

# 基于卷积神经网络和二进制 K-means 的图像快速聚类

柯圣财<sup>1</sup> 李弼程<sup>2</sup> 唐永旺<sup>1</sup> 吴志兵<sup>3</sup> 万建平<sup>4</sup>

(1. 解放军信息工程大学信息工程学院, 郑州, 450002; 2. 华侨大学计算机科学与技术学院, 厦门, 361021; 3. 江南计算技术研究所, 无锡, 214083; 4. 西安卫星测控研究中心, 西安, 710000)

**摘要:** 当前主流的图像聚类方法采用的视觉特征缺乏自主学习能力, 导致其图像表达能力不强, 而且传统的聚类算法计算复杂度较高, 聚类效率低, 难以适应大数据环境。针对这些问题, 本文提出了一种基于卷积神经网络和二进制 K-means 的图像快速聚类方法。首先, 利用卷积神经网络学习图像内容的内在隐含关系, 得到图像高阶特征, 增强特征的视觉表达能力和区分性; 然后, 利用哈希方法将高维图像特征映射为低维二进制哈希码, 并通过聚类中心构造多索引哈希表来加速寻找最近的聚类中心, 以降低时间复杂度; 最后, 利用二进制 K-means 完成二进制哈希码的快速聚类。在 ImageNet-1000 图像集上的实验结果表明, 本文方法能够有效地增强图像特征的表达能力和提高图像聚类效率, 性能优于当前主流方法。

**关键词:** 深度学习; 图像聚类; 卷积神经网络; 二进制 K-means; 多索引哈希

**中图分类号:** TP391 **文献标志码:** A

## Fast Image Clustering Based on Convolutional Neural Network and Binary K-means

Ke Shengcai<sup>1</sup>, Li Bicheng<sup>2</sup>, Tang Yongwang<sup>1</sup>, Wu Zhibing<sup>3</sup>, Wan Jianping<sup>4</sup>

(1. Institute of Information System Engineering, Information Engineering University, Zhengzhou, 450002, China; 2. College of Computer Science and Technology, Hua Qiao University, Xiamen, 361021, China; 3. Jiangnan Institute of Computing Technology, Wuxi, 214083, China; 4. Xi'an Satellite Control Center, Xi'an, 710000, China)

**Abstract:** Visual features used in state-of-the-art image clustering methods lack of independent learning ability, which leads to low image expression ability. Furthermore, the efficiency of traditional clustering methods is low for large image dataset. So, a fast image clustering method based on convolutional neural network and binary K-means is proposed in this paper. Firstly, a large-scale convolutional neural network is employed to learn the intrinsic implications of training images so as to improve the discrimination and representational power of visual features. Secondly, hashing is applied to map high-dimensional deep features into low-dimensional hamming space, and multi-index hash table is used to index the initial centers so that the nearest center lookup becomes extremely efficient. Finally, image clustering is accomplished efficiently by binary K-means algorithm. Experimental results on ImageNet-1000 dataset indicate that the proposed method can effectively enhance the expression ability of image features, increase the image clustering efficiency and has better performance than state-of-the-art methods.

**Key words:** deep learning; image clustering; convolutional neural network; binary K-means; multi-index hashing

## 引 言

随着大数据时代的到来,互联网图像资源迅猛增长,如何对大规模图像资源进行有效的聚类管理以满足用户快速获取有用信息的需求亟待解决。传统的图像聚类方法都是以视觉底层特征(如 GIST<sup>[1]</sup>、SIFT<sup>[2]</sup>、SURF<sup>[3]</sup>等)表示图像内容。虽然这些特征在图像处理领域表现出优良的性能,但是生成这些特征时固定的编码步骤和浅层的网络结构使得描述子缺少学习能力,限制了其图像内容表达能力,难以适应多样的图像数据。另外,经典的聚类算法(如 K-means<sup>[4]</sup>、K-medoids<sup>[5]</sup>等)对特征点进行聚类需要多次迭代高维近似邻计算,随着数据量的增加聚类效率会急剧下降,因而难以应用于大规模图像资源的有效聚类管理。

为学习得到大规模图像数据的内在隐含关系,增强特征的图像表达能力,Hinton等<sup>[6-8]</sup>针对深度置信网络(Deep belief network, DBN)提出非监督贪婪逐层训练算法,解决了深度学习(Deep learning)模型优化困难的问题,并将深度学习成功应用于图像处理领域中,为提取更加有效的图像特征提供了新思路。Tang等<sup>[9]</sup>在 Hinton 的研究基础上将 DBN 的第一层采用稀疏化连接,同时利用降噪算法进一步提高了 DBN 输出特征对噪声的鲁棒性。Nair等<sup>[10]</sup>则将 DBN 顶层改用三阶玻尔兹曼机(Boltzmann machine, BM)结构,并用于三维目标特征提取,该特征对目标旋转变化的鲁棒性。为学习得到三维目标的高阶特征,Bu等<sup>[11]</sup>利用几何词袋模型(Geometric bag-of-words)得到低阶 3-D 形状描述子之间的几何关系并生成中层特征,最后通过 DBN 对中层特征学习得到 3-D 目标的高阶特征表示。Zhang等<sup>[12]</sup>针对高分辨率卫星图像内容复杂、特征提取计算大等问题,利用显著度分析对卫星图像进行分割,再采用稀疏自编码方法对图像块进行编码,最后通过两层的卷积神经网络(Convolutional neural network, CNN)生成特征。Lee等<sup>[13]</sup>将 CNN 中的权值共享机制引入 DBN 中,构建了卷积深度置信网络(Convolutional deep belief network, CDBN),该模型能从未标注的自然图像中学习有效的高阶特征表示。He等<sup>[14]</sup>通过在 CNN 的卷积层和全连接层加入空间金字塔融合(Spatial pyramid pooling, SPP)层,直接对不同大小的图像进行学习并生成多尺度特征。通过深度学习生成的图像特征维数较高,存在维数灾难问题<sup>[15,16]</sup>,当图像数据规模较大时,若采用经典的聚类算法进行聚类就会使聚类效率会急剧下降,难以适用于大规模数据。

K-means<sup>[4]</sup>是经典的聚类算法,其时间复杂度为  $O(knd)$ ,其中,  $k$  为聚类中心数目,  $n$  为特征点数目,  $d$  为特征维数。为实现高维特征数据的快速聚类,Philbin等<sup>[17]</sup>将 KD-Tree 引入 K-means 中提出了 KDK-means 算法,利用 KD-Tree 对聚类中心构建索引目录,加速寻找最近的聚类中心以提高聚类效率。Nister等<sup>[18]</sup>提出了层次 K-means(Hierarchical K-means, HKM),将时间复杂度降为  $O(nd \log k)$ ,但是该方法忽略了特征维数  $d$  对聚类效率的影响。为此,研究者们提出基于降维的聚类方法,如主成分分析(Principal component analysis, PCA)<sup>[19]</sup>、自组织特征映射(Self-organizing feature map, SOFM)<sup>[20]</sup>等,主要思路是利用降维算法对高维特征数据进行降维,再用聚类算法对降维后的特征点进行聚类。为克服聚类中心初始化对聚类结果的不利影响,Frey等<sup>[21]</sup>提出近邻传播聚类方法(Affinity propagation clustering, APC),APC 算法可在其迭代过程中自动确定聚类数目  $k$ ,并不断搜索合适的聚类中心,但对于比较松散的聚类结构,该算法生成的聚类中心数目过多,从而不能给出准确的聚类结果。此外,还有研究者提出了核聚类方法<sup>[22,23]</sup>,利用核函数将输入数据映射到高维空间中,使得输入数据中未显现的特征突显出来,从而能够在高维空间中进行更好地聚类。上述聚类方法在一定程度上提高了聚类效率,当特征点数目  $n$  较大时仍有较好的表现,但是当  $n$  和  $k$  均很大时,聚类效率会急剧下降,这严重限制了其在大规模数据集上的应用。

综上所述,为增强视觉特征的图像内容表达能力,提高图像聚类效率,本文提出一种基于 CNN 和二进制 K-means 的图像快速聚类方法。首先,引入 CNN 对训练图像进行学习,利用其特殊的网络结构隐

式地学习得到图像数据的高阶表示,生成具有更强区分性和表达能力的深层特征;然后,引入主成分分析迭代量化方法(Principal component analysis iterative quantization, PCA-ITQ)方法<sup>[24]</sup>对深层特征进行哈希编码,降低特征维数,利用多索引哈希技术对聚类中心构造索引目录,加速寻找最近的聚类中心;最后,利用二进制 K-means 对哈希码进行聚类,完成图像快速聚类。

## 1 基于卷积神经网络的图像深层特征提取

CNN 由 Fukushima<sup>[25]</sup> 首次提出,是第一个真正成功训练多层网络结构的学习算法,并被广泛应用于解决如何提取学习图像数据的深层特征问题<sup>[13,14]</sup>。其基本思想是将图像的局部感知区域作为网络的输入,信息再依次传输到不同的层,每层通过一个数字滤波器去获取对平移、旋转和缩放具有不变性的显著特征。用于图像深层特征提取的 CNN 结构<sup>[26]</sup> 如图 1 所示。

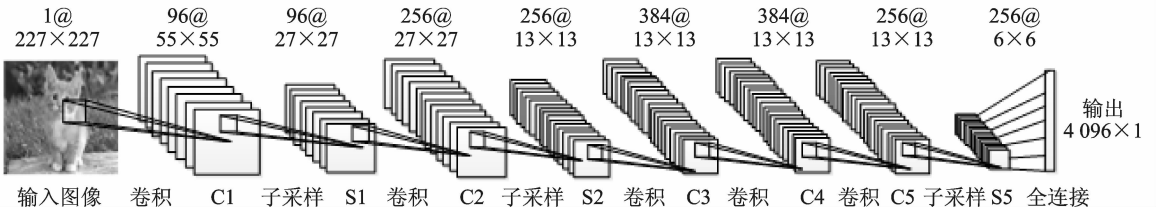


图 1 用于图像深层特征提取的卷积神经网络结构图

Fig. 1 Illustration of the architecture of CNN used for image deep feature extraction

从图 1 可以看出, CNN 是由多个卷积层(C层)和子采样层(S层)组成的多层深度神经网络,卷积层后一般都会跟着子采样层,每层都是由多个二维平面组成。该 CNN 的输入图像大小为  $227 \times 227$ , 输出为  $4096 \times 1$  的图像深层特征,一共包含 5 个卷积层和 5 个子采样层。在卷积层,前一层的特征图  $x_i^{(l-1)}$  与可学习的卷积核  $K_{ij}$  进行卷积,卷积的结果经非线性函数  $g(\cdot)$  生成这一层的特征图  $y_j^{(l)}$ , 具体为

$$y_j^{(l)} = g\left(\sum_{i \in M_j} K_{ij} \otimes x_i^{(l-1)} + b_j\right) \quad (1)$$

式中:  $y_j^{(l)}$  为第  $l$  个卷积层  $C_l$  的输出,  $\otimes$  代表卷积运算,  $b_j$  为偏置,卷积核  $K_{ij}$  可与前一层的一个或多个特征图确定卷积关系,  $M_j$  代表输入特征图集合,常用的非线性函数有  $g(x) = \tanh(x)$  和  $g(x) = (1 + e^{-x})^{-1}$ , 与上述非线性函数相比,  $g(x) = \max(0, x)$  能有效提高训练效率<sup>[27]</sup>。卷积层生成的特征图大小  $h_l$  为

$$h_l = \frac{h_{l-1} + 2 \times \rho_l - z_l}{\lambda_l} + 1 \quad (2)$$

式中:  $h_{l-1}$  为第  $l-1$  层特征图的大小,  $z_l$  表示第  $l$  层卷积核的大小,  $\lambda_l$  表示卷积核移动步长,  $\rho_l$  表示卷积运算时对前一层特征图边缘补零的列数。在子采样层,文献[23]研究表明相对于传统的无重叠采样,使用重叠采样不仅能提高特征的准确性,还可以防止训练阶段出现过拟合。因此,这里采用重叠采样方法对特征图进行最大值采样,采样区域为  $3 \times 3$ , 采样步长为 2 个像素。

卷积神经网络的训练主要分前向传播和后向传播两个阶段:

(1) 前向传播阶段。从训练样本中选取一个样本  $(X, Y_p)$ ,  $X$  从输入层经逐级变换传送到输出层,计算相应的实际输出,即

$$O_p = F_n(\cdots(F_2(F_1(X_p W^{(1)}) W^{(2)}) \cdots) W^{(n)}) \quad (3)$$

(2) 后向传播阶段,也称误差传播阶段。计算实际输出  $O_p$  与对应理想输出  $Y_p$  的误差为

$$E_p = \frac{1}{2} \sum_{j=1}^m (y_{pj} - o_{pj})^2 \quad (4)$$

将误差  $E_p$  反向逐层后推得到各层的误差,并按最小化误差方法调整神经元权值,当总误差  $E \leq \epsilon$  时,完成该批次训练样本的训练。当所有批次训练完成后,将图像输入卷积神经网络中,图像数据逐级通过各个网络层后,在输出端即可得到图像的深层特征。

## 2 二进制 K-means 快速聚类算法

由于通过 CNN 得到图像的深层特征维数较高,而且数据类型为浮点型小数,直接对其进行聚类操作会消耗大量的存储空间和时间,因此本文采用二进制 K-means (Binary K-means, BK-means) 方法<sup>[28]</sup>实现对图像特征的快速聚类,其基本思想是利用 PCA-ITQ 方法<sup>[24]</sup>保持高维图像特征的相似性,并映射为二进制哈希码  $H = \{\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n\} \in \{-1, +1\}^q$ ,用降维后的哈希码作为图像的特征表示实现快速聚类,得到  $k$  个聚类中心  $C = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k\}$ 。

PCA-ITQ 方法先对原始空间的数据集进行 PCA 降维处理,再通过对降维后的空间进行旋转,将数据点映射到二进制超立方的顶点上,使得对应的量化误差最小,从而将高维图像特征映射到低维汉明空间中,同时保持了原始特征之间的相似性,能有效降低后续聚类算法的计算复杂度。利用 PCA-ITQ 方法生成紧致的二进制哈希码后,针对二进制数据设计目标函数为

$$\begin{aligned} \min_{\mathbf{c}_j} \sum_i^n \sum_j^k \|\mathbf{h}_i - \mathbf{c}_j\|_2^2 \\ \text{s. t. } \mathbf{c}_j \in \{-1, +1\} \end{aligned} \quad (5)$$

式(5)类似于 K-means 算法的目标函数,可采用最大期望 (Expectation maximization, EM) 算法进行求解。当所有的图像都已分配到最近的聚类中心后,仅考虑聚类中心  $\mathbf{c}_j$  时,可将目标函数变换为

$$\min_{\mathbf{c}_j} \sum_i^p \|\mathbf{h}_i - \mathbf{c}_j\|_2^2 = \sum_i^p \|\mathbf{h}_i\|_2^2 + \sum_i^p \|\mathbf{c}_j\|_2^2 - \sum_i^p \mathbf{h}_i \mathbf{c}_j^T \quad (6)$$

式中:由于  $\mathbf{h}_i$  和  $\mathbf{c}_j$  均为二进制数,因此  $\sum_i^p \|\mathbf{h}_i\|_2^2$  和  $\sum_i^p \|\mathbf{c}_j\|_2^2$  均为常量,则式(6)可由最小化问题等价转换为最大化问题,即

$$\begin{aligned} \max_{\mathbf{c}_j} \sum_i^p \mathbf{h}_i \mathbf{c}_j^T = \max_{\mathbf{c}_j} \left( \sum_i^p \mathbf{h}_i \right) \mathbf{c}_j^T \\ \text{s. t. } \mathbf{c}_j \in \{-1, +1\} \end{aligned} \quad (7)$$

先计算式(7)中属于聚类中心  $\mathbf{c}_j$  的所有样本  $\mathbf{h}_i$  的连和  $\mathbf{m}_j = \sum_i^p \mathbf{h}_i$ ,可得  $\mathbf{c}_j$  解析形式的最优解为

$$c_{j\ell} = \text{sign}(m_{j\ell}) = \begin{cases} +1 & m_{j\ell} \geq 0 \\ -1 & m_{j\ell} < 0 \end{cases} \quad (8)$$

式中: $q$  为样本  $\mathbf{h}_i$  的维数。更新聚类中心  $\mathbf{c}_j$ ,BK-means 快速聚类算法的具体实施步骤如下所示。

输入:图像深层特征集  $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ ,聚类中心个数  $k$ ,哈希码位数  $q$ 。

预处理:利用 ITQ-PCA 方法将图像深层特征映射生成二进制哈希码,得到  $H = \{\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n\} \in \{-1, +1\}^q$ 。

训练:初始化:从  $H$  中随机选择  $k$  个样本,构成聚类中心的初始值  $C^{(0)} = \{\mathbf{c}_1^{(0)}, \mathbf{c}_2^{(0)}, \dots, \mathbf{c}_k^{(0)}\}$ ,设置最大迭代次数 maxIter;

for  $i=1, \dots, \text{maxIter}$

//第一步:

根据文献[28]中的多索引哈希方法为聚类中心集  $C^{(i-1)}$  建立多索引哈希表  $M$ ;

for  $j=1, \dots, n$

$\text{idx}_j = M. \text{Lookup1NN}(\mathbf{h}_j)$ ;

end for

//第二步:更新聚类中心集  $C$

for  $j=1, \dots, k$

    计算聚类中心  $\mathbf{c}_j^{(i-1)}$  中样本的连和  $\mathbf{m}_j$ ;

    更新聚类中心  $\mathbf{c}_j^{(i)} = \text{sign}(\mathbf{m}_j)$ ;

end for

更新聚类中心  $C^{(i)} = \{\mathbf{c}_1^{(i)}, \mathbf{c}_2^{(i)}, \dots, \mathbf{c}_k^{(i)}\}$

end for

输出:聚类中心集  $C = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k\}$ , 图像所属类别  $ID = \{\text{id}x_1, \text{id}x_2, \dots, \text{id}x_n\}$

### 3 基于卷积神经网络和二进制 K-means 的图像快速聚类方法

基于 CNN 和 BK-means 的图像快速聚类方法流程图如图 2 所示。首先,利用训练图像集对 CNN 参数进行训练,训练完成后提取图像深层特征;然后,引入 PCA-ITQ 算法对深层特征进行哈希编码,并利用 BK-means 算法对二进制哈希码进行聚类得到聚类中心;最后,通过对聚类中心构造多索引哈希表快速查找得到图像所属类别  $id$ ,完成图像的快速聚类。

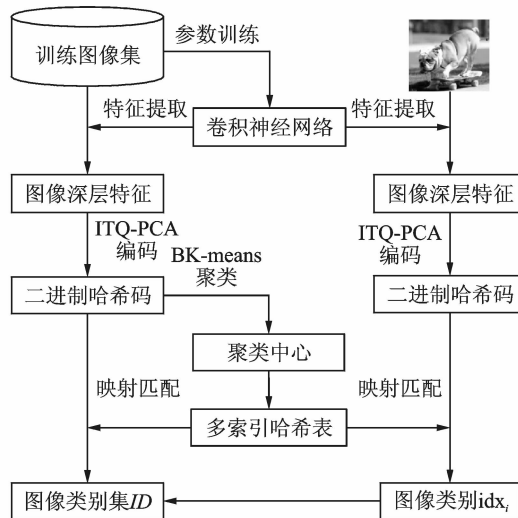


图 2 基于卷积神经网络和二进制 K-means 的图像快速聚类方法流程图

Fig. 2 Flow chart of fast image clustering based on CNN and Binary K-means

## 4 实验设置与性能评价

### 4.1 实验设置

为验证本文方法的有效性,在 ImageNet-1000 图像集<sup>[30]</sup>上对其进行了评估。ImageNet-1000 图像集是由 Li 主持构建的大规模图像数据库,其类别按照 WordNet 构建,包含 1 000 个类别,每个类别中至少包含 1 000 张图像,共计 120 万张图像。实验采用具有 5 层卷积层和 5 层子采样层结构的卷积神经网络,其结构如图 1 所示。各层卷积核大小  $Z = \{z_1 = 11, z_2 = 5, z_3 = z_4 = z_5 = 3\}$ , 移动步长  $\Delta = \{\lambda_1 = 4, \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 1\}$ , 特征图边缘补零列数  $P = \{\rho_1 = 0, \rho_2 = 2, \rho_3 = \rho_4 = \rho_5 = 1\}$ , 学习速率为 0.01, 子采样层采用重叠采样方法对特征图进行最大值采样,采样区域为  $3 \times 3$ , 采样步长为 2 个像素。

实验硬件配置是内存分别为 6 GB 和 16 GB 的 GPU 设备 GTX Titan 和 Intel Xeon CPU 服务器。

图像聚类性能指标采用平均查准率均值(Mean average precision, MAP)和平均聚类时间,分别定义为

$$\text{MAP} = \frac{\sum_{i=1}^k \frac{n_{i\max}}{N_i}}{k} \times 100\% \quad (9)$$

$$\text{平均聚类时间} = \frac{\text{多次聚类总时间}}{\text{聚类次数}} \quad (10)$$

式中: $k$ 表示聚类中心数目, $c$ 为图像类别数, $N_i$ 为分配到第*i*个聚类中心的图像数目, $n_{i\max}$ 为 $N_i$ 中每个类别图像数目的最大值。

## 4.2 实验结果分析

为验证 CNN 生成特征的有效性,将其与当前的一些主流图像特征在 ImageNet-1000 图像集上进行了实验比较,包括 GIST 特征<sup>[1]</sup>、SIFT 特征<sup>[2]</sup>和 SURF 特征<sup>[3]</sup>。首先从 ImageNet-1000 图像集中随机选取 50 类,100 类,500 类以及全部 1 000 类图像;然后,对这些图像分别提取 GIST 特征、SIFT 特征、SURF 特征和 CNN 深层特征;最后,利用 K-means 对特征点分别以  $k=c, 10c$  ( $k$ 为聚类中心数目, $c$ 为图像类别数)进行聚类,得到的实验结果如图 3 所示。

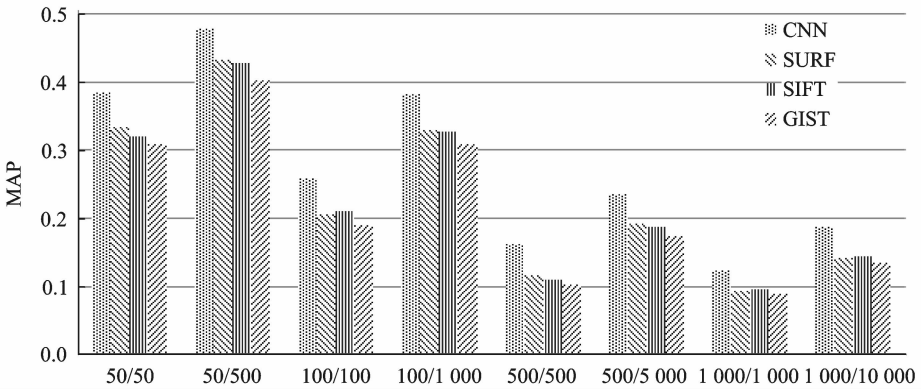


图 3 不同图像特征在 ImageNet-1000 上聚类准确率的对比

Fig. 3 Mean average precision comparison of different image features on ImageNet-1000 database

从图 3 可以看出,当图像类别较少时,利用各个特征进行聚类均能取得较高的 MAP 值;随着图像类别的增加,各特征取得的 MAP 值逐渐下降,而且在同一数据集下, $k=10c$  比  $k=c$  时的 MAP 高。实验结果表明:基于 CNN 提取深层特征进行聚类的 MAP 值明显高于其他方法,说明利用 CNN 提取的图像深层特征具有更强的区分性和表达能力。这是因为 GIST、SIFT 和 SURF 特征提取步骤固定,不具有自主学习能力,适应性不强,使得其图像表达能力受限,而 CNN 模拟大脑皮层对输入数据进行分层处理,用每一层神经网络提取原始数据不同水平的特征,使得在神经网络较高层中学习得到的特征更能体现输入数据的本质信息,而且其深层的网络结构能有效地挖掘图像内在隐含关系,从而增强了特征的图像表达能力。

为验证 BK-means 快速聚类算法的有效性,实验将 BK-means 算法与 K-means<sup>[4]</sup>、K-medoids<sup>[5]</sup>和 KDK-means<sup>[17]</sup>在对 ImageNet-1000 图像集上的聚类准确率和时间消耗进行了对比,结果分别如图 4 和表 1 所示,其中输入数据为图像深层特征经 PCA-ITQ 编码后的 256 bit 哈希码。

从图 4 中不难看出,K-means 算法的 MAP 值在各个情况下均是最高的,其他算法都是在 K-means 算法的基础上放宽了某些限制,以聚类准确率的少量降低换取聚类速率的提高。BK-means 的 MAP 值仅次于 K-means,仅仅低 1%~2%,优于 KDK-means 和 K-medoids;KDK-means 在聚类中心较少时取

得较好的聚类效果,随着图像类别的增加 KDK-means 聚类性能下降最快;K-medoids 算法严格要求选取原始数据点作为聚类中心,与实际中心会有偏离,导致聚类准确率下降。

对比表 1 可知,K-means 算法的聚类时间最长,当图像数据较少时,如 50 类图像/50 个聚类中心,聚类时间与其他算法相当,但是随着图像类别的增加,K-means 算法的聚类时间急剧增长,而 BK-means、KDK-means 和 K-medoids 聚类时间明显少于 K-means。K-medoids 以二进制的数据点为聚类中心,以 Hamming 距离作为距离度量,可减少距离计算的时间开销,提高了聚类效率,聚类时间远少于 K-means,但是该算法更新聚类中心时需要计算该类中数据点到其他所有点的距离,当类内数据点较多时,聚类时间较长,因此 K-medoids 算法不适用于类内数据较多的聚类。KDK-means 算法利用 KD-tree 对聚类中心构建索引,加快了最近聚类中心寻找过程,从而提高了聚类效率,但随着图像数据的增加,其聚类准确率下降最快,难以适应大数据环境。当聚类中心较少时(如 50 类图像/50 个聚类中心),由于建立多索引哈希表增加了时间消耗,使得 BK-means 聚类时间与 K-means 相当,随着聚类中心的增加,K-means 算法聚类时间急剧增长,而 BK-means 聚类时间随着图像数据线性增长,聚类时间远少于 K-means。结合图 4 和表 1 数据可以看出,BK-means 聚类准确率与 K-means 相当(仅仅低 1%~2%),而聚类效率相比于 K-means 有了大幅度提高,且性能优于 K-medoids 和 KDK-means 算法。

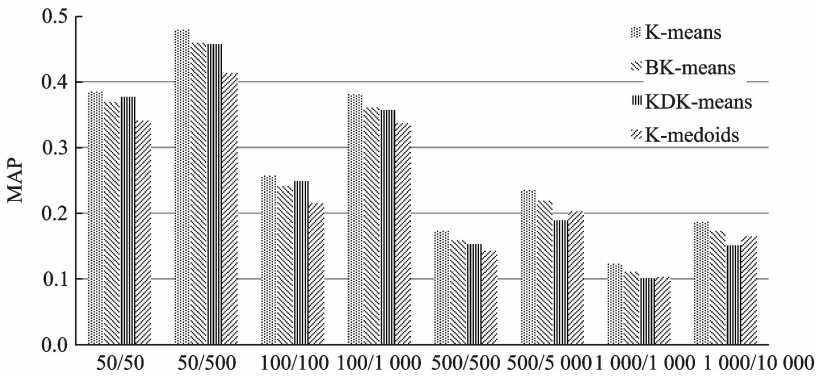


图 4 不同方法在 ImageNet-1000 上聚类准确率的对比

Fig. 4 Mean average precision comparison of different methods on ImageNet-1000 database

表 1 不同方法在数据集 ImageNet-1000 的聚类时间对比

Tab. 1 Image clustering times of different methods on ImageNet-1000 database

s

方法	50/50	50/500	100/100	100/1 000	500/500	500/5 000	1 000/1 000	1 000/10 000
K-means	5.4	20.3	15.5	71.1	206.4	1 654.8	709.7	6 085.1
K-medoids	8.1	4.8	17.3	10.9	108.8	56.7	213.5	109.4
KDK-means	6.3	6.7	13.4	13.6	68.2	68.8	138.6	141.3
BK-means	5.7	5.8	12.1	12.4	59.6	60.3	123.2	124.8

为进一步验证本文方法(记为 CNN+BK-means)的有效性,将其与 K-means 方法<sup>[4]</sup>、文献[31]中的 AP 方法以及文献[32]中的谱聚类(Spectral clustering, SC)方法在 ImageNet-1000 图像集上进行了实验对比,得到的实验结果如表 2 所示。从表 2 可以看出,CNN+BK-means 方法的 MAP 值均高于其他主流方法,而且聚类时间最短;SC 和 AP 方法的 MAP 值略低于 CNN+BK-means;K-means 的 MAP 值最低,但其聚类时间仅高于 CNN+BK-means,远少于 SC 和 AP 方法,其中 AP 方法的聚类时间最长,随着图像数据的增加,SC 和 AP 方法的聚类时间急剧增长,当图像数据量较大时,SC 和 AP 方法因内存消耗

过大、聚类时间太长无法完成聚类。K-means 方法通过多次迭代近邻计算寻找最近聚类中心,对初始聚类中心的选择敏感且容易陷入局部极值,算法复杂度为  $O(nkd)$ ,尤其当  $n$  和  $k$  的值均较大时,需要消耗大量的计算时间和存储空间;SC 方法构造一个描述图像之间相似度的矩阵,通过矩阵谱分析理论导出聚类对象的新特征,利用新特征进行聚类不易陷入局部最优解,有效提高了聚类准确性,但需要求解相似矩阵的特征值与特征向量,算法复杂度为  $O(n^2d+n^3+nk^2)$ ,随着图像数量  $n$  的增加,聚类时间急剧增加,难以对大规模数据集聚类;AP 方法为使聚类结果不受初始点选择的影响,将所有图像作为候选的聚类中心,但由于在迭代过程中需要计算相似矩阵、信任矩阵和响应度矩阵,因此计算复杂度较高,无法对大规模数据进行有效聚类;CNN+BK-means 方法通过 CNN 对 ImageNet-1000 图像集进行学习,不需要任何精确数学表达式就能学习得到大量输入与输出之间的映射关系,能够有效挖掘图像的内在隐含关系,增强特征的图像表达能力,而且 BK-means 利用哈希方法将图像高维特征映射为低维二进制哈希码,同时对聚类中心构建多索引哈希表,加速寻找最近的聚类中心,将时间复杂度从  $O(nkd)$  降低到  $O(nq)$ ,使得 CNN+BK-means 优于当前主流方法。最后,图 5 给出了本文方法在 ImageNet-1000 图像集上的部分聚类结果。

表 2 不同方法在数据集 ImageNet-1000 的聚类性能对比

Tab. 2 Comparison of image clustering results for different methods on ImageNet-1000 database

方法	K-means		Spectral Clustering(SC)		Affinity Propagation(AP)		CNN+BK-means	
	MAP/%	Time/s	MAP/%	Time/s	MAP/%	Time/s	MAP/%	Time/s
50/50	31.63	236	35.23	9 681	35.61	436 335	36.82	212
50/500	41.51	848	43.48	21 677	44.56	1 425 378	45.77	219
100/100	20.21	732	22.14	36 758	—	—	25.16	457
100/1 000	31.74	3 265	33.97	114 042	—	—	36.04	471
500/500	10.89	10 256	—	—	—	—	15.86	2 367
500/5 000	18.32	87 625	—	—	—	—	21.78	2 439
1 000/1 000	9.23	34 722	—	—	—	—	11.08	4 804
1 000/10 000	14.42	301 438	—	—	—	—	17.23	4 926

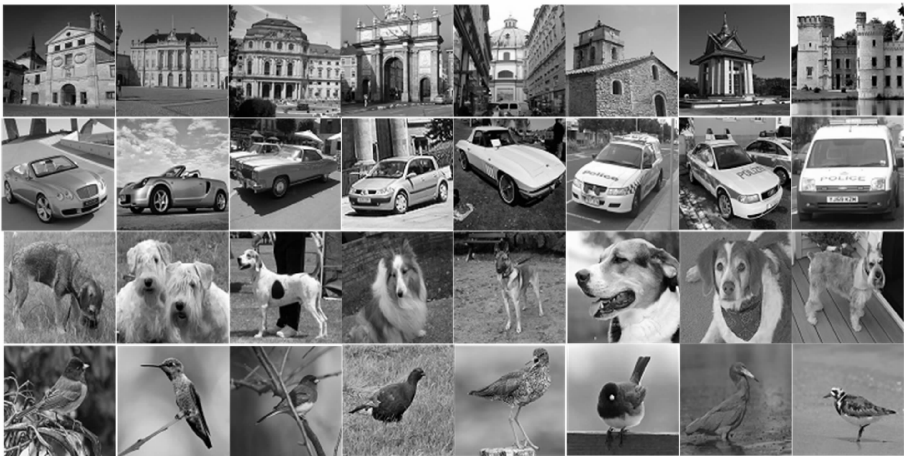


图 5 CNN+BK-means 在 ImageNet-1000 图像集上的部分聚类结果

Fig. 5 Some sample clustering results of CNN+BK-means on ImageNet-1000 database



## 5 结束语

针对当前图像特征图像表达能力不强、聚类算法计算复杂度较高和聚类效率低等问题,本文提出了一种基于 CNN 和 BK-means 的图像快速聚类方法。首先,引入卷积神经网络模仿大脑皮层对图像数据分层处理,充分挖掘图像内容的内在隐含关系,并得到图像数据的高阶特征表示,有效增强了图像特征的区别性和表达能力;然后,引入 BK-means 方法,通过 ITQ-PCA 将高维图像特征映射为低维二进制哈希码,利用二进制数据的特性提出简单有效的目标函数,并对聚类中心构建多索引哈希表加速寻优过程,实现了对大规模图像集的快速聚类。实验结果表明本文方法能有效增强特征的图像表达能力、提高图像聚类效率、性能优于当前主流方法。如何加快 CNN 的训练速度,有效防止出现过拟合现象是本文的下一步研究方向。此外,如何降低初始聚类中心选择对聚类结果的影响也是今后亟待解决的问题。

### 参考文献:

- [1] Oliva A, Torralba A. Modeling the shape of the scene a holistic representation of the spatial envelope [J]. *International Journal in Computer Vision*, 2001, 42(3):145-175.
- [2] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [3] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features [J]. *Computer Vision and Image Understanding*, 2008, 110(3): 346-359.
- [4] Hartigan J A, Wong M A. A K-means clustering algorithm[J]. *Applied Statistics*, 1979,28(1):100-108.
- [5] Park H S, Jun C H. A simple and fast algorithm for K-medoids clustering[J]. *Expert Systems with Applications*, 2009, 36(2): 3336-3341.
- [6] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets[J]. *Neural Computation*, 2006, 18(7): 1527-1539.
- [7] Bengio Y, Lamblin P, Dan P, et al. Greedy layer-wise training of deep networks[C]// *International Conference on Neural Information Processing Systems*. [S.l.]: ACM, 2006,19:153-160.
- [8] Hinton G E. To recognize shapes, first learn to generate images[J]. *Progress in Brain Research*, 2007, 165: 539-547.
- [9] Tang Yichuan, Eliasmith C. Deep networks for robust visual recognition[C]// *Proceedings of the 27th International Conference on Machine Learning*. Haifa, Israel: ACM Press,2010:1055-1062.
- [10] Nair V, Hinton G E. 3D object recognition with deep belief nets[C]// *Proceedings of the 24th Annual Conference on Neural Information Processing Systems*. Vancouver, Canada: ACM Press, 2009:1339-1347.
- [11] Bu Shuhui, Liu Zhenbao, Han Junwei, et al. Learning high-level feature by deep belief networks for 3-D model retrieval and recognition[J]. *IEEE Transactions on Multimedia*,2014,16(8):2154-2167.
- [12] Zhang F, Du B, Zhang L. Saliency-guided unsupervised feature learning for scene classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 53(4): 2175-2184.
- [13] Lee H, Grosse R, Ranganath R, et al. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations[C]// *Proceedings of the 26th International Conference on Machine Learning*. New York: ACM Press, 2009: 609-616.
- [14] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015,37(9):1-14.
- [15] 杨阳, 张文生. 基于深度学习的图像自动标注算法[J]. *数据采集与处理*, 2015, 30(1): 88-98.  
Yang Yang, Zhang Wensheng. Image auto-annotation based on deep learning[J]. *Journal of Data Acquisition and Processing*,2015,30(1): 88-98.
- [16] Ilya V, Vladimir P. Curse of dimensionality in pivot-based indexes[C]//*Proceedings of International Workshop on Similarity Search and Applications*. Ottawa, Canada:IEEE Computer Society Press, 2009:39-46.
- [17] Philbin J, Chum O,Isard M, et al. Object retrieval with large vocabularies and fast spatial matching[C]// *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Minneapolis, USA: IEEE Computer Society Press, 2007:1-8.

- [18] Nister D, Stewenius H. Scalable recognition with a vocabulary tree[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE Computer Society Press, 2006:2161-2168.
- [19] Goes J, Zhang T, Arora R, et al. Robust stochastic principal component analysis[C]// Proceedings of the 17th International Conference on Artificial Intelligence and Statistics. Reykjavik, Iceland: IEEE Computer Society Press, 2014: 266-274.
- [20] Goswami A K, Jain R, Tripathi P. Automatic segmentation of satellite image using self organizing feature map (SOFM): An artificial neural network (ANN) approach[J]. International Journal of Advanced Research in Computer Science, 2014, 5 (8):92-97.
- [21] Frey B J, Dueck D. Clustering by passing messages between data points[J]. Science, 2007, 315(5814): 972-976.
- [22] Mark G. Mercer kernel-based clustering in feature space[J]. IEEE Transactions on Neural Networks, 2002, 13(3): 780-784.
- [23] Lu Y, Wang L, Lu J, et al. Multiple kernel clustering based on centered kernel alignment[J]. Pattern Recognition, 2014, 47(11): 3656-3664.
- [24] Gong Y C, 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.
- [25] Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position[J]. Biological Cybernetics, 1980, 36(4):193-202.
- [26] Alex K, Ilya S, Geoffrey E H. ImageNet classification with deep convolutional neural networks [C]// Proceedings of Advances in Neural Information Processing Systems. Lake Tahoe, USA: ACM Press, 2012: 1106-1114.
- [27] Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines[C]// Proceedings of International Conference on Machine Learning. Haifa, Israel: ACM Press,2010:807-814.
- [28] Gong Y, Pawlowski M, Yang F, et al. Web scale photo hash clustering on a single machine[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015: 19-27.
- [29] Norouzi M, Punjani A, Fleet D J. Fast search in hamming space with multi-index hashing[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE Computer Society Press, 2012: 3108-3115.
- [30] Jia D, Wei D, Richard S, et al. ImageNet: A large-scale hierarchical image database[C]// Proceedings of Computer Vision and Pattern Recognition. Miami, USA: IEEE Computer Society Press, 2009:248-255.
- [31] Zhang Y, Zhang H. Image clustering based on SIFT-affinity propagation[C]// International Conference on Fuzzy Systems and Knowledge Discovery (FSKD). Xiamen, China: IEEE Computer Society Press, 2014: 358-362.
- [32] Zbib H, Mouysset S, Stute S, et al. Unsupervised spectral clustering for segmentation of dynamic PET images[J]. IEEE Transactions on Nuclear Science, 2015,62(3):840-850.

## 作者简介:



**柯圣财** (1991-), 男, 硕士研究生, 研究方向: 图像理解与分析、图像分类及图像检索, E-mail: keshengcai0705@163.com。



**李弼程** (1970-), 通信作者, 男, 教授, 博士生导师, 研究方向: 智能信息处理, E-mail: lbclm@163.com。



**唐永旺** (1981-), 男, 讲师, 研究方向: 智能信息处理。



**吴志兵** (1967-), 女, 高级工程师, 研究方向: 大数据分析 & 处理、高性能计算。



**万建平** (1987-), 男, 工程师, 研究方向: 光学图像处理。

