

DOI:10.11992/tis.201603028  
网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.TP.20160513.0926.030.html>

# 基于卷积神经网络和哈希编码的图像检索方法

龚震霆<sup>1,2</sup>, 陈光喜<sup>1,2</sup>, 任夏荔<sup>1,2</sup>, 曹建收<sup>1,2</sup>

(1. 桂林电子科技大学 计算机与信息安全学院, 广西 桂林 541004; 2. 广西高校图像图形智能处理重点实验室, 广西 桂林 541004)

**摘 要:**在图像检索中,传统的基于人工特征的检索方法并不能取得很好的效果。为此提出一种结合卷积神经网络和以前最好水准的哈希编码策略的图像检索方法。鉴于近几年卷积神经网络在大量的计算机视觉任务上的巨大进步,该方法首先使用在 ILSVRC 数据集上预训练过的 VGGNet-D 网络模型对实验图像数据集提取卷积特征来得到图像的深层表示,再采用以前最好水准的哈希策略将这些深层表示进行编码,从而得到图像的二进制码,最后再进行快速图像检索。在两个常用的数据集 Caltech101 和 Caltech256 上的实验结果表明,本文方法的 5 个策略相比于以前最好水准的相应的图像检索策略在“精度-召回率”和“平均正确率值-编码位数”两个指标上能获得更优异的性能,证明了本文方法在图像检索上的有效性。

**关键词:**图像检索;人工特征;卷积神经网络;卷积特征;哈希编码

**中图分类号:**TP391   **文献标志码:**A   **文章编号:**1673-4785(2016)03-0391-10

中文引用格式:龚震霆,陈光喜,任夏荔,等.基于卷积神经网络和哈希编码的图像检索方法[J]. 智能系统学报, 2016, 11(3): 391-400.

英文引用格式:GONG Zhenting, CHEN Guangxi, REN Xiali, et al. An image retrieval method based on a convolutional neural network and hash coding[J]. CAAI transactions on intelligent Systems, 2016, 11(3): 391-400.

## An image retrieval method based on a convolutional neural network and hash coding

GONG Zhenting<sup>1,2</sup>, CHEN Guangxi<sup>1,2</sup>, REN Xiali<sup>1,2</sup>, CAO Jianshou<sup>1,2</sup>

(1. School of Computer and Information Security, Guilin University of Electronic Technology, Guilin 541004, China; 2. Guangxi Colleges and Universities Key Laboratory of Intelligent Processing of Computer Images and Graphics, Guilin 541004, China)

**Abstract:** For image retrieval, traditional retrieval methods based on artificial features are not effective enough. Hence, we propose an image retrieval method, which combines a convolutional neural network and previous state-of-the-art hash coding strategies. In view of the great progress that convolutional neural networks have made in a large number of computer vision tasks in recent years, this method first uses the model "VGGNet-D" pre-trained on the ILSVRC's dataset to extract the convolutional features from experimental image datasets to get the deep representations of images, then adopts previous state-of-the-art hash coding strategies to encode the deep representations to obtain the binary codes, and, finally, performs a quick image retrieval. The experimental results on the commonly used Caltech101 and Caltech256 datasets show that this method's five strategies, compared with the previous state-of-the-art image retrieval strategies, can obtain better, indeed excellent, performance in both the "Precision-Recall" and "mean Average Precision-Number of bits" metrics, proving the effectiveness of the proposed method in image retrieval.

**Keywords:** image retrieval; artificial features; convolutional neural network; convolutional features; hash coding

收稿日期:2016-03-17. 网络出版日期:2016-05-13.

基金项目:国家自然科学基金项目(61462018);广西学位与研究生教育  
改革和发展专项课题(JGY2014060);广西数字传播与文化软  
实力中心开放项目(ZFZD1408008);广西高校图像图形智能  
处理重点实验室开放基金项目(LD15042X).

通信作者:龚震霆.E-mail:gongxs7@163.com.

基于内容的图像检索是通过对图像内容的分析来搜索相似的图像,近些年一直被多媒体研究者广泛地研究。随着该研究领域的发展,虽然出现了许多新技术,但由于机器获得的低水平图像像素和人

类接受的高水平语义概念之间存在语义鸿沟问题,因此在当前基于内容的图像检索研究中从人类的感知方面把像素水平的信息和语义信息联系起来仍然是最具有挑战的问题<sup>[1-2]</sup>。近些年,研究者们使用了一些人工视觉特征去表示图像<sup>[3-4]</sup>,但基于这些人工特征的图像检索方法的性能一直不是很好。

这些挑战来源于人工智能的根本难题。而机器学习是一个很有前景的技术,可以解决这个长期的挑战。深度学习作为机器学习发展的一个分支,其动机在于建立模拟人脑进行分析学习的神经网络,通过组合低层特征形成更加抽象的高层特征,模仿人脑的机制来解释数据,如图像、声音、信号和文本,近些年已得到广泛应用<sup>[5-7]</sup>。

卷积神经网络(convolutional neural network, CNN)是人工神经网络的一种,由于它的权值共享结构和生物神经网络更类似,可以减少权值的数量,从而降低网络模型的复杂度,现已成为深度学习中的一个研究热点。目前, CNN 已经应用到视频中的人体动作识别信号重构、人脸等目标检测识别、图像分类等各个领域<sup>[8-11]</sup>。

鉴于 CNN 的这些进步,我们可以在图像检索中利用它的优点来提高检索性能。

## 1 相关工作

最近邻搜索是机器学习算法的一个基本步骤。近些年,网上可使用的数据快速增加,而且数据很容易就达到几百或几千维度,因此,在这样一个巨大的数据集上进行穷极线性搜索是不可行的。但是,在一些应用中,寻找近似最近邻(approximate nearest neighbor, ANN)却是十分有效的,其中基于树的方法和哈希方法是两个流行的框架。而哈希方法由于在速度和存储方面的优势,它作为一种 ANN 方法广泛用于大规模图像的检索中。哈希学习是学习一种保持图像相似性的紧密位表示,使得相似的图像能够匹配到相近的二进制哈希编码。

目前,主要流行一些非监督和监督的哈希方法<sup>[12-18]</sup>。非监督的哈希方法使用非标签数据来学习一组哈希函数<sup>[12-13,17]</sup>。其中,最具有代表性的是使用随机映射构造哈希函数的局部感知哈希(locality sensitive Hashing, LSH)<sup>[12]</sup>,它意在把相似的数据匹配到相近的二进制编码的概率最大化。然而, LSH 通常需要使用较长的哈希编码才能达到令人满意的精度,这导致了较大的存储空间需求和召回率普遍偏低的问题。谱哈希(spectral Hashing, SH)<sup>[13]</sup>

是另外一个具有代表性的方法,其利用非线性函数沿着数据的主成分分析(principal component analysis, PCA)方向通过设定阈值来产生二进制编码。

之后,大量的研究证明,使用监督信息可以提高哈希编码的学习性能。特别的,文献<sup>[14,16,18]</sup>在学习利用数据的标签信息来生成有效的哈希函数。

但是,图像检索中的这些哈希方法都是首先对图像提取人工视觉特征,而人工特征不需要获得图像的相似性,因此这通常可能影响这些哈希方法的效果,从而降低了图像检索的性能。

随着 2012 年 Alex 等<sup>[19]</sup>的 CNN 模型在 ILSVRC 数据集上训练 120 万数量的带标签图像获得了更高的图像分类准确率,最近几年深度卷积特征得到广泛研究,并在计算机视觉工作上取得了很大的突破<sup>[20-23]</sup>。2014 年, Xia 等<sup>[24]</sup>提出一种监督哈希方法 CNNH 和 CNNH+,该方法首先把训练图像数据成对的语义相似度矩阵因式分解成近似哈希编码,然后利用这些近似哈希编码和图像标签训练一个深度卷积网络,取得了更好的性能,但是, CNNH 和 CNNH+ 中的矩阵分解会带来额外的错误,使得训练目标偏离。2015 年, Guo 等<sup>[25]</sup>提出一种直接基于 CNN 的哈希方法 CNNBH,利用阈值 0 把一个全连接层的激活值二值化为二进制结果,从而得到哈希编码。同时,文献<sup>[26]</sup>提出一种简单但是非常有效的深度学习框架,该框架在文献<sup>[19]</sup>的深度 CNN 模型基础上添加一个隐藏层来同时学习特定领域的图像特征表示和一组类哈希函数,并取得了最好的检索性能。

如今这些优秀的基于 CNN 的检索方法虽然获得了巨大的进展,但是它们并没有把以前最好水准的哈希编码策略联系起来。

## 2 CNN 和哈希编码相结合的方法

在图像检索中,如果开始对图像提取的特征不出色的话,那么后面采用很高水准的哈希编码策略进行编码检索也不一定能取得很好的效果。因此,本文提出一种 CNN 和以前最好水准的哈希策略相结合的方法。我们采用文献<sup>[28]</sup>在 IMAGENET Large-scale visual recognition challenge (ILSVRC)数据集上预训练过的 VGGNet-D 网络模型对目标图像集进行特征提取,获得图像的深层卷积特征表示,然后对这些表示分别采用 LSH<sup>[12]</sup>、SH<sup>[13]</sup>、SKLSH<sup>[15]</sup>、ITQ<sup>[16]</sup>、PCA-RR<sup>[16]</sup>、DSH<sup>[18]</sup> 6 种哈希编码策略进行编码得到哈希码,最后进行快速检索。将这些结合

策略分别命名为 LSH-VC、SH-VC、SKLSH-VC、ITQ-VC、PCA-RR-VC 和 DSH-VC。

2.1 VGGNet-D 网络模型

最近几年,CNN 模型在一些目标检测和图像识别中已经获得了令人印象深刻的结果。文献[28]使用很小卷积核的网络架构,对增加深度进行了全面的评估;实验表明,把网络的卷积层深度增加到 16~19 层时可以显著提升性能,设计的网络在 ImageNet Challenge 2014 的定位和分类项目上分别获得了第一名和第二名的成绩。此外,文献[28]的模型特征对许多不同的计算机视觉任务和数据集有很强的泛化能力,能够相比或者超过建立在较浅图像特征上的更复杂的识别方法。

本文选用文献[28]中优异的 VGGNet-D 网络模型。VGGNet-D 的结构及每层参数设置如表 1。

表 1 VGGNet-D 网络模型结构及参数

Table1 VGGNet-D model's structure and parameters			
conv1	conv2	conv3	conv4
64×3×3 st. 1, pad 1	64×3×3 st. 1, pad 1× 2 pooling	128×3×3 st. 1, pad 1	128×3×3 st. 1, pad 1× 3 pooling
conv5	conv6	conv7	conv8
256×3×3 st. 1, pad 1	256×3×3 st. 1, pad 1	256×3×3 st. 1, pad 1× 2 pooling	512×3×3 st. 1, pad 1
conv9	conv10	conv11	conv12
512×3×3 st. 1, pad 1	512×3×3 st. 1, pad 1× 2 pooling	512×3×3 st. 1, pad 1	512×3×3 st. 1, pad 1
conv13	FC14	FC15	FC16
512×3×3 st. 1, pad 1× 2 pooling	4 096×7×7 dropout	4 096×1×1 dropout	1 000×1×1 softmax

VGGNet-D 模型由 13 卷积层(conv1~13)和 3 个全连接层(FC14~16)构成。表 1 中,卷积层的第一行参数表示卷积滤波器的数量和局部感受野的大小;“st.”表示卷积的步幅,“pad”表示空间填充;“x2 pooling”表示 max-pooling 下采样。接着,FC14 和 FC15 使用 dropout<sup>[19]</sup>方法来调整某些隐含层节点的权重不工作,最后的 FC16 是多路 softmax 分类器。其中,VGGNet-D 的激活函数使用矫正线性单元(rectification linear unit,ReLU)。

2.2 VGGNet-D 模型的训练学习

和文献[19]一致,VGGNet-D 网络模型使用批量梯度下降法在 ILSVRC 数据集上进行训练学习;使用的超参数:批量大小、动量、权重衰减的  $l_2$  惩罚系数、dropout 比率和初始学习率分别设置为 256、0.9、0.0005、0.5 和 0.01,而学习率在验证准确率停止增加时减小为原来值的十分之一;网络的每一层的权重由均值为 0 和方差为 0.01 的高斯分布来初始化。具体的,该模型的训练学习是最小化其整体代价函数。这里假定有一组数量为  $m$  的训练样本集  $\{(x^{(l)}, y^{(l)}), \dots, (x^{(m)}, y^{(m)})\}$ ,对于整体样本,网络整体代价函数可表示为式(1):

$$J(W,b)=\left[\frac{1}{m}\sum_{i=1}^m\left(\frac{1}{2}\parallel h_{w,b}(x^{(i)})-y^{(i)}\parallel^2\right)\right]+\frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(W_{ji}^{(l)})^2\tag{1}$$

式中: $h_{w,b}(x)$ 是神经网络模型, $W_{i,j}^{(l)}$ 是第  $l$  层第  $j$  单元与第  $l+1$  层第  $i$  单元之间的连接权重, $b$  是隐藏层神经元的偏置项;式(1)右边是一个减小权重幅度的规则化项,可以起到防止过拟合的作用, $\lambda$  来调整代价函数中前后两项的相对重要性。求解式(1)最小值采用著名的批量梯度下降法最优化算法,而计算  $J(W,b)$  对  $W$  和  $b$  的偏导数时,使用反向传导算法。

2.3 本文策略和以前最好水准的哈希策略

针对图像检索中重要的哈希编码部分,下面我们介绍本文方法的 6 种策略的核心技术。

1)LSH-VC:和 LSH<sup>[27]</sup>一样,首先使用高斯随机矩阵将实数输入空间  $R^d$  中的数据点随机映射到一个  $R^t$  空间中,其中  $t$  是一个很小的超常量,接着使用球分割方法把这个  $t$  维实数空间分割成多个单元,然后用哈希函数返回含有数据映射点的球单元的索引,度量标准选用  $l_1$  范数。

2)SH-VC:和 SH<sup>[13]</sup>一样,基于量化沿着数据 PCA 方向计算得到的解析特征函数值。

假设数据集中样本点  $\mathbf{x}_i, \mathbf{x}_j$  属于输入特征空间  $R^d, \{\mathbf{y}\}_{i=1}^n$  表示  $n$  个样本点的长度为  $k$  的哈希码,  $\mathbf{W}_{n \times n}$  是一个关联矩阵,由于输入空间数据之间的欧式距离和相似度相关,那么输入特征空间用高斯核度量相似度  $W(i,j)=\exp(-\parallel \mathbf{x}_i-\mathbf{x}_j\parallel^2/\varepsilon^2)$ ,参数  $\varepsilon$  规定空间  $R^d$  中相似数据的距离,相似的近邻间的平均汉明距离为  $\sum_{ij} W_{i,j} \parallel \mathbf{y}_i-\mathbf{y}_j\parallel^2$ ,这样对图像特征向量的哈希编码过程可以表示为式(2)的优化问题:



$$\begin{aligned}
& \text{minimize: } \sum_{ij} W_{ij} \|y_i - y_j\|^2 \\
& \text{s.t. } y_i \in \{-1, 1\}^k \\
& \sum_i y_i = 0 \\
& \frac{1}{n} \sum_i y_i y_i^T = I
\end{aligned} \quad (2)$$

式中:约束条件  $\sum_i y_i = 0$  要求哈希码的每一位有 50% 的概率是 0 或 1,同时约束条件  $\frac{1}{n} \sum_i y_i y_i^T = I$  要求哈希码每位之间不相关。这个优化问题等价于一个平衡图分割问题,是一个 NP-hard 问题,对相似图的拉普拉斯矩阵的特征向量特征值的分析可为该问题提供一个松弛解。首先,使用 PCA 找出数据的主成分,然后沿着每一个 PCA 方向使用矩形近似来计算  $k$  个最小的单维度解析特征函数,最后在 0 处阈值化解析特征函数,从而获得二进制编码。

3) SKLSH-VC: 和 SKLSH<sup>[15]</sup> 一样,基于随机映射,适应任意分布的样本数据,在映射过程中不仅保留了原始数据之间的位置关系,而且考虑了生成的哈希码之间归一化的汉明距离的上下限。

考虑  $D$  维实数空间  $R^D$  上的一个核函数  $K(\cdot, \cdot)$  对于所有数据点  $x, y \in R^D$  满足下列条件:

①  $K(\cdot, \cdot)$  具有平移不变性,例如  $K(x, y) = K(x - y)$ 。

②  $K(\cdot, \cdot)$  是归一化的,例如  $K(x - y) \leq 1$ , 而且  $K(x - x) = K(0) = 1$ 。

③ 对于任何实数  $\alpha \geq 1$ ,  $K(\alpha x, \alpha y) \leq K(x - y)$ 。

高斯核函数  $K(x, y) = \exp(-\gamma \|x - y\|^2 / 2)$  或者拉普拉斯核函数  $K(x, y) = \exp(-\gamma \|x - y\|_1)$  是满足上面条件的两个著名样例,我们构建一个映射函数  $F^n$ ,先将原始  $R^D$  空间中的数据映射到低维空间,得到数据的随机傅里叶特征,确保平移不变核值没有变化,再对随机傅里叶特征映进行随机二值量化,从而得到二进制哈希码。

4) ITQ-VC、PCA-RR-VC: 和 ITQ、PCA-RR<sup>[16]</sup> 一样,对数据集进行 PCA 降维处理,问题转化为将该数据集中的数据样本点映射到一个以零为中心的二维超立体的顶点上,不同的量化误差得到对应该数据集的不同的二进制编码。

假设有一组数据点  $\{x_1, x_2, \dots, x_n\}$ ,  $x_i \in R^d$  形成数据集矩阵  $X \in R^{n \times d}$ , 且以 0 为中心,目标是学习二进制哈希编码矩阵  $B \in \{-1, 1\}^{n \times c}$ ,  $c$  是编码长度,这样整个编码过程可以表示为  $B = \text{sgn}(X\tilde{W})$ ,  $\tilde{W}$  是映射矩阵。

对随机生成的矩阵进行奇异值分解 (singular value decomposition, SVD) 得到对应的  $c \times c$  的正交矩

阵来初始化旋转矩阵  $R$ ,  $W$  是数据 PCA 方向的矩阵,PCA-RR-VC 策略采用映射矩阵  $\tilde{W} = WR$ , 将经过 PCA 处理后的数据集  $V$  乘以随机正交矩阵  $R$  进行旋转变换;而 ITQ-VC 策略中使用的映射矩阵  $\tilde{W}$  中的  $R$  是经过优化的量化误差最低的正交矩阵,这个优化问题,即找出最优的正交旋转矩阵  $R$  和与之对应的哈希编码矩阵  $B$ ,采用交替迭代的求解方法来解决。

5) DSH-VC: 和 DSH<sup>[18]</sup> 一样,通过  $k$  均值聚类的量化结果,寻找  $r$ -adjacent 组,即利用数据的几何结构来指导哈希函数映射的选择。

假设有  $n$  个训练数据样本  $x_1, x_2, \dots, x_n \in R^d$ ,  $L$  是数据哈希编码的长度,参数  $\alpha$  控制样本被分组的数量。首先使用  $k$  均值算法迭代  $p$  次生成中心点分别为  $\mu_1, \mu_2, \dots, \mu_{\alpha L}$  的  $\alpha L$  个分组  $S = \{S_1, S_2, \dots, S_{\alpha L}\}$ , 其中  $\mu_i = \frac{1}{|S_i|} \sum_{x \in S_i} x$ ,  $i = 1, 2, \dots, \alpha L$ , 定义分组的  $r$  最近邻矩阵  $W$  为式(3):

$$W_{ij} = \begin{cases} 1, & \mu_i \in N_r(\mu_j) \text{ or } \mu_j \in N_r(\mu_i) \\ 0, & \text{否则} \end{cases} \quad (3)$$

式中:  $N_r(\mu_i)$  表示中心  $\mu_i$  的最近邻集合。当且仅当  $W_{ij} = 1$  时,组  $S_i$  和组  $S_j$  为  $r$ -adjacent 组(相比于随机映射,选取能够很好分开两个 adjacent 组的映射),在生成所有的  $r$ -adjacent 组后,对每一对 adjacent 组产生映射  $w$  和截距  $t$ ,接着利用带有权值的中心计算所有候选映射的熵,按降序对熵值进行排序,最后使用前  $L$  个映射并根据哈希映射函数式(4)来生成二进制编码:

$$h(x) = \begin{cases} 1, & w^T x \geq t \\ 0, & \text{否则} \end{cases} \quad (4)$$

式中  $h(x)$  是样本空间  $R^d$  中样本  $x$  的哈希函数。

### 3 实验结果与分析

本文将在 Caltech101<sup>[17]</sup> 和 Caltech256<sup>[18]</sup> 2 个著名的图像集上对 LSH-VC、SH-VC、SKLSH-VC、ITQ-VC、PCA-RR-VC、DSH-VC 等 6 种策略和 LSH、SH、SKLSH、ITQ、PCA-RR、DSH 这些以前最好水准的图像哈希检索策略进行对比实验。

我们通过精度-召回率 (Precision-Recall) 和平均正确率值-编码位数 (mAP-Number of bits) 两种标准对本文方法进行性能评估。实验中,本文使用数据集第 50 个最近邻的平均距离作为阈值来判断一个查询到的数据是否和查询数据属于同一类。

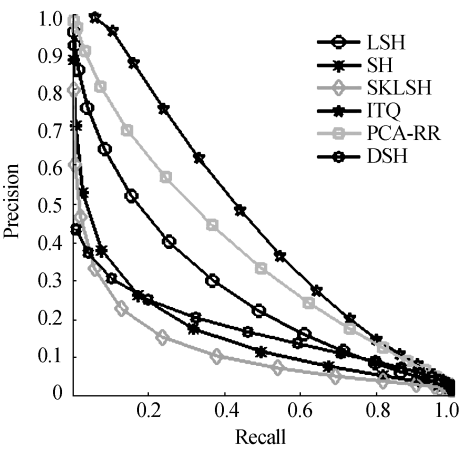
实验环境: 64 位 Win8 系统, CPU 主频 2.50 GHz, 8 Gbyte 内存, MATLAB2014a。

#### 3.1 Caltech101

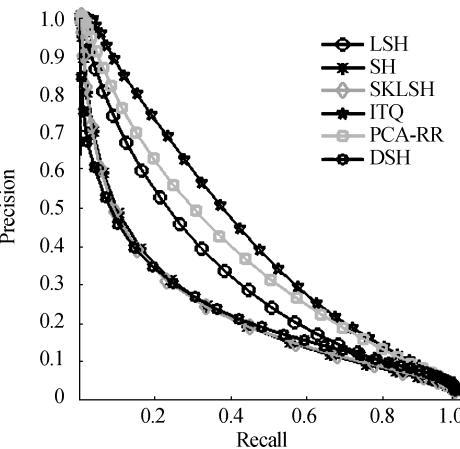
Caltech101 数据集包含 8 677 张图片,共 101 个

类别,包括家具、动物、运动器材、车辆等,每一类图像的数量最少是 31,最多是 800。其中,图片的大小各不相同。为了公平的对比,实验中,所有策略都是随机选取 1 000 张作为测试图像。

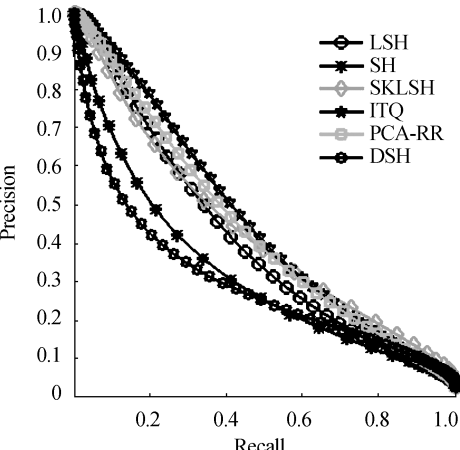
1)我们采用常用的 32、64、128 和 256 等 4 种哈希编码位数在 Precision-Recall 这个标准上进行实验分析,实验结果如图 1 和图 2。



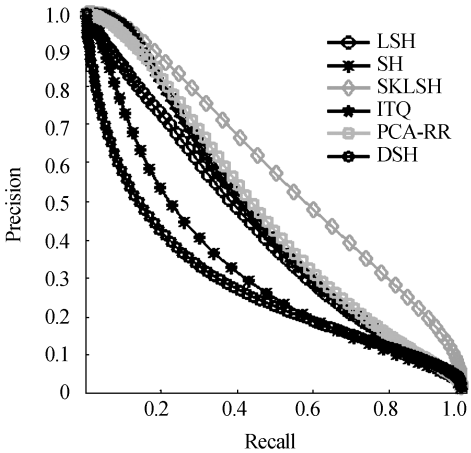
(a) 32 位编码时 Precision-Recall 曲线



(b) 64 位编码时 Precision-Recall 曲线



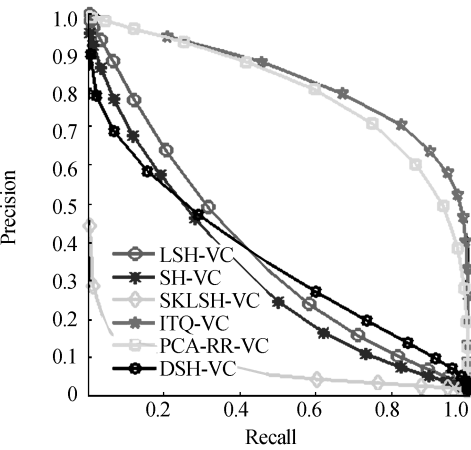
(c) 128 位编码时 Precision-Recall 曲线



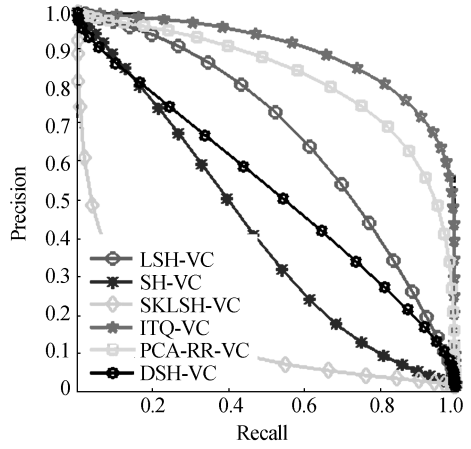
(d) 256 位编码时 Precision-Recall 曲线

图 1 以前最好水准的策略在 Caltech101 上的 Precision-Recall 曲线

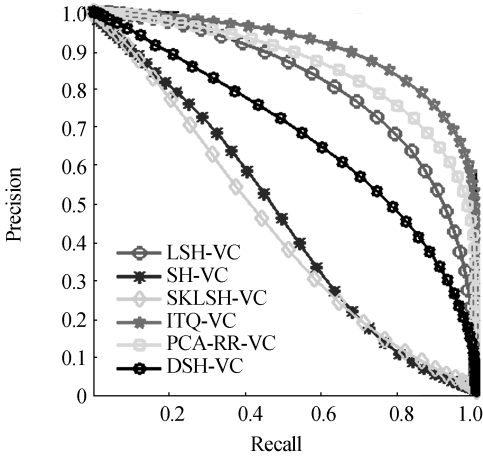
Fig.1 The Precision-Recall curve of the previous state-of-the-art strategies on Caltech 101



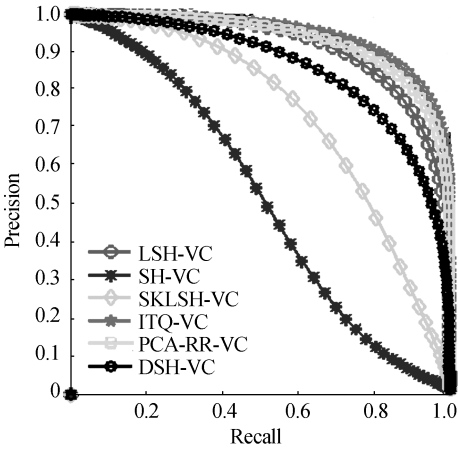
(a) 32 位编码时 Precision-Recall 曲线



(b) 64 位编码时 Precision-Recall 曲线



(c) 128 位编码时 Precision-Recall 曲线



(d) 256 位编码时 Precision-Recall 曲线

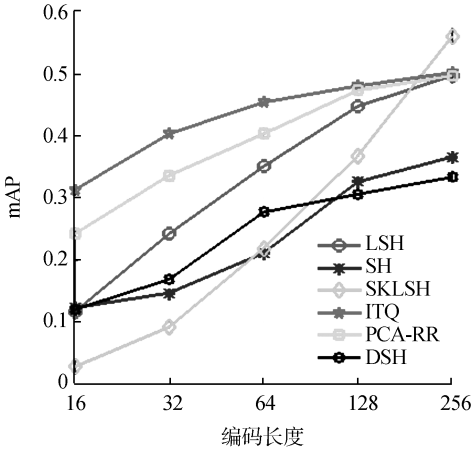
图 2 本文方法 6 种策略在 Caltech101 上的 Precision-Recall 曲线  
Fig.2 The Precision-Recall curve of the six strategies on Caltech 101

对比图 1 和图 2,我们清楚地看到,本文的 6 种图像检索策略相比于以前最好水准的对应的策略有更优的 Precision-Recall 曲线(64 位编码时 SKLSH-VC 略差于 SKLSH);其中 LSH-VC、ITQ-VC、PCA-RR-VC 和 DSH-VC 的优势更明显。依赖数据的策略(ITQ、ITQ-VC、PCA-RR、PCA-RR-VC 和 SH、SH-VC)在 gist 和 CNN 特征下随着编码位数的增加都会遇到性能瓶颈,而独立于数据的 SKLSH 和 SKLSH-VC 策略就没有这种限制。

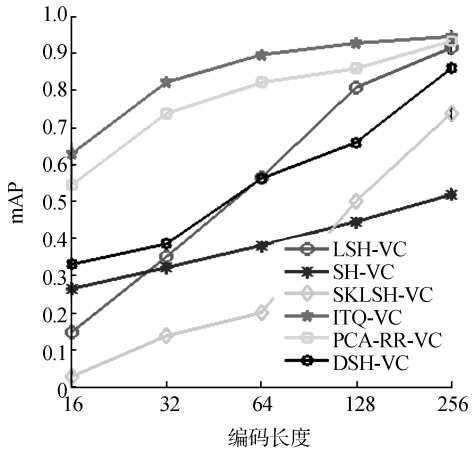
此外,图 2 表明 ITQ-VC 和 PCA-RR-VC 在编码位数较小时,相比于本文其他 4 种策略在相同召回率情况下的准确率有更大的优势。

2)mAP 是反映一个方法在全部相关数据上性能的单值指标。利用该方法检索出来的相关数据的排位越靠前,mAP 就可能越高。本文方法在 mAP-Number of bits 标准上的实验结果如图 3。

图 3 中,使用 CNN 特征的本文 6 种策略的 mAP 值明显高于以前最好水准的对应的使用 gist 特征的检索策略(16 和 64 位编码时 SKLSH-VC 略差于 SKLSH);同样的,LSH-VC、ITQ-VC、PCA-RR-VC 和 DSH-VC 的优势更明显。其中 ITQ-VC 在不同的编码位数时一直拥有最高的 mAP 值,相比于 ITQ 策略分别提升了 103.0%、105.7%、97.2%、93.4%、89.3%,效果显著。



(a) gist 特征下不同编码位数时 mPA 值



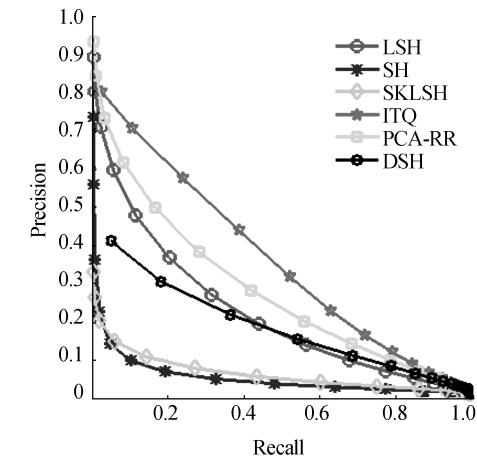
(b) cnn 特征下不同编码位数时 mPA 值

图 3 所有策略在 Caltech101 上的 mAP 值  
Fig.3 The mean average precision of all strategies on Caltech 101

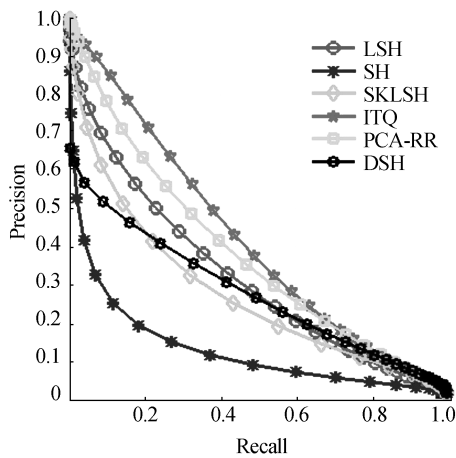
3.2 Caltech256

Caltech256 数据集包含 29 780 张图片,共 256 个类别,相比于 Caltech101,类与类之间和每类中图像主体位置变化更明显,十分适合于图像检索实验。为了公平地对比,实验中,所有策略同样都是随机选取 1 000 张作为测试图像。

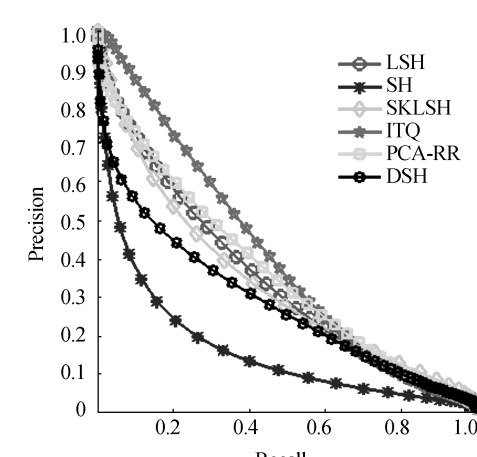
1)和 3.1 节一样,在 Precision-Recall 这个标准上我们采用 32、64、128 和 256 这 4 种编码位数进行实验分析,实验结果如图 4、5。



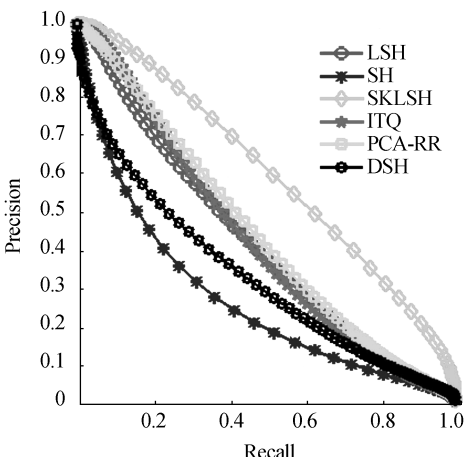
(a) 32 位编码时 Precision-Recall 曲线



(b) 64 位编码时 Precision-Recall 曲线



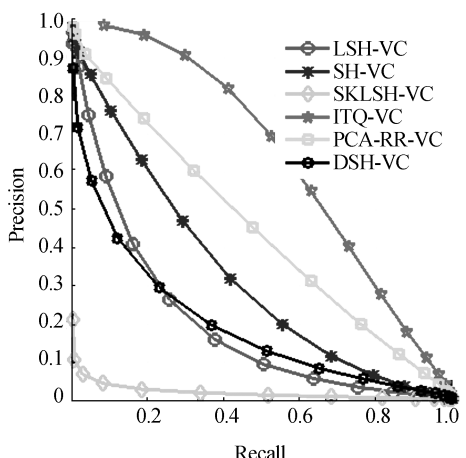
(c) 128 位编码时 Precision-Recall 曲线



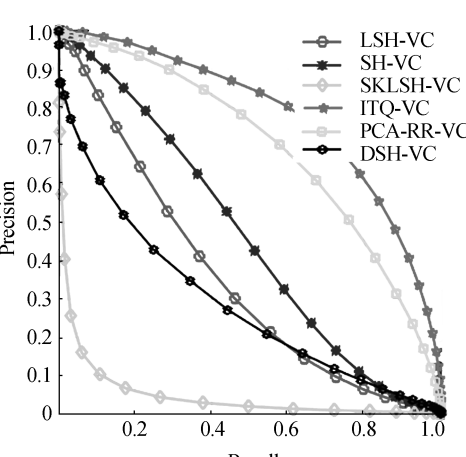
(d) 256 位编码时 Precision-Recall 曲线

图 4 以前最好水准的策略在 Caltech256 上的 Precision-Recall 曲线

Fig.4 The Precision-Recall curve of the previous state-of-the-art strategies on Caltech 256

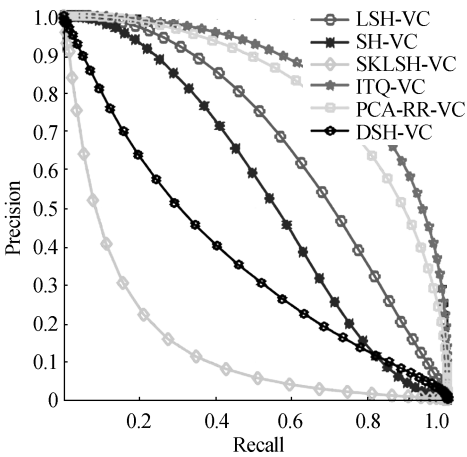


(a) 32 位编码时 Precision-Recall 曲线

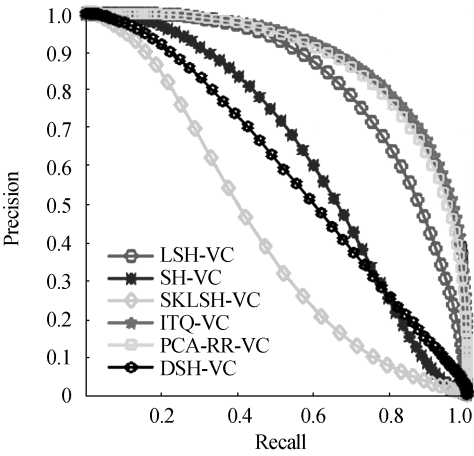


(b) 64 位编码时 Precision-Recall 曲线





(c) 128 位编码时 Precision-Recall 曲线



(d) 256 位编码时 Precision-Recall 曲线

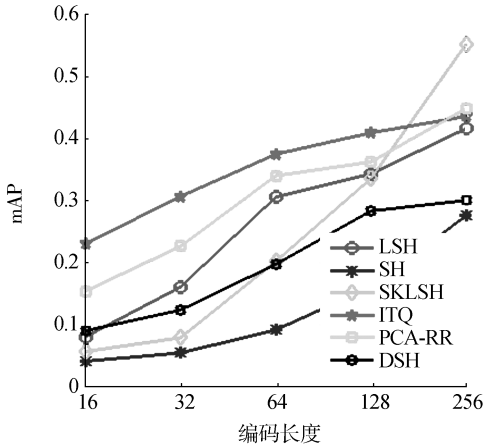
图 5 本文方法 6 种策略在 Caltech256 上的 Precision-Recall 曲线

Fig.5 The Precision-Recall curve of the six strategies on Caltech 256

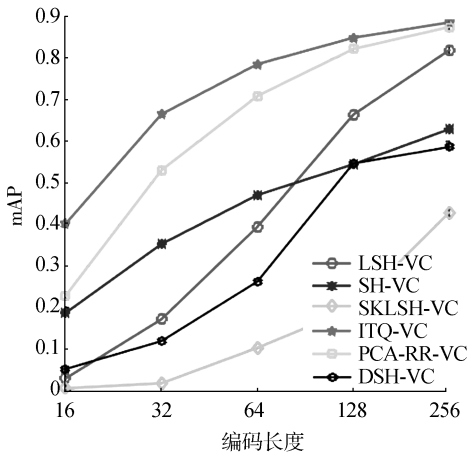
对比图 4、5, 可以看到本文方法的 5 种策略相比于以前最好水准的对应策略具有更好的 Precision-Recall 曲线(SKLSH-VC 除外)。随着编码位数增加, ITQ-VC、PCA-RR-VC 和 SH-VC 似乎达到了性能的上限, 而 LSH-VC 在 CNN 特征下却不受这个限制, 性能提升迅速。同时, 可以清楚地看到结合 CNN 特征的 ITQ-VC 检索策略在 4 种不同的编码位数下都具有最优的 Precision-Recall 曲线。

2) 在 mean Average Precision-Number of bits 标准上的实验结果如图 6。分析图 6, 相比以前最好水准的对应策略, 使用 CNN 特征的本文 5 种策略获得了更高的 mAP 值(SKLSH-VC 除外)。ITQ-VC 在所有编码位数下一直有最高的 mAP 值, 分别为 0.401、0.665、0.785、0.849 和 0.886, 在 ITQ 上提高了 74.2%、116.8%、109.5%、107.4% 和 103.3%, 效果显

著;结合前面的实验中 ITQ-VC 一直有最好的性能表现, 可以得知基于人工特征的性能较好的编码策略在深层卷积特征下依旧能够获得较好的性能。



(a) Gist 特征下不同编码位数时 mAP 值



(b) cnn 特征下不同编码位数时 mAP 值

图 6 所有策略在 Caltech256 上的 mAP 值

Fig.6 The mean Average Precision of all strategies on Caltech 256

在 Caltech101 和 Caltech256 数据集上的两组实验充分说明了使用 CNN 的卷积特征进行哈希编码检索能够获得一定程度的性能提升。

## 4 结束语

在图像检索上, 本文提出一种结合 CNN 和以前最好水准的哈希编码策略的有效方法。首先, 采用 VGGNet-D 网络模型对目标图像集提取图像的深层特征表示, 再使用以前最好水准的哈希编码策略把这些特征编码成二进制码。实验结果显示, 本文方法的 LSH-VC、SH-VC、ITQ-VC、PCA-RR-VC 和 DSH-VC 策略相比于以前最好水准的对应策略获得了更高的性能。同时本文方法表明, 当 CNN 这种最先进



的技术在图像检索领域迅猛发展的时候,一些先前经典的哈希编码策略仍不能忽略,这将在图像检索应用上具有一定的参考价值。另外,我们还有若干问题有待解决,如卷积特征并不是对所有的哈希编码策略都有效,这些问题有待进一步研究。

## 参考文献:

- [1] SMEULDERS A W M, WORRING M, SANTINI S, et al. Content-based image retrieval at the end of the early years [J]. *IEEE transactions on pattern analysis and machine intelligence*, 2000, 22(12): 1349-1380.
- [2] WAN Ji, WANG Dayong, HOI S C H, et al. Deep learning for content-based image retrieval: a comprehensive study [C]//*Proceedings of the 22nd ACM international conference on multimedia*. Orlando, USA, 2014: 157-166.
- [3] LOWE D G. Distinctive Image features from scale-invariant keypoints [J]. *International journal of computer vision*, 2004, 60(2): 91-110.
- [4] BAY H, TUYTELAARS T, VAN GOOL L. SURF: speeded up robust features [M]//LEONARDIS A, BISCHOF H, PINZ A. *Computer vision-ECCV 2006*. Berlin Heidelberg: Springer, 2006: 404-417.
- [5] SARIKAYA R, HINTON G E, DEORAS A. Application of deep belief networks for natural language understanding[J]. *IEEE/ACM transactions on audio, speech, and language processing*, 2014, 22(4): 778-784.
- [6] LANDECKER W, CHARTRAND R, DEDEO S. Robust sparse coding and compressed sensing with the difference map[C]//*Proceedings of the 13th European conference on computer vision*. Zurich, Switzerland, 2014: 315-329.
- [7] GRAVES A, MOHAMED A R, HINTON G. Speech recognition with deep recurrent neural networks[C]//*Proceedings of the IEEE international conference on acoustic speech and signal processing*. Vancouver, British Columbia, Canada, 2013: 6645-6649.
- [8] BRUNA J, SZLAM A, LECUN Y. Signal recovery from pooling representations[J]. *Eprint Arxiv*, 2013: 307-315.
- [9] LI Haoxiang, LIN Zhe, SHEN Xiaohui, et al. A convolutional neural network cascade for face detection[C]//*Proceedings of the 2015 IEEE conference on computer vision and pattern recognition*. Boston, Massachusetts, USA, 2015: 5325-5334.
- [10] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//*Proceedings of the 2014 IEEE conference on computer vision and pattern recognition*. Columbus, OH, USA, 2014: 580-587.
- [11] LIN Min, CHEN Qiang, YAN Shuicheng. Network in network [C]//*Proceedings of international conference on learning representations*. Banff, Canada, 2014.
- [12] GIONIS A, INDYK P, MOTWANI R. Similarity search in high dimensions via hashing[C]//*Proceedings of the 25th international conference on very large data bases*. San Francisco, CA, USA, 1999: 518-529.
- [13] WEISS Y, TORRALBA A, FERGUS R. Spectral hashing [C]//*Proceedings of conference on neural information processing systems*. Vancouver, British Columbia, Canada, 2008: 1753-1760.
- [14] KULIS B, DARRELL T. Learning to hash with binary reconstructive embeddings[C]//*Advances in neural information processing systems 22: 23rd annual conference on neural information processing systems 2009*. Vancouver, British Columbia, Canada, 2010: 1042-1050.
- [15] RAGINSKY M, LAZEBNIK S. Locality-sensitive binary codes from shift-invariant kernels[C]//*Advances in neural information processing systems 22: conference on neural information processing systems 2009*. Vancouver, British Columbia, Canada, 2009: 1509-1517.
- [16] GONG Yunchao, LAZEBNIK S, GORDO A, et al. Iterative quantization: a procrustean approach to learning binary codes for large-scale image retrieval[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2013, 35(12): 2916-2929.
- [17] NOROUZI M, FLEET D J. Minimal loss hashing for compact binary codes[C]//*Proceedings of the 28th international conference on machine learning*. Bellevue, WA, USA, 2011: 353-360.
- [18] JIN Zhongming, LI Cheng, LIN Yue, et al. Density sensitive hashing[J]. *IEEE transactions on cybernetics*, 2014, 44(8): 1362-1371.
- [19] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//*Advances in Neural Information Processing Systems 25*. Lake Tahoe, Nevada, USA, 2012.
- [20] DONAHUE J, JIA Yangqing, VINYALS O, et al. DeCAF: a deep convolutional activation feature for generic visual recognition [C]//*Proceedings of the 31st international conference on machine learning*. Beijing, China, 2014: 647-655.
- [21] ZEILER M D, FERGUS R. Visualizing and Understanding Convolutional Networks[C]//*Proceedings of the 13th European conference on computer vision*. Zurich, Switzerland, 2014: 818-833.
- [22] SERMANET P, EIGEN D, ZHANG Xiang, et al. OverFeat: integrated recognition, localization and detection using convolutional networks[J]. *Eprint Arxiv*, 2013.
- [23] RAZAVIAN A S, AZIZPOUR H, SULLIVAN J, et al. CNN features off-the-shelf: an astounding baseline for recognition [C]//*Proceedings of the 2014 IEEE conference*

on computer vision and pattern recognition workshops. Columbus, OH, USA, 2014: 512-519.

[24] XIA Rongkai, PAN Yan, LIU Cong, et al. Supervised hashing for image retrieval via image representation learning[C]//Proceedings of the 24th AAAI conference on artificial intelligence. Qubec City, Canada, 2014: 2156-2162.

[25] GUO Jinma, LI Jianmin. CNN Based Hashing for Image Retrieval[Z]. arXiv: 1509. 01354v1, 2015.

[26] LIN K, YANG H F, HSIAO J H, et al. Deep learning of binary hash codes for fast image retrieval[C]//Proceedings of the 2015 IEEE conference on computer vision and pattern recognition workshops (CVPRW). Boston, Massachusetts, USA, 2015: 27-35.

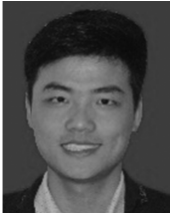
[27] ANDONI A, INDYK P. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions[C]//Proceedings of IEEE 54th annual symposium on foundations of computer science. Berkeley, CA, USA, 2006: 459-468.

[28] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. Eprint Arxiv, 2014.

[29] LI Feifei, FERGUS R, PERONA P. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories [J]. Computer vision and image understanding, 2007, 106 (1): 59-70.

[30] GRIFFIN G, HOLUB A, PERONA P. Caltech-256 object category dataset[R]. CaltechAUTHORS: CNS-TR-2007-001, 2007.

作者简介:



龚震霆,男,1991年生,硕士研究生,主要研究方向为计算机视觉、机器学习。



陈光喜,男,1971年生,博士生导师,主要研究方向为可信计算、图像处理。主持完成国家自然科学基金项目2项、广西省科学基金及企业开发项目多项。获桂林市科技进步三等奖1项、广西教学成果奖一等奖1项。发表学术论文30余篇,主编教材1部。



任夏荔,女,1992年生,硕士研究生,主要研究方向为计算机视觉、深度学习。

2016 年控制论、机器人与控制国际会议  
2016 International Conference on Cybernetics,  
Robotics and Control (CRC 2016)

2016 International Conference on Cybernetics, Robotics and Control will be held in Hong Kong during August 19-21, 2016.

The idea of the conference is providing a platform for scientists, scholars, and engineers from all over the world to present ongoing researches, and to foster research relations between the Universities and the Industry. We are looking forward to your participation.

It features invited keynote speakers as well as peer-reviewed paper presentations. The conference is completely open (one needs to register first), you will not have to be an author or a discussant to attend.

The covered key topics will include (but will not be limited to):

- Modern Advanced Control Strategies;
- Data Analysis, Prediction & Model Identification;
- Decision Making and Information Retrieval;
- Control System Application;
- Database System;
- Human-Machine Systems;
- Robotics and Automation;
- Multimedia and Communication Systems;
- Hybrid Systems.