

一种基于卷积神经网络的多特征服装图像检索方法

吴帆¹, 邓作杰¹, 尚书妃²

(1. 湖南工程学院 计算机与通信学院, 湘潭 411104; 2. 湖南工程学院 管理学院, 湘潭 411104)

摘要:在服装图像检索问题上,由于受服装样式个性化和背景因素的影响,传统的神经网络不能准确地提取出图像细节特征,会导致检索不理想.本文在残差神经网络的基础上,利用多尺度串联图像特征,提出了一种多尺度特征的最大池化方法.该方法能输出固定维度的特征向量,同时大幅度减少了池化层的模型参数.基于这种新型池化方法再融合哈希函数的相似度信息,提出了一种新型服装图像检索算法.实验证明,这种算法能明显提升检索的准确度.

关键词:图像检索;残差神经网络;池化;哈希算法

中图分类号:TP312 **文献标识码:**A **文章编号:**1671-119X(2020)03-0048-06

DOI:10.15987/j.cnki.hgbjz.2020.03.009

0 引言

伴随着电子商务蓬勃发展,服装图像数据量呈突飞猛进的增长趋势,使用互联网在线上对服装图像进行快速又准确地检索受到广泛关注.为了帮助客户在电商平台上快速准确地搜索自己想要的服装,大量的服装图像搜索引擎软件层出不穷.传统的通过语义文字等进行查找图像的方式已经不能满足时代的发展潮流,并且基于颜色、纹理、轮廓等主观性的低层视觉特征进行检索时常伴随着语义鸿沟的现象.其后发展起来的“以图搜图”在一定程度上提升了服装图像的检索精度,但不同的人对同一幅图片往往会有不同的理解,这种主观性以及不确定性会影响检索的效果.针对这些问题,大量的人员对此进行了研究.Yang等^[1]人通过特征点的信息进行全局表述得到了一个SCSPM模型,其稀疏编码在进行量化过程中依赖于手工设计的底层特征,然而其在进行特征信息提取时耗时且困难.薛培培等^[2]在图片颜色特征和尺度不变特征基础上通过支持向量机来提高图像的检索率,但是该计算

量繁琐冗长.Tang等^[3]提出了进行矩阵分解训练后的跨模态哈希算法,在监督的基础上,此算法克服了小数据集对特征信息表达不够的问题,使图文数据对的标签信息在大规模的数据集中所达到的效果也比较乐观,但是其忽略了跨模态数据之间的语义一致性,使得检索出来的结果不如人意.

为了解决图像检索精度不理想和匹配时间较长等问题,本文提出了一种新型基于卷积神经网络的多特征服装图像检索方法.首先,通过神经网络的最后一个卷积层作为特征层用来捕获服装图像样式、图案、轮廓等特征信息;再利用多尺度特征新型最大池化层输出固定维度的特征向量;最后融合哈希函数的相似度信息来降低算法运行时间,有效地提升检索的准确度.

1 融合残差思想的神经网络

随着深度的加深,卷积神经网络能准确地提取出图像细节特征,与此同时,深层次的卷积神经网络常伴随着梯度消失的问题,在反向传播的过程中,导致网络特征参数或权值更新缓慢,因此本文

收稿日期:2020-01-18

基金项目:湖南省教育厅重点资助项目(18A343).

作者简介:吴帆(1997-),女,硕士研究生,研究方向:服装图像的检索及应用.

通过融合残差学习的思想来优化卷积神经网络模型性能,如图1所示,相对于普通的卷积神经网络^[4],引入快捷连接,有效避免了网络不断加深过程中导致的梯度消失等问题,并且在恒等映射过程中快捷连接不会产生额外的参数和增加计算复杂度.

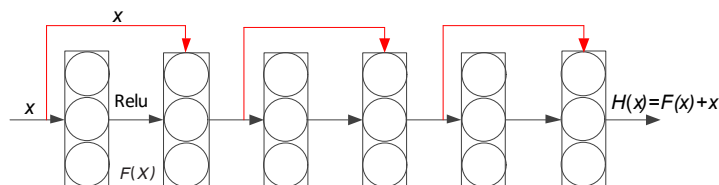


图1 ResNet结构示意图

为了实现对大规模服装图像的快速、高效检索,本文采用ResNet-34这个预测模型作为提取服装图像的卷积子网络^[5],将原始的ResNet-34最后一个卷积层后的平均池化层和全连接层去掉,将其替换成适应哈希函数需求的一些层.

2 一种基于多尺度特征的新型最大池化层

传统的卷积神经网络把图像输入全连接层时,图像的维度必须是固定的^[6],因此在输入全连接层

前,图像会被裁剪或者缩放^[7],在固定其维度等操作过程中容易丢失特征信息.因此,本文设计了一种新型池化方法,使不同维度的图像经过池化后得到固定的维度输出.这种池化方法示意图如图2所示,是在最后一个卷积层后连接该池化层.其具体过程如下:首先把最后一个卷积层输出的图像信息复制4份,然后按照不同尺度进行划分区域的计算公式如下所示:

$$H(x) = F(x) + x \quad (1)$$

其中 $F(x)$ 表示残差, $H(x)$ 表示的是网络的映射输出.

$$a = 2\min(W, H)/(l + 1) \quad (2)$$

其中,按照尺度均匀进行划分区域 $l = 1, 2, 3, 4$.

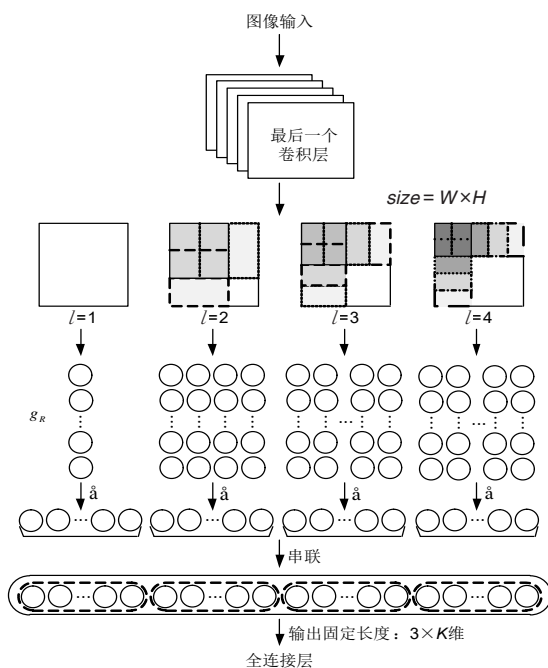


图2 一种基于多尺度特征的新型最大池化层示意图

划分完区域后进行最大池化计算,设最后一个卷积层输出的 k 个特征信息图尺寸为 $W \times H$,表示为 $X=(X_i), i=1,2,3 \cdots, k$,则方形区域 $R \subseteq [1, W] \times [1, H]$ 的特征向量定义如下:

$$g_R = [g_{R,1} \cdots g_{R,i} \cdots g_{R,k}]^T \quad (3)$$

其中, $g_{R,i} = \max_{p \in R, X_i(p)}$ 表示在特征通道 R 区域第 i 个位置中的最大值。

特征图片按照几种不同的尺寸被最大池化后,会相应得到 K 个区域特征向量 g_R ,通过交叉匹配的方式把区域中相同尺度的向量归为一列,最后将这些区域所有尺度向量串联成一列 $4 \times k$ 维的多尺度向量。

表1给出了一个本文池化过程的例子, $l=1 \times 1 \cdots, l=4 \times 4$ 为池化的4种尺寸,最后一个卷积层输出256个尺寸为 6×6 的特征图,将这些区域的向量分别简单相加为4列长度为256维度的特征向量,最后串联为一列 $4 \times 256 = 1024$ 维度的多尺度特征向量,通过该池化后,不用裁剪图片尺寸也能得到固定维度的向量,解决了图像信息损失的问题。

表1 池化结构示例

l	1	2	3	4
type	max pool	max pool	max pool	max pool
inputs	6×6	6×6	6×6	6×6
kernel size	6	4	2	3
stride	6	2	2	1

3 损失函数

图像经过哈希层后将高维特征映射到低维空间中,然后生成相应的二进制哈希编码,最后用海明距离计算查询图片哈希编码到其他图片的哈希编码距离,同类别图像中,汉明距离越小,不同类别的图像中,汉明距离越大^[8]。如图3所示,本文使用两种损失函数得出两种分类损失,在优化损失函数过程中学习哈希函数的参数,从而快速的进行图片检索。假设哈希函数集合为 $\{f_1, f_2, \cdots, f_n\}$,得出多个编码为 $f: R^d \rightarrow \{-1, 1\}$ 的哈希函数,通过最优化目

标函数策略学习哈希函数的参数 w_1, w_2, \cdots, w_b , 一对点 (x_i, x_j) 。

3.1 铰链损失函数

方差损失函数定义如下:

$$l(f(x_i), f(x_j); w) = (f(x_i)^T f(x_j) - bs_{ij})^2 \quad (4)$$

为了提高算法的高效性并且控制量化误差,需要对其进行优化:

$$\min_{w \in R^{a \times b}} J(W) = \sum_{ij} l(f(x_i), f(x_j); W) = \|F^T F - bS\|_F^2 \quad (5)$$

使用梯度下降的方法进行迭代,每迭代一次便随机选取一对点,则更新的权重公式如下:

$$W^{t+1} = W^t - \eta \nabla_w l(f(x_i), f(x_j); W^t) \quad (6)$$

为了使更新的汉明距离误差更小,基于平方误差损失的更新采用铰链损失函数替代。

引进铰链损失函数:

$$L_h(f(x_i), f(x_j)) = \begin{cases} \max(0, d_H - (1-\alpha)bs_{ij}) & s_{ij}=1 \\ \max(0, \alpha b - d_H) & s_{ij}=-1 \end{cases} \quad (7)$$

其中, $d_H = \|f(x_i) - f(x_j)\|_1$, $\alpha \in [0, 1]$ 。

3.2 Softmax 分类损失函数

为了更好地利用图像的标签信息,本文还结合了Softmax 分类损失函数^[9],该损失函数可以使同种类的距离缩小,不同种类的距离变大,引用度量学习机制,将图像组成三元组^[10],记 (x_i^a, x_i^p, x_i^n) , x_i^p, x_i^a 是同类别, x_i^n, x_i^a 是不同类别图像,图像特征表达式 $f(x)$ 需经过L2规范化,满足 $\|f(x)\|_2 = 1$,那么要求满足如下关系:

$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2 \quad \forall (f(x_i^a), f(x_i^p), f(x_i^n)) \in T \quad (8)$$

其中 α 是阈值参数, T 是所有三元组的集合,损失函数表达式为:

$$L_t(x_i^a, x_i^p, x_i^n) = \sum_i \max \{0, \|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha\} \quad (9)$$

因此,本文算法的损失函数 L 表示为:

$$L = wL_h + (1-w)L_t \quad (10)$$

参数 w 是权重系数。

图3给出了一个损失函数示意图,图像经多尺度特征的最大池化层输出后连接隐含层fc1(节点数为500)和哈希层fc2(节点数等于哈希长度 q),通

过哈希层,得出中间特征向量后分为两路进行特征向量传输.其具体过程如下:

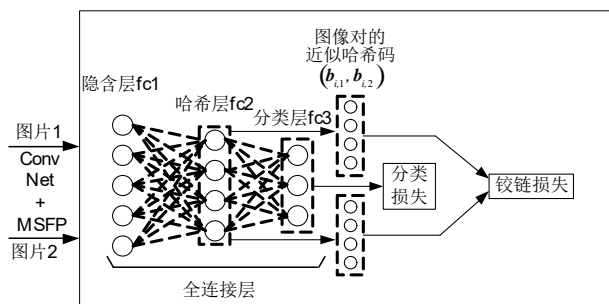


图3 损失函数示意图

(1)第一路,将经过哈希层出来的图像特征中间向量作为近似哈希二进制编码输入铰链损失函数层。

(2)另外一路,将经过哈希层出来的图像特征中间向量传输到分类层fc3(节点数为图像类的数量),再计算三元组损失函数。

4 多特征服装图像检索方法

由于普通卷积神经网络特征提取能力较差,不能准确地提取出图像细节特征,会导致检索不理想,本文基于融合残差思想的神经网络、多尺度特征的新型最大池化层和损失函数设计了一种基于卷积神经网络的多特征服装图像检索方法,如图4所示,其具体步骤如下:

步骤1:本文融合残差学习的思想来优化CNN网络模型性能,采用最后一个卷积层提取服装图像样式、图像、轮廓等特征信息。

步骤2:为了保证图片维度的固定性,特征信息的完整性,采用了多尺度融合池化层替代了传统的池化层的方法,进而避免了图像特征信息损失的现象。

步骤3:将最后一个卷积层后的平均池化层和全连接层去掉,将其替换成适应哈希函数需求的一些层,神经网络提取到的服装特征之后,再融合哈希函数计算图像间的相似度,以此来提升检索的准确度。

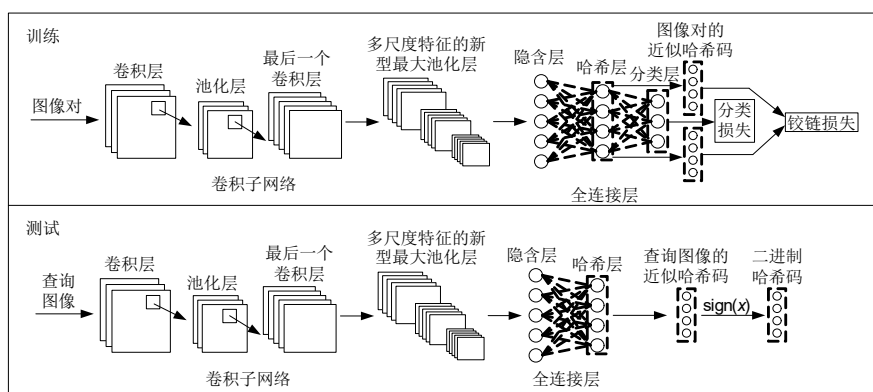


图4 基于卷积神经网络的多特征服装图像检索流程

5 实验结果和分析

本文的硬件环境是 Windows10, CPU 频率为 2GHz, 内存 4GB, Python3.59 软件, 仿真实验基于开源深度学习框架 TensorFlow 实现, 在训练使权重系数 α 、 β 和 λ 取值均为 1, 三元组损失函数学习速率的初始值取 0.01, 动量取值 0.9, 权重衰减系数取值 0.0002. 实验所用数据集源于香港中文大学开放的

一个 large-scale 数据集, 总共 289222 张衣服图片, 213642 张图片用作训练集, 75580 张图片用作测试集, 主要对服装 3 类语义属性进行预测, 图片分为 11 个类别, 每个类别代表一种数字(数字 0~10), 如表 2 所示。

表2 服装的语义属性

样式	长袖, 短袖, 无袖
轮廓	圆领, V 领, 翻领, 立领
图案	纯色, 印花, 条纹, 格子

5.1 评估分类器的性能

在性能的比较上,实验所选择的KSH、CNNH、DLBHC等算法都具有代表性,所选择的这些对比算法都是基于卷积神经网络算法和哈希算法的主流算法^[11].可由实验数据分析得出:本文提出的算法相比无监督哈希算法的特征提取更加灵活,与基于普通的卷积神经网络的算法相比,本文所提的算法将图像的相似度信息和分类信息充分地结合了起来,使检索效果得到大幅度提高.在数据集上48位编码时得到的PR曲线如图5所示.

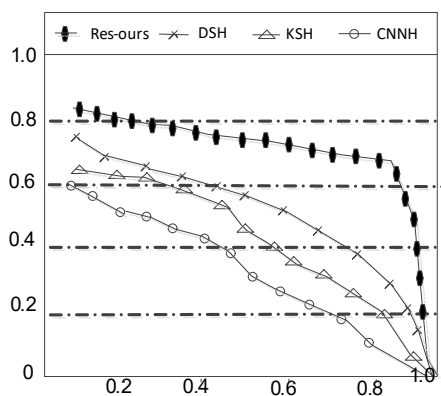


图5 48位编码时的PR曲线

5.2 池化方法测试

为了验证基于多尺度特征的新型最大池化方法的有效性,本文与SPP池化方法进行了对比,除了池化层结构不同,其他网络结构均保持不变,表3是实验对比的结果.

表3 不同池化方法的MAP对比

方法	不同哈希编码长度模型下的MAP			
	12位	24位	36位	48位
Our Way	0.91	0.83	0.79	0.74
SPP	0.85	0.81	0.75	0.72

经过两种方法的对比可知,本文的池化方法跟SPP池化方法都是在区域上进行多种尺度的特征信息提取^[12],但是本文的池化方法使各个区域相同的尺度向量通过简单的交叉匹配归为了一类 K 维特征向量,最后将这些区域所有尺度向量串联成一系列固定维度的多尺度特征向量,这种池化操作不

仅大幅度减少了池化层的模型参数,而且还保证了图像特征信息向量维度的固定,大幅度提高了检索精度.

6 结语

本文利用残差神经网络在图像特征提取上具有的层次性,准确地提取出图像细节特征,再利用多尺度特征新型最大池化层输出固定维度的特征向量,同时减少网络模型的参数,并融合哈希算法来找到同类特征信息的最大相似度,以此提高检索效率和准确度.后续我们将研究如何有效地进行三维服装图像检索.

参考文献

- [1] TANG J, WANG K, SHAO L. Supervised Matrix Factorization Hashing for Cross-Modal Retrieval [J]. IEEE Transactions on Image Processing, 2016, 25(7): 3157-316.
- [2] Yang J, Yu K, Gong Y, et al. Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification [A]. Computer Vision and Pattern Recognition, CVPR 2009[C]. IEEE Conference on, 2009(3): 1794-1801.
- [3] 薛培培, 郭延辉. 基于图像内容和支持向量机的服装图像检索方法研究[J]. 移动通信, 2016, 40(2): 79-82.
- [4] Zhuang W, Shanshan Jiang, Xiaolei Zhou, Yuanyuan Wang, Yuanyuan Zuo, Zhewei Wu, Lei Liang, Qi Liu. Application of Image Retrieval Based on Convolutional Neural Networks and Hu Invariant Moment Algorithm in Computer Telecommunications [J]. Computer Communications, 2020, 150(C): 729-738.
- [5] Lucas Pascotti Valem, Daniel Carlos Guimar, es Pedrone-tte. Unsupervised Selective Rank Fusion for Image Retrieval Tasks[J]. Neurocomputing, 2020, 85(6): 182-199.
- [6] He Kaiming, Zhang Xiangyu, Ren Shaoqing, Sun Jian. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [7] Nicolas oria, David B. Blumenthal, SÈbastien Bougleux, Luc Brun. Improved Local Search for Graph Edit Distance

- [J]. Pattern Recognition Letters, 2020, 98(5):19-25.
- [8] Leulseged Tesfaye Alemu, Marcello Pelillo. Multi-feature Fusion for Image Retrieval Using Constrained Dominant Sets[J]. Image and Vision Computing, 2020, 94(7): 21-23.
- [9] Simone Bianco, Marco Buzzelli, Gianluigi Ciocca, Ramondo Schettini. Neural Architecture Search for Image Saliency Fusion[J]. Information Fusion, 2020;57(C):89-101.
- [10] Liu Deyin, Liang Chengwu, Zhang Zhiming, Qi Lin, Lovell Brian C. Exploring Inter-Instance Relationships within the Query Set for Robust Image Set Matching [J]. Sensors (Basel, Switzerland), 2019, 19(22): 65-70.
- [11] 龚安,周静红,李华昱.基于图案的多层次跨场景服装检索[J].计算机与数字工程,2019,47(10):2556-2560.
- [12] Dariusz Dereniowski, Dorota Osula, Paweł Rządewski. Finding Small-width Connected Path Decompositions in Polynomial Time [J]. Theoretical Computer Science, 2019,794(74): 85-100.

A Clothing Image Retrieval Method Based on Convolutional Neural Network for Multi-feature

WU Fan¹, DENG Zuo-jie¹, SHANG Shu-fei²

(1. College of Computer and Communication, Hunan Institute of Engineering, Xiangtan 411104, China;

2. College of Management, Hunan Institute of Engineering, Xiangtan 411104, China)

Abstract: Due to the effect of the personalization of clothing style and background factor, the traditional neural network cannot extract image details accurately in the problem of image retrieval, which leads to unsatisfactory for image retrieval. To solve this problem, this paper builds a residual neural network, and proposes a new type of pooling method based on multi-scale feature. It can reduce the pooling level of mode parameters greatly. At the same time, it can output the information of feature vector for image dimension unchanged. A clothing image retrieval method based on convolutional neural network for multi-feature is proposed. It makes full use of the validity and hierarchy of deep convolutional neural network in image feature extraction, and the similarity information of hash function. The experiment results show that this method can improve the retrieval accuracy significantly.

Keywords: image retrieval; residual neural network; pooling; hash algorithm