.
- [9] Chang H, Yeung D Y, Xiong Y. Super-resolution through neighbor embedding [J]. Proc Computer Vision & Pattern Recognition, 2004, 1: 275-282.
- [10] Yang J, Wright J, Huang T S, et al. Image super-resolution via sparse representation. [J]. IEEE Transactions on Image Processing a Publication of the IEEE Signal Processing Society, 2010, 19(11): 2861-2873.
- [11] Timofte R, De V, Gool L V. Anchored neighborhood regression for fast example-based super-resolution [C]. IEEE International Conference on Computer Vision, IEEE Computer Society, 2013: 1920-1927.
- [12] Gao J, Guo Y, Yin M. Restricted boltzmann machine approach to couple dictionary training for image super-resolution [C]. Image Processing (ICIP), 2013 20th IEEE International Conference on. IEEE, 2013: 499-503.
- [13] Cui Z, Chang H, Shan S, et al. Deep network cascade for image super-resolution [M]. Computer Vision-ECCV 2014, Springer International Publishing, 2014: 49-64.
- [14] Dong C, Chen C L, He K, et al. Learning a deep convolutional net-

- work for image super-resolution [M]. Computer Vision-ECCV 2014, Springer International Publishing 2014: 184-199.
- [15] Ledig C, Theis L, Huszar F, et al. Photo-realistic single image super-resolution using a generative adversarial network [C]. IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, 2017: 105-114.
- [16] Wang Zhao-hui, Chen Long, Jiao Bin-liang. Image super-resolution reconstruction algorithms based on nonlocal self-similarities and classification dictionaries learning [J]. Journal of Chinese Computer Systems 2015, 36(7): 1617-1619.
- [17] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]. International Conference on Neural Information Processing Systems, Curran Associates Inc. 2012: 1097-1105.
- [18] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. ArXiv Preprint arXiv: 1409.1556 2014.
- [19] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C]. Computer Vision and Pattern Recognition, IEEE 2015: 1-9.
- [20] 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 2015: 1026-1034.
- [21] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors [J]. Nature, 1986, 323(6088): 533-536.
- [22] Lian Qiu-sheng, Shi Bao-shun, Chen Shu-zhen. Research advances on dictionary learning models, algorithms and applications [J]. Acta Automatica Sinica 2015, 41(2): 240-260.
- [23] Engan K, Aase S O, Husoy J H. Method of optimal directions for frame design [C]. IEEE International Conference on Acoustics, Speech and Signal Processing, Proceedings, IEEE, 1999: 2443-2446.
- [24] Aharon M, Elad M, Bruckstein A. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation [J]. IEEE Transactions on Signal Processing 2006, 54(11): 4311-4322.
- [25] Pati Y C, Rezaiifar R, Krishnaprasad P S. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition [C]. Signals, Systems and Computers, Conference Record of the Twenty-Seventh Asilomar Conference on. IEEE 2002: 40-44.
- [26] Zhang J, Zhao C, Xiong R, et al. Image super-resolution via dual-dictionary learning and sparse representation [C]. IEEE International Symposium on Circuits and Systems, IEEE, 2012: 1688-1691.
- [27] Chen Hua-hua, Jiang Bao-lin, Liu Chao, et al. Image super-resolution reconstruction based on residual error [J]. Journal of Image and Graphics 2013, 18(1): 42-48.

附中文参考文献:

- [16] 王朝晖, 陈龙, 焦斌亮. 基于图像非局部自相似性与分类字典学习的超分辨率重建算法 [J]. 小型微型计算机系统, 2015, 36(7): 1617-1619.
- [22] 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 [J]. 自动化学报, 2015, 41(2): 240-260.
- [27] 陈华华, 姜宝林, 刘超, 等. 基于残差的图像超分辨率重建 [J]. 中国图象图形学报, 2013, 18(1): 42-48.