

融合颜色特征和深度特征服装图像检索算法

侯媛媛 何儒汉* 刘军平

(湖北省服装信息化工程技术研究中心 湖北 武汉 430200)

摘要 针对残差网络直接应用于服装图像检索时,检索出的图片常存在款式相似但颜色差异较大的问题,提出一种融合传统颜色特征和基于 ResNet50 的深度特征的服装图像检索方法。充分利用残差网络在图像特征提取上的有效性和层次性,融合服装图片的深层网络特征和颜色特征作为最终特征向量,进行相似性度量,得到检索结果。实验结果表明,该方法能结合图片多种特征信息,提高了其检索准确率,具有较强的鲁棒性和实用性,优于其他主流的检索方法。

关键词 服装图像检索 残差网络 深层网络特征 颜色特征

中图分类号 TP3 文献标志码 A DOI:10.3969/j.issn.1000-386x.2020.10.031

FUSION COLOR FEATURE AND DEPTH FEATURE CLOTHING IMAGE RETRIEVAL ALGORITHM

Hou Yuanyuan He Ruhan* Liu Junping

(Engineering Research Center of Hubei Province for Clothing Information, Wuhan 430200, Hubei, China)

Abstract When the residual network is directly applied to clothing image retrieval, the images retrieved often have similar styles but different color differences. Therefore, we propose a clothing image retrieval method combining traditional color features and depth features based on ResNet50. It made full use of the effectiveness and hierarchy of residual network in image feature extraction, and integrated deep network features and color features of clothing images as the final feature vector to measure the similarity and get the retrieval results. The experimental results show that the method can combine multiple feature information of images, and improve the retrieval accuracy. It has strong robustness and practicability, which is superior to other mainstream retrieval methods.

Keywords Clothing image retrieval Residual network Deep network feature Color feature

0 引言

伴随电子商务行业的快速发展,服装行业作为其中重要一部分,其数据量不断增多。为了处理海量的服装图像数据,一种全新的线上服装搜索模式被用户所使用——“以图搜图”,其核心是图像检索技术。服装图像检索作为服装智能推荐、服装搜索等应用的核心,具有广泛的市场应用前景。服装展现出当代人的一种潮流趋势及品味,大量的语义及细节信息蕴含其

中,服装的色彩搭配及款式是其重要的语义信息,纹理和材质等表现其细节信息。然而服装不规则的外形、柔软的材质、较强的可变形性以及光照的敏感性等,使服装图像检索快速精准的实现成为了一个极具挑战的难题。基于内容的图像检索(Content-based Image Retrieval, CBIR)^[1-5]是服装图像检索领域研究的主流方法。其过程可大致分为两类:①提取数据库图的表示向量;②将待检索图片的表示向量与库向量中的每个向量进行距离度量,以最近邻方式进行搜索。一幅图片表示向量的好坏直接决定了图像检索算法的优

劣。早期的研究主要是基于颜色、形状和纹理三大视觉特征,将其单一特征^[6-7]或聚合特征^[8]作为服装图像的全局表示向量。黄冬艳等^[9]首先对待检索图片做了联合分割处理,然后提取出分割后的服装图片的 Bundled 和颜色特征,最后与特征库中的服装图像特征进行相似性度量。陶彬娇等^[10]将分块加权颜色直方图与 Grabcut 图像分割算法相结合,提升了对复杂背景的服装图片的检索效果,优于分别单独使用这两种方法。葛俊等^[11]进行两次检索,首先利用颜色直方图,再使用局部二值模式(Local Binary Patterns, LBP)算子对第一次的返回结果进行二次检索。Gupta 等^[12]获得服装图片的颜色和纹理的特征,通过计算向量距离得出检索结果。陈倩等^[13]计算颜色直方图的特征值占比,然后将其进行降序排列,去掉占比较小的值,则服装图片的主要颜色被保留并用于图像检索。近几年,随着深度学习的兴起,其在处理计算机视觉任务上有一系列重大的突破,此方法利用深度神经网络从大量的训练数据集中提取深层特征,发现训练的图片信息中隐藏的特征。由于卷积神经网络(Convolutional Neural Networks, CNN)^[14]具备较强的非线性表示能力,能学习到图片更深层次的信息,所以在图像分类^[15-17]、图像分割^[18]和目标检测^[19-21]等领域都表现出较好的性能。如今基于卷积神经网络的服装图像检索方法成为了该研究的主流方向。

早期 CNN 在图像检索方面的应用,主要是提取深度网络的全连接层作为图片的特征向量^[22-23]。林城龙^[24]提出了一个更加轻量的深度卷积神经网络模型(Lighten-VGGNet),采用多任务分类方法对层次化标注后的服装图像数据分类,使得网络在分类时所提取的特征拥有对更多细化服装类别属性的表示能力,从而使服装图像分类准确率得以提升。然而只提取全连接层无法保持空间结构,且该特征更多的是表示全局信息,丢失了服装图片的局部特征信息,导致检索的平均精度均值(mean Average Precision, mAP)较低,所以融合多种服装图片特征,成为服装图像检索的热门研究方法^[25-29]。陈媛媛等^[25]基于深度卷积神经网络定位服装关键点,将关键点局部区域的特征与全局特征进行融合。Huang 等^[26]提出双路神经网络模型,其不仅提取全连接层特征,同时对靠前的卷积特征层下采样,然后将结果与全局特征组合在一起。Liu 等^[27]设计了目前最大的标注服装图像数据库 DeepFashion(超过 80 万幅图片,50 个细粒度类别和 1 000 个属性),其将 VGG 16^[15]最后一层卷积分为三个分支,一分支进

行特征点可见性预测以及位置回归,另外两个分支分别提取图片的局部特征(有助于应对服饰变形和遮挡问题)和全局特征。Li 等^[28]采用多任务深度学习框架来学习表示,提出了用于不平衡学习的多重深度卷积神经网络。陈彦杰^[29]利用了 Faster RCNN^[20]的框架,首先定位服装的位置,然后利用关键点定位对服装进行特征点(landmark)定位,抽取各特征点附近的服装特征作为局部特征,最后和全局分支的特征进行特征融合作为最终的特征表示。

虽然上述方法融合了局部和全局特征且都在一定程度上提高了检索的 mAP,但是检索出的服装款式相似而颜色差异较大。在服饰上,色彩起到了视觉醒目的作用,人们首先看到的是颜色,其次才是其样式、纹理等,所以颜色在服装图片分类检索中是一个重要属性,是评判检索优劣的重要指标之一。而通过深度网络获取的特征涵盖的颜色信息较少,所以本文提出一种新的服装图像检索方法,即以 ResNet50^[17]为基础网络,提取出其深层特征和颜色特征,融合全局特征和局部特征进行检索,综合考虑服装图片的多种特征。在一个较大规模的服装图像数据集 deepfashion^[27]的一个子集上进行,综合利用 ResNet50 提取图像的特征:深层网络特征和颜色特征,利用 ResNet50 全局平均池化(Global Average Pooling, GAP)层降低模型的参数数量,与平均池化后的原图对比获取颜色特征,融合颜色特征和深层网络特征,获得较好的特征表达能力,有益于取得较好的检索准确率。

1 基于深度网络的多特征融合服装图像检索框架

基于深度网络的多特征融合服装图像检索包括特征提取和相似性度量两个过程,如图 1 所示。特征提取过程中,先将数据集中的图片输入到预训练好的网络模型中,提取出通过网络层输出的深层特征,再使用聚合方法融合其他的特征信息作为图像的全局特征表示,存入到特征库中;相似性度量过程,是将待检索的服装图片输入到与数据集相同的神经网络中,并使用同种聚合方法,获取待查询服装图片的全局特征向量,通过比较待查询图片特征向量与特征库中向量的距离,来进行相似度的排序,按距离升序排列返回检索结果。特征提取方法的好坏直接影响了检索的准确率,这是本文研究的重点。

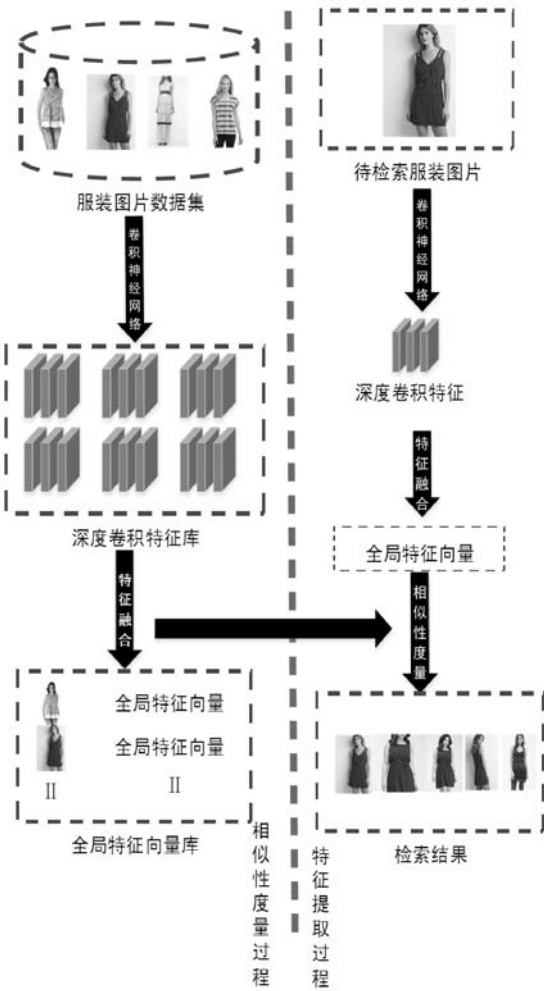


图1 基于深度网络的多特征融合服装图像检索框架

2 融合颜色特征和残差网络深度特征的方法

2.1 模型设计

针对直接提取全连接层的特征进行服装图像检索而带来的颜色效果不鲜明的问題,充分提取服装图片的特征,提高其检索的平均准确率,提出一种融合颜色特征和深度特征的服装图像检索方法。因实验数据集较大,网络层的加深有利于训练大型数据集,但有时深层网络训练时出现过拟合效果反而没有浅层网络好。针对这种情况,残差网络^[17]随之出现在各种计算机视觉领域被广泛应用,解决了网络层越多效果反而越不明朗的退化现象以及网络层增加梯度爆炸等问題,且 ResNet50^[17]网络自身融合了多层网络结构,具有一定的层次性。所以本文实验预训练网络模型以其为基础,去掉其最后的两层全连接层,增加 FC1 层和 FC2 层,分别输出 512 维和 30 维(本实验数据集选取了 30 种类别的图片)的特征向量。融合颜色特征的深度网

络模型如图 2 所示。

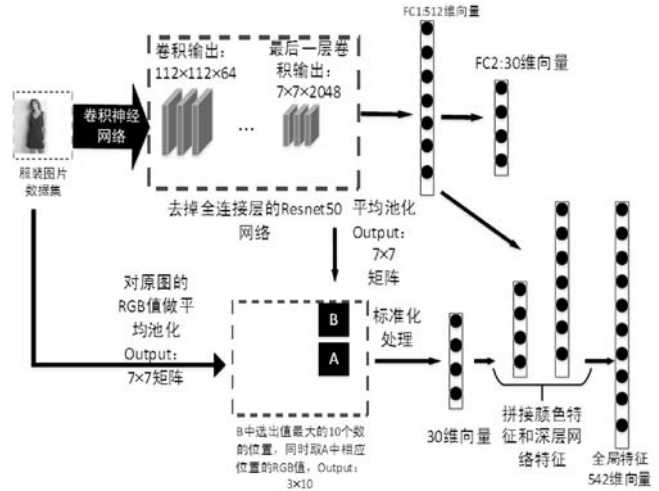


图2 融合颜色特征的深度网络模型

2.2 方法概述

- (1) 将训练数据集以 224×224 大小输入到 Res-Net50 为基础的网络模型中。
- (2) 融合深层特征和颜色特征得到图片全局特征,并建立特征库。提取 FC1 层 512 维向量作为深度特征向量;网络层的最后一层卷积层提取特征图进行平均池化,得到一个矩阵,记录最大的 10 个值的位置,对原图的 RGB 同样做平均池化得到矩阵,根据以上选出的 10 个位置,取出原图池化后矩阵相应位置的 RGB 值,得到一个 10×3 的二维矩阵,将其标准化后转换为一个 30 维的向量,作为图片的颜色特征;将得到的两种向量进行串行合并,输出 542 维向量作为图片最终的全局特征表示,保存在特征库中。

(3) 使用 K-means 算法对特征库中的向量进行聚类,设置类簇数 $K = 30$ (服装数据集类别数维 30)。

(4) 用相同方式获取待检索图片的全局特征向量。

(5) 依次计算聚类中心的向量与待检索图片向量的距离(为负数),得到与目标图片特征向量距离最近的 N 个结果对应的图片作为检索结果显示。 N 可取值为 5、10、20 三个参数。

2.3 深层网络特征提取

向深度卷积网络输入一个 $C \times H \times W$ 的三维张量, C 表示特征的通道数,初始值为 3; H 和 W 分别表示输入图片的高和宽的像素大小。在图 2 中,图片经过多层卷积操作,每一层卷积的输入为上一层的输出,计算方法如下:

$$X_j^l = \text{ReLU} \left(\sum_{i \in M_j^n} X_i^{l-1} \times k_{ij}^l + b_j^l \right) \quad (1)$$

式中: X_j^l 表示第 l 层的第 j 个运算结果; M_j^n 表示第 n 层第 j 个输出特征图所对应的多个输入特征图的索引集合; b_j^l 表示偏置项,其被所有输入特征图共享; k_{ij}^l 表示第 l 层一个 $i \times j$ 大小的卷积核;ReLU 为激活函数,公式如下:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ x & x > 0 \end{cases} \quad (2)$$

经过卷积后的特征图维数较高、参数较多,增加了计算的时间和难度,所以池化层作为一个重要操作出现在卷积操作之后,其不仅减小了网络模型的计算难度,而且提升了模型的泛化能力。残差网络采用了平均池化(mean-pooling)对邻域内特征点求平均值,池化操作公式如下:

$$X_j^l = f(\beta_j^l \text{down}(X_j^{l-1}) + b_j^l) \quad (3)$$

式中:down()为池化操作过程,池化操作输出结果乘以权重 β 加上偏置 b 之后,再经过激活函数 $f()$ 输出最终结果。

如图 2 所示,以 ResNet50^[17] 为主网络模型,其最后的卷积层输出若干个 7×7 的卷积结果,去掉其最后的两层全连接层,加入两个全连接层:FC1 层和 FC2 层,分别输出 512 维和 30 维(本实验数据集选取了 30 种类别的图片)的特征向量。对网络模型进行微调,使用已有的 ResNet50 模型及参数,来训练全连接层的参数,得到一个拟合服装图片数据集的网络模型,提取 FC1 层的 512 维的向量,作为本文的深层网络特征。FC1 层融合了 ResNet50 网络中三个模块的网络特征,使提取的深层特征具有一定的有效性和层次性。

2.4 服装颜色特征提取

网络模型的最后一层卷积层中有 2 048 个 7×7 卷积,因此可以直接利用卷积层后的特征图提取信息较强位置的 RGB 值。经过几十次卷积操作之后保留的特征图的信息更加具有针对性,对其进行平均池化后得出 7×7 的矩阵,在其中选择最大的 10 个值,并记录其位置。然后将原图的 RGB 值做平均池化,池化为 7×7 的矩阵,根据上一步得到的 10 个值最大的位置,选出图片中相对应的 10 个 RGB 值得到 3×10 维的颜色特征,将该矩阵标准化处理,z-score 标准化公式如下:

$$z = \frac{x - \mu}{\sigma} \quad (4)$$

式中: x 表示真实值; μ 表示矩阵的平均值; σ 表示标准差。之后将矩阵转换为一个 30 维向量。最后将全连接层得到的 512 维向量与 30 维颜色向量进行并行连

接,得到一个 542 维的特征向量,作为最终的特征表示。

3 实验

3.1 数据和参数准备

为了验证本文所提出方法的效果,本实验选用了 Category and Attribute Prediction Benchmark 作为数据集,其包含 20 多万数量级的 50 种类别的服装图片,本实验从其子集中抽取 6 万训练集、2 万测试集和 2 万验证集进行试验,其中包含 30 种类别的图片。实验采用 Python 编译并实现。

特征库中所有图像的深度特征均是基于 Pytorch 框架的网络层所提取的,这些网络参数是在 ImageNet 数据集上预先训练好的。

3.2 评价指标

在检索策略中,通常采用平均精度均值(mean Average Precision, mAP)来判断检索效果的优劣,一次检索完成后,与待检索图片相似的图片数量占检索出图片总数量的比例即为检索准确率。定义如下:

$$\text{准确率} = \frac{\text{检索出的相关图片数}}{\text{检索出的总图片数}} \times 100\% \quad (5)$$

$$\text{mAP} = \frac{\text{多次准确率的和}}{\text{检索的次数}} \times 100\% \quad (6)$$

3.3 实验结果与分析

为判断融合服装颜色特征和深层网络特征图像检索方法的性能优越性,本文展开了一系列对比实验,分别以 ImageNet 上预训练的网络模型:VGG16、GoogleNet、ResNet50 为基础进行微调,并提取倒数第二层全连接层的特征,然后以 ResNet50 网络结构为基础提取两种特征并进行融合(ResNet50 + Color)。以四种方法进行实验对比,同时都采用 K-means 聚类进行检索,根据实验检索返回的前 5、10、20 幅图片来计算准确率,实验结果如表 1 所示。可以看出,在服装数据集上,选择的几种卷积神经网络模型中,ResNet50 的效果更好。所以本文选择以 ResNet50 为基础模型,并进行微调,同时增加了颜色特征,本文改进方法的实验结果明显优于没有增加颜色特征的方法,相比之下检索 mAP 在 $N=5, 10, 20$ 时分别提升了 4.45%、6.85%、2.49%。因为融合的多种特征比单一的特征更能表达图片的信息,所以本文提出的方法能得到更优的检索结果。

表 1 算法模型的检索 mAP 比较 %

算法模型	$N = 5$	$N = 10$	$N = 20$
VGG16	82.32	78.26	57.22
GoogleNet	85.67	80.75	65.64
ResNet50	88.37	82.35	68.83
ResNet50 + Color	92.82	89.24	71.32

因服装数据集类别较多,所以从 30 种类别中抽取了 8 种,在四种不同的算法下进行实验,每种服装图像 $N = 5$ 时的检索准确率如图 3 所示。从单个类别的服装来看,由于每种类别服装图片数量不同造成准确率有所波动,但是本文方法的平均检索准确率依然高于其他方法。

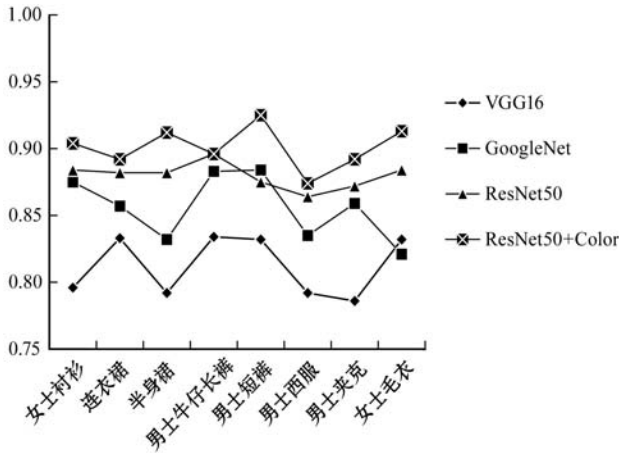
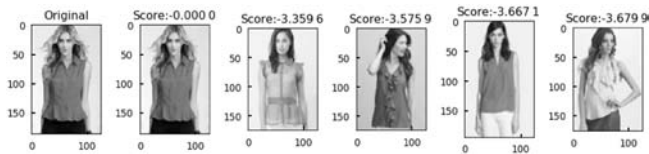
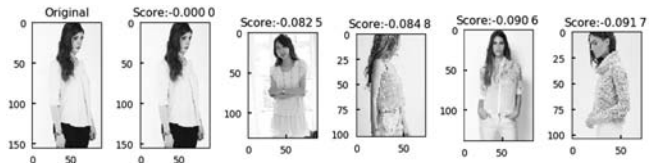


图 3 不同方法下每类服装的 mAP ($N = 5$)

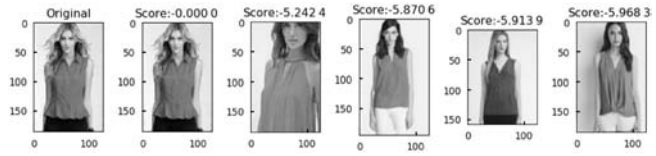
ResNet50 与 ResNet50 + Color 的特征提取检索对比结果如图 4 所示,其中:Score 表示待检索图片向量与检索出的图片向量的距离计算结果,结果取负数,Score 为 0 表示两幅图片相似度最高,负数越大表示相识度越低。可以看出,只提取全连接层特征,服装的款式虽然相似,但是颜色差异较大;(c)和(d)融合了两种特征之后,不论是服装的颜色还是款式,都能得到预期的效果,所以融合两种特征明显提高了服装图像检索的 mAP,比单一的特征更能表示图片信息。



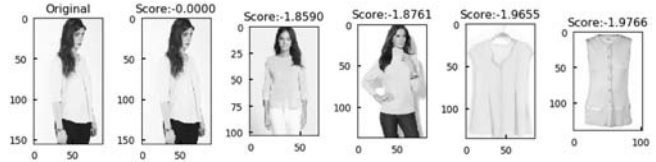
(a) 只提取 ResNet50 全连接层特征检索结果



(b) 只提取 ResNet50 全连接层特征检索结果



(c) 融合颜色和深层特征的检索结果



(d) 融合颜色和深层特征的检索结果

图 4 ResNet50 与 ResNet50 + Color 特征提取检索对比图

表 2 为 $N = 5$ 时只提取 ResNet50 全连接层特征和提取两种特征并进行融合检索的准确率和时间对比。可以明显看出,融合了颜色特征和深层网络特征之后进行检索虽然提高了检索精度,但是检索时间增加了,原因是使用两种特征融合增加了特征表示向量的维度,所以在进行相似性度时耗时要长一些。

表 2 ResNet50 与 ResNet50 + Color 方法检索的 mAP 与时间比较

算法模型	mAP/%	时间/s
ResNet50	88.37	0.032
ResNet50 + Color	92.82	0.053

4 结 语

本文提出的融合颜色特征和深度特征服装图像检索算法,通过 ResNet50 预训练网络模型提取服装图片自身深层网络特征和颜色特征,将两个特征向量进行并行连接,最终通过 K-means 对特征进行聚类检索,从而提高检索效率,减小时间开销。实验对比结果显示,本文算法的 mAP 明显要高于直接使用 ResNet50 深度网络提取的单一特征算法,且显示的检索图片效果款式和颜色相似性明显,检索时间虽然比单一特征提取耗时长,但是相差并不大,对检索没有造成很大的影响。深度学习需要大量的数据量支持,后续可以增加服装图片量或者换其他数据集进行实验验证,优化网络模型,以进一步提高检索的准确率。

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] Lew M S, Sebe N, Djeraba C, et al. Content-based multimedia information retrieval: State of the art and challenges [J]. *ACM Transactions on Multimedia Computing Communications and Applications*, 2006, 2(1): 1-19.
- [3] Liu Y, Zhang D S, Lu G J, et al. A survey of content-based image retrieval with high-level semantics [J]. *Pattern Recognition*, 2007, 40(1): 262-282.
- [4] Zhang L N, Wang L P, Lin W S. Generalized biased discriminant analysis for content-based image retrieval [J]. *IEEE Transactions on Systems, Man and Cybernetics Part B*, 2012, 42(1): 282-290.
- [5] Ji W, Wang D Y, Hoi S C H, et al. Deep learning for content-based image retrieval: A comprehensive study [C]// *ACM International Conference on Multimedia*, 2014.
- [6] Lu Z Y, Shen L, Yuan Y L, et al. Semantic classification based clothing image retrieval [C]// *2012 International Conference on Audio, Language and Image Processing*. IEEE, 2012: 876-883.
- [7] Wang X W, Zhang T, Tretter, D R, et al. Personal clothing retrieval on photo collections by color and attributes [J]. *IEEE Transactions on Multimedia*, 2013, 15(8): 2035-2045.
- [8] Hou A, Zhao L Q, Shi D C. Garment image retrieval based on multi-features [C]// *2010 International Conference on Computer, Mechatronics, Control and Electronic Engineering*. IEEE, 2010: 194-197.
- [9] 黄冬艳, 刘骊, 付晓东, 等. 联合分割和特征匹配的服装图像检索 [J]. *计算机辅助设计与图形学学报*, 2017, 29(6): 1075-1084.
- [10] 陶彬娇, 陈倩, 潘中良, 等. 服装检索中分块加权颜色直方图方法的改进 [J]. *激光杂志*, 2017, 38(8): 97-101.
- [11] 葛俊, 于威威. 一种基于加权颜色形状特征和 LBP 服装图像检索方法 [J]. *现代计算机(专业版)*, 2018(19): 35-40.
- [12] Gupta M, Bhatnagar C, Jalal A S. Clothing Image Retrieval Based On Multiple Features for Smarter Shopping [J]. *Procedia Computer Science*, 2018, 125: 143-148.
- [13] 陈倩, 陶彬娇, 潘中良, 等. 主颜色提取算法及其在服装图像检索中的应用 [J]. *华南师范大学学报(自然科学版)*, 2019, 51(1): 116-124.
- [14] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
- [15] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. *Advances in Neural Information Processing Systems*, 2012, 25(2): 1110-1111.
- [16] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions [C]// *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2015: 1-9.
- [17] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C]// *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016: 770-778.
- [18] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 39(4): 640-651.
- [19] He K M, Zhang X Y, Ren S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 37(9): 1904-1916.
- [20] Ren S Q, He K L, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [21] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection [C]// *2016 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2016: 779-788.
- [22] Babenko A, Slesarev A, Chigorin A, et al. Neural codes for image retrieval [C]// *European Conference on Computer Vision*. Springer, 2014.
- [23] Lin K, Yang H F, Liu K H, et al. Rapid clothing retrieval via deep learning of binary codes and hierarchical search [C]// *Proceedings of the 5th ACM International Conference on Multimedia Retrieval*. ACM, 2015: 498-502.
- [24] 林城龙. 基于多任务卷积神经网络的服装图像分类与检索 [D]. 北京: 北京化工大学, 2018.
- [25] 陈媛媛, 李来, 刘光灿, 等. 基于关键点的服装检索 [J]. *计算机应用*, 2017, 37(11): 3249-3255.
- [26] Huang J S, Feris R S, Chen Q, et al. Cross-domain image retrieval with a dual attribute-aware ranking network [C]// *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015: 1062-1070.
- [27] Liu Z W, Luo P, Qiu S, et al. DeepFashion: Powering robust clothes recognition and retrieval with rich annotations [C]// *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016: 1096-1104.
- [28] Li R F, Feng F X, Ahmad I, et al. Retrieving real world clothing images via multi-weight deep convolutional neural networks [J]. *Cluster Computing*, 2017(8): 1-12.
- [29] 陈彦杰. 基于深度网络的在线服装检索及推荐 [D]. 北京: 北京邮电大学, 2018.