软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2015,26(11):2930−2938 [doi: 10.13328/j.cnki.jos.004898]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

基于 Pivots 选择的有效图像块描述子*

谢博鋆[1,2], 朱 杰[1,3], 于 剑[1]

[1](交通数据分析与挖掘北京市重点实验室(北京交通大学),北京 100044)

[2](河北省机器学习与计算智能重点实验室(河北大学),河北 保定 071002)

[3](中央司法警官学院 信息管理系,河北 保定 071000)

通讯作者: 于剑, E-mail: jianyu@bjtu.edu.cn

摘 要: 设计图像块特征表示是计算机视觉领域内的基本研究内容,优秀的图像块特征表示能够有效地提高图像分类、对象识别等相关算法的性能.SIFT(scale-invariant feature transform)和 HOG(histogram of oriented gradient)是人为设计图像块特征表示的优秀代表,然而,人为设计图像块特征间的差异往往不能足够理想地反映图像块间的相似性.核描述子(kernel descriptor,简称 KD)方法提供了一种新的方式生成图像块特征,在图像块间匹配核函数基础上,应用核主成分分析(kernel principal component analysis,简称 KPCA)方法进行特征表示,且在图像分类应用上获得不错的性能.但是,该方法需要利用所有联合基向量去生成核描述子特征,导致算法时间复杂度较高.为了解决这个问题,提出了一种算法生成图像块特征表示,称为有效图像块描述子(efficient patch-level descriptor,简称 EPLd).算法建立在不完整 Cholesky 分解基础上,自动选择少量的标志性图像块以提高算法效率,且利用 MMD(maximum mean discrepancy)距离计算图像间的相似性.实验结果表明,该算法在图像/场景分类应用中获得了优秀的性能.

关键词: 标志性图像块;不完整 Cholesky 分解;核描述子;有效图像块描述子;MMD 距离

中图法分类号: TP391

中文引用格式: 谢博鋆,朱杰,于剑.基于 Pivots 选择的有效图像块描述子.软件学报,2015,26(11):2930−2938. http://www.jos.org.cn/1000-9825/4898.htm

英文引用格式: Xie BJ, Zhu J, Yu J. Efficient patch-level descriptor for image categorization via patch pivots selection. Ruan Jian Xue Bao/Journal of Software, 2015,26(11):2930−2938 (in Chinese). http://www.jos.org.cn/1000-9825/4898.htm

Efficient Patch-Level Descriptor for Image Categorization via Patch Pivots Selection

XIE Bo-Jun[1,2], ZHU Jie[1,3], YU Jian[1]

[1](Beijing Key Laboratory of Traffic Data Analysis and Mining (Beijing Jiaotong University), Beijing 100044, China)

[2](Key Laboratory of Machine Learning and Computational Intelligence (Hebei University), Baoding 071002, China)

[3](Department of Information Management, Central Institute for Correctional Police, Baoding 071000, China)

Abstract: Designing patch-level features is essential for achieving good performance in computer vision tasks, such as image classification and object recognition. SIFT (scale-invariant feature transform) and HOG (histogram of oriented gradient) are the typical schemes among many pre-defined patch-level descriptors, but the difference between artificial patch-level features is not good enough for reflecting the similarities of images. Kernel descriptor (KD) method offers a new way to generate features from match kernel defined over image patch pairs using KPCA (kernel principal component analysis) and yields impressive results. However, all joint basis vectors are involved in the kernel descriptor computation, which is both expensive and not necessary. To address this problem, this paper presents an efficient patch-level descriptor (EPLd) which is built upon incomplete Cholesky decomposition. EPLd automatically selects a small number of image patches pivots to achieve better computational efficiency. Based on EPLd, MMD (maximum mean discrepancy) distance

* 基金项目: 国家自然科学基金(61370129, 61375062, 61300072); 高等学校博士学科点专项科研基金(20120009110006); 河北省教育厅青年基金(QN2015099); 河北省社会科学基金(HB15TQ013)

收稿时间: 2015-05-28; 修改时间: 2015-07-14; 定稿时间: 2015-08-26

is used for computing similarities between images. In experiments, the EPLd approach achieves competitive results on several image/scene classification datasets compared with state-of-the-art methods.

Key words: patch pivot; incomplete Cholesky decomposition; kernel descriptor; efficient patch-level descriptor; MMD distance

1 图像特征表示方法概述

设计图像的特征表示是计算机视觉中一项非常基本的研究内容,图像的分类、检索、标注等工作都是以提取图像特征为初始步骤,好的特征表示可以在相关图像分析中取得更佳的效果.因此,图像特征的设计与构造,直接影响算法的性能.而如何定义一个好的图像特征却是非常困难的:一方面,设计的图像特征对于同一类别下图像之间的变化(比如尺度、光照变化、对象位置变化等)要有足够的鲁棒性;另一方面,设计的图像特征要具备足够的判别性来处理不同类别间图像的变化.

近年来,研究者提出了大量的底层特征用于各种图像分析任务,其中最具有代表性的是基于梯度朝向直方图的 SIFT(scale-invariant feature transform)[1]和 HOG(histogram of oriented gradient)[2].尽管这类特征取得了一定意义的成功,但研究者发现,这类单一的底层特征并不足以在某些应用上达到更好的效果,因此提出了一类中间层的图像特征表示方法.其中,BoW(bag of words)[3]是这类图像特征表示方法的典型代表,该方法在场景分类中获得了较好的性能.BoW 算法生成图像特征表示分为 3 个过程:图像底层特征的获取、学习过完备字典和计算图像的码字直方图表示.然而,BoW 方式并没有考虑特征向量在图像空间上的位置关系,使得其特征描述能力并没有达到最大化.为了弥补这一缺陷,空间金字塔匹配(spatial pyramid matching,简称 SPM)[4]方法通过在一幅图像的不同层次上计算码字直方图,形成了一个 BoW 多层特征,将 BoW 模型与图像空间进行合理融合.然而,由于 SPM 方法利用直方图交核函数来度量两幅图像间的相似度,导致无法产生低维度的图像特征表示,而且需要完整计算训练集图像间相似度的 Gram 矩阵,因此,其算法复杂度为 $O(n^2)$(其中,n 为训练集中图像的个数).为了解决这一问题,有效匹配核算法(efficient match kernel,简称 EMK)[5]在码字间相似性的基础上构造了一个低维特征映射空间,整个图像的特征可以表示为码字映射在这个低维特征空间后的平均,且可以采用线性 SVM 方法训练分类器,在图像分类应用中获得了非常不错的效果.然而,有效匹配核算法仍然依赖于人为定义的图像局部特征(如 SIFT 或 HOG),只不过是通过计算有限维空间的局部线性特征表示来推出整体图像的线性特征.

Bo 等人扩展了有效匹配核算法并提出了核描述子(kernel descriptor,简称 KD)[6]方法.这种方法只需定义任意两个局部图像块之间的相似性,且该相似性函数满足核函数定义.由于每个核函数都隐性定义了一个映射,它将图像块映射为再生核希尔伯特空间(reproducing kernel Hilbert space,简称 RKHS)中一个非常高维的向量,这样,核函数可以表示为 RKHS 中两个高维向量的内积,通过核主成分分析(kernel principal component analysis,简称 KPCA)[7]算法,可以由核函数推出图像块特征的有限维线性表示.这种低维空间中的表示就称为核描述子,并且采用 EMK 算法将其推广到整个图像的特征表示.

尽管核描述子方法的设计思想较为新颖,但仍然存在计算复杂度过高这一缺陷,限制了其在大规模图像数据库上的应用.事实上,在 KPCA 方法的离线阶段,所有联合基向量对之间的相似性都需要计算,这是非常耗时的.更重要的是:在线阶段计算一个新图像块的特征映射时,该图像块与所有联合基向量之间的相似性也是需要计算的,而这实际上是不需要的.Xie 等人[8]通过使用不完整 Cholesky 分解替代 KPCA 算法,成功地解决了这个问题,并且通过迭代,应用不完整 Cholesky 分解算法表示整个图像特征[9].但文献[8,9]中,通过不完整 Cholesky 分解得到的标志联合基向量并没有对应实际的图像块,因此,其产生的特征判别能力并没有最大化地得到利用.Wang 等人提出了有监督的核描述子方法[10],该方法利用训练集中的图像类标来辅助设计底层图像块特征.尽管他们利用该特征取得了不错的分类效果,但这个算法运行过程中需要大量有类标的图像,并且对象优化函数求解过程复杂,时间复杂度过高.

除了上述生成图像底层特征表示的方法以外,另外一类构成图像特征的方法基于深度学习理论.2006 年,Hinton 等人[11,12]提出了用于深度信任网络(deep belief network,简称 DBN)的无监督学习算法,DBN 的多层结构,使得它能够学习得到层次化的特征表示,实现自动特征抽象,文献[12]将 DBN 模型成功用于手写数字识别应用

上.Bengio 等人在文献[13]中提出了基于自编码器(auto-encoder)[14]的深度学习网络,在手写数字识别图像数据库上得到了类似的实验结果.另外,文献[15−17]提出了一系列基于稀疏编码[18]的深层学习网络,在图像应用中取得了一定的成功.LeCun 等人[19,20]用误差梯度设计并训练卷积神经网络(convolutional neural network,简称CNN),其在图像分类,特别是手写体字符识别应用中得到优越的性能.在此基础上,Krizhevsky 等人[21]将 CNN 模型应用到分类大规模 ImageNet 图像数据库[22],更加充分地显示了深度学习模型的表达能力.尽管在深度学习模型下获得的图像特征有很强的判别表示能力,但其要求计算机硬件条件较高,单机环境下很难实现.除此之外,更加详细地介绍图像特征描述子领域的综述可以参考文献[23].

本文在大数据时代背景下,为了能够快速得到图像块的线性特征表示,提出了有效图像块描述子(efficient patch-level descriptor,简称 EPLd)方法.该方法在不完整 Cholesky 分解基础上,可以自动地进行图像块筛选,对于求解新图像块的线性特征表示,只需计算它和一小部分基图像块的相似性就足够了.有了图像块的特征表示之后,一幅图像就对应着一个图像块特征的集合,该集合可以看作是特征空间中基于某个分布的样本集,这样,两幅图像之间的差异可以看作两个分布的距离.本文采用基于高维概率分布的 MMD 距离[24]进行估算,进而计算两幅图像间的相似性.

本文首先介绍核描述子方法,然后给出有效图像块描述子算法的具体实现过程以及如何利用MMD距离计算两幅图像的相似性,并在几个著名的图像分类数据库上进行实验,最后给出工作的结论和展望.

2 核描述子方法简介

核描述子方法是对图像像素点属性(梯度/形状/颜色+位置)基础上生成的联合基向量应用KPCA方法,从而计算新图像块的有限维特征表示.为了方便叙述,本文采用像素点的梯度属性来介绍核描述子方法.

首先,考虑两个图像块间的核函数采用梯度匹配核函数,则两个图像块之间的相似度函数可由公式(1)给出:

$$K_{grad}(P,Q)=\sum_{z\in P}\sum_{z'\in Q}\tilde{m}_z\tilde{m}_{z'}k_o(\tilde{\theta}_z,\tilde{\theta}_{z'})k_p(z,z') \tag{1}$$

其中,P 和 Q 分别代表两个不同的图像块,z 表示像素点在图像块中的相对位置(标准化为[0,1]),$\tilde{m}_z$ 和 $\tilde{\theta}_z$ (梯度向量 $\tilde{\theta}_z=[\sin(\theta_z)\ \cos(\theta_z)]$,$\theta_z$为对应像素点的梯度角度)分别表示像素点位置 z 的梯度幅值和朝向,k_o和 k_p 各自为梯度朝向和像素点位置的高斯核函数.

之后,图像块 P 的低维特征映射可以通过公式(2)计算得到:

$$\bar{F}^t_{grad}(P)=\sum_{i=1}^{d_o}\sum_{j=1}^{d_p}\alpha^t_{ij}\left\{\sum_{z\in P}\tilde{m}(z)k_o(\tilde{\theta}(z),x_i)k_p(z,y_j)\right\} \tag{2}$$

其中,$\{x_i\}_{i=1}^{d_o}$ 和 $\{y_j\}_{j=1}^{d_p}$ 分别代表梯度朝向和像素点位置的基向量,这些基向量分别在各自合理取值范围内通过均匀采样得到;d_o和 d_p是对应基向量的采样个数;α^t_{ij} 是对联合基向量 $\{\phi_o(x_1)\otimes\phi_p(y_1),...,\phi_o(x_{d_o})\otimes\phi_p(y_{d_p})\}$ (⊗代表Kronecker 张量积)应用 KPCA 算法后计算得到的第 t 维的映射系数.

通过公式(2)可以看到,核描述子方法的主要缺陷有以下3点:(1) 算法计算复杂度高,因为需要对 $d_o\times d_p$ 维的联合基向量形成的Gram矩阵计算特征值分解,如果联合基向量的维度过高或者个数过多,KPCA算法甚至无法实施;(2) 对联合基向量进行 KPCA 获得的 α^t_{ij} 并不是稀疏的,这也就意味着在计算新图像块的特征表示时,需要和所有的联合基向量进行在线计算,所以算法需要存储全部的联合基向量;(3) 算法无法进行特征选择,即,并不知道联合基向量中哪些样本最具代表性.

3 有效图像块描述子算法

针对核描述子方法的 3 点不足之处,文献[8]解决了其主要缺陷的第一、第二两点,但是文献[8]在本质上仍然使用联合基向量,所以没有明确地进行特征选择,即,找出哪些图像块是最具代表性的,使得其特征表示能力并没有达到最大化.为了更加完善地解决核描述子方法的缺陷,本文提出了一种新的图像块特征表示方法,称为有效图像块描述子.该方法基于对图像块相似度矩阵执行不完整 Cholesky 分解[25,26].

总体上来说,有效图像块描述子算法由两部分构成:

1) 首先从训练图像集中均匀抽取足够的图像块,然后在这些图像块形成的 Gram 矩阵上执行不完整 Cholesky 分解算法.如果设定 N 代表图像块的个数,M 代表分解后矩阵的秩,通常情况下,$M<<N$.这样做的好处有两点:首先,在分解过程中只需要按需计算 $O(MN)$个 Gram 矩阵元素的值;其次,对 Gram 矩阵执行 Cholesky 分解的时间复杂度为 $O(M^2N)$,远远低于 KPCA 算法的 $O(N^3)$.

2) 经过第 1 步分解步骤之后,选择出了 M 个最具代表性的基图像块,新图像块的特征表示仅仅通过 $O(M)$次计算就可以得到.算法的具体步骤将在以下部分详细介绍.

3.1 Gram矩阵的低秩近似

半正定的 Gram 矩阵 K 可以分解为 GG^T,所以不完整 Cholesky 分解的目标就是找到一个矩阵 $\tilde{G}$,其大小为 $N\times M$,使得 $\tilde{G}\tilde{G}^T$ 在 M 足够小的情况下近似 K.

在执行不完整 Cholesky 分解算法的过程中,选择出 M 个最具代表性的基图像块,利用所有图像块和这 M 个基图像块之间的相似性,可以近似恢复 Gram 矩阵 K.这里,M 的值是可以通过算法在线确定的,由算法中提前给定的近似精度参数ε来控制.

关于不完整 Cholesky 分解的详细执行过程可以参考文献[26],其中,作为输入参数的 Gram 矩阵 K 实际上是按需计算的,即,算法执行过程中需要用到哪两个训练图像块间的相似度,就按照公式(1)计算得到.算法执行后,就得到了一些具有代表性的基图像块,用向量 P 保存基图像块的索引序号,同时得到了矩阵 $\tilde{G}$, 使得 $\tilde{G}\tilde{G}^T\approx K$.

3.2 构造图像块特征映射算法

一旦获得了 $N\times M$ 的矩阵 $\tilde{G}$, 新图像块的特征(有效图像块描述子)就可以由 $\tilde{G}$ 构造.其中,新图像块特征维度大小由 M 确定,每一维度 i 的值可由新图像块与 $P(i)$所指示的基图像块间相似性 $K(newpatch,P(i))$恢复得到,因为 $K(newpatch,P(i))=G_{\text{new}}(i)\tilde{G}(P(i),i)+\sum_{j=1}^{i-1}G_{new}(j)\tilde{G}(P(i),j)$, 其中,$G_{\text{new}}$ 表示新图像块的特征向量.详细的有效图像块描述子构造过程参见算法 1.

算法 1. 有效图像块描述子算法(生成一个新图像块的特征表示).

输入: $\tilde{G}$:针对图像块间相似性矩阵执行不完整 Cholesky 分解得到(参见第 3.1 节);

P:执行不完整 Cholesky 分解得到基图像块的索引序号且按重要性排序;

M:执行不完整 Cholesky 分解后选择出重要基图像块的个数;

S:M 维向量,其每一维就是新图像块与选择出的一个重要基图像块之间的相似性值(由公式(1)计算得到).

for i=1:M //计算新图像块第 i 维的特征值:

$$G_{new}(i)=[S(i)-\sum_{j=1}^{i-1}G_{new}(j)\tilde{G}(P(i),j)]/\tilde{G}(P(i),i).$$

输出:G_{new}:新图像块的 M 维特征表示.

通过算法 1 可以看到:选择出的 M 个最具代表性的基图像块可以看成是一系列局部图像块的非线性滤波器,将每个新图像块和这些基图像块进行相似性度量的过程,也可看成是对这个新图像块进行特征提取的过程.

另外,针对图像块相似度矩阵执行不完整 Cholesky 分解往往可以保证获得精度非常高的低秩近似,且分解过程中只与某些训练样本(图像块)有关.也就是说,利用这些训练样本就可以很好地近似恢复相似度矩阵,所以训练集中的图像块具有不同程度的重要性.因此,我们称重要性最高的前 M 个图像块为“最具代表性”的基图像块.为了更加形象地展示这些重要的基图像块,我们在 Scene-15 图像库上提取了最重要的前 16 个基图像块,如图 1 所示(每个图像块由其像素点的梯度幅值来表示).可以看到,每个图像块都包含了丰富的边缘和纹理信息.

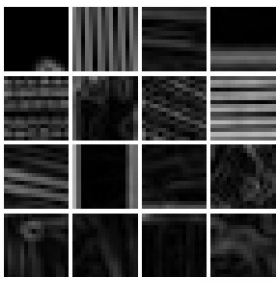

Fig.1 Top-16 pivot patches automatically selected from the Scene-15 data set
图 1 由 Scene-15 图像库上自动选择出的前 16 个基图像块

本文提出的有效图像块描述子算法不只继承了文献[8]的有效性,而且很好地解决了核描述子算法中的第 3 点缺陷,最大限度地发挥了图像块特征的判别能力.

4 利用 MMD 距离计算图像间的相似性

基于算法 1,每一个图像块都可以用有效图像块描述子来表示.一幅图像通过稠密采样确定很多关键点,每一个关键点都对应着一个局部的图像块,因此,一幅图像就对应着一个局部特征的集合.假定图像 I_1 包含 m 个图像块,则其特征集合可以表示为 $F_p(patch_{p1},patch_{p2},\ldots,patch_{pm})$,图像 I_2 包含 n 个图像块,其特征集合表示为 $F_q(patch_{q1},patch_{q2},\ldots,patch_{qn})$.$F_p$ 可以看作特征空间中来自分布 p 的一个样本集,同样,F_q 也可以看作是来自分布 q 的样本集.这样,图像 I_1 与 I_2 之间的差异性就可以由 p 和 q 两个分布的距离表示.当然,这两个概率分布之间的距离只能通过这两个样本集进行估算.为此,本文采用基于高维概率分布的 Maximum Mean Discrepancy(MMD)距离[24]进行估算.MMD 距离可以看作是将两个概率分布,通过非线性核函数映射到再生核希尔伯特空间(RKHS)后均值的距离.对于上述分布 p 和 q 的 MMD 距离估计可由公式(3)计算:

$$MMD(F_p,F_q)=\left[\frac{1}{m^2}\sum_{i,j=1}^{m}k(patch_{pi},patch_{pj})-\frac{2}{mn}\sum_{i,j=1}^{m,n}k(patch_{pi},patch_{qj})+\frac{1}{n^2}\sum_{i,j=1}^{n}k(patch_{qi},patch_{qj})\right]^{\frac{1}{2}} \tag{3}$$

其中,核函数 k 采用非线性的高斯核,参数σ通过交叉验证方式进行验证.

单纯地利用公式(3),并没有考虑局部特征在整幅图像上的空间分布信息.为了解决这个问题,本文首先采用空间金字塔方法将整幅图像进行逐层划分;然后,在两幅图像每个层次对应的小图像上计算它们之间的 MMD 距离;最终,将所有层次的 MMD 距离按照其对应层次的权重进行汇总求和,然后度量两幅图像 I_1 与 I_2 之间的差异性.

5 实 验

本文使用像素点的梯度、形状和颜色属性分别构造基于梯度的有效图像块描述子(EPLd-G)、基于形状的有效图像块描述子(EPLd-S)和基于颜色的有效图像块描述子(EPLd-C).为了测试有效图像块描述子算法的性能,分别在 3 个著名的图像分类数据库(Scene-15[3,4,27],Caltech-101[28]和 UIUC-8[29])上做了实验.

在接下来的实验中,计算 3 个不同类型的有效图像块描述子都是首先将图像按照固定比率缩放到不超过 300×300 像素点;特别地,在计算 EPLd-G 和 EPLd-S 时,将缩放后的图像中的像素点的灰度值标准化为[0,1]范围.图像块通过每隔 8 个像素点的稠密采样方式从训练集图像中进行抽取,大小为 16×16 像素点.EPLd-All 是将 EPLd-G,EPLd-S 和 EPLd-C 这 3 个描述子串接起来形成的.训练线性 SVM 分类器使用 LIBLINEAR[30],其中,图像间的相似性利用 MMD 距离来定义.在计算 MMD 时,将图像按照 1×1,2×2 和 3×3 分为 3 个层次来汇总求和,尺度参数σ在不同的数据库上利用交叉验证方法确定.所有的实验均重复 10 次,每次的训练集和测试集都随机抽取确定,将 10 次分类准确率的平均值和方差记录下来.实验中的其他参数从公平比较的角度考虑,与文献[6,8]

设置相同.

5.1 Scene-15

Scene-15 场景数据库包含 4 485 张图片,这些图片分属 15 个类别,有室内场景和室外场景,每一个类别包含 200 张~400 张图片不等.按照惯例,从每个类别中随机抽取 100 张图片作为训练,剩余图片作为测试.在算法中设置 Pivots 的个数为 200,即,利用不完整 Cholesky 分解选出 200 个最具代表性的基图像块来构造维度为 200 的有效图像块描述子.实验结果列在表 1 中(其中,KD 代表核描述子方法[6],EKD 代表有效核描述子方法[8],EPLd 代表本文提出的有效图像块描述子方法),EPLd 方法获得在这个数据库上的最佳分类准确率(87.0%).另外,EPLd 方法在所有 4 种不同情况(梯度、形状、颜色和上述 3 种属性的汇总)下的性能均超过了文献[6,8].

Table 1 Comparison of the classification accuracy on Scene-15

表 1 Scene-15 上的分类准确率对比

算法	梯度	形状	颜色	综合
KD[6]	81.6±0.6	79.8±0.5	38.5±0.4	81.9±0.6
EKD[8]	82.7±0.6	81.0±0.5	49.1±1.2	86.3±0.4
EPLd	**82.9±0.5**	**81.3±0.6**	**50.0±0.6**	**87.0±0.7**

在实验中,除了测试分类准确率来体现 EPLd 的判别能力,还通过不同维度下测试分类准确率来体现 EPLd 的有效性.我们发现,在特征维度只有 50 维的情况下也获得了接近最优分类准确率的性能,这充分体现出 EPLd 算法的有效性和健壮性.事实上,通过表 2 可以看到:特征维度从 50 维增加到 300 维,分类准确率并没有得到明显的提升.造成这一现象的原因是,不完整 Cholesky 分解容易获得高质量的低秩近似.表 2 中的数据表明:即使是 50 维的低秩近似也足以体现 Gram 矩阵中的关键信息,而这些关键信息直接决定了分类的性能.在后面的实验中,从算法效率的角度考虑都使用了 100 维的特征表示.

Table 2 EPLd classification accuracy on Scene-15 by changing the number of pivots from 50 to 300

表 2 EPLd 算法在 Scene-15 图像库上的分类准确率,其中,EPLd 的特征维度从 50 变化到 300

特征维度	EPLd-G	EPLd-S	EPLd-C
50 维	81.8±0.8	79.8±0.5	49.1±0.9
100 维	82.1±0.6	80.9±0.6	49.8±1.2
200 维	82.9±0.5	81.3±0.6	50.0±0.6
300 维	82.8±0.7	81.2±0.7	50.0±0.5

5.2 Caltech-101

Caltech-101 图像数据库包含 9 144 张图片.这 9 144 张图片隶属于 101 个对象类别外加一个背景类别,每个类别中的图片在 31 张~800 张不等.表 3 中,将 EPLd 与其他有代表性的描述子算法进行了对比.同样根据惯例,每个类别随机挑出 30 张图片进行训练,从剩余图片中挑选不超过 50 张进行测试.可以看到:EPLd 算法达到了最佳的分类准确率(77.1%),甚至在仅仅使用梯度属性的情况下(EPLd-G)也达到了非常不错的分类效果(73.7%).

Table 3 Comparing the classification accuracy of seven algorithms on Caltech-101

表 3 7 种不同的算法在 Caltech-101 上的分类准确率对比

算法	分类准确率
Shabou 等人[31]	73.23±0.81
NBNN[32]	73.0
LLC[33]	73.4±0.5
Jia 等人[34]	75.3±0.7
KD-All[6]	74.5±0.8
EKD-All[8]	76.9±0.5
EPLd-G	73.7±0.6
EPLd-All	**77.1±0.7**

5.3 UIUC-8

UIUC-8 图像数据库包含 1 579 张图片,这 1 579 张图片隶属于 8 个运动类别,每个类别下包含图片 137 张~250 张不等.按照惯例,随机从每个类别中抽取 70 张图片进行训练,从剩余图片中挑选 60 张进行测试.分类准确率结果列于表 4 中.通过表 4 可以看到,EPLd-All 非常接近最佳分类准确率(87.2% vs. 87.23%).

Table 4 Comparing the classification accuracy of four algorithms on UIUC-8

表 4 4 种不同的算法在 UIUC-8 上的分类准确率对比

算法	分类准确率
Liu 等人[35]	84.56±1.5
Shabou 等人[31]	**87.23±1.14**
EKD-All[8]	87.1±1.4
EPLd-All	87.2±1.3

在实验部分的最后,本文对比了构造 3 种不同描述子(EPLd vs. KD vs. EKD)的计算效率.其中,最耗时的是形状特征,一幅标准图像(最大 300×300 分辨率,图像块大小为 16×16 像素点,图像块间隔 8 个像素点)上的 EPLd-S 与 EKD-S 描述子在 Matlab 环境下计算需要耗时 2s,而 KD-S 需要耗时 2.5s.对于梯度特征,EPLd-G 与 EKD-G 描述子耗时 0.9s,KD-G 耗时 1s.以上对比结果列在表 5 中.

Table 5 Comparing the computational efficiency of three descriptors

表 5 3 种不同描述子的计算效率对比

算法	形状(s)	梯度(s)
KD[6]	2.5	1
EKD[8]	2	0.9
EPLd	2	0.9

表 5 中的对比结果是在生成 100 维特征情况下得到的,如果提高特征的维度,EPLd 与 EKD 的计算效率提升相对于 KD 会表现得更加明显.另外一点需要指出的是:EPLd 与 EKD 的计算耗时虽然基本相同,但 EPLd 描述子的特征判别能力相对于 EKD 描述子要强很多,这一点通过在 3 个图像数据库上的实验对比结果可以得到印证.所以,综合考虑,EPLd 描述子无论在计算效率还是在判别能力上都要优于 EKD 和 KD 描述子.

6 结束语

本文提出了有效图像块描述子算法.该算法的主要思想是:通过不完整 Cholesky 分解对图像块的相似性进行逆向学习,选出具有代表性的基图像块.这些基图像块可以看成是一系列的非线性滤波器,将每个新图像块和这些基图像块进行相似性度量的过程,就是对这个新图像块进行特征提取的过程.另外,本文还设计了更为优秀的基于局部特征的整体图像相似性度量,也就是利用 MMD 距离计算两幅图像之间的相似性,该相似性度量方式不仅能够反映局部图像特征之间的相似性,而且能够有效地利用特征的空间分布信息,从而最大限度地提高整体图像相似性度量的精确度和敏感度.实验结果显示:EPLd 方法相对于 KD 方法和其他一些代表性的方法,在 3 个著名图像分类数据库上都获得了非常不错的性能.

References:

[1] Lowe DG. Distinctive image features from scale-invariant keypoints. Int'l Journal of Computer Vision, 2004,60(2):91−110. [doi: 10.1023/B:VISI.0000029664.99615.94]

[2] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proc. of the CVPR. 2005. 886−893. [doi: 10.1109/CVPR.2005.177]

[3] Li FF, Perona P. A bayesian hierarchical model for learning natural scene categories. In: Proc. of the CVPR. 2005. 524−531. [doi: 10.1109/CVPR.2005.16]

[4] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proc. of the CVPR. 2006. 2169−2178. [doi: 10.1109/CVPR.2006.68]

[5] Bo LF, Sminchisescu C. Efficient match kernel between sets of features for visual recognition. In: Proc. of the NIPS. 2009. 135−143.

[6] Bo LF, Ren XF, Fox D. Kernel descriptors for visual recognition. In: Proc. of the NIPS. 2010. 244−252.

[7] Schölkopf B, Smola A, Müller KR. Nonlinear component analysis as a kernel eigenvalue problem. Neural computation, 1998,10(5): 1299−1319. [doi: 10.1162/089976698300017467]

[8] Xie BJ, Liu Y, Zhang H, Yu J. Efficient kernel descriptor for image categorization via pivots selection. In: Proc. of the 20th IEEE Int'l Conf. on Image Processing (ICIP). 2013. 3479−3483. [doi|: 10.1109/ICIP.2013.6738718]

[9] Xie BJ, Liu Y, Zhang H, Yu J. Efficient image representation for object recognition via pivots selection. Frontiers of Computer Science, 2015,9(3):383−391. [doi: 10.1007/s11704-015-4182-7]

[10] Wang P, Wang JD, Zeng G, Xu WW, Zha HB, Li SP. Supervised kernel descriptors for visual recognition. In: Proc. of the CVPR. 2013. 2858−2865. [doi: 10.1109/CVPR.2013.368]

[11] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science, 2006,313(5786):504−507. [doi: 10.1126/science.1127647]

[12] Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. Neural Computation, 2006,18(7):1527−1554. [d: 10.1162/neco.2006.18.7.1527]

[13] Bengio Y, Lamblin P, Popovici D, Larochelle H. Greedy layer-wise training of deep networks. Advances in Neural Information Processing Systems, 2007,19:153−160.

[14] Bourlard H, Kamp Y. Auto-Association by multilayer perceptrons and singular value decomposition. Biological Cybernetics, 1988, 59(4-5):291−294. [doi: 10.1007/BF00332918]

[15] Coates A, Ng AY, Lee H. An analysis of single-layer networks in unsupervised feature learning. In: Proc. of the Int'l Conf. on Artificial Intelligence and Statistics. 2011. 215−223.

[16] Lee H, Ekanadham C, Ng AY. Sparse deep belief net model for visual area V2. In: Proc. of the Advances in Neural Information Processing Systems. 2008. 873−880.

[17] Ngiam J, Koh PW, Chen Z, Bhaskar S, Ng AY. Sparse filtering. In: Proc. of the Advances in Neural Information Processing Systems. 2011. 1125−1133.

[18] Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, 1996,381(6583):607−609. [doi: 10.1038/381607a0]

[19] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-Based learning applied to document recognition. Proc. of the IEEE, 1998, 86(11):2278−2324. [doi: 10.1109/5.726791]

[20] LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, Jackel LD. Backpropagation applied to handwritten zip code recognition. Neural Computation, 1989,1(4):541−551. [doi: 10.1162/neco.1989.1.4.541]

[21] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Proc. of the Advances in Neural Information Processing Systems. 2012. 1097−1105.

[22] Deng J, Dong W, Socher R, Li LJ, Li K, Li FF. ImageNet: A large-scale hierarchical image database. In: Proc. of the CVPR. 2009. 248−255. [doi: 10.1109/CVPR.2009.5206848]

[23] Huang KQ, Ren WQ, Tan TQ. A review on image object classification and detection. Chinese Journal of Computers, 2014,37(6): 1225−1240 (in Chinese with English abstract). [doi: 10.3724/SP.J.1016.2014.01225]

[24] Gretton A, Borgwardt KM, Rasch M, Schölkopf B, Smola AJ. A kernel method for the two-sample-problem. In: Proc. of the Advances in Neural Information Processing Systems. 2006. 513−520.

[25] Fine S, Scheinberg K. Efficient svm training using low-rank kernel representation. Journal of Machine Learning Research, 2001,2: 243−264.

[26] Bach FR, Jordan MI. Kernel independent component analysis. Journal of Machine Learning Research, 2002,3:1−48.

[27] Oliva A, Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope. Int'l Journal of Computer Vision, 2001,42(3):145−175. [doi: 10.1023/A:1011139631724]

[28] Li FF, Fergus R, Perona P. Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. Computer Vision and Image Understanding, 2007,106(1):59−70. [doi: 10.1016/j.cviu.2005.09.012]

[29] Li LJ, Li FF. What, where and who? Classifying events by scene and object recognition. In: Proc. of the ICCV. 2007. 1−8. [doi: 10.1109/ICCV.2007.4408872]

[30] Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ. Liblinear: A library for large linear classification. Journal of Machine Learning Research, 2008,9:1871−1874.

[31] Shabou A, Borgne HL. Locality-Constrained and spatially regularized coding for scene categorization. In: Proc. of the CVPR. 2012. 3618−3625. [doi: 10.1109/CVPR.2012.6248107]

[32] Boiman O, Shechtman E, Irani M. In defense of nearest-neighbor based image classification. In: Proc. of the CVPR. 2008. 1−8. [doi: 10.1109/CVPR.2008.4587598]

[33] Wang JJ, Yang JC, Yu K, Lv FJ, Huang T, Gong YH. Locality-Constrained linear coding for image classification. In: Proc. of the CVPR. 2010. 3360−3367. [doi: 10.1109/CVPR.2010.5540018]

[34] Jia YQ, Huang C, Darrell T. Beyond spatial pyramids: Receptive field learning for pooled image features. In: Proc. of the CVPR. 2012. 3370−3377. [doi: 10.1109/CVPR.2012.6248076]

[35] Liu LQ, Wang L, Liu XW. In defense of soft-assignment coding. In: Proc. of the ICCV. 2011. 2486−2493. [doi: 10.1109/ICCV.2011.6126534]

附中文参考文献:

[23] 黄凯奇,任伟强,谭铁牛.图像物体分类与检测算法综述.计算机学报,2014,37(6):1225−1240. [doi: 10.3724/SP.J.1016.2014.01225]

谢博鋆(1981－),男,河北保定人,博士生,主要研究领域为机器学习,计算机视觉.

朱杰(1982－),男,博士生,主要研究领域为机器学习,对象识别,图像分类.

于剑(1969－),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为机器学习,聚类分析,图像处理.