基于 CNN 与 ELM 的二次超分辨率重构方法研究

张 静[1),2),3)] 陈益强[1),3)] 纪 雯[1),3)]

[1)](中国科学院计算技术研究所移动计算与新型终端北京市重点实验室 北京 100190)

[2)](中国科学院计算技术研究所网络数据科学与技术重点实验室 北京 100190)

[3)](中国科学院大学 北京 100049)

摘 要 为了实现将低分辨率图像重构为高分辨率图像,弥补高、低分辨率图像间信息损失,文中提出了卷积神经网络与极限学习机结合的二次超分辨率重构方法.首先通过基于深度学习的超分辨率重构优化方法,快速训练端对端的卷积神经网络重构模型,学习结构化的图像信息;然后采用像素级的特征提取,并采用极限学习机模型对图像进行高频分量的补充,通过二次重构获得具有更好视觉效果的高分辨率图像.实验结果表明,文中的优化方法将原有卷积神经网络重构模型的训练效率提高了 3 个数量级,重构效果在主观和客观评估中均优于当前代表性的超分辨率重构方法.

关键词 超分辨率重构;深度学习;图像处理;卷积神经网络;极限学习机

中图法分类号 TP18 **DOI 号** 10.11897/SP.J.1016.2018.02581

Two-Tie Image Super-Resolution Based on CNN and ELM

ZHANG Jing[1),2),3)] CHEN Yi-Qiang[1),3)] JI Wen[1),3)]

[1)](*Beijing Key Laboratory of Mobile Computing and Pervasive Device, Institute of Computing Technology, Chinese Academy of Sciences, Beijing* 100190)

[2)](*Key Laboratory of Network Data Science & Technology, Institute of Computing Technology, Chinese Academy of Sciences, Beijing* 100190)

[3)](*University of Chinese Academy of Sciences, Beijing* 100049)

Abstract With the rapid proliferation of information technology, there is a growing requirement for high quality images and videos. High-resolution images can offer more abundant details, which can not only satisfy people's need for visual effect, also lay a solid foundation of implementing other visual analysis tasks. Image super-resolution is proven to be an effective method to provide high-resolution images. The key point of image super-resolution is to find the mapping relation and complementation information between low and high quality images and search the feasible solution space using this ill-posed problem. In order to reconstruct a high-resolution image from a low-resolution one, complementary information between low and high quality images, we propose a two-tie image super-resolution method combining CNN (Convolutional Neural Networks) and ELM (Extreme Learning Machines). At first, we establish an end-to-end CNN reconstruction model using an improved deep learning method, which can learn the structural image information. Then, we perform pixel-level feature extraction, where we use the high-frequency information learned by ELM to complement the lost component, thus fine-visual high-resolution images can be obtained after the second-time reconstruction. The main work and contributions of this paper

收稿日期:2016-12-04;在线出版日期:2017-04-21.本课题得到国家自然科学基金(61572466,61472399,61572471)、中国科学院科研装备研制项目(YZ201527)、北京市自然科学基金(4162059)资助.**张 静**,女,1990 年生,硕士,主要研究方向为图像处理、机器学习. E-mail: zhangjing2013@ict.ac.cn. **陈益强**(通信作者),男,1973 年生,博士,研究员,博士生导师,中国计算机学会(CCF)会员,主要研究领域为普适计算、人机交互等. E-mail: yqchen@ict.ac.cn. **纪 雯**,女,1976 年生,博士,研究员,博士生导师,中国计算机学会(CCF)会员,主要研究方向为信息编码与多媒体通信网络等.

are as follows: (1) An improved image super-resolution method based on deep learning. We make the following improvements on existing deep learning based high-resolution methods. First, the training data of CNN are processed according to their respective structural features. We utilize ISODATA algorithm to conduct clustering on the images after Sobel filtering in order to obtain two classes of training image sets, one of them is more complex and the other tends to be smooth. Then, we combine pre-training and fine-tuning strategies to train the network. In this work we use complicated images for pre-training and the whole training data set for fine-tuning. In the end, we make use of smaller scale parameters network to increase the training speed of model. Experimental results show that our improved model achieves the same super-resolution construction effect while only takes one thousandth iteration times compared to the original model[26], making the training phase more efficient. (2) A framework combining CNN and ELM to perform rapid two-tie reconstruction. To improve the image quality after CNN reconstruction, we perform pixel-wise feature extraction on those images. We train the ELM model with a smaller upscale factor than the global zoom factor and get the high-frequency components of low-resolution images. After that, we combine those components with the results from CNN based on their weights. Thus the two-tie image reconstruction can be implemented to get the ultimate high-resolution images. In addition, we also develop a demo which is capable of making visual improvements on original low-resolution text images based on the proposed method, and it can be deployed as a function of a remote immersive interaction system to break the limitation of low-resolution camera sensor, and perform the transmission of high-resolution text images. We perform sufficient experimental to demonstrate characteristics of proposed method and the results show that, comparing with the original work, our improved model make the training phase of CNN model more efficient, and the proposed method achieves better performance on majority of datasets compared to the state-of-the-arts.

Keywords super-resolution; deep learning; image processing; CNN; ELM

1 引 言

随着信息技术的发展,人们对图像与视频的质量要求不断提升.高分辨率图像可以提供更多更丰富的细节信息,不仅可以满足人们对视觉效果的追求,同样也是实施其他视觉任务的良好基础.图像的超分辨率重构技术(Image Super-Resolution,SR)是为了满足对高分辨率图像的需求而产生的一种改善图像质量的有效方法[1-2],其核心是通过图像处理技术将低分辨率(Low Resolution,LR)的图像重构成高分辨率(High Resolution,HR)的图像.在图像的实际应用中,通常会对其进行下采样等降质操作减轻存储及传输的负担,而这些降质过程是不可逆的,故超分辨率重构本质上是个病态问题,则找到其合适的可行解,即找到高、低分辨率图像之间的映射关系,是超分辨率重构技术的关键.

纵观图像超分辨率重构的发展,针对 SR 技术的研究主要经历了从频域转向空域的过程和从无训练样本转为有训练样本的过程.SR 研究领域最早的工作[3,4]是将图像重构问题转到频域来进行截止频率以上信息的恢复,但基于频域算法通常对噪声较为敏感.基于非均匀插值的算法是最直观通用的空域超分辨率重构方法,这类方法通过在像素之间插入合适的点,以实现图像分辨率的提升.然而,简单的非均匀插值算法倾向于使图像变得平滑,后续的研究工作[5,6]通过在插值过程中引入更多的图像先验来克服这个问题,但基于插值的方法对图像的视觉表观复杂度较为敏感,导致对于有许多纹理或阴影的图像,这类方法容易产生类似水彩画的效果.基于重建的方法的思路是对一系列低分辨率图像进行一致性约束,再结合自然图像的先验知识进行求解[7,8],这类算法通常能在先验信息充足的条件下获得较好的重构效果,而图像序列数目不足会导致重构效果的退化.

上述传统 SR 研究工作通常需要较强约束,且

对放大系数较为敏感，难以实现在有限条件下的高质量重构. 近年来，基于机器学习的超分辨率重构技术发展迅速，打破了超分辨率重构结果对放大系数敏感不足的问题，尤其是稀疏编码理论的发展为 SR 方法带来了许多启发[9-15]，深度学习模型在超分辨率重构领域的运用为重构效果带来了进一步的提升. 但现有的基于深度学习的超分辨率重构算法模型复杂度较高，需要估计的参数规模较大且通常初始化为随机值，通常需大量的迭代训练才能学习到有意义的参数；同时用于 SR 训练的自然图像的数据量不宜过小，否则会影响模型的泛化能力. 故基于深度学习的 SR 算法训练过程效率较低，在普通计算机上难以实现有效的训练. 在一些需要针对特殊领域图像进行超分辨率重构的应用中，模型需根据图像重构的视觉特点与实际需求进行相应调整，而昂贵的训练代价限制了此类改进措施的实施.

本文针对深度学习超分辨率模型训练效率低的问题展开了研究，主要研究工作和创新点如下：

(1) 提出了基于深度学习的超分辨率重构快速训练方法，实现了在普通计算机上进行高质量的 SR 模型训练. 在现有的 SRCNN 模型基础之上，本文对其训练过程进行了如下优化：① 对训练数据进行基于图像结构特征分类的预处理，划分为较复杂和原始数据两个集合；② 利用分类结果，采用预训练与调优结合的方法训练模型，在预训练阶段采用较复杂数据集刺激神经元更快速学习到有意义的参数值，再利用原始数据集进行调优进一步训练和调整参数，以防止重构时出现瑕疵；③ 设计较小参数规模的 CNN 模型，SRCNN 模型在大量迭代训练情况下模型仍未达到收敛，且大部分参数因未得到有效的训练而处于无序状态，可推测适当的减少参数规模可提升参数更新效率. 实验结果表明本文的优化方法在迭代次数仅为 SRCNN 模型千分之一时便取得了相近的重构效果.

(2) 提出了 CNN 与 ELM 结合的二次超分辨率重构方法. 在 CNN 超分辨率优化算法的基础上，本文对小于全局放大系数的高、低分辨率图像进行像素级跨放大系数的特征提取，应用训练效率更高的 ELM 模型学习低分辨率图像所缺失的高频分量，并将该分量与 CNN 重构后的结果进行加权叠加，实现对输入图像的二次重构，获得了优于传统方法的图像重构效果.

本文第 2 节介绍相关工作；第 3 节介绍基于 CNN 的超分辨率重构优化方法；第 4 节介绍 CNN 与 ELM 的联合的二次超分辨率重构模型；第 5 节对本文方法进行实验评估和分析；第 6 节对本文研究工作进行总结.

2　相关工作

根据训练数据来源的不同，基于机器学习的 SR 方法可划分为基于外部样例的方法和基于内部样例的方法.

2.1　基于内部样例的 SR 方法

基于内部样例的超分辨率重构方法的理论依据是图像样例的纹理倾向于在原图像或其跨比例版本图像中重复出现，通过在自身图像和其生成的其他样例中进行搜索，通常可以找到与样例本身特征较一致的素材. 文献[16-18]均利用图像样例的内部相似性，基于自身图像进行超分辨率重构；Freedman 和 Fattal[19] 利用跨比例生成的图像构造了更多的训练数据；Cui 和 Chang 等人[20] 提出了基于深度学习方法的 DNC 模型，采用对不同放大系数的内部样例块序列进行相似块搜索和加权组合. 但这类重构方法局限于图像自身或跨比例图像中均搜索不到可辨别的重复样例的情况，泛化能力较为受限.

2.2　基于外部样例的 SR 方法

基于外部样例的超分辨率重构方法通常用一系列通用的样例基元去预测高分辨率图像丢失的信息(高频分量). Chang 等人[21] 应用流形学习中局部线性嵌入的基本思想，在高、低分辨率两个流形空间中建立相似性关系，将低分辨率图像块空间的局部几何特征映射到高分辨率空间，生成由图像邻域块线性组合的目标图像. Chen 和 Qi[22] 采用低秩矩阵恢复和联合学习的方法，将原始高、低分辨率图像块特征的低秩分量映射到统一空间中，进而完成基于邻域嵌入的重构. Qiao 等人[23] 应用支持向量回归模型(SVR)来预测图像所丢失的高频分量，An 和 Bhanu 等人[24] 采用类似的思路，用极限学习机(ELM)训练回归模型进行高频分量的预测和补充，取得了较好的重构效果. ELM 模型在解决超分辨率重构问题上的优势在于模型仅通过矩阵乘法和求逆运算来进行参数求解，与 SVR[23] 等迭代求解的方法相比具有更高的训练效率，相同时间内能够处理的训练数据量远大于 SVR，使得其具有更好的泛化能力.

基于稀疏字典的超分辨率重构是 SR 领域的研究热点，Yang 等人在文献[9]中最早提出了基于稀疏编码的自适应选择最相关邻域的策略，可有效避

免传统学习方法在超分辨率重构时出现的过拟合或欠拟合的问题;文献[10]基于压缩感知理论,联合训练了对应高分辨率图像和低分辨率图像的双字典;文献[11]中高、低分辨率图像用两个特征空间进行表示,将基于稀疏编码的 SR 方法拓展到更为普适的情况;He 等人[12]允许高低分辨率的稀疏表达系数之间存在映射关系;Timofte 等人[13]结合了稀疏字典学习与邻域嵌入(NE)的策略,提出了固定邻域回归方法;Zhu 等人[14,15]允许图像块变形,以及设定稀疏字典元素只包含奇异基元(如单独的边缘结构),使得学习到的稀疏字典更具有表达能力. 基于稀疏编码的超分辨率重构方法通常能够获得较好的视觉效果,但在实际应用中处理效率较低.

2.3 基于深度学习的 SR 方法

随着深度学习在计算机视觉领域的不断发展,基于深度学习的超分辨率重构方法近年来获得了较多关注. Gao 等人[25]将深度受限玻尔兹曼机(RBM)模型应用于图像的超分辨率重构,通过 RBM 模型学习高低分辨率图像共享的稀疏表达系数;Cui 等人[20]采用联合局部自动编码机(CLA)模型对层内的重构进行约束,在全局范围内将所有的栈式 CLA 模型级联起来,用反向传导算法来进行全局范围内误差的抑制,训练深层的网络;Dong 等人[26]在自然图像数据集上训练了端对端的卷积神经网络模型来进行基于外部样例的超分辨率重构,称为 SRCNN 模型,该模型仅采用 3 层网络结构便取得了较好的重构效果,但其训练的代价较为高昂. Wang 等人[27]利用了 DNC 与 SRCNN 模型的各自优势,提出了一种内部样例与外部样例结合的深度学习模型,并将模型分解成一些专用的子模型分别进行数据分块处理和训练.

基于深度学习的超分辨率重构方法通常能取得非常好的重构效果,但质量提升的同时也伴随着模型复杂度的增加,效率问题成为超分辨率重构在实际场景中广泛应用的技术壁垒,本文以保证重构质量的前提下提升模型训练效率为出发点展开研究,提出了一种结合 CNN 和 ELM 的二次超分辨率重构方法.

3 基于 CNN 的超分辨率优化方法

3.1 卷积神经网络介绍

卷积神经网络(Convolutional Neural Networks,CNN)是一种前馈人工神经网络,是多层感知机的变形,在本世纪初被广泛地应用[28,29]. 卷积神经网络的二维拓扑结构与输入图像的结构较为一致,能够在训练过程中隐式的提取特征,权值共享与局部感受野的设计可以使模型变得简单且具有位移、尺度不变性的优势,故非常适合处理视觉问题.

卷积神经网络的每一层都由多个二维平面组成,而每个平面由多个独立神经元组成. 主要包括输入层、卷积层、池化层和输出层. CNN 中卷积层执行特征提取的操作,池化层功能为求局部平均和二次提取特征. 通常每一个卷积层后都会接着一个池化层,进而级联起来构成整个深度网络. 其典型结构如图 1 所示.

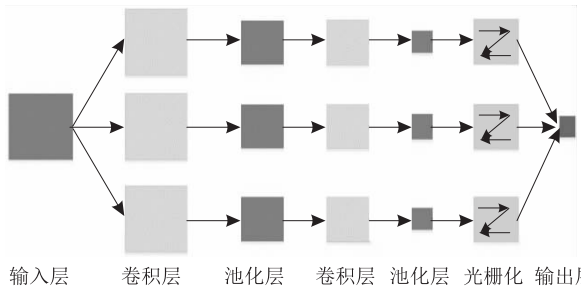


图 1 卷积神经网络的平面展开结构

3.2 基于 CNN 的超分辨率重构模型

设 X 为高分辨率图像 HR,Y 为 X 先后经过下采样和上采样操作所获得的与 X 尺寸相同的低分辨率图像 LR,本文研究的目标为由 LR 图像重构得到超分辨率图像 $F(Y)$,即通过卷积神经网络模型的学习到映射 F,使得 $F(Y)$ 尽可能地与 HR 图像 X 相同. 在学习映射模型 F 的过程中,本文的 CNN 网络模型借鉴了 SRCNN 的结构,采用 3 层卷积神经网络架构,特征层次间为全连接模式. 由于超分辨率重构需尽可能恢复图像经过降质后所损失的信息,故本文的 CNN 模型不设池化层,以防止损失掉更多的图像细节信息. 本文采用校正线性单元(ReLU)作为模型的唯一激活函数,其形式为

$$y=\max(0,\boldsymbol{x}) \tag{1}$$

与 SRCNN 一致,本文的三层网络模型对低分辨率图像的处理对应如下 3 种操作:

(1)块提取和表达. 由输入图像 Y 生成的图像块中提取多个特征块,并将每个块表达为一个高维向量,这些向量组合成了第一层的特征映射图.

(2)非线性映射. 将上一层获取的高维向量映射为另一个高维向量. 每一个映射后的向量概念性的表达着一个高分辨率的图像块,这些向量组成了第二层特征映射图.

(3) 重构. 将上一层输出的高分辨率特征表达整合，生成最后的高分辨率图像，通过使该图像与 X 间差值最小化来训练整个网络.

上述操作超分辨率重构模型中的位置及网络整体的架构如图 2 所示. 具体的，下文将逐一描述每一层次操作的执行方法与意义.

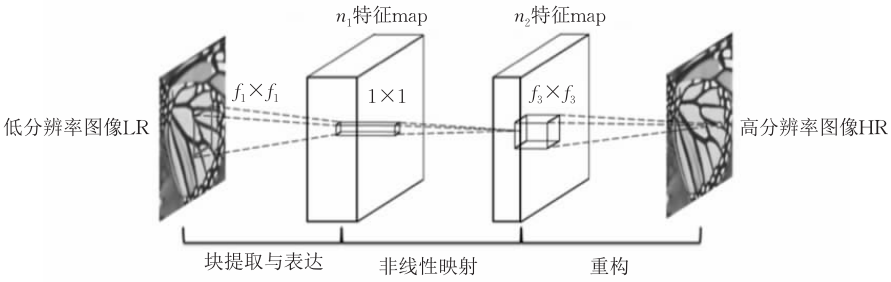


图 2　基于 CNN 的超分辨率重构算法模型

(1) 在块提取与表达阶段，设该层所执行的操作为 F_1，根据前向传播公式，输入层低分辨率图像块 Y 经过 F_1 运算表达为

$$F_1(Y)=\max(0,W_1\times Y+\boldsymbol{B}_1) \tag{2}$$

上式中 W_1 和 $\boldsymbol{B}_1$ 分别表示滤波器和偏置，此时 W_1 的大小为 $c\times f_1\times f_1\times n_1$，$c$ 为 YCbCr 色彩空间中图像的通道数，f_1 为卷积层滤波器的大小，n_1 为滤波器的个数，$\boldsymbol{B}_1$ 为 n_1 维的向量. 该卷积层的含义为：W_1 应用了 n_1 个核大小为 $c\times f_1\times f_1$ 的滤波器在输入的低分辨率图像上，获得了 n_1 个特征映射图.

(2) 在非线性映射阶段，上一层对于每个 patch 的输出为 n_1 维的特征映射图，该层的作用为将这些特征映射图进行非线性的映射，输出 n_2 维的特征映射图，相当于将 n_2 个空间大小为 1×1 的滤波器作用在上一层的特征映射图上. 其运算表达为

$$F_2(Y)=\max(0,W_2\times F_1(Y)+\boldsymbol{B}_2) \tag{3}$$

上式中 W_2 的大小为 $c\times n_1\times 1\times 1\times n_2$，$\boldsymbol{B}_2$ 的大小为 n_2 维. 每个该层的输出的特征映射图概念性地表达着可以组合成高分辨率图像的各个频段的信息.

(3) 在重构阶段，由于训练数据为有重叠区的图像块，通常会对重叠区域执行平均化操作来生成最终的整张图像，平均化操作可被视作采用均值滤波器对图像进行卷积，故重构阶段设计卷积层表达形式如下：

$$F_3(Y)=W_3\times F_2(Y)+\boldsymbol{B}_3 \tag{4}$$

其中，W_3 的大小为 $c\times n_2\times f_3\times f_3\times n_3$，$\boldsymbol{B}_3$ 的大小为 c 维. W_3 应用了 n_3 个核大小为 $c\times f_3\times f_3$ 的滤波器作用在输入的特征映射图上，所对应的操作是对图像特征映射图的重叠区域执行取均值操作，使得输出层可以重组为完整的图像.

参数 $\theta=\{W_1,W_2,W_3,\boldsymbol{B}_1,\boldsymbol{B}_2,\boldsymbol{B}_3\}$，本文通过构建损失函数模型，最小化 $F(Y;\theta)$ 与原始图像 X 之间的误差来进行 CNN 的参数估计. 给定一系列高分辨率训练数据 $\{X_i\}$ 和与其对应降质后的低分辨率训练数据 $\{Y_i\}$，本文采用均方误差作为损失函数，其形式为

$$L(\theta)=\frac{1}{n}\sum_{i=1}^{n}\|F(Y_i;\theta)-X_i\|^2 \tag{5}$$

其中 n 为训练样本数，采用随机梯度下降法与 BP 算法来训练整个网络，最小化上述损失方程.

CNN 模型训练的整体流程如图 3 所示.

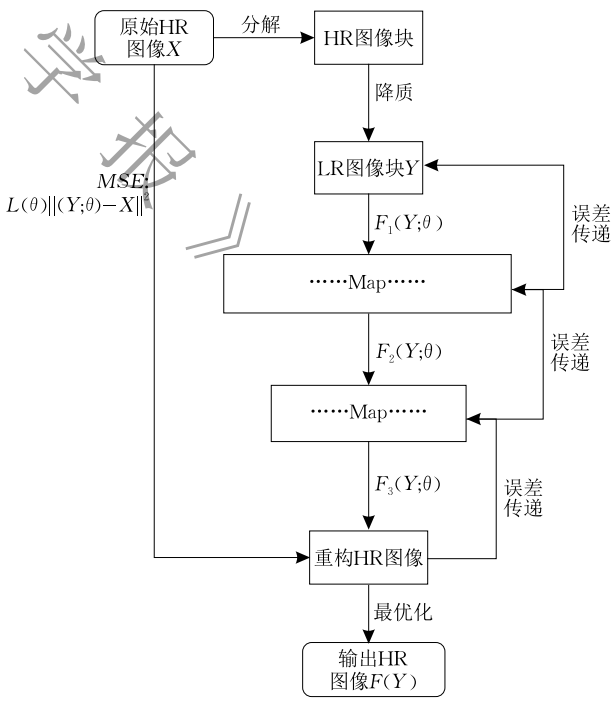


图 3　基于 CNN 的 SR 模型训练流程

3.3　超分辨率重构模型的优化方法

3.3.1　优化策略分析

SRCNN 算法中由于需要训练的参数规模较大，模型在 GTX 770 型 GPU 上训练 3 天仍远未达

到收敛,但随着迭代次数的增加,其训练准确度和重构效果一直保持着提高的趋势.然而,为取得高重构效果花费的额外计算资源和计算时间是实际应用中需要权衡和考虑的.

在超分辨率重构问题中,图像的边缘结构是目标重构信息的重要组成部分,对超分辨率重构效果影响很大.按照图像纹理的相似性进行子类别划分并分别建模,是提升超分辨率重构模型效果的一个直观的想法.文献[27]中采用 K-Means 聚类方法对训练图像集进行分类,结果表明在合理选择 K 值的情况下,子模型划分策略可以对训练结果带来准确度提升.但该方法在测试时同样需对测试图像进行与训练图像相同的分块和聚类操作,以确定每一图像块所对应的子模型,最后将图像块拼接为完整的图像.由于不同子模型训练得到的效果有所差异,在重叠区域可能会损失掉一部分重构出来的高频分量,也可能引入瑕疵.

本文参考子模型划分思路,对训练集进行基于边缘相似度的聚类.但本文并不采取针对不同训练集分别建模的策略,而是学习对于所有图像通用的网络模型,以避免上述子模块划分方法的弊端.本文方法着重考虑学习效率中的提升,目标是通过类别划分来减少模型参数中的规模,在保证较好的重构效果的情况下减少训练所需的时间.

在深度学习网络参数训练时通常采用一定范围内的高斯随机参数作为模型的初始化参数,而随机的初值可能会导致模型在训练时陷入局部最优解或无法收敛的问题.许多深度学习模型的训练都包括预训练和调优两个过程,经过预训练的模型一定程度上相当于有监督的学习模型,可有效避免过拟合的情况,模型表达具有良好分布,且模型参数具有一定稀疏性.

基于上述分析,本文对超分辨率重构模型采取了如下优化:

(1)采用基于图像结构特征的 ISODATA 算法对训练数据集进行聚类;

(2)针对不同类别的图像训练集,采用预训练与调优结合的方法进行模型训练;

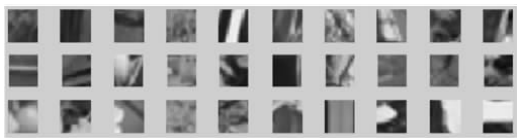
(3)减小 CNN 网络的参数规模,以加速学习效率.

3.3.2 模型优化方法

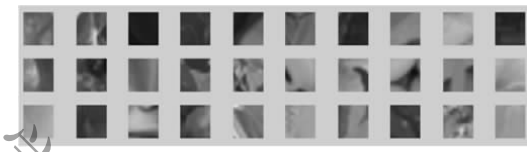
与传统的 K-Means 聚类方法相比,ISODATA 聚类方法的过程是可控的,意味着聚类过程可进行参数的调节,可避免类别中样本数量过少而导致训练数据不充足和重构效果变差的情况.

在对图像训练数据集进行聚类时,基于图像块像素灰度值的聚类策略[27]受图像块整体灰度的影响大于纹理边缘等结构特征的影响.然而,恢复纹理边缘区域损失的高频分量是超分辨率重构问题的关键.故本文采取仅对图像纹理边缘区域的聚类策略,首先采用 Sobel 算子进行滤波,再对滤波后的特征图执行分块操作,采用 ISODATA 算法对这些特征图块进行聚类,然后找到每一类中特征图块所对应的原始高、低分辨率图像对,最终得到图像块数据集.

在实验中本文将训练集分为两类,与上文所述的设想一致,两类训练集间的重要区别在于,第一类中图像块边缘区域较多,纹理较为复杂,第二类中的图像块边缘区域较少,整体上更为平滑.两类图像(均为低分辨率)的视觉效果如图 4 所示.



(a)较复杂数据集



(b)较平滑数据集

图 4

由于 CNN 的结构特性使得其参数通常具有一定的稀疏性,侧重于学习到图像的结构性特征,如边缘、拐角等.纹理较为复杂,边缘区域多的样本在训练时倾向于对网络的神经元产生更多的刺激,易于学习到具有更多提取边缘特征效果的参数.这是由于复杂样本在下采样过程中损失的信息更多,在重构过程中损失 $L(\theta)$ 值更大,从而在反向传导误差过程中 ΔW 和 ΔB 的变化更为剧烈,故而能够加大参数更新的幅度,起到够提升学习效率的作用.

根据上述结论,为了让模型可以重构出较多的结构特征,快速学习到较优的网络参数,本文采用 ISODATA 分类后的复杂训练数据集对 CNN 模型进行预训练;同时,为了避免全部采用复杂训练集训练网络导致在自然图像重构时产生冗余信息和瑕疵,本文利用全部图像数据集对模型进行调优(Fine-tune),使重构的图像更贴近真实的自然图像.

SRCNN 模型在大量迭代(8×10^8 次)训练情况下模型仍未达到局部最优(继续迭代时误差仍继续

减小)，说明模型仍未收敛，结束迭代后 SRCNN 中大部分滤波器处于无结构的状态，说明模型没有得到充分的训练. 为了提升模型训练效率，使得模型参数的学习更充分，本文将卷积神经网络的规模进行调整，采用更为轻量级模型进行超分辨率的重构，称为 mini CNN. 具体的，本文将模型前两层的滤波器数 n_1 和 n_2 减小为 SRCNN 模型中的三分之二.

本文第 5 节实验与分析部分将对上述优化策略的有效性进行评估，结果表明本文方法在迭代次数仅为 SRCNN 的千分之一时便取得了与其相当的重构效果，证明上述改进极大地提升了训练效率.

4　基于 CNN 与 ELM 的二次重构方法

4.1　ELM 极限学习机简介

ELM(Extreme Learning Machine)[30] 是一种单隐层前馈神经网络，其特点是输入层与隐层间的参数的初始化取随机值. ELM 学习速度快，应用方式简单，训练精度高，可处理二分类、多分类以及回归等问题. 本文采用 ELM 来进行图像的超分辨率重构，与其他回归模型(如文献[23]采用的 SVR 模型)相比，ELM 模型复杂度低，在可接受的时间内能够处理更多的训练数据，泛化能力及重构效果都更为出色.

ELM 模型的结构如图 5 所示.

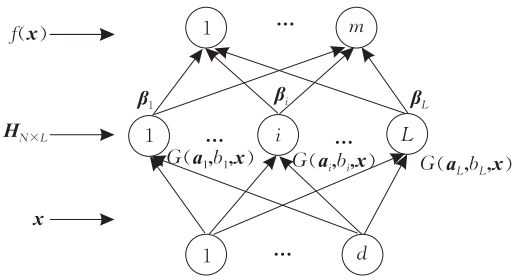


图 5　ELM 模型的结构示意图

ELM 包括 3 个层次：输入层 $\boldsymbol{x}$，隐层 $\boldsymbol{H}_{N\times L}$，输出层 $\boldsymbol{f}(\boldsymbol{x})$. d 个输入节点对应着输入层的 d 维特征向量($\boldsymbol{x}\in R^d$，$\boldsymbol{x}=(x_1,\cdots,x_d)^{\mathrm{T}}$). 特征向量 $\boldsymbol{x}$ 在隐层被映射为向量$(G(a_1,b_1,\boldsymbol{x}),\cdots,G(a_L,b_L,\boldsymbol{x}))^{\mathrm{T}}$. $G(a_i,b_i,\boldsymbol{x})$ 是第 i 个加性隐节点的输出，其计算公式如下：

$$G(a_i,b_i,\boldsymbol{x})=g(a_i\cdot\boldsymbol{x}+b_i),\ a_i\in R^d,\ b_i\in R \tag{5}$$

其中 $g(\boldsymbol{x})$ 表示激活函数. 隐层的 L 维的特征向量在经过线性变换后可得到一个 m 维的向量 $\boldsymbol{f}(\boldsymbol{x})$，$m$ 个输出节点相当于输出层的 m 个类别. 其公式如下：

$$\boldsymbol{f}(\boldsymbol{x})=\sum_{i=1}^{L}\beta_i G(a_i,b_i,\boldsymbol{x}),\ \beta_i\in R^m \tag{6}$$

ELM 的训练集合为$\{(\boldsymbol{x}_j,\boldsymbol{t}_j)\mid \boldsymbol{x}_j\in R^d,\boldsymbol{t}_j\in R^m,j=1,\cdots,N\}$，其中 $\boldsymbol{x}_j$ 是特征向量，$\boldsymbol{t}_j$ 是 $\boldsymbol{x}_j$ 的标签. 在训练阶段，每个实例 $\boldsymbol{x}_j$ 作为一个输入层的向量传入 ELM 中，$\boldsymbol{t}_j$ 作为输出层的期待输出结果，对每一个标签向量 $\boldsymbol{t}_j$ 都有一个输入实例 $\boldsymbol{x}_j$ 与之相对应. 在回归问题中，t 为实数，$\boldsymbol{t}_j$ 的值直接代表了模型对于输入 $\boldsymbol{x}_j$ 的响应. 隐层节点的参数值$(a_i,b_i,i=1,\cdots,L)$是随机初始化的，变量$(\beta_i,\cdots,\beta_L)$可以由如下公式计算得出：

$$\boldsymbol{H}\boldsymbol{\beta}=\boldsymbol{T} \tag{7}$$

其中：

$$\boldsymbol{H}=\begin{bmatrix}G(a_1,b_1,x_1) & \cdots & G(a_L,b_L,x_1)\\ \vdots & \ddots & \vdots\\ G(a_1,b_1,x_N) & \cdots & G(a_L,b_L,x_N)\end{bmatrix}$$
$$=[h(x_1),\cdots,h(x_N)]^{\mathrm{T}}_{L\times N},$$
$$h(\boldsymbol{x})=\begin{bmatrix}G(a_1,b_1,\boldsymbol{x})\\ \vdots\\ G(a_L,b_L,\boldsymbol{x})\end{bmatrix}_{L\times 1},$$
$$\boldsymbol{\beta}=\begin{bmatrix}\boldsymbol{\beta}_1^{\mathrm{T}}\\ \vdots\\ \boldsymbol{\beta}_L^{\mathrm{T}}\end{bmatrix}_{L\times m},\ \boldsymbol{T}=\begin{bmatrix}\boldsymbol{t}_1^{\mathrm{T}}\\ \vdots\\ \boldsymbol{t}_L^{\mathrm{T}}\end{bmatrix}_{N\times m} \tag{8}$$

$\boldsymbol{\beta}^*$ 的最小二乘解可以利用 MP 广义逆解析得到，并且具有最小范数：

$$\boldsymbol{\beta}^*=\boldsymbol{H}^{\dagger}\boldsymbol{T} \tag{9}$$

最终可得到 ELM 的输出方程：

$$\boldsymbol{f}(\boldsymbol{x})=\boldsymbol{\beta}^{*\mathrm{T}}h(\boldsymbol{x}) \tag{10}$$

本文中 ELM 的训练及测试的过程如图 6 所示.

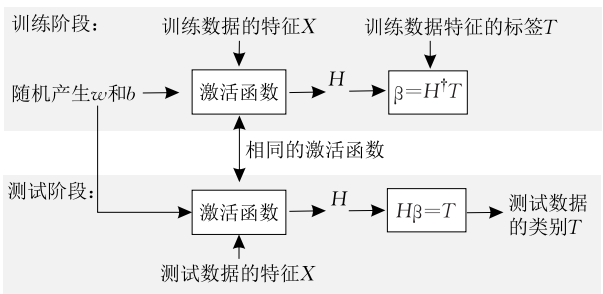


图 6　ELM 的训练及测试过程

4.2　基于 ELM 的图像高频分量学习

基于 CNN 的超分辨率重构方法的原理与稀疏编码是一脉相承的[26]，即关注图像的结构性特征，但其未考虑到图像的每一个像素点与其他像素点的亮度变化关系. 为此本文提出了二次重构的方法，即

通过像素级(pixel-wise)的 ELM 训练对 CNN 输出结果进行高频分量的补偿.具体的,本文对图像中每个像素进行特征提取,通过 ELM 对图像进行跨放大系数的二次重构.跨放大系数是指:设通过整体 SR 模型的目标放大系数为 k(即训练数据经过 k 倍下采样),在第二次重构时用放大系数小于 k 的数据进行 ELM 训练,以使得二次重构时所补充的高频分量更接近 CNN 重构后图像所缺失的部分,可以学到更精细的图像细节,使得最终获得的图像具有更好的视觉效果.

将 ELM 训练中不同的放大系数所学习到的高频分量进行可视化显示,效果如图 7 所示.

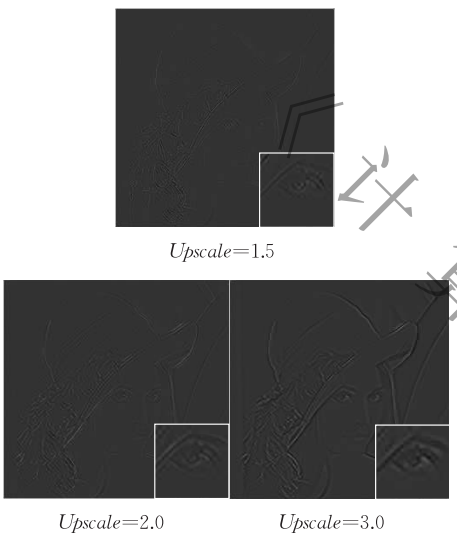


图 7 不同放大系数的 ELM 模型学习到的高频分量

从图 7 中可以看出,放大系数较小的 ELM 模型所学习到的高频分量能够表现出更细腻的细节信息(见各局部放大图),图像整体像素值差异较小;而高倍放大系数的模型学习到的边缘轮廓粒度较粗,整体像素值差异较大.

采用 ELM 中学习高频分量并进行二次重构的具体步骤如下:

(1)高频分量的提取.设高分辨率训练图像为 I_{HR},将 I_{HR} 进行 $Upscale<k$ 的下采样,再通过插值方法(bicubic)进行相应的放大,产生与 I_{HR} 相同尺寸的低分辨率图像 I_{LR},再计算图像的高频分量 I_{HF}:

$$I_{HF}=I_{HR}-L_{LR} \tag{11}$$

(2)像素级特征提取.设 $P_{i,j}$ 为低分辨率图像 I_{LR} 中位置坐标为(i,j)的像素,提取以 $P_{i,j}$ 为中心的 8 个相邻像素的灰度值,与 $P_{i,j}$ 的灰度值共同组成 9 维的特征向量,在对 $P_{i,j}$ 的 x,y 方向分别进行一阶和二阶偏导的计算,获得 5 维导数特征向量,与之前的 9 维向量共同组合为 14 维的 ELM 输入特征.

对图像 I_{LR} 中的所有像素进行求导的结果为 $\left(\frac{\partial I_{LR}}{\partial x},\frac{\partial I_{LR}}{\partial y},\frac{\partial I_{LR}}{\partial x^2},\frac{\partial I_{LR}}{\partial y^2},\frac{\partial I_{LR}}{\partial xy}\right)$.

(3)ELM 训练.对于训练图像中的每个像素 $P_{i,j}$ 将获取的特征向量作为 ELM 的输入,I_{HF} 中坐标为(i,j)的值作为其对应的标签,设置 ELM 的隐层节点数和激活函数,执行训练过程,得到模型参数.

(4)图像重构.设本文中基于 CNN 重构后的图像为 I_{SR},对 I_{SR} 进行与步骤(2)中相同的特征提取,输入到已训练的 ELM 模型中,获得模型的输出 I_{SHF},继而对 I_{SR} 进行高频分量补充,得到最终的重构结果 I_{ELMSR}:

$$I_{ELMSR}=I_{SR}+\gamma I_{SHF} \tag{12}$$

式中 γ 为平衡 I_{SR}、I_{SHF} 两项相对重要性的参数.

ELM 进行二次重构的流程如图 8 所示.

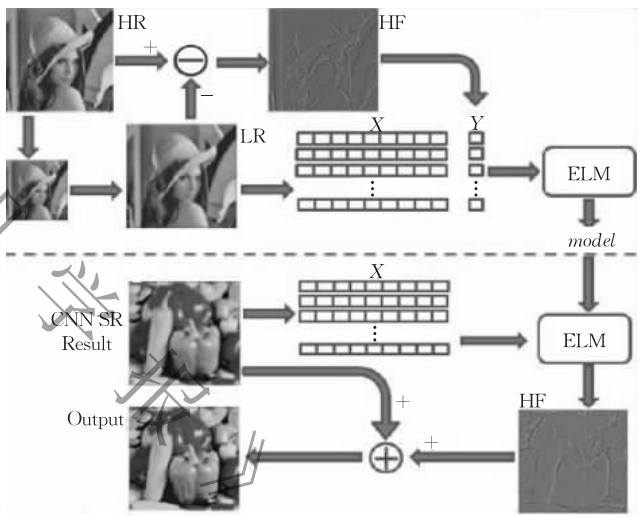


图 8 采用 ELM 进行高频补充过程[24]

4.3 基于 CNN 和 ELM 的二次重构算法整体流程

本文所提出的超分辨率重构方法结合了对图像结构特征的学习的 CNN 模型和对图像像素级特征的学习 ELM 模型的优势,将 CNN 重构出来的图像输入到已训练的 ELM 模型中进行二次重构,获得最终的 SR 图像,本文方法的整体框架如图 9 所示.

图 9 详细的描述了本文算法训练及测试的整体流程.值得注意的是,在训练过程中,本文的两个算法模型 CNN 和 ELM 是相互独立的.具体的,在执行 CNN 训练时采用放大系数为 k 的高、低分辨率图像块样本对,通过 ISODATA 分类后进行预训练和调优结合的训练过程.而在 ELM 的训练时,高、低分辨率训练数据样本为整张图像,且放大系数小于 k,以提升图像重构的效果.

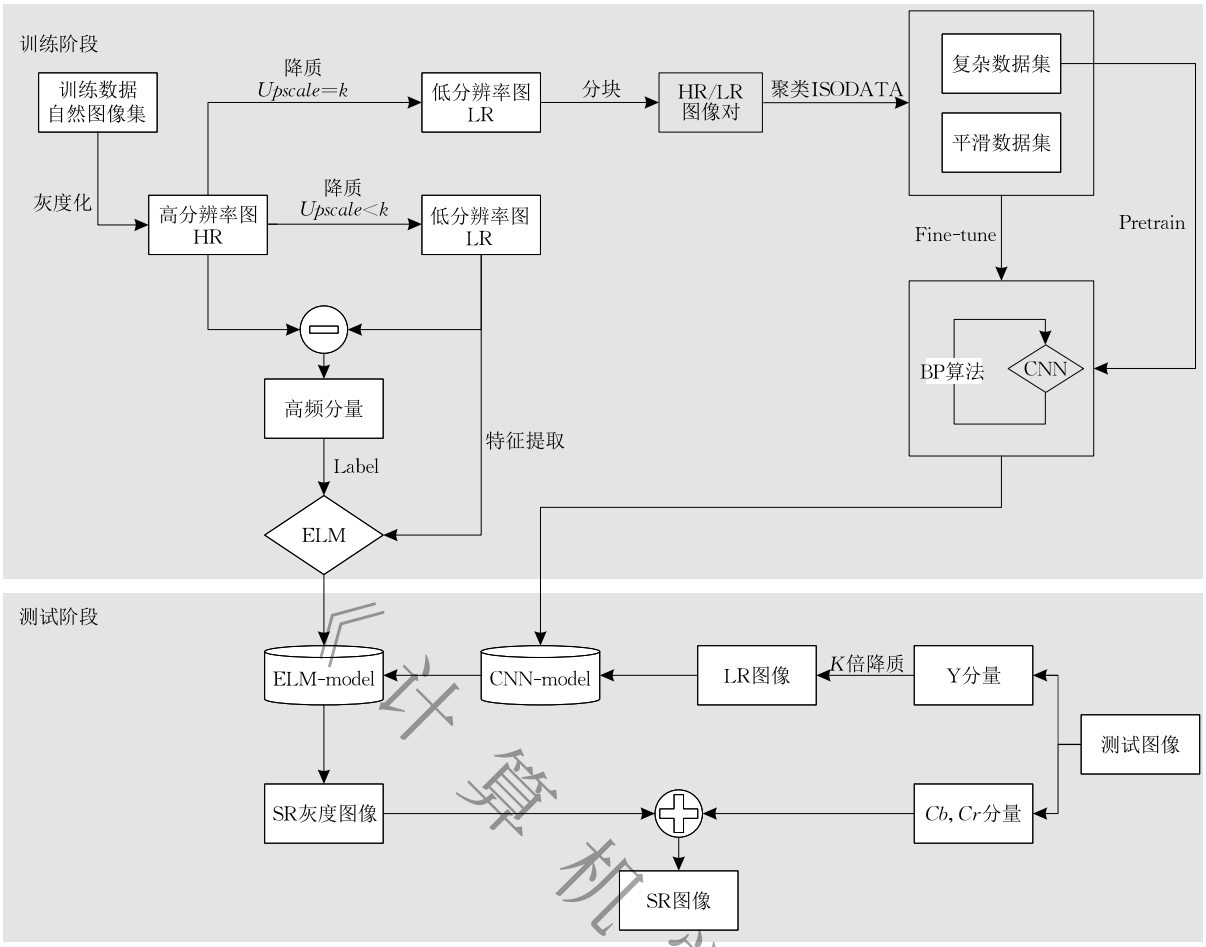


图 9　CNN＋ELM 算法的整体流程

4.4　算法应用场景讨论

在实际应用中，通常情况下超分辨率重构模型作为离线模型仅需训练一次，这种场景下 SR 应用运行时间（测试效率）更为重要. 本文方法能够实现在普通计算机上进行高质量 SR 模型的训练，但二次重构的运行时间要大于单独模型的运行时间. 如果具备高性能的计算资源，则可利用本文优化后的 CNN 模型增加迭代次数完成训练，能够得到更好的重构效果（与相同条件下 SRCNN 模型相比），由于本文的 CNN 模型参数规模更小，故测试阶段的运行速度也更快. 如果用户对重构质量仍然不够满意，希望能进一步提高，则可以采用本文提出的二次重构方法再进行处理，代价是运行速度变慢. 基于 CNN 的 SR 和基于 ELM 的 SR 两个步骤均为可选，实际场景中可以灵活的配置.

5　实验与分析

5.1　实验环境、训练集与测试集

本文的评估实验在 Intel Core（TM）i5-3435 的 CPU 上进行，主频为 3.10 GHz，内存为 12 GB. 卷积神经网络搭建在深度学习框架 Caffe 上，Ubuntu14.04 操作系统，ELM 模型的训练以及整体测试过程在 Matlab 平台上进行.

在训练数据处理方面，本文获取低分辨率图像块的策略与一般的 SR 研究一致，首先对图像进行下采样，进而用 Bicubic 插值法进行 k 倍放大，获取到与原始图像的图像相同尺寸的低分辨率图像.

本文的训练集与文献[9-11]和 SRCNN 方法的训练集一致，包括 91 张自然图像（ELM 模型训练时使用了其中的一部分）. 在采用卷积神经网络进行超分辨率重构的训练时，训练数据为高低分辨率的图像块，以步长为 14 像素的滑动窗在自然图像集中进行截取，获得具有重叠区域训练数据，其中低分辨率图像块大小为 32×32 像素，为了避免卷积神经网络模型中的边缘效应，本文仅考虑每个块的中间区域，取高分辨率图像块大小为 28×28 像素.

本文的卷积神经网络 mini CNN 模型中各项参数为

$$c=1,\ n_1=48,\ f_1=9,\ n_2=24,\ n_3=1,\ f_3=5.$$

本文方法的测试集也采用超分辨率重构研究中常用的公开测试集 Set5 与 Set14.

由于超分辨率重构仅对 YCbCr 颜色空间的亮度通道,即 Y 通道敏感,故本文训练过程中仅考虑亮度通道,生成最终的重构图像时再与其他通道信息合并,但本文模型也可方便的拓展到多个通道.

5.2 实验效果

5.2.1 客观评估效果

为衡量本文所提算法的效果,对放大系数 k 为 2、3、4 倍的测试图像进行了客观指标(*PSNR* 和 *SSIM*)评估实验,与目前具有代表性的几种 SR 方法进行了对比,结果如表 1 和表 2 所示.

表 1 测试集 Set5 中的对比实验结果

k	*PSNR*			*SSIM*		
	2	3	4	2	3	4
bicubic	33.66	30.39	28.42	0.929	0.868	0.810
SC	—	31.42	—	—	0.882	—
NE+LLE	35.77	31.84	29.61	0.949	0.895	0.840
ANR	35.83	31.92	29.69	0.949	0.896	0.841
A+	**36.54**	32.59	**30.28**	0.954	0.908	**0.860**
SRCNN	36.34	32.39	30.09	0.950	0.900	0.858
本文	36.46	**32.60**	30.26	**0.954**	**0.909**	**0.860**

表 2 测试集 Set14 中的对比试验结果

k	*PSNR*			*SSIM*		
	2	3	4	2	3	4
bicubic	30.23	27.54	26.00	0.868	0.773	0.701
SC	—	28.31	—	—	0.795	—
NE+LLE	31.76	28.60	26.81	0.899	0.807	0.733
ANR	31.80	28.65	26.85	0.900	0.809	0.735
A+	32.28	29.13	27.32	0.905	0.818	0.749
SRCNN	32.35	29.00	27.50	**0.906**	0.821	0.751
本文	**32.41**	**29.24**	**27.55**	**0.906**	**0.821**	**0.759**

在对比算法中,bicubic 为插值类超分辨率重构方法中经典的二次三项插值,也是本文方法和进行对比实验的其它几种算法的原始图像预处理算法;SC[10]为基于稀疏编码的超分辨率重构算法中具有代表性的训练联合稀疏字典的方法;NE+LLE[21]为基于外部样例学习的超分辨率重构方法;ANR 和 A+[13]为近年提出的基于外部样例学习的超分辨率重构算法,同样应用了稀疏编码及邻域嵌入的思想;SRCNN 为近年来采用深度学习框架进行超分辨率重构的代表性方法,也是本文提出的二次重构策略中第一阶段的原型算法,也是本文模型的重要对比方法.

由上述实验结果的两项评估指标可以发现,本文所提的算法在大部分测试集上均取得了优于其他方法重构效果,且对于放大系数较大的数据集的重构质量提升尤为明显.但是在放大系数较小时,如 $Upscale=2$ 时,本文在 Set5 中的结果会略逊于文献[13]中的结果.其原因在于,本文在二次重构时的高频分量的补充对于放大系数较大的情况会较为有效,由于放大系数较小的图像其本身已获得较好的重构,在二次重构进一步补充高频分量的同时也会引入一些不必要的信息,从而限制了重构的效果提升.

5.2.2 主观评估效果

本文的算法与其他超分辨率重构方法的主观实验效果对比如图 10、图 11 和图 12 所示.

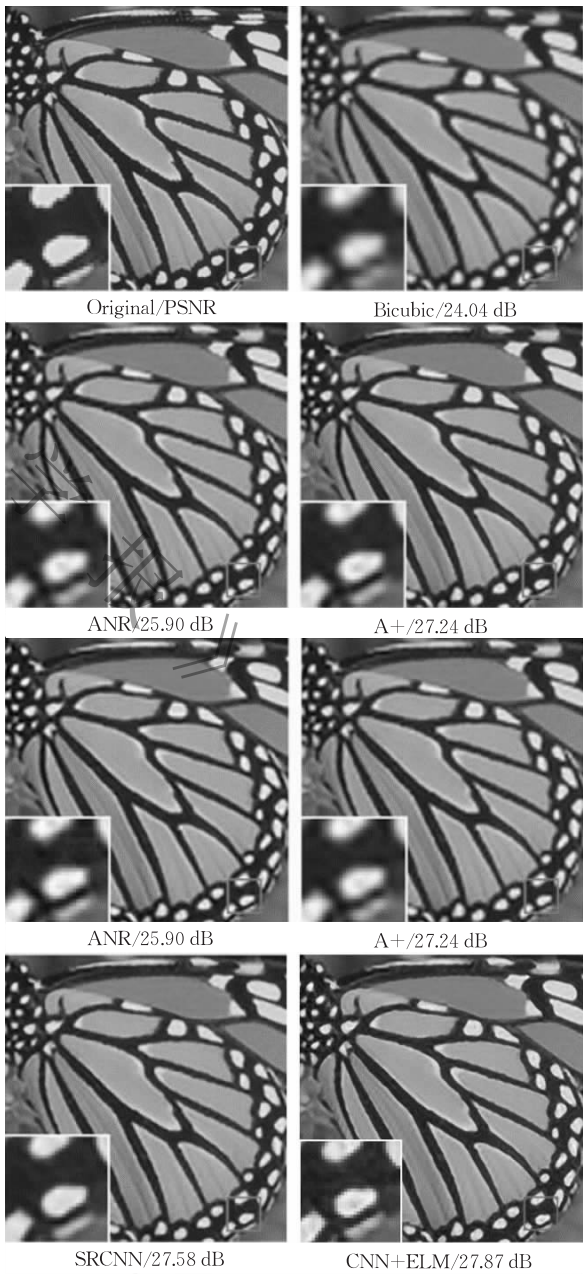


图 10 Set5 数据集中的图像 Butterfly($Upscale=3$)

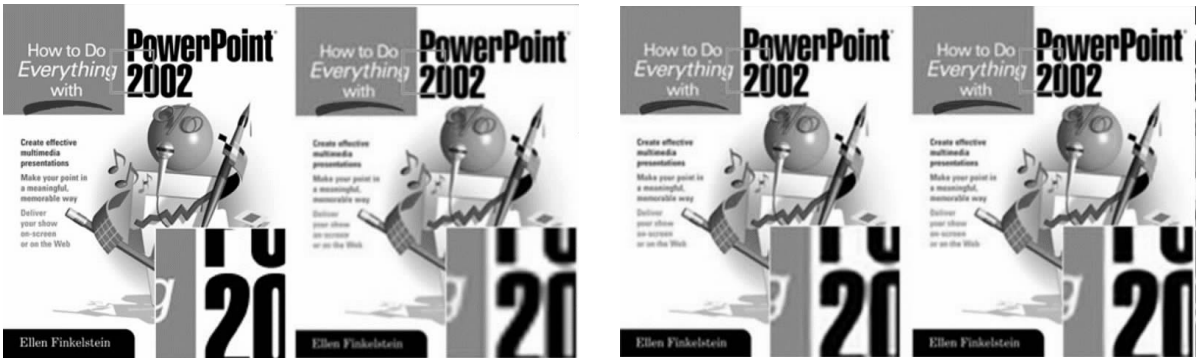


图 11　Set14 数据集中的图像 PPT3(*Upscale*＝3)

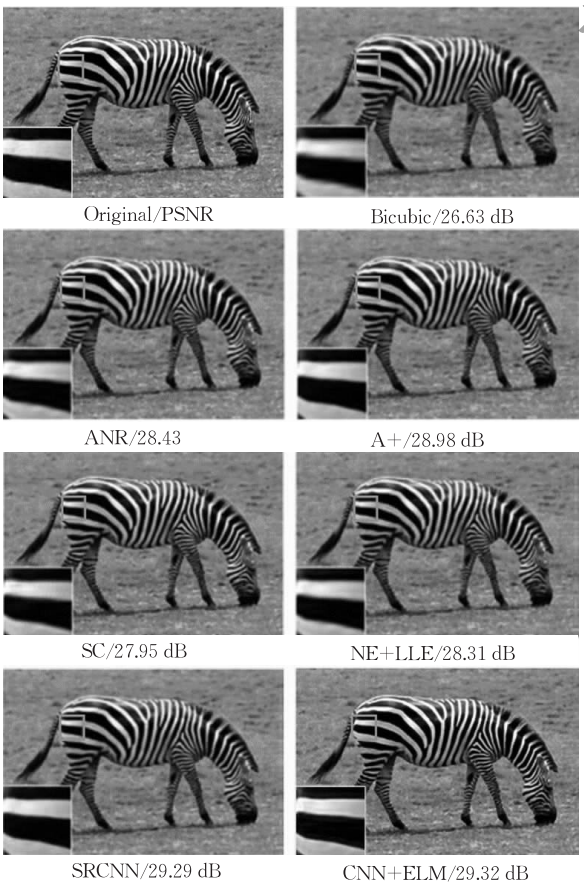


图 12　Set14 数据集中的图像 Zebra(*Upscale*＝3)

上述结果表明，本文所提的方法能够恢复出更清晰分明的图像边缘和更为丰富的细节信息，具有较好的视觉效果.

5.3　模型结构对超分辨率重构的影响

5.3.1　单一模型与二次重构

本文创新性的采用了二次重构方式对图像进行超分辨率的质量提升，但是，独立的卷积神经网络模型与基于 ELM 回归的模型都可以处理超分辨率重构问题. 为了验证二次重构的有效性，本小节对 SRCNN 方法和 ELM 方法，与本文提出的 CNN＋ELM 的二次重构方法进行了对比实验.

在重构效果方面其对比如图 13 所示.

在执行效率方面，采用相同训练集情况下，ELM 模型在本文实验环境中训练时长约为 0.3 时，本文的 CNN＋ELM 模型训练时长约为 144 时，SRCNN 模型在 GPU 上训练了约 72 时，在本文实验环境中 SRCNN 训练过程是无法在可接受时间内完成的. 但 CNN＋ELM 模型在测试过程中性能逊于采用单一模型.

由上述实验结果可以看出，二次重构的策略在超分辨率效果上优于单一的重构模型，并且实现了在普通计算机上进行高质量的 SR 模型训练.

ELM/28.87 dB CNN/29.78 dB ELM+CNN/30.12 dB

Bicubic/28.35 dB Ground Truth/PSNR

图 13 单一模型与二次重构主观效果对比

5.3.2 CNN 模型的优化

在采用卷积神经网络进行超分辨率重构时，本文采取了三项在 SRCNN 基础上的改进，实现了快速的模型训练. 其中，第一项为对训练数据进行基于结构特征的分类；第二项为利用分类后的数据进行预训练、调优结合的模型训练方式；第三项将 CNN 模型的参数规模缩减. 为了验证三项优化策略的有效性，设计实验如下：

(1) 在预训练过程中，三种对比方法分别为 SRCNN 方法，ISODATA 分类后的复杂数据集与 SRCNN 相同参数规模的方法(原始 CNN)，以及复杂数据集和减小参数规模后的方法(mini CNN)，以 Set5 中 Butterfly 图放大系数为 3 倍的重构结果作为对比，其重构后的 *PSNR* 指标如表 3 所示.

表 3 预训练阶段三种方法效果对比

迭代次数	SRCNN	分类+原始 CNN	分类+mini CNN
168 500	25.07 dB	25.49 dB	25.32 dB
379 000	26.11 dB	26.73 dB	26.71 dB
414 176	26.11 dB	**26.94** dB	**26.94** dB

由表 3 可以看出，本文采取的 ISODATA 分类策略在用复杂训练集进行预训练时参数的学习效率更高，同等迭代次数下重构效果优于 SRCNN 的方法. 具有小参数规模的模型 mini CNN 虽然在开始时重构效果逊于具有大规模参数的模型，但随着迭代次数的增加，其性能与大规模参数逐步接近，说明了在卷积神经网络模型上的训练远未达到收敛，本文的训练集也不会使模型过拟合，减小参数规模有助于提高训练效率.

在预训练迭代约 41 万次后，本文采用全部图像集对原始 CNN 和 mini CNN 网络分别进行调优，其各迭代阶段的 Butterfly 图像客观指标对比如表 4 所示.

表 4 Fine-tune 阶段参数规模对重构的影响

迭代次数	原始 CNN	mini CNN
100 000	26.96 dB	27.01 dB
200 000	26.99 dB	27.23 dB
450 000	27.21 dB	**27.55** dB

由上述实验可知，在采用全部训练集图像块调优阶段，大规模参数的网络随着迭代次数的增加重构效果有所提升，但提升速度较慢，而小规模参数的 mini CNN 仍保持较高的学习效率，最终重构质量好于采用大规模参数的原始 CNN.

最终的主观重构质量对比如图 14 所示.

SRCNN/27.58 dB CNN/27.21 dB mini CNN/27.55 dB

图 14 三种训练方式的最终重构效果

图 14 中从左到右依此为：SRCNN 方法迭代 8×10^5 的效果，CNN 采用大规模参数 Fine-tune 450 000 次的效果，减小参数规模的 mini CNN Fine-tune 450 000 次的效果.

由本小结的实验结果可以看出，本文的采取的优化 CNN 训练方法在相同迭代次数时对图像的重构效果优于 SRCNN 的方法，在迭代次数相差多个数量级(近 1000 倍)的情况下，仍能取得与 SRCNN 方法相近的结果，可推测本文方法若在高性能计算机上进行与 SRCNN 相同时间训练可获得更好的重构结果.

5.3.3 两种重构模型的结合方式

上文验证了二次重构对 SR 重构效果具有提升作用，但两种模型的结合方式对重构的效果也同样起着至关重要的作用. 通常情况下二次重构的一个直观的思路是级联模式，即以上一步的重构图像和原始图作为下一步重构的训练数据，目的是直接学习到上一步重构中所缺失的信息. 而本文采取两个模型独立训练的方式，可以与前一步共享原始训练图. 为了验证本文方法的有效性，设计了如下步骤的实验：

方案 1. 级联训练策略：

(1) 设原始图像集为 I_{HR}，经过降质后获得低分辨率图像集 I_{LR}，对 I_{LR} 采用 CNN 模型进行重构后输出图像集 I_{SR1}；

(2) 将 I_{SR1} 作为 ELM 的低分辨率训练数据，对 I_{SR1} 进行特征提取作为 ELM 模型的输入，获取高频分量 $I_{HF1}=I_{HR}-I_{SR1}$ 作为 ELM 训练模型的响应，进行模型的训练.

(3) 以同一 CNN 模型重构后的其他图像作为级联策略 ELM 模型的测试图像，重构结果为 I_1.

方案 2. 独立训练策略：

(1) 对于原始图像集 I_{HR}，采取与级联策略采取相同的降质方法获得低分辨率图像集 I_{LR}，对 I_{LR} 采用 CNN 模型进行重构后输出图像集 I_{SR2}；

(2) 直接将 I_{HR}，I_{LR} 作为 ELM 模型的训练数据，对 I_{LR} 进行特征提取作为模型输入，高频分量 $I_{HF2}=I_{HR}-I_{LR}$ 作为 ELM 模型训练的响应，进行模型训练.

(3) 以同一 CNN 模型重构后的其他图像作为级联策略 ELM 模型的测试图像，重构结果为 I_2.

为保证单一变量，本实验不进行训练数据放大系数的调整，与 CNN 模型进行等倍数的 ELM 模型训练. 级联训练策略与独立训练策略所得到的实验结果如图 15 所示，以 256×256 像素的 Lena 图作为测试数据，$Upscale=2$.

级联训练策略 PSNR: 34.77 dB　　独立训练策略 PSNR: 35.20 dB　　Bicubic对比 PSNR: 32.69 dB

图 15　模型结合方式的对比实验结果

由上述实验结果可以看出，本文的独立训练策略测试结果优于级联的训练方法. 其原因可能为基于结构性特征的重构本身便会产生误差，而以此为训练数据的二次重构会导致误差的进一步传播，致使重构的结果更差. 故在采取二次重构进行图像质量提升时，需考虑两个模型的结合方式是否合理.

5.3.4　二次重构时权重参数的影响

式(12)表达了本文提出的二次重构算法的核心思想，即将 CNN 和 ELM 两算法的输出结果进行叠加. 式中 γ 为平衡两项结果相对重要性的参数，其取值对最终的重构结果有着重要的影响. 本文以 Set5 中各个测试图像重构效果的 $PSNR$ 加和平均值作为评估指标，固定其他的参数，对参数 γ 的影响效果进行评估，结果如图 16 所示.

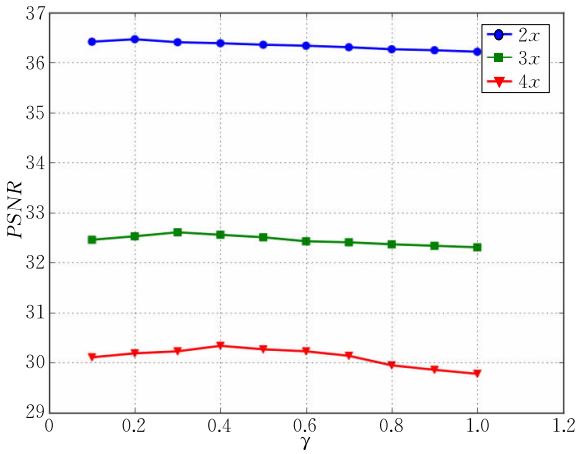


图 16　参数 γ 对重构效果的影响

由上述实验结果可以发现，γ 过大或过小都会导致重构的效果变差. 而当放大系数越大时，重构效果最好的点出现的位置越靠右. 说明当放大系数越大时图像损失的信息越多，而二次重构所补充的高频分量的相对会显得越来越重要.

5.3.5　ELM 训练时放大系数的影响

由于本文的 CNN 模型与 ELM 模型是独立进行训练的，故其训练过程相互不干扰，由图 7 效果可知，放大系数较小的 ELM 模型所学习到的高频分量能够表现出更多的细节信息. 由于 CNN 重构后已补充了图像大部分的缺失的信息(中高频)，ELM 训练时放大系数小的所补充的高频率的部分更多，即 CNN 重构后仍缺失的部分. 对于全局放大系数 k，本文采用小放大系数 $ELM\text{-}Upscale \leq k$ 的训练数据对 ELM 模型进行训练，并设计了不同放大系数对图像重构的影响实验来验证本文方法的有效性. 为了使最终的 ELM 模型的 $ELM\text{-}Upscale$ 具有固定的值，本实验将重构时每组的 γ 值调到最优，以 Set5 中的各个测试图像重构效果的 $PSNR$ 加和平均值作为评估指标，得到结果如图 17 所示.

由上述实验结果可以看出，ELM 训练时参数 $ELM\text{-}Upscale$ 取值小于全局放大系数时可以获得更好的重构效果，验证了本文在二次重构时这一改进的有效性. 但是过大或过小的放大系数都会对重构产生不利的影响，若不对权重参数 γ 进行调整，可能会导致重构效果比单独的重构模型更差.

总体而言，当 $ELM\text{-}Upscale$ 取值为 1.5～2 之间时都可以得到较好的重构效果. 本文所述的方法中

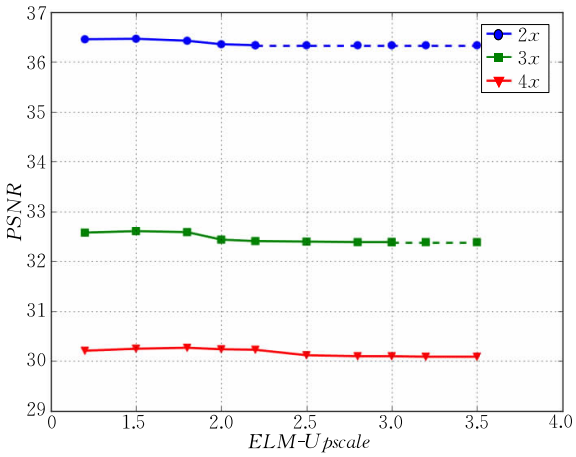


图 17 *ELM-Upscale* 对重构效果的影响

在全局放大系数为 2 和 3 时，取定 *ELM-Upscale* 的值为 1.5，全局放大系数为 4 时取定 *ELM-Upscale* 的值为 1.8.

5.3.6 ELM 训练时隐层节点个数的影响

在 ELM 算法中，唯一需要调节的参数即隐层节点的个数，设为 *NodeNum*，隐层节点数越多，ELM 的学习能力越强. 本文采用的 ELM 训练数据集远大于隐层节点的个数，故不存在过拟合的问题. 为了探究隐层节点数对重构效果的影响，本文针对不同隐节点数的 SR 重构进行了实验，考虑到计算机内存的限制，本文采用了较少数目的训练数据集（约 20 万像素点）. 以 Set5 中的各个测试图像重构效果 *PSNR* 加和平均值作为评估指标，固定其他的指标参数，得到结果如图 18 所示.

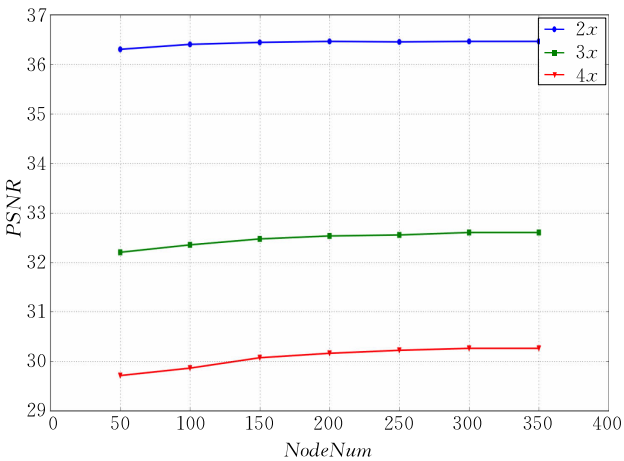


图 18 不同隐节点个数对重构效果的影响

对上述实验结果进行分析可以得知，对于较小放大系数的测试图像，如 2 倍下采样的图像，当隐层节点数达到 150 以上时，继续增大隐节点的个数对最终的重构效果影响不大，图像趋于一条直线. 当放大系数较大时，如对测试数据进行 4 倍下采样，ELM 隐层节点数的增加会对重构结果的提升有较大的促进作用. 故而对于高放大系数的超分辨率重构问题，在一定的内存限制下，可采取减少训练数据，增加隐节点个数的策略以达到重构效果的提升.

5.3.7 模型结构及参数总体分析

本文所提出的二次重构模型涉及了诸多参数，包括 CNN 和 ELM 模型规模，两种模型的结合方式，模型测试时两项的相对权重，模型训练过程的优化方式，以及 ELM 节点个数. 本节的前几小节对各项参数对重构效果的影响给出了具体的实验，在实验及实际应用场景中需要针对不同放大系数的图像分别进行 CNN 及 ELM 模型的训练及各项参数（γ, *ELM-Upscale*, *NodeNum*）的调节. 总体而言，采用模型独立训练方式的二次重构方法，相同放大系数的图像共享一致的重构模型及参数时能够得到最佳的重构效果.

5.4 基于 SR 技术的高清图文传输应用

在远程视频交互系统中，如视频会议系统中，由于受限于摄像设备及网络资源，纸质文本资料通常无法有效的共享和传达，而在用户需要进行纸质资料共享时往往需采用额外的设备（扫描仪等），转成电子版后进行远程协同共享，使得纸质资料共享的需求不能得到便捷高效的满足.

为了解决上述问题，本文将基于 CNN＋ELM 的超分辨率重构模型集成到远程视频交互系统中，作为机物协同交互中实体材料的高清传输功能，可实现利用现有设备（摄像头）来进行实体资料的共享，尤其是满足实际需求较多的文本材料的传输需求. 其应用场景和如图 19 所示.

远程交互系统中高清实物图像传输功能可通过摄像头直接对文本资料进行拍摄采集，利用 SR 技术对图像进行质量的提升，能够使得原本不清晰的图像和文字变得清晰起来.

为了验证超分辨率重构功能在远程视频交互系统中提高文字图像质量的有效性，本文采用沉浸式交互系统中常用的视觉传感器，微软公司 Kinect 的彩色摄像头对纸质文本材料进行拍摄，采用打印着不同字体及字号的文本材料作为实验素材，并将拍摄图像转化为灰度图. 利用本文提出的基于 CNN＋ELM 的二次重构模型处理后，获得的主观效果对比如图 20 所示（可放大后查看）.

本实验中采用的 Kinect 摄像头的分辨率为 640×480，在本文进行实验的普通计算机上处理一帧图像的时间为 39 s，通常认为该时长在用户可接

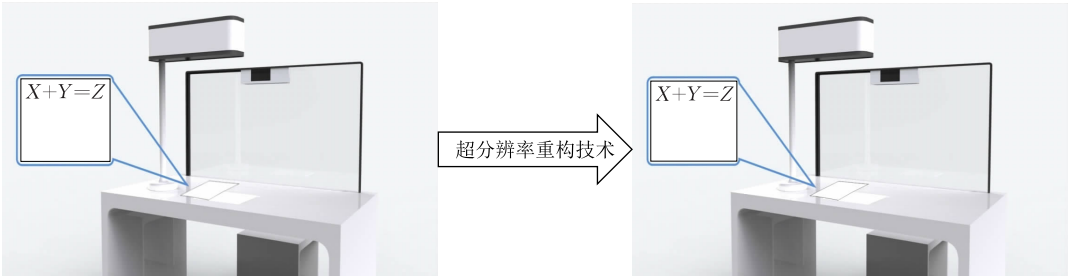
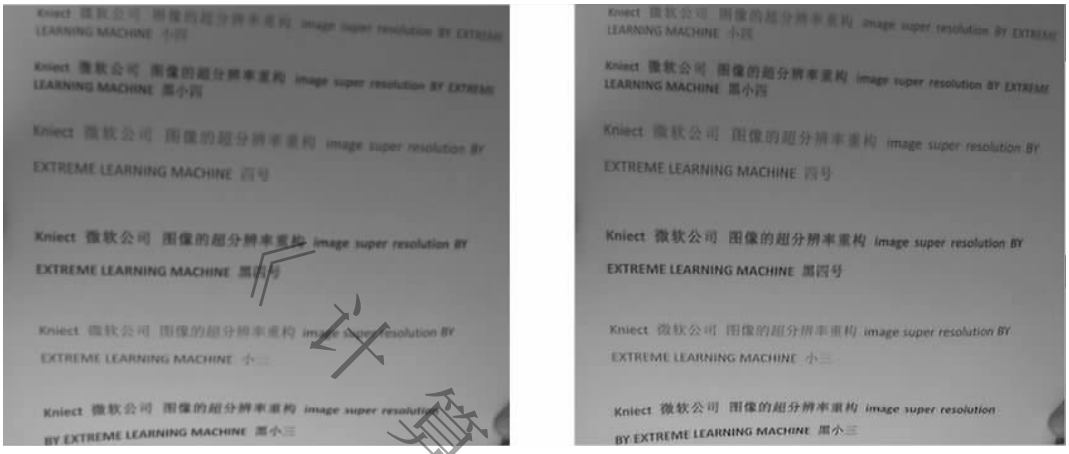


图 19　SR 技术在远程交互系统中的应用



原始灰度图　　超分辨率重构后

图 20　本文模型对纸质文本材料的重构效果

受的范围内.

由上述实验结果可以看出，本文所述超分辨率重构方法处理后对较细笔画的文字具有边缘提升的效果，使得文字更为易于辨认，说明本文的 SR 方法可以在一定程度上解决远程视频交互系统中高清图文传输现存的问题.

6　总　结

本文提出了一种基于卷积神经网络与极限学习机的二次超分辨率重构策略，在训练数据的预处理、模型的训练方法、网络模型的规模、ELM 的训练方式和两种模型的结合方式等多个关键点进行改进和创新，实现了具有较高训练效率和较优视觉效果的联合超分辨率重构方法，并在实际应用场景中验证了有效性. 较之 SR 领域目前取得领先成果的其他方法，本文的 SR 方法训练效率更高，重构效果的提升显著. 与此同时，本文所提出逐步提高图像质量的二次重构框架可以拓展到与其他各类算法的优化组合(如 SC+ELM)，方便解决更多不同需求的实际应用问题.

本文在超分辨率重构的图像质量提升方面取得了一定的研究成果，但是像素级的特征提取在测试时会耗费较多的时间. 故本文下一步将在保证重构质量的前提下着重提升算法的测试效率. 为此，拟对图像集进行感兴趣区域的提取，如对图像进行分块，仅对具有明显灰度变化的图像块进行训练/重构，过滤掉较平滑的区域. 此外，本研究将考虑把 CNN+ELM 整体模型在 Caffe 上实现，使得两种学习模型更紧密的结合，进一步优化超分辨率重构模型训练及测试过程.

参　考　文　献

[1] Huang T S, Tsai R. Multi-frame image restoration and registration. Advances in Computer Vision and Image Processing, 1984, 1: 317-339

[2] Goodman J W. Introduction to Fourier optics. Physics Today, 1996, 22(4): 97-101

[3] Katsaggelos A K, Lay K T, Galatsanos N P. A general framework for frequency domain multi-channel signal processing. IEEE Transactions on Image Processing, 1993, 2(3): 417-420

[4] Ji H, Fermuller C. Robust wavelet-based super-resolution reconstruction: Theory and algorithm. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(4): 649-660

[5] Sun J, Xu Z, Shum H. Image super-resolution using gradient profile prior//Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). Anchorage, USA, 2008: 1-8

[6] Giachetti A, Asuni N. Real-time artifact-free image upscaling. IEEE Transactions on Image Processing, 2011, 20(10): 2760-2768

[7] Irani M, Peleg S. Improving resolution by image registration. CVGIP: Graphical Models & Image Processing, 1991, 53(3): 231-239

[8] Schultz R R, Stevenson R L. Video resolution enhancement. Proceedings of SPIE, 1995, 2421(1): 23-34

[9] Yang J, Wright J, Huang T S, et al. Image super-resolution as sparse representation of raw image patches//Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). Anchorage, USA, 2008: 1-8

[10] Yang J, Wright J, Shuang T. Image super-resolution via sparse representation. IEEE Transactions on Image Processing, 2010, 19(11): 2861-2873

[11] Yang J, Wang Z, Lin Z. Coupled dictionary training for image super-resolution. IEEE Transactions on Image Processing, 2012, 21(8): 3467-3478

[12] He L, Qi H, Zaretzki R, et al. Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution//Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). Portland, USA, 2013: 345-352

[13] Timofte R, De V, Gool L V. Anchored neighborhood regression for fast example-based super-resolution//Proceedings of the IEEE International Conference on International Conference on Computer Vision (ICCV). Sydney, Australia, 2013: 1920-1927

[14] Zhu Y, Zhang Y, Yuille A, et al. Single image super-resolution using deformable patches//Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, USA, 2014: 2917-2924

[15] Zhu Y, Zhang Y, Bonev B, et al. Modeling deformable gradient compositions for single-image super-resolution//Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA, 2015: 5417-5425

[16] Yang C, Huang J, Yang M, et al. Exploiting self-similarities for single frame super-resolution//Proceedings of the 10th Asian Conference on Computer Vision. Queenstown. New Zealand, 2010: 497-510

[17] Glasner D, Bagon S, Irani M, et al. Super-resolution from a single image//Proceedings of the IEEE International Conference on International Conference on Computer Vision (ICCV). Kyoto, Japan, 2009: 349-356

[18] Mairal J, Bach F, Ponce J, et al. Non-local sparse models for image restoration//Proceedings of the IEEE International Conference on Computer Vision (ICCV). Kyoto, Japan, 2009: 2272-2279

[19] Freedman G, Fattal R. Image and video upscaling from local self-examples. ACM Transactions on Graphics, 2011, 30(2): Article12, 1-11

[20] Cui Z, Chang H, Shan S, et al. Deep network cascade for image super-resolution//Proceedings of the IEEE International European Conference on Computer Vision (ECCV). Zurich, Switzerland, 2014: 49-64

[21] Chang H, Yeung D, Xiong Y, et al. Super-resolution through neighbor embedding//Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). Washington, USA, 2004: 275-282

[22] Chen Xiao-Xuan, Qi Chun. Single-image super-resolution via low-rank matrix recovery and joint learning. Chinese Journal of Computers, 2014, 37(6): 1372-1379(in Chinese)
(陈晓璇, 齐春. 基于低秩矩阵恢复和联合学习的图像超分辨率重建. 计算机学报, 2014, 37(6): 1372-1379)

[23] Qiao Jianping, Liu Ju, Zhao Caihua. A novel SVM-based blind super-resolution algorithm//Proceedings of the 2006 IEEE International Joint Conference on Neural Network Proceedings. Vancouver, Canada, 2006: 2523-2528

[24] An L, Bhanu B. Image super-resolution by extreme learning machine//Proceedings of the IEEE International Conference on Image Processing (ICIP). Orlando, USA, 2012: 2209-2212

[25] Gao J, Guo Y, Yin M, et al. Restricted Boltzmann machine approach to couple dictionary training for image super-resolution //Proceedings of the IEEE International Conference on Image Processing (ICIP). Melbourne, Australia, 2013: 499-503

[26] Dong C, Changeloy C, He K. Learning a deep convolutional network for image super-resolution//Proceedings of the IEEE International Conference on European Conference on Computer Vision (ECCV). Zurich, Switzerland, 2014: 184-199

[27] Wang Z, Yang Y, Wang Z, et al. Self-tuned deep super resolution//Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition Deep Learning for Computer Vision Workshop (CVPR DeepVision). Boston, USA, 2015: 1-8

[28] LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series//Arbib M A ed. The Handbook of Brain Theory and Neural Networks. Cambridge, USA: MIT Press, 1998: 255-258

[29] Lécun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11):2278-2324

[30] Huang G B, et al. Extreme learning machine: Theory and applications. Neurocomputing, 2006, 70(1): 489-501

ZHANG Jing, born in 1990, M. S. Her research interests include image processing and machine learning.

CHEN Yi-Qiang, born in 1973, Ph. D., professor, Ph. D. supervisor. His main research interests include human computer interaction and ubiquitous computing.

JI Wen, born in 1976, Ph. D., professor, Ph. D. supervisor. Her main research interests include information coding and multimedia communication network.

Background

High-resolution images are capable of offering more abundant details, not only satisfy people's need for visual effect, also lay a solid foundation of implementing other visual analysis task. Image super-resolution is proved to be an effective method providing high-resolution images. The very essential basic of this technology is performing image reconstruction on low-quality images using image processing techniques to generate high-quality ones. Whereas the image deterioration is irreversible due to down-sampling during the process of transfer and storage, image super-resolution an ill-posed problem. While, the key point of image super-resolution is to find the mapping relation and complementation information between low and high quality images in order to search the feasible solution. Many other method tend to learn the mapping function between high-resolution and low-resolution imagines by building different models, but as the reconstruction quality becoming better, the training time and computing consumption become larger. Therefore, this paper proposes an image super-resolution method that can improve the efficiency of training largely as while as achieving better reconstruction quality. Proposed method take advantages of the original model of CNN (Convolutional Neural Networks) and ELM (Extreme Learning Machines), implement a two-tie super-resolution model which manage to complete training process on normal computer. Using our method, Fine-visual high-resolution images can be constructed without GPU and other external computing device.

This work was supported by the National Natural Science Foundation of China (No. 61572466, No. 61472399, No. 61572471), the Chinese Academy of Sciences Research Equipment Development Project under Grant No. YZ201527, the Natural Science Foundation of Beijing No. 4162059. The project aims to promote the development of multimedia and human computer interactive technology. The team has published some high quality papers in related area.