计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2019,28(3):229−234 [doi: 10.15888/j.cnki.csa.006826]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于深度学习的服装图像检索方法①

陈　双, 何利力, 郑军红

(浙江理工大学 信息学院, 杭州 310018)
通讯作者: 陈　双, E-mail: cscmyk@outlook.com

摘　要: 为实现面向大规模服装图像集的图像快速精准检索, 突破当前常规检索方法的局限性, 本文提出了一个新的深度学习模型: Fashion-16 服装图像检索模型. 采用先分类再类内检索的思想, 基于 VGG-16 模型强大的图像特征提取能力, 以卷积神经网络 softmax 分类器进行分类, 对同一类别下采用局部敏感哈希的思想进行近似最近邻的查找, 实现了针对服装类别属性的图像检索模型修正. 实验结果表明, 模型具有良好的稳定性、精确率及检索速度, 有其实用价值与研究意义.

关键词: 服装图像检索; 深度学习; 特征提取; Softmax 分类器; 局部敏感哈希

引用格式: 陈双,何利力,郑军红.基于深度学习的服装图像检索方法.计算机系统应用,2019,28(3):229–234. http://www.c-s-a.org.cn/1003-3254/6826.html

Clothing Image Retrieval Method Based on Deep Learning

CHEN Shuang, HE Li-Li, ZHENG Jun-Hong

(School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: In order to achieve fast and accurate image retrieval for large-scale clothing image sets and break through the limitations of current conventional retrieval methods, this study proposes a new deep learning model: Fashion-16 clothing image retrieval model. Based on the idea of first classification and intra-class retrieval, based on the powerful image feature extraction ability of VGG-16 model, the convolutional neural network Softmax classifier is used for classification, and the nearest neighbor search is performed for the idea of locally sensitive hashing under the same category. An image retrieval model correction for clothing category attributes is implemented. The experimental results show that the model has good stability, accuracy, and retrieval speed, and has practical value and research significance.

Key words: clothing image retrieval; deep learning; feature extraction; Softmax classifier; locality sensitive hashing

随着服装电子商务的飞速发展, 互联网上服装图像数据量急剧增长, 如何从海量的图像库中进行快速精准检索成为了近几年研究的热点[1].

目前服装图像检索的常规方法有两类, 一类是基于文字的图像检索 (TBIR), 通过对服装图像的文字描述进行语义式匹配; 另一类是基于图像内容的图像检索 (CBIR), 从图像的颜色、纹理等方面进行特征提取, 实现“以图搜图”[2]. 但这两类方法都具有一定的局限性, 文字描述所进行的人工语义标签十分繁琐, 且具有主观性; 而内容特征不能全面地反映图像丰富的视觉特征, 机器从低级的可视化特征得到的相似性与人从高级的语义特征得到的相似性间存在着巨大的“语义鸿沟”[3], 造成检索的效果不佳. 为此, 在深度学习技术与图像处理技术飞速发展的当下, 借助深度学习强大的特征提取能力, 直接对图像进行处理, 消除不同底层特征带来的影响, 进行服装图像检索研究.

① 基金项目: 浙江省科技厅 (重大) 项目 (2015C03001)
Foundation item: Major Program of Science and Technology Bureau of Zhejiang Province (2015C03001)
收稿时间: 2018-09-28; 修改时间: 2018-10-23; 采用时间: 2018-10-30; csa 在线出版时间: 2019-02-22

基于深度学习的方法在图像分类、图像检索方面具有独特的优越性. AlexNet 模型[4]与 VGG 模型[5]成功地验证了深度卷积神经网络在学习图像特征表示上的能力. 而对于有着纹理、款式等特有视觉特征的服装图像的检索, 目前仍处于探索阶段. 如基于深度学习进行对服装图片的自动标注[6], 以及着重于深度卷积神经网络的层次多任务服装分类等[7], 它们主要借助于深度学习的图像特征表达能力来进行研究. 面对大规模服装图像, 利用深度卷积神经网络学习训练样本的近似哈希编码得到哈希构造函数, 采用 CNN 与哈希方法相结合的算法[8]能有效提高图像检索的速度.

本文主要进行 3 方面的研究. 1) 基于 Fashion-MNIST 数据集建立卷积神经网络模型, 进行服装类别标签分类. 2) 基于 VGG-16 预训练模型, 对服装数据集进行特征提取, 并映射成哈希编码, 建立服装特征哈希索引库, 实现图像的快速检索. 3) 综合以上两个模型, 以爬取的服装图像进行分类训练, 建立大规模服装数据集, 提出一种新的 Fashin-16 神经网络模型, 实现基于深度学习的先分类再类内检索的服装图像检索, 并通过实验分析与对比实验的设计验证其检索效果.

1 研究方法

1.1 研究环境与预处理

实验环境: 基于 Keras 深度学习框架, Tensorflow 做为后端.

预训练数据: 采用 Fashion-MNIST 数据集进行预训练. Fashion-MNIST 是德国研究机构 Zalando Research 发布的一个服装图像数据集[9], 含有 60 000 个训练样本和 10 000 个测试样本, 包括 10 个类别标签: T 恤, 裤子, 套头衫, 裙子, 外套, 凉鞋, 衬衫, 运动鞋, 包, 踝靴.

实验数据: 使用爬虫爬取服装图像与网上相关服装数据集, 获取了总计 325 820 张服装图像, 关联对应的服装类别标签, 建立大规模服装图像数据集作为实验样本集, 并将样本集随机分为三批, 20 万个样本作为训练数据, 进行模型的训练, 10 万个样本进行模型的参数调优, 剩下的样本用来衡量最优模型的性能.

标签选择: 为实现高效精准的分类, 本文参考 Fashin-MNIST 数据集的类别标签, 采用单标签的方法. 考虑缺少多标签之间的关联, 泛化能力不足[10], 后期将对 Fashion-16 模型进行调整, 实现多标签的分类检索.

预处理: 对于输入预训练模型与进行检索的图像, 为减少图像冗余信息, 去除背景、光照、多主体等因素的影响, 需对实验数据进行预训练格式标准化, 主要进行去均值与归一化:去均值是指对图像的每个数据点进行均值消除, 移除图像的平均亮度, 消除数据的直流分量; 归一化是指令 x_train=x_train/255, 使样本值处于[0, 1]之间, 减少各维度数据取值范围的差异带来的干扰[11].

图 1 Fashion-MNIST 及抓取图像部分数据集

1.2 基于卷积神经网络的服装类别标签分类模型

服装图像中含有丰富的服装特有属性信息, 如颜色、花纹、袖子的长短等. 本文从服装的类别进行研究, 采用卷积神经网络的非线性映射能力与自学习能力, 根据服装图像与类别标签, 自动学习服装类别标签特征, 以网络的高层语义激活值表示服装的类别标签特征, 实现服装高效精准分类, 构建基于卷积神经网络的服装类别标签分类模型[12]. 建立卷积神经网络结构如表 1 所示.

表 1 服装类别标签分类模型结构

神经网络层	Input	Output
conv2d_1	(None, 28, 28, 1)	(None, 28, 28, 16)
conv2d_2	(None, 28, 28, 16)	(None, 28, 28, 32)
max_pooling2d_1	(None, 28, 28, 32)	(None, 14, 14, 32)
conv2d_3	(None, 14, 14, 32)	(None, 14, 14, 64)
conv2d_4	(None, 14, 14, 64)	(None, 14, 14, 128)
max_pooling2d_2	(None, 14, 14, 128)	(None, 7, 7, 128)
dropout_1	(None, 7, 7, 128)	(None, 7, 7, 128)
Faltten_1	(None, 7, 7, 128)	(None, 6272)
Dense_1	(None, 6272)	(None, 256)
Dropout_2	(None, 256)	(None, 256)
Dense_2	(None, 256)	(None, 10)

卷积层通过卷积核与输入的相互作用进行特征提取, 池化层弱化位置信息并过滤不重要的高频信息, 形成更抽象的特征, 逐层提取并组合成完备的描述特征, 保证图像的局部关联性与空间不变性[13]. 服装类别标签分类模型采用卷积层和池化层两次交替, 对 Fashion-MNIST 28×28 的输入图像采用 3×3 的卷积核, $f(x)=$

$\max(0,x)$ 的 ReLU 激活函数, 每两次卷积后进行一次 2×2 的 Max pooling 池化.

在全连接层展平像素前引入两层 Dropout 层, 以 0.25 与 0.5 的阈值随机去除权重, 这一过程虽然降低了训练速度, 但提高了网络的泛化能力, 防止过度拟合.

在最后一层输出层中使用 Softmax 函数进行分类, 采用 Logistic 回归代价函数做为 Softmax 分类器的代价函数, 给出样本对每一类别的概率 $p\left(y^{(i)}=j|x^{(i)}\right)$, 获得类别标签[14]. 在表达中, 令模型的参数为 $\theta_1,\theta_2,\cdots,\theta_k$, 采用归一化使所有的概率和为 1, 对所有输入的列向量 $h_0\left(x^{(i)}\right)$, 有:

$$h_0\left(x^{(i)}\right)=\begin{bmatrix} p\left(y^{(i)}=1|x^{(i)};\theta\right) \\ p\left(y^{(i)}=2|x^{(i)};\theta\right) \\ \vdots \\ p\left(y^{(i)}=k|x^{(i)};\theta\right) \end{bmatrix}=\frac{1}{\sum_{j=1}^{k}e^{\theta_j^{\mathrm{T}}x^{(i)}}}\begin{bmatrix} e^{\theta_1^{\mathrm{T}}x^{(i)}} \\ e^{\theta_2^{\mathrm{T}}x^{(i)}} \\ \vdots \\ e^{\theta_k^{\mathrm{T}}x^{(i)}} \end{bmatrix}$$

通过模型训练, 对预训练数据进行参数 fine-tune, 使模型效果达到预定的分类准确率, 形成以 4 层卷积层初步提取服装图像特征, 2 层池化层提取其主要特征, 2 层全连接层进行特征汇总, Softmax 分类层进行分类预测, 最终返回一个包含 10 个类别的对应概率的一维矩阵, 概率的最大值即该图像的服装类别, 实现基于卷积神经网络的服装类别标签分类模型.

1.3 基于 VGG-16 的图像检索模型

实验中用到的 VGG-16 模型是深度卷积神经网络 VGGNet 的一种, 是由 16 层神经网络构成的经典模型, 采用由 ImageNet 预训练的权重. 采用小核堆叠的思想, 反复堆叠 3×3 的小型卷积核和 2×2 的最大池化层, 包含 13 个卷积层、3 个全连接层, 对 224×224×3 的输入数据, 以多层卷积与池化进行特征提取[15]. 其整体结构如图 2 所示.

利用 VGG-16 的卷积层与池化层学习到的服装图像特征, 对模型进行调整, 在原模型中引入哈希层, 采用局部敏感哈希算法思想, 以随机超平面的方法构造哈希函数[16], 将高维的服装图像特征映射成二进制哈希码, 具有相同的二进制哈希码的样本保存在相同库中, 以此构建服装数据集的特征哈希索引库[17].

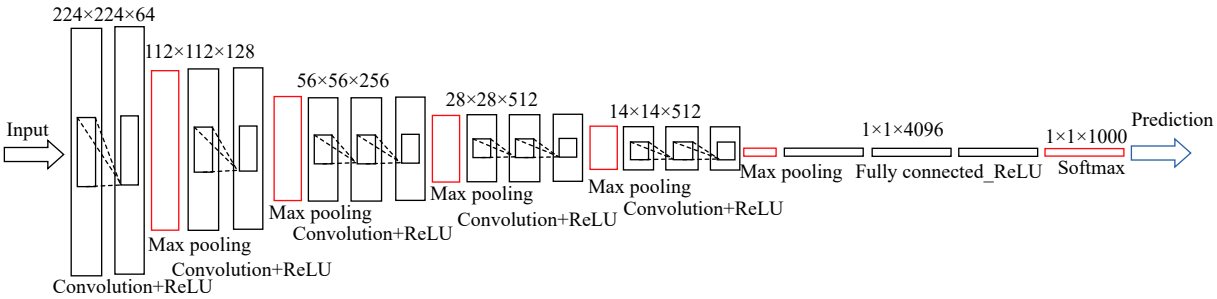

图 2 VGG-16 整体结构

在检索时, 对输入的图像进行同样的特征提取与映射, 对比得到与输入图像相似性高的样本所在的索引库, 将输入图像的二进制哈希码传入库中与库中的哈希码逐一进行相似性度量[18], 根据比较结果返回 20 个相似度最高的图片及对应的相似度, 实现服装图像的快速检索.

1.4 Fashion-16 服装图像检索模型

结合基于卷积神经网络的服装类别标签分类模型与基于 VGG-16 的图像检索模型, 本文提出了一种新的模型: Fashion-16 服装图像检索模型.

整体采用先分类再类内检索的思想. 借助于上述两个模型的特征提取能力与 Softmax 分类功能, 对训练样本进行先分类再根据类别进行特征信息的保存, 对于检索图像进行特征提取及局部敏感哈希进行近似最近邻的查找, 在相应类别的服装图像集中检索到按相似度降序图像[19]. 实现图像的精准分类与快速检索, 以分类优化检索.

首先采用 VGG-16 模型对爬取的服装图像样本集进行特征提取, 并映射成哈希编码. 然后对 VGG-16 模型的最后一个卷积层进行调整, 添加能处理服装类别标签的网络, 即卷积神经网络模型的 Softmax 分类层. 根据分类信息将训练模型信息存至 HDF5 文件, 分别

构造特征哈希索引库. 对测试样本进行相似性度量, 衡量模型的性能并进行调参.

对于一次完整的检索过程, 将待检索图像输入网络模型进行前向传播, 层层采样获得图像特征, 哈希编码后根据服装类别标签卷积神经网络模型 Softmax 分类器的结果传入对应的索引库进行近似最近邻查找, 返回按相似度排序的图像结果, 实现服装图像的精准分类和快速检索[20]. 模型整体构造如图 3 所示.

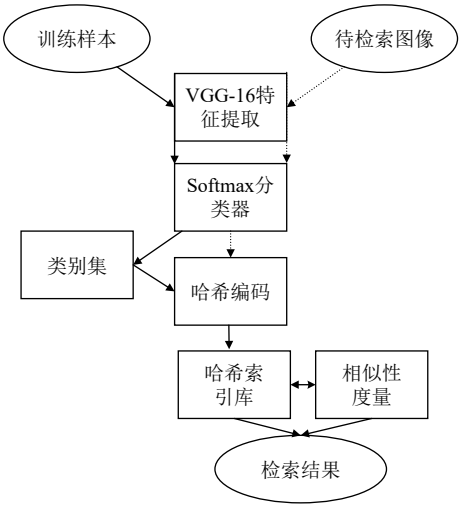

图 3 Fashion-16 模型整体架构

2 实验分析与对比

2.1 结果分析

根据实验设计, 对实验数据集进行 20 次迭代, 总计用时 7301 s, 得到损失率 (*Loss*) 与准确率 (*Accurancy*) 如图 4 所示.

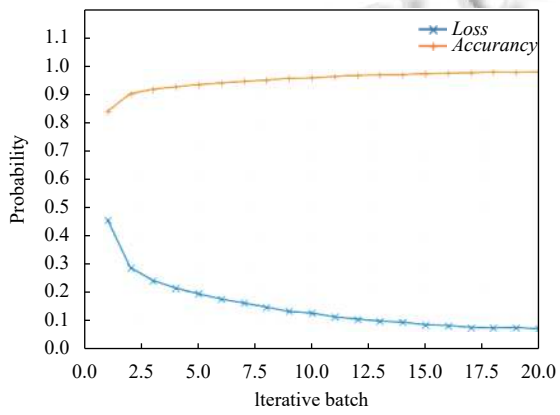

图 4 20 次迭代过程损失值与准确率变化情况

从图中可以发现, 前几次准确率上升、损失值下降速度较快, 后续趋于平缓, 表明实验在多次迭代后结果趋于稳定.

采用 Flask 进行 Web 实现, 得到服装图像分类检索页面样式如图 5 所示.

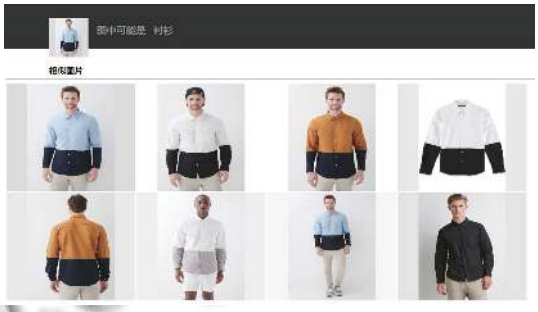

图 5 服装图像分类检索页面样式

2.1.1 分类精确度

考虑在图像领域常用的评价指标, 对于服装图像检索的精确率, 分类精确度直接影响检索精确度, 本文采用查准率对分类精确度进行度量, 定义为检索结果中正确图像数目 m 与返回图像数目 k 的比值, 计算公式为:

$$precision = \frac{m}{k} \times 100\%$$

对服装图像集 2 万个随机样本进行测试, 得到 Softmax 分类器的分类精确度为 92.71%. 实验表明, 检索对服装图像的类别具有良好的针对性, 能达到预期的检索效果.

2.1.2 稳定性

设计实验, 对于数据集大小从预实验的 Fashin-MNIST 数据集进行逐 50 000 数据量的增加, 对于验证数据实验生成的检索精确度如表 2 所示.

表 2 数据量与检索精确度

实验数据量	检索精确度
50 000	0.8841
100 000	0.9060
150 000	0.9222
20 000	0.9297
250 000	0.9355
300 000	0.9387

实验表明, 随着数据集的扩大, 检索精确度同步得到了提升, 而后续逐渐趋于稳定, 保证了模型面向超大规模服装图像集具有一定的稳定性.

2.1.3 检索速度

对于验证数据, 随机取 1000 次检索时间的平均值, 得到平均检索用时为 3.2416 s, 检索具有较好的检索速度, 能应对日常的图像检索需求.

这主要是采用类内检索缩小了检索范围, 并协同哈希方法在检索中计算速度和存储空间的优越性, 保证了模型优异的检索速度.

2.2 实验对比

2.2.1 SIFT 特征提取与卷积神经网络特征提取

特征提取方法选用基于内容的图像检索中比较著名的局部特征描述子 SIFT 特征, SIFT 由于对旋转、尺度交换以及一定的视角和光照变化等图像变化具有不变性, 可以获得较好的特征效果[21]. 而在本模型中, 使用预训练模型 VGG-16 进行服装图像特征的提取, 依靠多层卷积与池化, 层层采样, 得到的不同层次的特征, 同样具有良好的特征表达能力.

对两种特征提取方式得到的特征分别进行后续的分类与检索, 通过对验证数据集 20 000 个样本的实验, 得到各自检索结果准确率如表 3 所示.

表 3 检索结果准确率

特征提取方法	检索准确率
SIFT 特征	0.8940
卷积神经网络特征	0.9271

实验结果表明, 相比于 SIFT 特征, 使用 VGG-16 模型所进行的对低层次特征学习、抽象、组合形成的高层特征具有更好的图像特征表示能力.

2.2.2 直接检索与类内检索

先分类再进行类内检索的方法限定了检索的范围, 得到的检索结果与目标的类别相同, 避免了特征描述的偏差引起不同类别间的相似性过高, 相比直接检索具有更高的准确性. 而对于检索速度而言, 对不分类直接进行检索与分类后进行类内检索这两种情况下进行对比, 随机取各自 1000 次检索时间的平均值, 得到平均检索用时如表 4 所示.

表 4 平均检索用时

检索方法	平均检索用时 (s)
直接检索	3.6375
类内检索	3.2416

实验结果表明, 分类后进行类内检索相比直接检索在检索速度上有 10.88% 的提升, 并且可预期的对于更大规模的图像集, 因检索范围的限定, 检索速度的差别将会更大.

3 结语

本文提出了一个新的深度学习模型 Fashion-16 服装图像检索模型, 借助于卷积神经网络强大的图像特征提取能力, 采用先分类, 再类内检索的思想, 在类内以局部敏感哈希算法进行近似最近邻的查找. 通过实验, 验证了模型具有良好的分类精确度、稳定性与检索速度. 模型针对服装领域类别属性的修正优化, 能够达到较好的服装检索效果.

参考文献

1 包青平. 基于深度学习的服装图像分类与检索[硕士学位论文]. 杭州: 浙江大学, 2017.

2 刘兵, 张鸿. 基于卷积神经网络和流形排序的图像检索算法. 计算机应用, 2016, 36(2): 531–534. [doi: 10.3969/j.issn.1001-3695.2016.02.049]

3 Liu PZ, Guo JM, Wu CY, *et al*. Fusion of deep learning and compressed domain features for content-based image retrieval. IEEE Transactions on Image Processing, 2017, 26(12): 5706–5717. [doi: 10.1109/TIP.2017.2736343]

4 Oquab M, Bottou L, Laptev I, *et al*. Learning and transferring mid-level image representations using convolutional neural networks. 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. 2014. 1717–1724.

5 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556, 2015.

6 郑森烈. 基于深度学习的服装图片自动标注系统的设计与实现[硕士学位论文]. 广州: 中山大学, 2015.

7 林城龙, 胡伟, 李瑞瑞. 基于深度卷积神经网络的层次多任务服装分类. 中国体视学与图像分析, 2018, 23(2): 159–165.

8 Xia RK, Pan Y, Lai HJ, *et al*. Supervised hashing for image retrieval via image representation learning. Proceedings of the AAAI Conference on Artificial Intelligence. Québec City, Canada. 2014.

9 Xiao H, Rasul K, Vollgraf R. Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. arXiv:1708.07747, 2017.

10 陈飞, 吕绍和, 李军, 等. 目标提取与哈希机制的多标签图像检索. 中国图象图形学报, 2017, 22(2): 232–240.

11 杨宇. 基于深度学习特征的图像推荐系统[硕士学位论文]. 成都: 电子科技大学, 2015.
12 胡二雷, 冯瑞. 基于深度学习的图像检索系统. 计算机系统应用, 2017, 26(3): 8–19. [doi: 10.15888/j.cnki.csa.005692]
13 郑启财. 基于深度学习的图像检索技术的研究[硕士学位论文]. 福州: 福建师范大学, 2015.
14 彭天强, 栗芳. 基于深度卷积神经网络和二进制哈希学习的图像检索方法. 电子与信息学报, 2016, 38(8): 2068–2075.
15 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: 10.1109/TPAMI.2016.2577031]
16 刘英帆. 基于局部敏感哈希的近似最近邻查询研究[硕士学位论文]. 西安: 西安电子科技大学, 2014.
17 Nguyen VA, Do MN. Deep learning based supervised hashing for efficient image retrieval. 2016 IEEE International Conference on Multimedia and Expo. Seattle, WA, USA. 2016. 1–6.
18 Lin K, Yang HF, Hsiao JH, *et al*. Deep learning of binary Hash codes for fast image retrieval. 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Boston, MA, USA. 2015. 27–35.
19 梁倜. 基于深度学习特征提取和树与哈希混合索引的图像检索方法[硕士学位论文]. 广州: 华南理工大学, 2017.
20 张洪群, 刘雪莹, 杨森, 等. 深度学习的半监督遥感图像检索. 遥感学报, 2017, 21(3): 406–414.
21 Jain S, Zaveri T, Prajapati K, *et al*. Deep learning feature map for content based image retrieval system for remote sensing application. International Journal of Image Mining, 2016, 2(1). [doi: 10.1504/IJIM.2016.079113]