

MGTL-SAE 精细特征学习的图像资源快速检索*

高蕴梅 (常熟理工学院图书馆 江苏 215500)

摘要: [目的/意义] 数字图书馆图像资源检索服务需要准确和快速的响应。针对现有深度学习图像检索方法存在特征不够精细和检索速度慢等问题,提出了多组双层稀疏自动编码器(Multi-Group-Two-Layer Sparse Auto-Encoder, MGTL-SAE)精细特征图像资源快速检索方法。[方法/过程] 首先,GoogLeNet 预处理模型提取图像资源深度学习特征作为初始特征,并用密度峰值聚类算法划分初始特征集。其次,依据聚类类别数构建 MGTL-SAE 网络结构,聚类划分图像子集中随机抽取图像块初始化 MGTL-SAE 的均值和方差参数,梯度下降法估计 MGTL-SAE 网络参数以学习特征直到目标函数收敛,迭代密度峰值聚类算法和 MGTL-SAE 参数估计以计算精细特征直到特征基收敛。最后,在图像资源精细特征密度峰值聚类划分子集上快速检索查询样本图像。[结果/结论] 实验结果表明, MGTL-SAE 方法相较于 CaffeNet 和 GoogLeNet 等深度学习检索方法具有更高查准率、查全率和平均 mAP, 更快检索速度,为数字图书馆图像资源检索研究提供了新视角。

关键词: 基于内容的图像检索 稀疏自动编码 深度卷积神经网络 目标函数 精细特征学习

Quick Retrieval of Image Resources Based on MGTL-SAE Fine Feature Learning

Gao Yunmei (Library of Changshu Institute of Technology, Jiangsu, 215500)

Abstract: [Purpose /significance] Image resource retrieval for digital library needs an accurate and fast response. However, traditional image retrieval methods based on deep learning have the problems of long time for training model, unrefined feature and slow retrieval speed. We proposed Multi-Group-Two-Layer Sparse Auto-Encoder (MGTL-SAE) to learn refine features for fast image resource retrieval. [Method /process] Firstly, extracted deep learning features of the image resources with the pre-train model GoogLeNet as initial features which were clustered with the density peak(DP) clustering algorithm. Secondly, designed the network of MGTL-SAE using the number of classes of clusters, initialized the mean and variance parameters of MGTL-SAE with the image blocks selecting from the images according to the image resource sub-sets partitioned by the clusters, estimate the parameters of MGTL-SAE to learn features by the gradient descent algorithm until the cost function is converged, iterate the DP clustering algorithm and estimating the parameters of MGTL-SAE to compute the refined feature until the feature bases are converged. Lastly, retrieved the query sample image fast within the sub-set of image resources partitioned by the DP clusters of refined features.[Result /conclusion] Experiment results show that our method has higher precision, recall and mAP(mean Average Precision) and less retrieval time than the methods based on CaffeNet and GoogLeNet . It also provides a new perspective for the research of image resource retrieval in the digital library.

Keywords: content based image retrieval sparse auto-encoder deep convolutional neural network cost function refine feature learning

1 引言

随着图像采集以及存储技术的不断发展,数字图

书馆的图像资源正在快速增长,大量数字出版物以及珍贵资料以图片、视频等形式被收藏。国内外的公共和高校图书馆越来越重视图像资源建设,如美国数字

*本文系国家自然科学基金项目“基于多目标优化和栈式稀疏编码的肝脏肿瘤图像识别研究”(项目编号:61572239)、教育部人文社会科学研究青年基金项目“学术信息网络中的核心边缘结构测度研究”(项目编号:18YJC870011)和国家级大学生实践创新训练计划项目“基于 fasttext 深度学习的专利文本快速分类研究”(项目编号:201910333028)的研究成果。

图书馆、中国中医科学院图书馆、南京图书馆、首都图书馆和其他高校图书馆^[1-5]分别建立了多媒体图像资源集、古籍数字化图像集、地域文化或人物图像数据集、特色或专题图像数据集等。此外,近年来数字人文研究和实践积累了海量图片、视频、3D模型等视觉数据集^[6]。综上,图像资源检索已经成为数字图书馆服务的重要内容。

在中国知网(CNKI)数据库中,以主题=图像检索,选择重要期刊(EI收录、中文核心或CSSCI收录),发表时间不限制,于2019年10月12日检索到2615篇论文。在Web of Science(WoS)核心数据库,主题=image retrieval,发表时间不限制,于2019年10月12日检索到12770篇。国外最早发文在1963年,国内最早发文在1992年。萌芽阶段,国外(1963—1990年)和国内(1992—1998年)每年发文数量较少,2000年以前以基于文本的图像检索方法为主。高速发展阶段,国外在1992—2011年期间每年发文量从36篇增长到549篇,国内在1999—2005年期间每年发文量从10余篇增长到180余篇,从2006到2012年逐步回落到120余篇,从2000—2015年期间,基于人工定义和选择特征内容的图像检索是主流方法。国外自2012年开始,国内从2016年,逐步转入以基于深度学习的图像检索方法为主。

基于文本的图像检索方法由人工标注图像的内容、特征和属性进行检索,这种方法简单直观、检索的准确率高且速度快,但标注图像信息需要投入大量人力和财力,费时费力、效率低、容易出错。手工定义和选择特征的图像内容检索方法主要包括颜色、纹理、形状、空间位置等内容描述的全局或局部特征,用户不需要对检索图像对象进行精确文字描述,比较适合实际应用。但是全局特征描述的粒度较粗,往往判别能力较弱,局部特征不具备图像内容的自学习能力、普适性较差、区分能力有限,底层特征和高层语义之间存在“语义鸿沟”。得益于大规模标注数据集和GPU的计算能力,深度学习方法通过模拟人脑感知视觉逐层抽象学习数据特征,从而发现数据内在的结构信息,目前已经成为解决“语义鸿沟”问题的有效解决方法。

基于深度学习的图像检索方法可以分为两类:一类是针对图像数据集设计网络结构并训练网络参数学习深度特征的检索方法;另外一类是微调大规模图像数据集的预训练模型或结合预训练模型和传统人工定义特征的方法。

第一类方法主要适合有大量标注数据和计算资源的图像检索应用。针对深度学习的图像检索方法存在有些样本的检索准确率为零和检索时响应时间过长的问題,王志明等^[7]提出融合神经网络不同层的信息作为图像特征表示,使用深度哈希的方法将图像特征映射为二进制的哈希码;针对无指导深度哈希不依赖标签具有良好收缩性,但松散哈希优化,缺乏语义指导而存在图像检索性能有限问题,Cui等^[8]提出了具有可伸缩的标签深度哈希的图像检索方法;针对手绘草图与自然彩色图片之间的视觉差异和语义差距问题,于邓等^[9]提出了分层跨域匹配框架和多层深度融合卷积神经网络的手绘草图检索。第一类方法可以学习有效特征,具有良好检索性能,但是需要标注大量样本和昂贵的计算资源,且学习时间长。大规模图像集的有指导深度学习网络比较困难且耗时,因此,无须标签数据的无指导深度学习可能成为研究热点^[10]。

第二类方法充分利用从海量图像数据中训练过的预处理模型检索图像。Tzelepi等^[11]在ImageNet图像集训练过的CaffeNet深度学习模型基础上,提出了完全无指导、利用数据集标签或部分数据集相关信息、基于相关反馈三种重新训练深度学习网络的方法,以适应内存和计算能力受限的无人机、智能手机或其他嵌入式系统环境下的图像检索。针对像素级图像检索,深度学习网络的倒数几层比最后一层特征具有更好的检索性能,Joe等^[12]提出了从OxfordNet和GoogLeNet网络的不同层中提取卷积特征,采用VLAD(Vector of Locally Aggregated Descriptors)编码特征,取得了较好的检索结果。Liu等^[13]提出了集成GoogLeNet和DDBTC(Dot-Diffused Block Truncation Coding)特征的方法。Ge等^[14]利用AlexNet、VGGM、VGG16和GoogLeNet等预训练CNNs描述迁移到高分辨率遥感图像提取CNN(Convolution Neural Network)图像特征并综合浅层特征,有效改善了图像检索的性能。Wu等^[15]提出了同时捕获底层颜色特征、纹理特征与高层语义信息的多层特征融合方法,利用类别信息训练神经网络,综合了浅层和超深层学习特征检索图像。

尽管基于DCNN(Deep CNN)的图像检索取得了阶段性成果,但还存在一些问题:直接微调DCNN预训练模型参数用于特定图像集检索的特征不够精细,而且降低了检索准确率。此外,线性顺序扫描海量、高维和动态的图像资源深度学习特征数据库的检索策略不能

满足实时性需求。

为了提高图像资源学习特征的精度和检索速度,本文提出了一种多组双层稀疏自动编码器(Multi-Group-Two-Layer Sparse Auto-Encoder, MGTL-SAE)精细特征图像资源快速检索方法。主要创新和贡献在于利用图像学习特征的密度峰值聚类把相似图像划分到一组,为每组图像集构造一个SAE,通过聚类压缩组内图像信息使得目标函数更容易收敛,聚类划分图像子集构建的MGTL-SAE可以学习精细特征。此外,在图像资源MGTL-SAE精细特征的密度峰值聚类分子集范围内检索查询样本图像,缩小了搜索范围,提高了检索速度。

2 MGTL-SAE精细特征学习的图像检索方法

充分利用图像深度卷积神经网络、密度峰值聚类、稀疏自动编码等前沿技术,本文提出了MGTL-SAE图像检索框架,主要处理流程如图1所示,包括GoogLeNet模型^[16]提取图像资源的初始特征、图像特征密度峰值聚类、MGTL-SAE网络结构设计和参数估计等模块。

2.1 GoogLeNet模型提取初始特征

GoogLeNet深度卷积神经网络结构是2014年ILSVRC(ImageNet Large Scale Visual Recognition Challenge)图像识别竞赛的冠军模型,广泛应用于医学图像诊断^[17]和遥感图像分类^[18]等。GoogLeNet包括卷积层、池化层和Inception模块等共计22层,是在涵盖2万多个类别共计1400多万幅图片的Imagenet大规模数据集

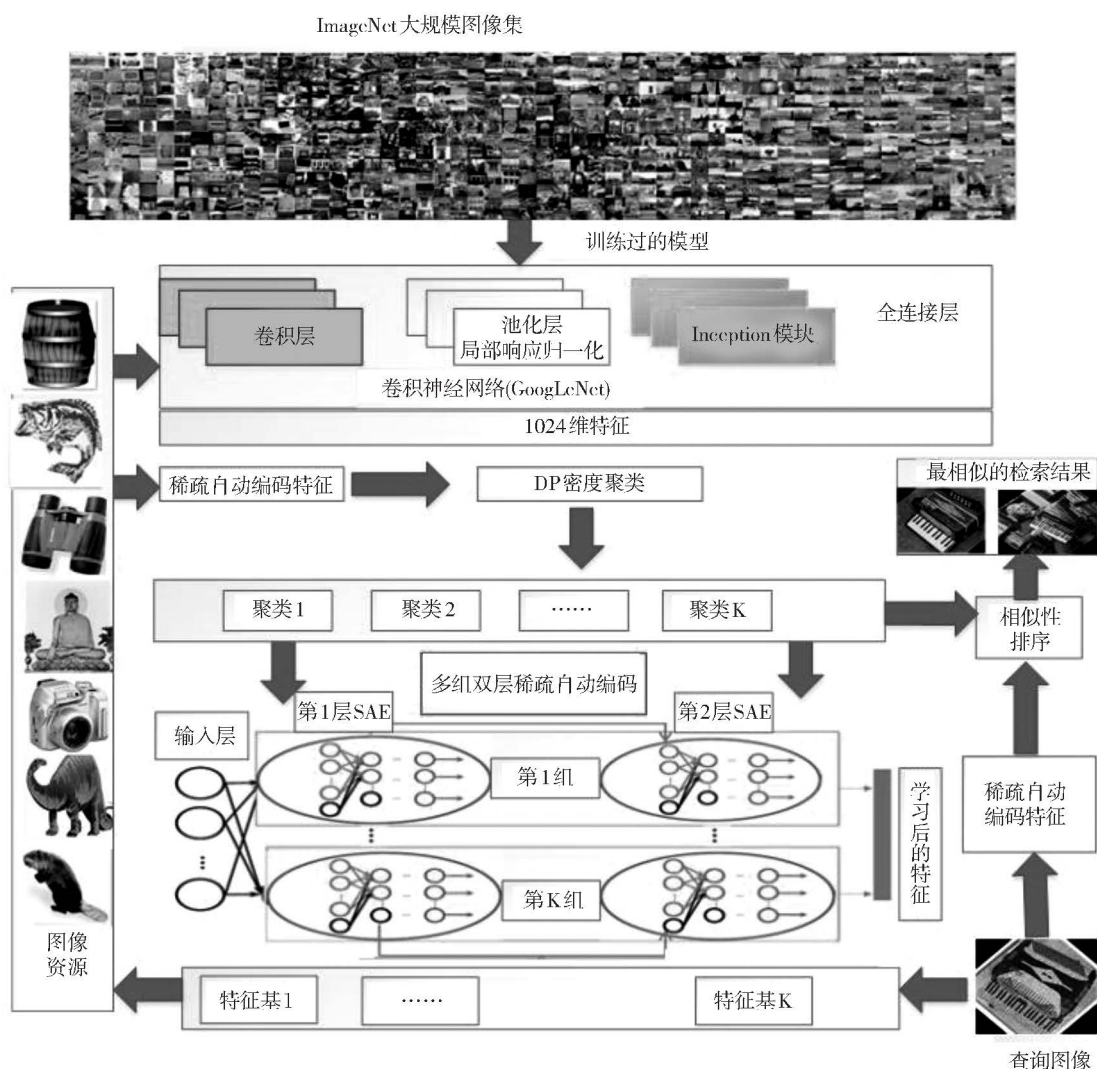


图1 MGTL-SAE图像检索方法框架

上训练过的深度学习模型。卷积层对输入图像不同位置的局部矩阵和卷积核矩阵权重对应位置的元素相乘之和作为输出图像对应位置的值。池化层采用向下采样策略从输入层某个位置的邻域中选取最大值、平均值或随机选像素值。Inception模块通过1×1的卷积降维,在多个尺寸上同时卷积再聚合。在相同计算量下提取到更多特征,提升训练结果。使用GoogLeNet模型从每幅图像资源中提取1024维深度学习特征。

2.2 图像特征密度峰值聚类

为了提取图像资源精细特征,本文利用密度峰值(Density Peaks, DP)聚类^[19]算法划分图像资源,为每个聚类划分图像资源子集单独建立一个稀疏自动编码器学习特征。DP聚类的基本思想是假设局部密度最大的数据点作为聚类划分的中心,每个中心吸引其周围密度较低的数据点,聚类划分中心点之间的距离相对较远。也就是说,满足数据点的局部最大密度以及与密度更大点之间的距离较远这两个条件确定聚类中心,从而得到任意形状的聚类划分^[20]。

假设图像集有 N 幅图像, X_1, X_2, L, X_N , GoogLeNet预训练模型从图像 X_i 中提取深度学习特征 $\zeta_i \in R^{1024} (1 \leq i \leq N)$ 作为初始特征。基于DP密度聚类思想,定义和计算每个数据点的局部密度和最小距离参数。

定义1 局部密度参数^[19],图像 X_i 的特征与其周围相邻图像 $X_j (1 \leq j \leq N)$ 的特征距离小于阈值 d_c 的图像数,计算公式为

$$\varphi_i = \sum_{j=1}^{n_i} x(d_{ij} - d_c) \quad (1)$$

其中函数 $x(x) = \begin{cases} 1 & x < 0 \\ 0 & x \geq 0 \end{cases}$, n_i 为相邻特征点个数, d_{ij} 表示特征 ζ_i 到 ζ_j 的欧氏距离。

定义2 最小距离参数^[19],图像 X_i 的特征到其他更高局部密度参数的图像 X_j 的特征最小距离,计算公式为

$$\delta_i = \min_{j: \varphi_j > \varphi_i} (d_{ij}) \quad (2)$$

定义3 聚类中心,在局部密度和最小距离的参数空间里,如果 $\varphi_i > \varphi_0$ 且 $\delta_i > \delta_0$,其中 φ_0 和 δ_0 分别为局部密度和最小距离阈值,对应的图像数据为聚类中心。定义3要求聚类中心的密度是局部最大,且聚类中心与其他密度更大的数据点之间的“距离”大于阈值。

尽管密度峰值聚类算法具有计算简单、可以发现任意形状的聚类划分,但是在计算局部密度和最小距

离时,阈值选取对聚类结果影响较大,目前有很多局部密度方法和网格化方法^[20]预处理,然后再使用密度峰值聚类算法中决策图的方法选取聚类中心。本文通过大量实验数据表明,聚类中心点距离局部密度参数轴和最小距离参数轴较远,非聚类中心数据距离参数轴较近。当阈值 φ_0 取局部密度参数最大值 φ_{\max} 的20%,阈值 δ_0 取最小距离参数最大值 δ_{\max} 的20%时,可以比较准确地识别聚类中心。因此,本文DP聚类参数阈值计算公式为

$$\varphi_0 = \frac{1}{5} \varphi_{\max}, \delta_0 = \frac{1}{5} \delta_{\max} \quad (3)$$

综上,图像资源深度学习特征的DP密度聚类算法如下。

算法1:基于参数阈值的图像深度学习特征密度聚类算法

输入:深度学习网络 ψ ,图像资源集 $\{X_1, X_2, L, X_N\}$

输出:聚类类别数 K ,聚类划分子集 $\{S_1, S_2, L, S_K\}$

(1)根据深度学习网络 ψ 计算图像 X_i 的特征 $\zeta_i = \psi(X_i) (1 \leq i \leq N)$;

(2)根据公式(1)和(2)计算局部密度参数 φ_i 和最小距离参数 $\delta_i (1 \leq i \leq N)$;

(3)根据公式(3)计算局部密度和最小距离参数阈值 φ_0 和 δ_0 ;

(4)选取满足 $\varphi_i > \varphi_0$ 且 $\delta_i > \delta_0$ 条件的图像数据作为聚类中心 $\{C_1, C_2, L, C_K\}$;

(5)计算图像特征与聚类中心的距离,按照距离最近的原则划分图像子集 $\{S_1, S_2, L, S_K\}$ 。

2.3 MGTL-SAE精细特征学习

自动编码器(Auto-Encoder, AE)可以自己从样本中学习一些代表原信息的主要成分的重要特征,发掘隐藏在巨大数据下的潜在解释性因素,找到代表原信息的主要成分。稀疏自动编码器(Sparse Auto-Encoder, SAE)通过在隐藏层节点中加入稀疏性限制达到学习数据特征的目的,从而获得有意义的特征表示。SAE广泛应用于病理图像分析、表情图像识别和高光谱图像分类等^[21]。但是传统SAE在整个图像集上训练,由于不同类别的图像子集特征差别很大,采用单一的稀疏自动编码网络结构稳定性差,存在收敛速度慢等问题。SAE对网络参数的初始值敏感,恰当的网络参数初始可以加速训练网络收敛。

为此,我们利用学习特征的密度峰值聚类结果的

类别数和聚类划分信息作为先验知识,根据图像资源特征的密度峰值聚类把相似图像划分到一组,通过聚类压缩组内图像信息使得目标函数更容易收敛,构造和设计了MGTL-SAE学习图像精细特征。

(1) 基于聚类类别数的MGTL-SAE网络结构设计

假设图像资源特征的密度峰值聚类为 K 类,每个聚类划分结果构造一组SAE作为基本单元,把 K 个SAE单元串联。为了使网络结构有更好的平衡性,每组拓展成2层SAE单元结构够并联构成,第1层采用Sigmoid激活函数,第2层采用ReLU激活函数,第1层的输出作为第2层的输入。综上,我们设计的MGTL-SAE网络结构如图2所示。

(2) 基于聚类划分的MGTL-SAE参数初始化

传统的SAE参数估计方法随机抽取满足均匀分布的数据初始化从输入层 L_1 到隐含层 L_2 的权重 $W_{n \times m}^{(1)}$ 、从隐含层 L_2 和输出层 L_3 之间的权重 $W_{m \times n}^{(2)}$,其中 n 为输入层和输出层的维数, m 为隐藏层的维数。没有利用训练数据本身的信息,导致SAE的学习收敛速度慢。

为了加快收敛速度,利用密度峰值聚类的图像块初始化网络参数。针对图2所示的MGTL-SAE网络结构,提出了基于聚类划分结果的参数初始化方法。依据聚类划分结果,从聚类划分子集 $S_i(1 \leq i \leq K)$ 的所有图像中抽取 l_i 个大小为 $\sqrt{n} \times \sqrt{n}$ 的图像块转换为 $[0, 1]$ 的灰度图像,并拉长为 nl_i 的向量,分别统计其均值 μ_i 和方差 σ_i 作为先验知识,生成高斯分布随机数矩阵初始化

参数。图像块的像素灰度均值和方差估计为

$$\mu_i = \frac{1}{nl_i} \sum_{k=1}^{nl_i} x_k (x_k \in S_i) \quad (4)$$

$$\sigma_i = \frac{1}{nl_i - 1} \sum_{k=1}^{nl_i} (x_k - \mu_i)^2 (x_k \in S_i) \quad (5)$$

x_k 表示图像块 S_i 的第 k 个像素灰度值。按照均值 μ_i 和方差 σ_i 生产大小为 $n \times m$ 的高斯分布随机数矩阵作为MGTL-SAE网络结果的第 i 组第 j 层参数权重 $W_{(i,j)n \times m}^{(1)}$ 和大小为 $m \times n$ 的高斯分布随机数矩阵作为权重 $W_{(i,j)m \times n}^{(2)} (j=1, 2)$ 网络权重分别为

$$W_{(i,j)n \times m}^{(1)} = \text{randn}(\mu_i, \sigma_i)_{n \times m} * 2r - I_{n \times m} * r \quad (6)$$

$$W_{(i,j)m \times n}^{(2)} = \text{randn}(\mu_i, \sigma_i)_{m \times n} * 2r - I_{m \times n} * r \quad (7)$$

其中常系数 $r = \sqrt{\frac{6}{n+m-1}}$, $\text{randn}(\mu_i, \sigma_i)$ 为满足均值 μ_i 和方差 σ_i 的正态分布随机数, $I_{n \times m}$ 为元素均为1的矩阵。

(3) 基于聚类划分的MGTL-SAE参数估计

为了估计MGTL-SAE模型参数,使用梯度下降法迭代更新计算参。在机器学习领域中,梯度下降算法是求解最优化问题的最重要和最基础方法。近年来,梯度下降算法已成为机器学习特别是深度学习研究的焦点,涌现出众多改进版本,大致分为动量、方差缩减、增量梯度和自适应学习率四种^[22]。本文选取梯度相反方向搜索并按固定步长搜索目标函数的最小值,参数更新公式的计算如下:

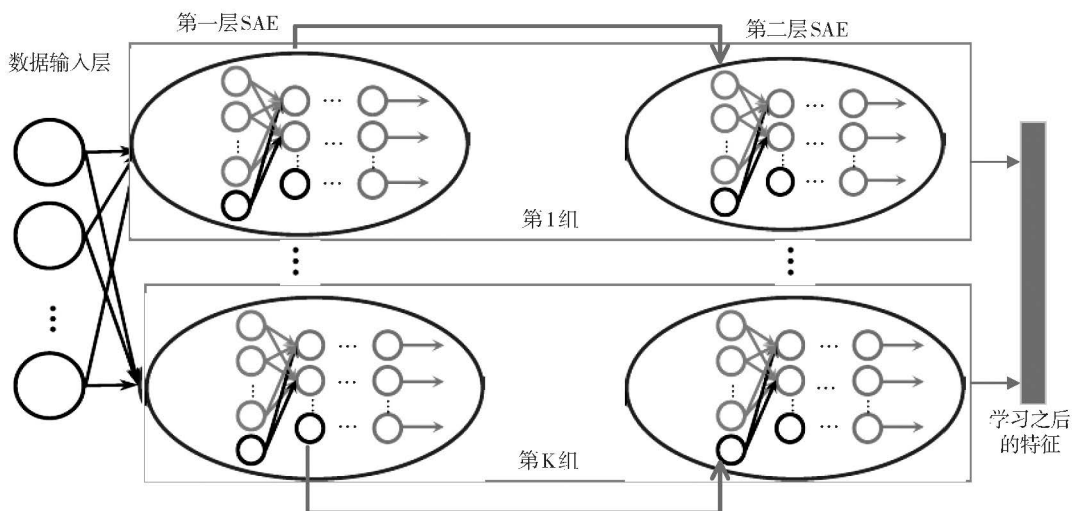


图2 MGTL-SAE网络结构

$$W_{ij}^{(l_2)} = W_{ij}^{(l_2)} - \alpha \frac{\partial J(W_{ij}^{(l_2)}, b_{ij}^{(l_2)})}{\partial W_{ij}^{(l_2)}} \quad (8)$$

$$b_{ij}^{(l_2)} = b_{ij}^{(l_2)} - \alpha \frac{\partial J(W_{ij}^{(l_2)}, b_{ij}^{(l_2)})}{\partial b_{ij}^{(l_2)}} \quad (9)$$

其中步长 α 为常系数, $W_{ij}^{(l_2)}, b_{ij}^{(l_2)}$ ($l_2=1,2$)分别表示MGTL-SAE的第*i*组第*j*层SAE单元结构权重和偏差,函数*J*表示目标函数, $\frac{\partial J(W_{ij}^{(l_2)}, b_{ij}^{(l_2)})}{\partial W_{ij}^{(l_2)}}$ 表示目标函数对权重参数方向的偏导, $\frac{\partial J(W_{ij}^{(l_2)}, b_{ij}^{(l_2)})}{\partial b_{ij}^{(l_2)}}$ 表示目标函数对偏差参数方向的偏导。

针对MGTL-SAE网络结构,提出了基于聚类划分的目标函数,即定义各个聚类对应的SAE结构单元的误差项、惩罚项和稀疏项之和,计算公式如下:

$$J = \sum_{i=1}^K \sum_{j=1}^2 \left(\frac{1}{n} \sum_{k=1}^n \left(\frac{1}{2} \|h_{w_{ij}, m_j}(x_k) - x_k\|_2^2 \right) + \frac{\gamma}{2} \sum_{l=1}^2 \|W_{ij}^{(l_2)}\|_2^2 + \beta \sum_{p=1}^m KL(\rho \| \rho_p) \right) \quad (10)$$

其中 $x_k \in S_i$ ($1 \leq i \leq K, 1 \leq k \leq n$), $h_{w_{ij}, m_j}(x)$ 表示输入数据 x_k 在第*i*组第*j*层SAE单元结构的重构输出值, γ 和 β 分别表示SAE单元结构的惩罚系数和稀疏系数, $KL(\rho \| \rho_p)$ 表示散度函数, ρ 是常数, ρ_p 是SAE单元结构第*p* ($1 \leq p \leq m$)个神经元素的平均激活函数。

MGTL-SAE模型参数估计通过迭代公式(8)、(9)和(10),直到目标函数公式(10)收敛,输出网络模型 $W_{ij}^{(l_2)}, b_{ij}^{(l_2)}$ ($l_2=1,2$)作为特征基,并通过特征基计算测试图像资源的稀疏自动编码特征。迭代算法1图像特征密度峰值聚类MGTL-SAE精细特征学习,直到特征基收敛,输出MGTL-SAE精细特征。

2.4 基于精细特征聚类划分的图像快速检索

图像资源精细特征的密度峰值聚类,把相似图像聚合成聚类划分,检索时首先找到与待查询图像最相似的聚类划分子集,然后在该划分子集内检索,缩小搜索范围,实现快速、准确地图像资源检索。采用算法1聚类分析图像资源集的MGTL-SAE精细特征,得到聚类中心集和聚类划分子集 $\{\mathcal{C}, \mathcal{S}, L, \mathcal{S}\}$ 。计算查询样本图像的MGTL-SAE精细特征,并按照定义1和2映射到局部密度参数和最小距离参数空间并标记 f_q 。计算 f_q 分别到所有聚类中心 $\{\mathcal{C}, \mathcal{C}, L, \mathcal{C}\}$ 的距离,选择距离最近的聚类中心 \hat{C} ($1 \leq i \leq K$)及其划分子集 \mathcal{S} ($1 \leq i \leq K$)作为检索对象。计算查询图像与图像子集 \mathcal{S} ($1 \leq i \leq K$)在密度聚类空间的

相似度,按照相似度降序返回检索结果。

3 实验结果及其分析

为了评价方法的有效性,MGTL-SAE方法与Tzeleni等^[11]提出的基于CaffeNet模型方法、Joe等^[12]提出的基于OxfordNet+GoogLeNet+VLAD模型方法和Liu等^[13]提出的基于GoogLeNet+DDBTC方法进行了比较实验。

3.1 评价指标

采用查准率(Precision)、召回率(Recall)、mAP(Mean Average Precision)等指标定量评价几种方法的性能。查准率指系统检索到相似图像的准确率,召回率是系统从图像集合中检出相关图像的成功率,两者取值在0和1之间。查准率和召回率的计算公式分别为

$$P = N_R / N_1 \quad (11)$$

$$R = N_R / N_2 \quad (12)$$

其中 N_R 是返回与查询图像相关的图像数, N_1 表示图像检索返回的图像数, N_2 表示在图像库中实际和检索图像相关的图像数。

用X轴表示精度,Y轴表示召回率,平均检索精度AP(Average Precision)为准确率-召回率(Precision-Recall, PR)曲线所围成的面积,定义为

$$AP = \frac{1}{N_R} \sum_{i=1}^{N_3} \frac{N_i}{i} t_i \quad (13)$$

其中 N_i 表示返回的前*i*个结果中与查询图像相关的图像数, N_3 是所有测试图像数。如果第*i*幅图像是相关的,则 $t_i=1$;否则 $t_i=0$ 。所有检索图像的AP平均值作为mAP,检索到相关图像越靠前,mAP就越高。

3.2 实验数据集

实验数据采用哥伦比亚大学的Coil-20图像集,加利福尼亚理工学院的Caltech-256图像集和自建随书光盘图像集Bookimages-30。Coil-20数据集有20个对象共计1440幅,每个对象旋转5度拍一张照片共有72幅图片,每幅图像大小为128×128。随机抽取400张图像作为测试集,其他图像作为训练集。Caltech-256数据集由256类图像,每类图像至少包含80张图像,共30607张图像。随机抽取600张图像作为测试集,其他图像作为训练集。我们图书馆自建随书光盘数据库Bookimages-30,已收录光盘10万余张,包含了大量图像资源,每幅图像包含多个目标且背景复杂。我们归类整理了30个类别,每类图像包含50张图像,共1500张图像。随机抽取500张图像作为测试集,其他图像作

为训练集。

3.3 实验方法和结果

本文实验采用硬件环境: Intel(R) Core(TM)i7-4770 CPU 和 4G 内存;软件环境:操作系统 win7 64 位, 编程语言 Matlab2014, GoogLeNet 预训练模型。定义 1 中距离阈值的计算方法: 计算对所有图像之间的特征欧氏距离, 按照从小到大的顺序排序, 选取前 2% 位置的距离作为阈值 d_c 。Coil-20、Caltech-256 和 Bookimages-30 三个数据集的距离阈值 d_c 、聚类划分的局部密度阈值 φ_0 和最小距离阈值 δ_0 分别如表 1 所示。

表1 MGTL-SAE 计算局部密度和最小距离的阈值

数据集	距离阈值 d_c	局部密度阈值 φ_0	最小距离阈值 δ_0
Coil-20	0.12	4	0.25
Caltech-256	0.35	2	0.31
Bookimages-30	0.28	3	0.23

以 Coil-20 数据集为例, 传统的 SAE 直接从所有图像集中学习特征, 从数据集中随机抽取 200 个图像块如图 3(a) 所示, 可见图像灰度对比、纹理等特征变换较大, SAE 特征学习迭代收敛速度慢。MGTL-SAE 方法首先对图像集采用密度峰值聚类, 再从聚类划分子集中随机抽取 200 个图像块如图 3(b) 所示, 对比图 3(a) 可见灰

度对比、纹理等特征变换较小。MGTL-SAE 从图像集的聚类子集中可以更有针对性地学习特征, 收敛速度快, 主要原因在于 MGTL-SAE 方法对图像 GoogLeNet 特征和稀疏自动编码特征聚类划分后, 相同特征信息的聚类划分压缩了大量图像信息, 目标函数更容易稳定。SAE 特征学习的特征基如图 3(c) 所示, MGTL-SAE 的部分特征基如图 3(d) 和 3(e) 所示, 对比可知图 3(d) 和 3(e) 的特征基包含了更多精细特征以体现图像聚类子集之间的细微特征差别, 而图 3(c) 的特征基包含了更多图像大类之间的粗粒度特征差别。

从检索结果来看, Tzelepi 等^[11-13]方法和 MGTL-SAE 方法在 Coil-20、Caltech-256 和 Bookimages-30 数据集的查准率和查全率 P-R 曲线结果对比分别如下页图 4、图 5 和图 6 所示, mAP 指标如表 2 所示。在 3 个数据集上, 4 种方法的查准率 P 都随着召回率 R 的增加而减少, 相同召回率 R 的情况下, MGTL-SAE 方法的查准率 P

表2 不同检索方法的mAP

数据集	Tzelepi	Joe	Liu	MGTL-SAE
Coil-20	0.268	0.156	0.289	0.512
Caltech-256	0.214	0.121	0.21	0.482
Bookimages-30	0.201	0.087	0.227	0.381

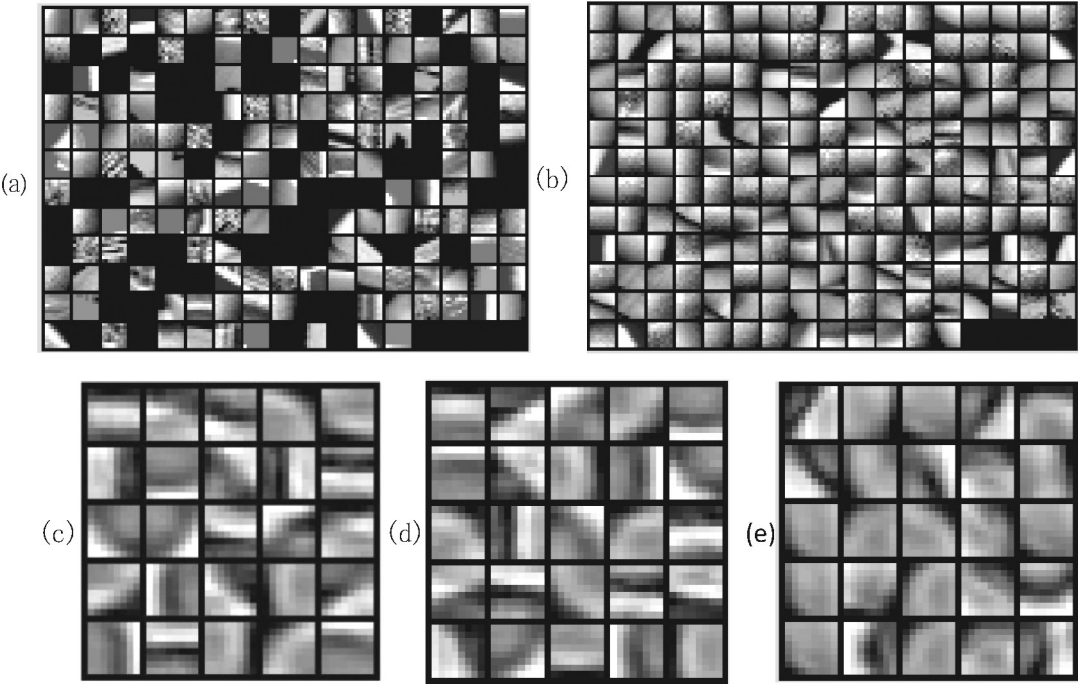


图3 Coil-20数据精细特征学习结果: (a)从数据集中随机抽取200个图像块; (b) 聚类划分子集中随机抽取200个图像块; (c)SAE的特征基; (d)和(e) MGTL-SAE的聚类子集特征基。

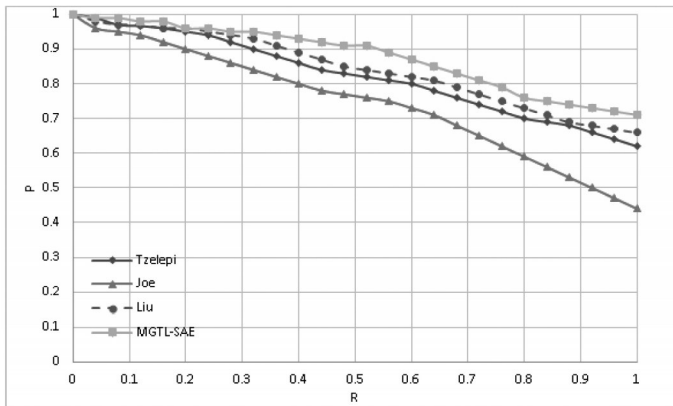


图4 不同方法在Coil-20的查准率和查全率对比

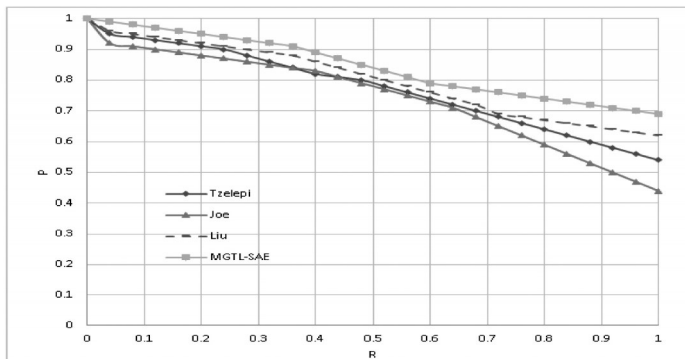


图5 不同方法在Caltech-256的查准率和查全率对比

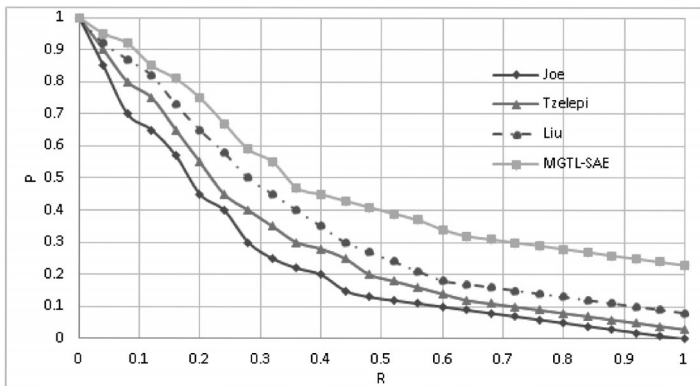


图6 不同方法在Bookimages-30的查准率和查全率对比

高于其他3种方法。由于Coil-20数据集的图像是没有背景的单个对象, Caltech-256数据集的图像是单个对象加背景, 而Bookimages-30的图像具有多个目标和复杂的背景, 所以图4和图5的P-R曲线高于图6。主要原因是Tzelepi等^[11]的CaffeNet模型方法、Joe等^[12]的OxfordNet + GoogLeNet + VLAD模型方法和Liu等^[13]的GoogLeNet+DDBTC模型方法分别针对全局数据集的

深度学习特征、多种深度学习特征的聚合、深度学习特征与底层特征的结合, 检索效果容易受到不同类别训练图像的干扰。而MGTL-SAE方法结合深度学习特征和密度峰值聚类算法应用于精细特征学习的每次迭代过程, 在聚类划分的局部数据集上有针对性地学习精细特征, 弥补了现有基于深度学习的图像检索算法在全局数据集学习特征的不足。实验结果表明, MGTL-SAE方法能更加准确地描述图像内容特征, 可以有效地提高数字图像资源的检索精度。

从时间性能角度看, 传统的SAE和MGTL-SAE方法在三个数据集上训练特征学习的时间如表3所示, MGTL-SAE的训练时间比SAE快10多倍, 主要原因是MGTL-SAE在图像特征相似性较大的聚类子集上学习, 目标函数更容易收敛。Tzelepi等^[11-13]方法和MGTL-SAE方法的平均检索时间如表4所示, Tzelepi等^[11]和Liu等^[13]方法的检索时间最长, MGTL-SAE方法检索时间最少。主要因为MGTL-SAE方法在密度聚类划分的子集上训练模型和检索图像, 分别加快了网络收敛的速度和缩小了搜索的范围。

表3 不同模型的训练时间(s)

数据集	SAE	MGTL-SAE
Coil-20	81.12	7.23
Caltech-256	1674.84	61.61
Bookimages-30	102.931	8.42

表4 不同方法的平均检索时间(s)

数据集	Tzelepi	Joe	Liu	MGTL-SAE
Coil-20	0.213	0.112	0.229	0.02
Caltech-256	3.137	2.415	3.256	0.23
Bookimages-30	0.201	0.102	0.231	0.02

经过实验分析, 可以发现本文的方法较之于传统的深度学习图像检索方法有着较大的优势。首先, 传统方法只考虑在全局图像资源集上深度学习特征和图像检索, 而MGTL-SAE方法利用了聚类划分图像子集的相似性可以有效压缩图像特征信息, 多组双层稀疏编码器学习更加精细特征。其次, MGTL-SAE方法的目标函数和特征基收敛得更快, 图像检索只需要在聚类划分子集完成, 结果更加精准和快速。本方法的适用性也非常广泛, 除了可以应用到

光盘资源的图像检索,还可以用于数字图书馆的地域文化图像数据库、专业特色图像数据库、人物图像数据库、记忆图像数据库、专题图像数据库和数字人文研究的视觉数据库检索。

4 结语

本文针对深度学习的图像检索方法存在计算量大、训练时间长、匹配检索速度慢等问题,提出了一种基于深度特征学习和密度峰值聚类的图像资源快速检索方法。研究了利用图像深度学习特征的密度峰值聚类结果设计多组双层稀疏自动编码器MGTL-SAE的网络结构,初始化参数和重构目标函数估计网络参数以学习图像资源精细特征。Coil-20、Caltech-256和Boo-kimages-30数据集上的实验验证了MGTL-SAE方法的有效性。然而本文研究仍然存在模型比较复杂,没有结合语义标注等不足,后续研究优化模型网络结构和参数估计方法,集合语义标注,面向数字图书馆的移动视觉检索具体应用,推进数字图书馆的图像和视频可视化资源等的相关研究和应用。

参考文献

- [1] 王伟平. 基于内容的图像检索技术研究进展及其在数字图书馆中的应用[J]. 图书情报工作, 2009, 53(5): 116-120.
- [2] 李敏, 邱润冬, 段青. 民国中医药文献数字资源服务平台图片库功能概述[J]. 中国中医药图书情报杂志, 2019, 43(5): 10-14.
- [3] 王国平. 公共图书馆自建特色数据库的实践与思考[J]. 河南图书馆学刊, 2015, 35(12): 25-27.
- [4] 史丽君. 古籍图像数据库建设常见问题及对策研究——以首都图书馆馆藏古籍珍善本图像数据库建设为例[J]. 图书馆工作与研究, 2016, 1(9): 62-66.
- [5] 郭瑞芳, 张会章. 我国高校图书馆特色图像数据库建设现状与构建原则[J]. 图书馆学研究, 2012(6): 50-54.
- [6] 曾子明, 秦思琪. 面向数字人文的移动视觉搜索模型研究[J]. 情报资料工作, 2018, 39(6): 23-30.
- [7] 王志明, 张航. 融合多层卷积神经网络特征的快速图像检索方法[J]. 计算机辅助设计与图形学学报, 2019, 31(8): 3567-3577.
- [8] Cui Hui, Zhu Lei, Li Jingjing. Scalable deep hashing for large-scale social image retrieval[J]. IEEE Transactions on Image Processing, 2020, 29(9): 1271-1284.
- [9] 于邓, 刘玉杰, 邢敏敏, 等. 包含跨域建模和深度融合网络的手绘草图检索[J]. 软件学报, 2019, 30(11): 3567-3577.
- [10] Afshan Latif, Aqsa Rasheed, Umer Sajid. Content-based image retrieval and feature extraction: a comprehensive review [J]. Mathematical Problems in Engineering, 2019; doi: 10.1155/2019/9658350.
- [11] Tzelepi M, Tefas A. Deep convolutional learning for content based image retrieval [J]. Neurocomputing, 2018, 275: 2467-2478.
- [12] Joe Yue-Hei Ng, Fan Yang, Larry S Davis. Exploiting local features from deep networks for image retrieval[C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2015: 53-61.
- [13] Liu P, Guo J M, Wu C Y, et al. Fusion of deep learning and compressed domain features for content based image retrieval[J]. IEEE Transactions on Image Processing, 2017, 26(12): 5706-5717.
- [14] Ge Yun, Jiang Shunliang, Xu Qingyong. Exploiting representations from pre-trained convolutional neural networks for high-resolution remote sensing image retrieval[J]. Multimedia Tools and Applications, 2018, 77(13): 17489-17515.
- [15] Wu Zebin, Yu Junqing. A multi-level descriptor using ultra-deep feature for image retrieval[J]. Multimedia Tools and Applications, 2019, 78(18): 25655-25672.
- [16] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C]. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), No. 15523970, IEEE, Boston, MA, USA.
- [17] 迟剑宁, 于晓升, 张艺菲. 融合深度网络和浅层纹理特征的甲状腺结节癌变超声图像诊断[J]. 中国图象图形学报, 2018, 23(10): 154-165.
- [18] 陈斌, 王宏志, 徐新良. 深度学习 GoogleNet 模型支持下的中分辨率遥感影像自动分类[J]. 测绘通报, 2019(6): 29-33.
- [19] Rodriguez A, Laio A. Clustering by fast search and find of density peaks[J]. Science, 2014, 344(6191): 1492-1496.
- [20] 徐晓, 丁世飞, 孙统风, 等. 基于网格筛选的大规模密度峰值聚类算法[J]. 计算机研究与发展, 2018, 55(11): 79-89.
- [21] 付晓, 沈远彤, 付丽华, 等. 基于特征聚类的稀疏自编码快速算法[J]. 电子学报, 2018, 46(5): 20-25.
- [22] 史加荣, 王丹, 尚凡华. 随机梯度下降算法研究进展[J/OL]. 自动化学报, 2019, <https://doi.org/10.16383/j.aas.c190260>.

[作者简介]高蕴梅,女,1982年生,常熟理工学院图书馆馆员。

收稿日期:2019-11-07