

面向图像超分辨率的上下文字典学习

于伟 姚鸿勋 孙晓帅 刘先明 许鹏飞

(哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001)

摘要 基于稀疏表示理论,提出了一种面向单张图片超分辨率的字典学习方法。通过对训练数据进行分类,期望在每一类训练数据训练字典的过程中,增强类内的上下文信息。与之前的面向图像分类的字典学习方法所不同的是,训练数据集由高分辨率图像块和对应的低分辨率图像块共同组成,这使训练得到的字典更适用于图像重构。利用有限的训练数据集,基于上下文的字典学习方法能够提高字典表示的拓展能力,消除由多重训练数据子集带来的冗余。

关键词 单张图片超分辨率,稀疏表示,上下文字典,图像块分类

中图分类号 TP37 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.10.020

Contextual Dictionary Learning for Super Resolution

YU Wei YAO Hong-xun SUN Xiao-shuai LIU Xian-ming XU Peng-fei

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

Abstract This paper proposed a novel dictionary learning method for single image super resolution based on sparse representation. We tried to utilize patch-level clustering to enhance the contextual information in atom learning stage. Unlike the previous dictionary learning works using the image classification, our training set is constructed from the high-resolution and low-resolution patch pairs labeled by different patch-level class, which is more appropriate for image reconstruction. This approach tried to promote the transfer ability of the dictionary which is built on a limited training set and can eliminate the atoms redundancy introduced by multiple training subsets.

Keywords Single image super resolution, Sparse representation, Contextual dictionary, Patch-level clustering

1 引言

数十年来,图像超分辨率技术逐渐在多个领域的应用中取得了长足的进步,如视频增强、数字监控和图像重构等等。对于给定的低分辨率图片,单张图片超分辨率技术旨在恢复其原始图片,即图片的高分辨率版本。然而,图像退化过程会导致大量的图像信息丢失,所引起的低分辨率图像的不适定性(ill-posed),也成为图像超分辨率过程所面临的最大挑战。为了解决这样的问题,利用一定的先验知识或引入相关的限制,是提高单张图片超分辨率技术的必要途径。

当前,性能最好的超分辨率方法一般是利用高分辨率图片和低分辨率图片之间的共生先验(co-occurrence prior)进行图像重构,即在图像重构之前,首先需要获取高分辨率图像块和相对应的低分辨率图像块之间的共生先验。这样的方法大多是基于学习的,并假设训练数据集中的高、低分辨率图像块能够有效地预测出输入图片所丢失的高频信息。如 Freeman 等人^[1]提出了基于样例的学习方法,即利用马尔科夫随机场构建图像和场景之间的关系模型。而 Sun 等人^[2]则将草图先验(sketch prior)引入到 Freeman 的方法中,并取得了令人满

意的结果。

随着压缩感知理论在计算机视觉领域内的应用的深入,稀疏表示技术在解决计算机视觉领域内多个问题中都体现出了强大的性能。为了解决高分辨率图片的恢复问题, Yang 等人^[3]利用稀疏表示技术构建了一个图像重构框架,其核心是一种新型的字典训练方法。首先,该框架分别对自然图像的高分辨率图像块和低分辨率图像块训练对应的字典,进而利用输入图像的图像块在低分辨率字典上的稀疏表示系数,结合高分辨率字典产生对应的高分辨率图像块,最终利用这些高分辨率图像块恢复出高分辨率图片。

同时,稀疏表示技术也引入到解决物体识别的问题当中,研究人员也提出了一系列字典训练方法。例如,基于费舍尔判别准则(Fisher Discrimination criterion), Yang 等人^[4]提出了 FDDL(Fisher Discrimination Dictionary Learning)算法,使训练得到的字典的每一个基都对应一个物体类别标签。求解对应该字典的稀疏表示系数时,增强系数的类间散度并削弱系数的类内散度,以提高不同物体对应的稀疏表示系数的可区分性。无独有偶, Zhou 等人^[5]则同样基于费舍尔判别准则提出了一个 JDL(Joint Dictionary Learning)算法。通过挖掘

到稿日期:2013-08-09 返修日期:2013-08-12 本文受国家自然科学基金项目(61071180),国家自然科学基金重点项目(61133003)资助。

于伟 男,博士,主要研究方向为计算机视觉、模式识别;姚鸿勋 女,博士,教授,博士生导师,主要研究方向为图像处理、模式识别、多媒体技术及自然人机交互, E-mail: h. yao@hit.edu.cn(通信作者);孙晓帅 男,博士,主要研究方向为计算机视觉、显著度检测;刘先明 男,博士,主要研究方向为计算机视觉、图像检索;许鹏飞 男,博士,主要研究方向为计算机视觉、文字检测。

一组视觉上相近的物体之间的关系, JDL 算法能够学习出一个适用于所有类别物体的共享字典以及多个仅适用于单个类别物体的特定字典, 以此来增强这一组字典的判别能力和表示能力。

很明显, 增强字典基之间的判别能力能够提高字典表示能力。然而为了提高字典的表示能力并获得更好的重构结果, 图像超分辨率技术中有几个问题是不能忽略的:

1. 在基于样例的字典训练算法中, 物体的类别^[4,5]和图像块的类别是不等价的。即使两张图片在物体级别上是视觉相关的, 但是组成图片的图像块并不一定具有很高的相关性, 例如, 来自同一动物不同部分的图像块就具有完全不同的纹理信息。

2. 相对于单字典来说, FDDL 提出的字典组的表示能力明显要更强一些^[5], 然而不可避免地, 不同字典之间存在着基的冗余。显然, 这些冗余的基可能仅属于若干个类别, 而不是所有类别。因此, 这些基重复属于对应的特定字典中, 而不属于对应所有类别的共享字典中。在字典训练过程中, 这些冗余的基就意味着需要消耗更多的计算资源。

3. 一般来说, 字典的学习仅仅依靠有限的训练数据, 以此得到一组对应大多数训练样本的完备基^[3]。因此, 当训练样本和输入图像差距较大时, 字典的性能也会随之下降。

考虑到以上 3 个方面, 本文提出了一个面向单张图片超分辨率的新型字典训练算法。与 FDDL 算法相比, 本文提出的字典训练方法学习得到单个字典, 字典中不同基的组合对应着不同类别, 而不是得到多个字典。同样, 在基于样例学习的字典训练算法中, 我们试图利用图像块聚类而非图像分类, 在基训练阶段增强字典的上下文信息。为了避免概念的误解, 本文所提到的上下文指的是相似图像块在特征空间中的局部关系, 而不是图像块在图片中的局部关系。相比于物体级别的分类, 我们相信, 图像块级别的聚类更有助于学习出具有特定表达能力的基。同时, 基于费舍尔准则增强图像块级别的上下文信息能够提高训练出来的基的判别能力, 换言之, 这样的基在表达能力上更具有针对性, 也更加全面。即使训练数据和测试数据差别较大, 由这样的基组成的字典在表达能力上也拥有更好的拓展能力。

2 图像超分辨率中的稀疏表示

对应于不同的字典, 高分辨率图像块和低分辨率图像块能够产生相应的稀疏表示系数, 而基于稀疏表示的图像超分辨率算法的核心思想在于找到同一图像块高分辨率版本和低分辨率版本的稀疏表示系数的映射关系。本章将对系数映射的学习过程和图像的重构过程进行数学上的描述。

2.1 超分辨率中的稀疏表示理论简介

对于给定的过完备字典 $D \in R^{n \times k}$, 其中 $k > n$, 向量 $x \in R^n$ 能够表示为字典 D 中包含的基的稀疏的线性组合:

$$x = D\alpha \text{ with } \|\alpha\|_0 \ll k \quad (1)$$

其中, $\alpha \in R^k$ 是稀疏表示系数向量, 由大部分的零分量和极少量的非零分量组成。在优化过程中, l_1 范式意味着目标函数的非凸可优化性, 而相关研究已经证明稀疏表示优化过程中

l_1 范式和 l_0 范式是等价的。根据文献[3]中的内容, 基于稀疏表示的图像超分辨率问题可以用如下公式进行表示:

$$\min \lambda \|\alpha\|_1 + \|D\alpha - x\|_2^2 \quad (2)$$

从认知的角度上来说, 字典 D 能够分为两个部分: D^d 和 D^c , 分别对应着具有判别力的基和共享基。因此, 式(2)的上边界可以估计为:

$$\begin{aligned} \|D\alpha - x\|_2^2 &= \|[D^d, D^c]\alpha - x\|_2^2 \\ &\leq \|D^d\alpha^d - x\|_2^2 + \|D^c\alpha^c - x\|_2^2 \end{aligned} \quad (3)$$

其中, α^d 和 α^c 分别对应具有判别力的基的系数和对应共享基的系数。

对于向量 x , D^d 中的基包含大部分有用的上下文信息, 这些信息有助于精确地重构图像细节。相较于 α^c 中的响应值, α^d 中的响应值也要相对高一些, 这就避免了过多地引入降低稀疏表示性能的无用分量。

然而在实际中, 找到一组满足所有样例的完备基是非常困难的, 因此, 超分辨率问题很难被完美地解决。但幸运的是, 优化一组与图像块上下文相关的紧凑且具有判别力的基相对更容易一些, 它也能在优化过程中更快速地收敛。

2.2 图像块级别的上下文字典训练

正如前面所分析的, 在图像块级别上训练具有判别力的基有利于增强字典的上下文信息。式(2)中, 第一项是稀疏优化项, 第二项是重构优化项。为了在图像块级别上训练出有利于图像重构的上下文字典, 需要对式(2)进行两方面的修改。首先, 在训练具有判别力的基的策略下, 重构优化项需要修改为:

$$\|D\alpha - x\|_2^2 + \|D^d\alpha^d - x\|_2^2 + \|D^c\alpha^c\|_2^2 \quad (4)$$

进而, 基于费舍尔判别准则, 增加一个优化项可以提高基的判别力, 这一项的优化主要是通过保证不同图像块类别的稀疏系数之间的可区分性, 来训练学习得到具有判别力的基^[5]。对于由图像块组成的训练数据集 X , 其对应的稀疏表示系数集合记为 A , 训练数据集可以在图像块级别上分为 c 个类别。对此, 该优化项由两部分组成: 第一部分最小化 A 的类内散度, 记为:

$$S_W(A) = \sum_{i=1}^c \sum_{j=1}^{n_i} (a_{i,j} - m_i)(a_{i,j} - m_i)^T \quad (5)$$

第二部分最大化 A 的类间散度, 记为

$$S_B(A) = \sum_{i=1}^c n_i (m_i - m)(m_i - m)^T \quad (6)$$

其中, m_i 代表第 i 个类别的训练数据对应的稀疏系数集合 A_i 的平均向量, m 代表所有数据对应的稀疏系数集合 A 的平均向量, $a_{i,j}$ 是稀疏系数集合 A_i 中的一个样本, n_i 则是该集合中样本的数量。

基于这样两个考虑, 本文提出的字典训练方法的目标函数可以分为 3 个部分: 稀疏表示优化项、重构优化项和判别力优化项。因此, 目标函数可以定义为:

$$\begin{aligned} D = \arg \min \{ &\lambda_1 \sum_{i=1}^c \sum_{j=1}^{n_i} \|a_{i,j}\|_1 + \lambda_2 \text{tr}(S_W(A)) - \text{tr}(S_B(A)) \\ &+ \sum_{i=1}^c \sum_{j=1}^{n_i} (\|D a_{i,j} - x_{i,j}\|_2^2 + \|D^d a_{i,j}^d - x_{i,j}\|_2^2 \\ &+ \|D^c a_{i,j}^c\|_2^2) \} \end{aligned} \quad (7)$$

其中, λ_1 和 λ_2 用来均衡 3 个优化项之间的惩罚程度。 λ_1 较大时, 表示向量的稀疏的惩罚将提高; λ_2 较大时, 表示向量判别程度的惩罚将提高; λ_1 和 λ_2 较小时, 稀疏表示的重构惩罚将提高。在实验过程中, λ_1 和 λ_2 依靠经验进行赋值。

2.3 目标函数的优化

为了优化式(7), 目标函数可以分为两个优化子问题: 固定字典 D , 升级稀疏表示系数集合 A ; 固定稀疏表示系数集合 A , 升级字典 D 。通过迭代优化这两个子问题, 直到得到最优化的字典 D 。固定 D 升级 A 的过程中, 利用迭代映射方法 (IPM, Iterative Projection Method) 对子问题进行优化; 固定

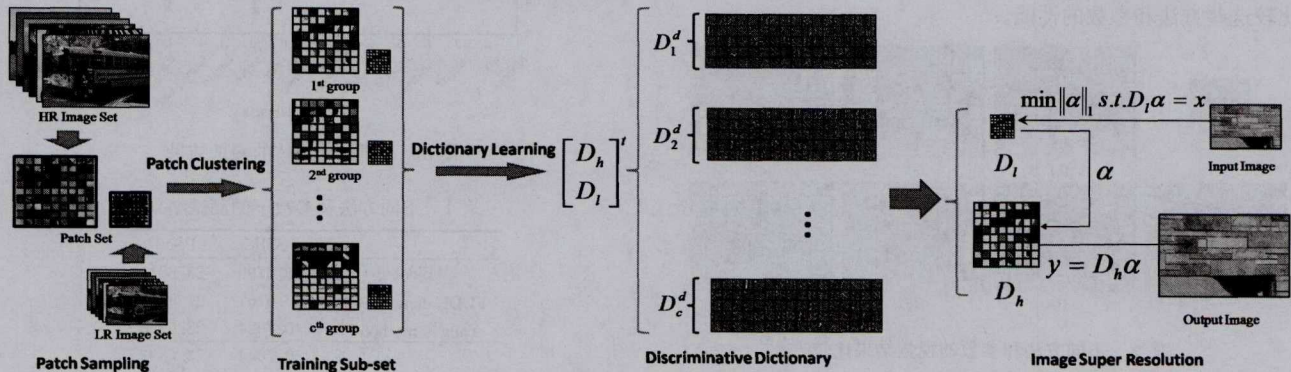


图 1 超分辨率算法框架示意图

为了整合上下文信息, 训练数据利用图像块级别的相似度采样获得, 而非物体级别的相似度。在基于上下文的图像块聚类阶段, 首先在训练图片集中随机采样出一组图像块训练数据集。进而基于特征空间的视觉相似度, 将这组数据集聚类成多个子训练数据集。图 2 显示了不同子训练数据集中的图像块在图片中的位置, 同一类别的图像块被染色成相同的颜色。

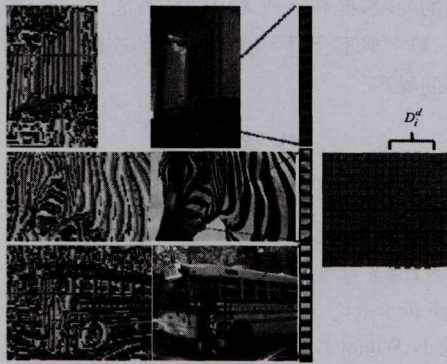


图 2 图像块级别的聚类示意图

图 2 主要展示了属于同一类别的图像块。虽然图 2 中展示的 3 组图像块采样自不同的物体, 但是很明显, 展示的图像块在本章训练得到的字典上均可以被同一组基所响应。这种图像块级别上的一致性提供了相关的先验知识, 能够更好地帮助视觉上相似的图像进行纹理重构和细节恢复。

正如 2.1 节所提到的, 图片的重构工作需要两个字典 D_h 和 D_l , 其中训练字典 D_h 是为了表示高分辨率图像块组成的训练数据集, 训练字典 D_l 是为了表示低分辨率图像块组成的训练数据集。在字典训练阶段, D_h 和 D_l 可以联合成 $D = [D_h; D_l]$ 进行统一训练。为了优化式(7), 目标函数可以划分

A 升级 D 的过程中, 利用二次规划算法 (Quadratic Programming algorithm) 对子问题进行优化。

3 基于图像块级别的上下文字典的图像超分辨率

基于以上章节的分析和讨论, 本节将提出一个上下文信息字典的训练方法, 即有效地利用上下文信息训练得到一组具有判别力的基构建字典, 进而利用该字典实现基于稀疏表示的图像超分辨率框架。该框架分为 3 个阶段: 基于上下文的图像块聚类、字典训练和图像超分辨率。图 1 显示了这个框架的整个过程。

为两个优化子问题^[4,5]: 首先在固定联合字典 D 的基础上更新稀疏表示系数集合 $\{a_{i,j}\}$; 进而在固定稀疏表示稀疏集合 $\{a_{i,j}\}$ 的基础上更新字典 D , 以此往复, 直到稀疏表示稀疏集合 $\{a_{i,j}\}$ 和字典 D 收敛到固定值。其中, 不同类别的图像块数据集分别训练出对应的基的集合, 并由这些基共同组合成字典 D 。

在图像超分辨率阶段, 每一个输入的低分辨率图像块都可以根据字典 D_l 计算得到一个稀疏表达系数 a , 而对应的高分辨率图像块可以通过这个稀疏表达系数 a 和字典 D_h 产生。将产生的高分辨率图像块组成高分辨率图片, 并加入重构限制, 得到的就是最终的超分辨率结果。

4 实验

实验中, 输入的低分辨率图像和对应的高分辨率图像之间的放大倍数设置为 3。低分辨率图片中每一个图像块的大小设置为 3×3 , 对应地, 高分辨率图片中每一个图像块的大小设置为 9×9 。训练数据集由 Caltech256 数据中 6 种不同的物体组成, 包括芙蓉花 (hibiscus)、尖塔 (minaret)、棕榈树 (palm-tree)、校车 (school-bus)、摩天楼 (skyscraper) 和斑马 (zebra), 每一个种类均包括 16 张高分辨率图片。为了训练字典 D_h 和 D_l , 每一张高分辨率图片都对应有一张下采样倍数为 $1/3$ 的低分辨率图片。

正如文献[3]所描述的, 观察者对彩色图片的亮度更为敏感, 也出于有效降低算法计算量的考虑, 本文实验同样仅对彩色图片的亮度通道进行超分辨率处理。实验中, 为了整合邻域信息以增强图像块在图片上的空间上下文特性, 在圆盘亮度通道上提取一阶梯度和二阶梯度, 作为待处理的低分辨率图片的表示。

在基于上下文信息的图像块聚类阶段,基于低分辨率图像块的特征,训练数据集被分为若干个子训练数据集。实验中,子训练数据集的数目分别设置为4、8、16和32。针对一个输入的低分辨率“台历”图片,图3展示了利用不同方法和参数产生的高分辨率结果。这些结果的放大倍数均为3,其中,(a)为输入的低分辨率图片,(b)为双三次插值的结果,(c)为基于FDDL字典的超分辨率结果,(d)为Yang^[3]方法的结果,(e)–(h)为本文方法的结果,其中子训练数据集的个数分别为4、8、16、32。输入图片为高分辨率图片下采样得到的低分辨率版本,通过观察日期部分和背景画面部分,可以从视觉上比较这些方法和参数的性能。

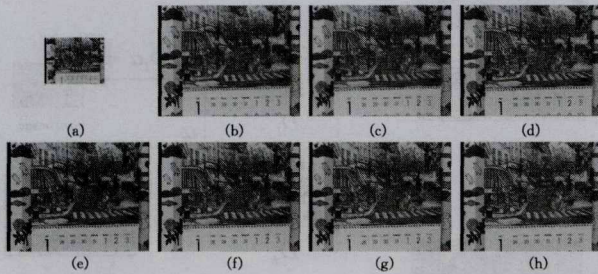


图3 不同方法和参数的视觉结果比较

为了量化实验性能,分别用SSIM和PSNR两个指标作为衡量图像重构结果好坏的标准。SSIM和PSNR均是用来衡量两张图片相似度的标准,重构结果越好衡量值也越大。为了测试算法的性能,我们从Caltech256中挑选16个类别的自然图片作为测试数据,包括动物、建筑、植物、自然风光等等。测试数据中一个样本的视觉比较结果如图4所示,该样本来自Caltech256中的天鹅(swan)类别。结果的放大倍数同样是3,其中,(a)为输入的低分辨率图片,(b)为双三次插值的结果,(c)为基于FDDL字典的超分辨率结果,(d)为Yang^[3]方法的结果,(e)–(h)为本文方法的结果,其中子训练数据集的个数分别为4、8、16、32。对于训练数据集中不存在的禽类,从图中可以看出不同结果的差距,能够看出本文的方法在字典表达能力上拥有很好的拓展能力。同时也避免了图像块冗余表示的引入,超分辨率得到的结果更加自然,噪声更少。

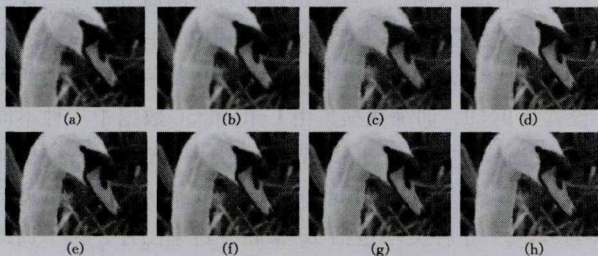


图4 测试数据集中一个样本视觉结果的比较

对于不同类别的测试数据集,图5展示了不同方法和参数下的SSIM平均值。表1则对比在不同的字典训练方法和参数下所有测试数据集得到的SSIM平均值和PSNR平均

值。从数值上来看,随着子训练数据数目的增多,图像重构的性能也随之提高。然而随着训练数据集的过度分类,性能将不会再有大幅度提升,如将子训练数据集数目设置为32时。

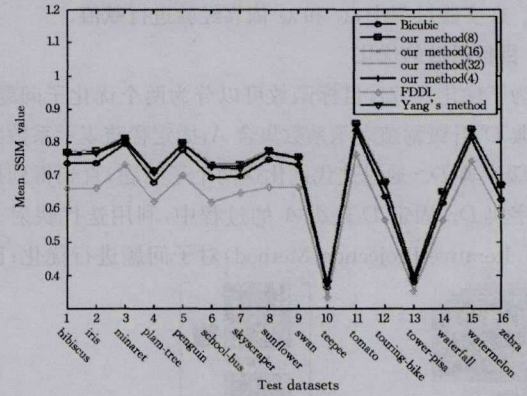


图5 不同类别物体的测试结果

表1 不同方法及参数下的测试结果

	SSIM	PSNR
Bicubic	0.6790	23.0011
FDDL dictionary ^[4]	0.6161	21.9158
Yang's method ^[3]	0.7024	23.6358
Our method		
4	0.7044	23.6832
8	0.7094	23.8636
16	0.7099	23.9006
32	0.7084	23.8358

结束语 基于稀疏表示的超分辨率方法能够得到与原图相近的重构结果,一个重要原因是这种方法能够利用训练得到的字典预测出低分辨率图片已经丢失但存在于训练数据中的细节部分。然而,也必须承认这样一个事实,重构图片的质量在很大程度上依赖于训练数据以及字典训练方法。通过增强基的判别力能够解决字典对训练样本数量低的类别表达能力弱的问题,提高字典表示的拓展能力。比较于多个字典的训练,单字典不会引入过多的冗余,在图片重构中也不会产生过多的噪声。

参考文献

- [1] Freeman W, Pasztor E, Carmichael O. Learning low-level vision [J]. IJCV, 2000, 40(1): 25-47
- [2] Sun J, Zheng N, Tao H, et al. Image hallucination with primal sketch priors [C]//CVPR. 2005: 36-43
- [3] Yang J, Wright J, Huang T, et al. Image super-resolution as sparse representation of raw image patches [C]//CVPR. 2008: 1-8
- [4] Yang M, Zhang L, Feng X, et al. Fisher discrimination dictionary learning for sparse representation [C]//IEEE ICCV. 2011: 543-550
- [5] Zhou N, Shen Y, Peng J, et al. Learning Inter-related Visual Dictionary for Object Recognition [C]//CVPR. 2012: 3490-3497
- [6] Chang H, Yeung D-Y, Xiong Y. Super-resolution through neighbor embedding [C]//CVPR. 2004