文章编号：2095-2163(2020)10-0014-05 中图分类号：TP181 文献标志码：A

跨场景时尚图像的在线提取

阿卜杜杰力力·热合麦提
（东华大学 计算机科学与技术学院，上海 201600）

摘 要：随着智能手机的普及，人们可以随时随地购物，但基于关键字的搜索很难准确检索特定服装款式。当看到想要的物品时，基于内容的在线检索方法可以在不知道确切的文本描述的情况下带来极大的便利性。然而，由于购物网站的图片是在专业的灯光、场景布局下拍摄的，而实时图像在背景、灯光等方面都有所不同，这使得最相似的物品很难匹配。如何在不同的拍摄角度下减少背景噪声干扰，获得准确率的检索结果是一个挑战。目前大规模的时尚图像数据集很容易获取图像特征，机器学习方法可以对数据进行预处理，消除背景干扰，提高不同角度下的检索精度。由于跨场景的不确定性，本文首先使用目标检测算法目标定位，找出需要检索的目标商品，再进行图片分割；其次，使用卷积神经网络对图片进行特征提取；最后，在图片数据库中找出与其最相似的数据图片。本文提出的不同的检索方式可以满足不同用户的不同需求，给予用户更好的体验。

关键词：跨场景；时尚图像；机器学习；特征提取；目标检测

Cross-scenario Online Retrieval of Fashion Images

Abudujielili Rehemaiti
（College of computer science and technology, Donghua University, Shanghai 201600, China）

【Abstract】 With the popularity of smartphones, people can shop anytime, anywhere, but in the past, keyword-based searches were difficult to accurately retrieve for specific clothing styles. When you see the clothes you want, content-based online retrieval method can bring great convenience without knowing the exact text description. However, since the image of the shopping website is shot in a professional lighting, scene layout, etc., and the real-time image is different in background, lighting, which makes it difficult to match the most similar item. how to reduce the background noise interference and obtain accurate retrieval results under different shooting angles is a challenge. Nowadays large-scale fashion image data set could be obtained and could be implemented in our research to retrieve abundant features, machine learning method could preprocess the data to remove background interference and improve the retrieval accuracy under different angles. This paper designs a fashion item recommendation system that can provide the best matching products in real time and accurately according to the pictures given. Our goal is to provide innovative models and methods as well as new technologies that contribute to research and future industrial applications.

【Key words】 Cross-scenario; Fashion image; Machine learning; Feature extraction; Target detection

0 引 言

随着时尚电商平台的普及和图像共享网站的快速发展，在线服装贸易市场巨大，时尚物品日益多样化，学术界和工业界越加关注相关的研究和应用。仅仅提供文字检索功能很难满足用户对物品更精细的查询要求；而且随着移动计算以及互联网的快速发展，街拍和社交网络分享逐日流行。基于图像的检索可以帮助用户通过街拍形式从品种繁杂、造型多样的物品中快速而准确地定位，大幅改善购物体验。不过街拍图像会受到背景、灯光、构图等影响，质量上远不如电商平台中较专业的图片，在对跨场景的图像进行匹配时存在难度，街拍图像的准确处理和特征精确提取是匹配的关键所在[1]。

街拍时尚物品查询模型一般分为以下步骤，用户拍照，通过在线电商 APP 上传，服务器对街拍图像和物品库图像进行相似性匹配，返回相似度最高的 top-k 个物品，如图 1 所示。衡量时尚图像的相似性是一个复杂的问题，时尚图像相似性不仅仅在于两个图像所对应的像素矩阵数值的相似程度，更重要的在于人眼所感知的两个物品的款式、色彩、花纹，以及其他细节是否相似。可以把特征分为两类：整体和局部。如两张关于 T 恤衫的图片，外表看上去很相似，但是其印花细节差别较大，或者两者的细节相似度很高，但是款式不同，只有同时满足整体（轮廓）和局部（细节）的物品才能被认为相似[2]。

跨场景的时尚物品的在线检索存在几个挑战：如何在保证整体和局部特征完整性的情况下满足实时的特征提取；如何在特征提取后进行快速的图像检索，找到相似度最高的 top-k 个物品[3]。

随着深度学习的快速发展，利用卷积神经网络相关方法进行图像特征提取和表示已经成为解决时尚领域图像检索和推荐的关键方法[4]。已有相关

作者简介：阿卜杜杰力力·热合麦提（1992-），男，硕士研究生，主要研究方向：大数据处理、机器学习、时尚图像等。

收稿日期：2020-05-21

方法注重时尚图片检索的准确性，包括结合视觉和非视觉特征的匹配[5]，结合多模态的查询[6]，基于图像分割和目标检测的匹配方法[7]；或者以适当降低精确度为代价，提高在线处理速度[8]。

图 1　基于街拍的在线时尚图像检索示例

Fig. 1　Examples of online fashion image retrieval based on street photography

本文在设计在线时尚物品检索框架时，采用目标检测算法来避免图片中目标物体大小带来的影响，采用图像分割算法来避免图片中复杂背景的影响，采用缓存提高匹配速度。框架既具有在线检索的实时性，也能兼顾检索的准确率，还能通过参数选择及时调整检索策略。

1　相关工作

1.1　图像处理方法

目前，国内外基于内容的图像搜索相关的技术有很多，主要可以分为传统方法和基于卷积神经网络的方法两种。基于内容的图像检索主要包括特征提取、相似性定义以及弥补语义鸿沟 3 个步骤。

在传统方法中，使用 SIFT、HOG、SURF 等方法来进行图像特征的提取，这些方法提取的还是图像像素级别的特征。如果提取的特征具有结构性的时，深度学习算法才能发挥作用。

随着深度学习的发展，通过神经网络提取特征得到了广泛的应用，以卷积神经网络为基础的一系列深度神经网络，诸如 AlexNet、VGGNet、GoogLeNet、ResNet 等，在计算机视觉领域也取得了卓越的成就。

1.2　图像特征提取

在基于内容的图像查询中，存在底层特征和上层理解之间的差异，主要原因是底层特征不能完全反映或者匹配查询意图。弥补这个鸿沟的技术手段主要有相关反馈（relevance feedback）、图像分割（image segmentation）和建立复杂的分类模型。根据用户对于查询结果进行评分来更新系统，为每对图片之间分配的相似性大小[9]，在相关反馈技术的基础上提出了核距离方法的应用，将每个图片向量映射到同一特征空间中，并使用 one-class SVM 核距离来计算 2 个向量之间的距离，最后建立 M-tree 来建立索引，用来查询最后需要输出的图片[10]。对图像进行像素级的分类，从而解决了语义级别的图像分割问题，与传统的 CNN 不同的是 FCN 可以接受任意尺寸的输入图像，并且最后输出的是一张已经标签好的图片[11]。开发了两阶段深度学习框架，根据输入图片推荐类似风格的时尚图像[12]。Hidayati S C 等人从社交网络大数据中学习服装样式和个体身材的兼容性，用来向用户推荐符合其身体属性的穿着服装[13]；周伟等人开发了新颖的时尚推荐模型，通过整合基于文本的产品属性和图像提取功能来匹配相似的产品[14]；Jaradat, Shatha 使用深度学习技术来分析深像素级语义分割和文本集成对推荐质量的影响[15]；陈婉玉等人提出了用于检索具有脸部形状特征的时尚照片的一种新的方法[16]。

由于推荐系统计算量庞大，又对实时性要求高，将计算过程大致分为两个部分，即离线部分和在线部分[17]。离线部分主要指系统离线构建索引库的流程，包括离线特征提取、索引构建、数据存储等环节，这些环节计算量非常大，需要定期执行一次，以保证系统的准确率；在线流程主要指用户提供一张图片，到最终返回推荐结果的过程，包含噪声去除、图像定位及分割、在线特征提取、图像匹配等过程。

2　服装图像检索系统框架（Framework for Fashion Image Retrieval System，FIRS）

跨场景时尚图像检索是一个复杂的课题，本文结合深度学习和度量学习等技术，提出了一个完整的时尚商品检索框架，包含了离线和在线二部分，通过二者的协作，极大地提高了检索的精度和效率，如图 2 所示。

（a）离线数据处理　　（b）时尚图片的在线检索

（a）Offline data processing　（b）Online retrieval of fashion images

图 2　服装图像检索系统框架

Fig. 2　The framework of clothing image retrieval system

（1）离线部分。离线部分是在线检索的基础，是决定在线检索精度的保障。离线部分包含 3 个模块。特征抽取模块：是一个图像表示学习的过程，本

文采用 fine-tuning 的方式,基于现有成熟的模型进一步训练以获得更好的表示;索引构建模块:检索的高效性离不开索引的构建,在特征抽取完成后,将图像与特征映射,从而构建索引;数据存储模块:数据存储方式是影响数据检索的一个重要因素,本文的数据采用 spark 分布式存储,从而提高整体检索效率。

(2)在线部分。在线检索存在二个问题:①噪声问题,是由用户上传图像的质量导致的;②多实例问题,是由于图像中存在多个实例引起的。本文使用目标检测和图像分割技术解决这些问题,并根据不同需求提供了简单泛化搜索,精细搜索,以及基于标签的搜索。

3 基于深度学习的离线数据处理

3.1 基于 ResNet 的学习图像特征学习

深度学习在特征提取和图像分类中应用非常广泛,在图像搜索时使用预训练的 ResNet 来提取特征已被证明非常有效[18]。具体方法为:将完整的 ResNet 的最后一层 SoftMax 层去掉,增加一个全连接层,将最后的输出特征向量修改成实际所需要的维度。该方法相较从零开始训练一个完整的神经网络,可以节省大量的资源,也避免了数据量不足和硬件条件不够所引发的问题。

3.2 索引构建

通过亚马逊数据集提供的源文件中的地址,下载图片并将图片命名为其 ASIN ID(即亚马逊商品统一编号),载入预训练 ResNet 模型,将每一张图片转换成所需要的输入格式(在这里定义的为 224 * 224 * 3)的彩色 RGB 格式。因为本系统选取的 100 万张图片是亚马逊总图片数的一个子集,所以在这里系统使用目录里的图片和源文件中的商品信息进行映射,并保存在一个二进制文件中。ASIN_Idx 保存编号的目录,ASIN_Data 保存编号所对应的商品信息。信息完成以后,进行批量的特征提取,将100 万张图片逐个放入预训练 ResNet 中,建立一个 ASIN到特征的映射,存放所有的提取之后的特征。

在图像存储方面,一般将图像特征量化成为数据存放在索引中,并存储在外部存储介质中,进行图像搜索的时候仅仅需要搜索索引表中的图像特征,按照匹配程度从高到低来查找类似的图像。对于图像尺寸分辨率不同的情况可以用下采样或者归一化的方法。

4 在线特征提取方法

4.1 图像预处理

因为商品的背景复杂,主体常常较小,所以为了减少大量背景干扰和多主体的影响,需要将搜索目标从图像中提取出来,这就涉及到了目标检测技术。

物体检测算法经历了传统的人工设计特征+浅层分类器的框架,到基于大数据和深度神经网络的 End-To-End 的物体检测框架,物体检测愈加成熟,本文使用其中最有代表性的 Faster-RCNN。

使用 Mxnet 框架下的 Faster R-CNN 进行目标检测的测试。目标检测的主要功能有 3 个:判断图片中物体属于前景还是背景,判断前景中每一个物体的类别(从预先保存的约 300 个类别中进行判断),将每一个类别的物体用方框圈起来,并返回方框中 4 个角的坐标。除了坐标以外,该模型还返回了判断属于该类别的概率(仅仅返回可能性大于0.5的类别),本文使用一张街拍——含有常见背景的人物服装照片来进行测试,如图 3 所示。

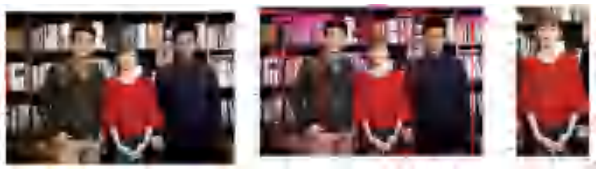

图 3 目标检测结果

Fig. 3 Target detection results

从测试结果可以看出,虽然有多个人物和物体出现在图像中,目标识别还是能够检测出其中概率最大的那一个,并且进行图像框定。但同时也看出现图像背景还是有其他物品存在,此时可以通过图像分割技术进一步处理。

对成功识别出的图像中最主要物体的轮廓识别和分割。在返回时需要选择分割的主体,即,返回主体的类别编号,本例返回以人物为轮廓进行分割的图片。使用 FCN 进行图像分割的效果明显,绝大部分背景干扰因素已经被去除,剩下的只有人物和其服装图片,如图 4 所示。在后续的图像索引中,能够获取更加匹配的目标图片。

图 4 FCN 图像分割后轮廓及其结果

Fig. 4 Results after image segmentation

4.2　基于最近邻的图像匹配算法

本文使用一种基于图的近似最近邻搜索的方法叫做 Hierarchical Navigable Small World graphs 算法。这个算法基于先前的 NSW(navigable small world)思想进行了优化,使用了一个跳表结构将多个多维向量按照图的结构划分为许多层,从顶到底构造一个层次化结构,最大的层数由以指数衰减的概率分布随机选择,在搜索时从最顶端的层级开始,逐步向下延展,如图 5 所示。这样的方式极大的提高了高召回率和高度集群数据的性能。性能评估表明这种方法在通用的空间向量的搜索中能够极大的优于先前开源的最先进的向量搜索方法。

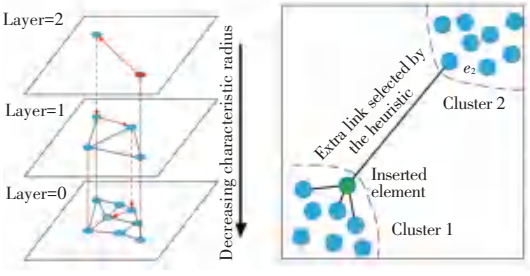

图 5　HNSWG 算法示意图

Fig. 5　HNSWG algorithm diagram

4.3　检索方法

4.3.1　简单检索和高级检索

简单检索提供了泛用性最强的图像搜索引擎,可以对全部 100 万张商品图片和信息进行搜索,分类非常广,能在近实时的基础上完成 Top10 的搜索。

而精细检索提供了最精细的自动目标检测和轮廓剪切,对于需要精细度较高的商品,如服装等,可以选择这种搜索方式,自动按照最精确的方式搜索,搜索时间会稍长,但是也能满足一般的实时性要求。

4.3.2　基于标签的检索

在许多情况下,有的商品虽然图案相似、形状相近,但是它们在本质上并不是同一类型的商品,这时候仅仅使用特征提取可能会出现误判,所以本系统使用了另一种搜索方式——基于目标检测和倒排索引的图像搜索方法。

倒排索引(Inverted index),也常被称为反向索引、置入档案或反向档案,是一种索引方法,被用来存储在全文搜索下某个单词在一个文档或者一组文档中的存储位置的映射,是文档检索系统中最常用的数据结构。通过倒排索引,可以根据单词快速获取包含这个单词的文档列表。倒排索引主要由两个部分组成即“单词词典”和“倒排文件”。倒排索引的主要形式为一条记录的水平反向索引(或者反向档案索引)包含每个引用单词的文档的列表。

使用商品检测出的目标标签当作文档,将所有的分类标签当作单词,建立目标图片集的完整倒排索引数据库,将每一张用户上传的图片中检测出来的词汇视为新的文档,检测以后同样查倒排索引,就可以得到总图片的一个子集。

在电商的图像搜索中同样可以使用这种思路进行图像搜索。在搜索中,文档就相当于是一张图片进行目标检测中所有元素(如一张人的图片中可以有帽子、衬衫、领带、皮包等元素),而单词就相当于是搜索目标商品的图片识别出的关键词,往往是单一的物品,搜索过程如图 6 所示。在实际使用中,系统会自动识别用户上传的图片并进行初步的目标识别,在图片集创建的文档中搜索和匹配。

图 6　倒排索引的搜索过程

Fig. 6　Search process of inverted index

5　实验结果与评价

5.1　数据集与实验环境

亚马逊是网络上最早的电子商务公司,经过数十年的积累,其商品涵盖各个种类达数百万之巨,是电商数据集的重要来源。本文使用由 Julian McAuley 团队整理的 Amazon product data,根据其唯一的 ASIN ID 便可获取对应图像。

实验运行的硬件平台,CPU 为 Intel(R) Core(TM) i5-8400 CPU @ 2.80GHz,GPU 为 NVIDIA GeForce GTX 1050,内存为 20G, 系统为 CentOS 7.4.1708。接口语言为 python,代码运行在 Mxnet 平台。

5.2　实验结果

在系统中,实时性是非常重要的一项指标,实验测试了普通搜索和精细搜索的所需时间。测试结果见表 1。即使是精细的搜索,系统也能够完全达到在一秒内返回所需要的结果。

表 1　搜索所需时间

Tab. 1　Search time

	特征提取	最近邻搜索	目标检测	图像分割	服务器响应
平均时间/ms	70	124	21	1.99	<200

6　结束语

本文提出了一个实时图像检索系统,用来检索基于亚马逊部分数据集的时尚单品。采用三种检索方式即简单搜索、精细搜索、基于标签的搜索。由于

跨场景的不确定性,首先使用目标检测算法来进行目标定位,找出需要检索的目标商品,再进行图片分割,将目标商品单独分割出来;然后使用卷积神经网络对图片进行特征提取;最后在图片数据库中找出与其最相似的数据图片。本文提出的不同的检索方式可以满足不同用户的不同需求,给予用户更好的体验。将会在以后的研究中加入文本信息,实现多模态检索,用文本信息来弥补图片信息没有满足的部分,实现更高精度的检索。

参考文献

[1] YAN S. Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set [C]// Computer Vision & Pattern Recognition. IEEE, 2012.

[2] LIU Z, LUO P, QIU S, et al. DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations [C]// Computer Vision & Pattern Recognition. IEEE, 2016: 1096-1104.

[3] AGARWAL D, BASU K, GHOSH S, et al. Online Parameter Selection for Web-based Ranking Problems [C]// KDD, 2018. 23-32.

[4] LAENEN K, ZOGHBI S, MOENS M. Web Search of Fashion Items with Multimodal Querying [C]// WSDM, 2018. 342-350.

[5] Iqbal, Murium, Adair Kovac, and Kamelia Aryafar. " A Multimodal Recommender System for Large - scale Assortment Generation in E - commerce." arXiv preprint arXiv: 1806. 11226 (2018).

[6] LIU Yujie. How to Wear Beautifully? Clothing Pair Recommendation [J]. Journal of Computer Science and Technology, 2018, 33(3): 522-530.

[7] Rui, Yong. Relevance feedback: a power tool for interactive content-based image retrieval [J]. IEEE Transactions on circuits and systems for video technology, 1998, 8(5): 644-655.

[8] CHEN K, LUO J. When fashion meets big data: Discriminative mining of bestselling clothing features [C]// WWW, 2017: 15-22.

[9] HIDAYATI S C, HSU C C, CHANG Y T, et al. What Dress Fits Me Best? Fashion Recommendation on the Clothing Style for Personal Body Shape [C]// ACM Multimedia Conference on Multimedia Conference. ACM, 2018: 438-446.

[10] ZHOU, WEI. " Fashion recommendations using text mining and multiple content attributes." (2017).

[11] JARADAT, SHATHA. Deep cross-domain fashion recommendation [C]//Proceedings of the Eleventh ACM Conference on Recommender Systems, ACM, 2017.

[12] BRACHER, CHRISTIAN, SEBASTIAN HEINZ, ROLAND VOLLGRAF. " Fashion DNA: merging content and sales data for recommendation and article mapping." arXiv preprint arXiv: 1609. 02489 (2016).

[13] ANDREEVA, ELENA, et al. Extraction of Visual Features for Recommendation of Products via Deep Learning [C]// International Conference on Analysis of Images, Social Networks and Texts. Springer, Cham, 2018.

[14] KRIZHEVSKY, ALEX, ILYA SUTSKEVER, GEOFFREY E. Hinton. Imagenet classification with deep convolutional neural networks [C]// Advances in neural information processing systems. 2012.

[15] TUINHOF H, PIRKER C, HALTMEIER M. Image - Based Fashion Product Recommendation with Deep Learning [C]// International Conference on Machine Learning, Optimization, and Data Science. Springer, Cham, 2018: 472-481.

[16] CHEN Wanyu, CHEN Jialin, CHEN Lianggee. On - the - fly fashion photograph recommendation system with robust face shape features [C]// 2014 IEEE International Conference on Consumer Electronics (ICCE). IEEE, 2014.

[17] ZHANG Y, PAN P, ZHENG Y, et al. Visual search at Alibaba [C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2018: 993-1001.

[18] HAN X, WU Z, JIANG Y, et al. Learning Fashion Compatibility with Bidirectional LSTMs [J]. acm multimedia, 2017: 1078 - 1086.

(上接第 13 页)

(4)全 IP 网络适合。现如今的 IP 化趋势让终端与物联网应用有着更为简捷的设置,仅仅需要遵循互联网模式去开发就可以了。

(5)核心网不主动释放连接。LTE 中心网并无主动释放功能,让永久在线业务获得更多保障,让“信令风暴”问题得以消除。在 LTE 网络内,仅仅是终端或 eNodeB 等装置能够经 NAS 消息去对中心网予以通知,然后才能够和终端形成连接。如果说终端与中心网的 NAS 层处于附着状态,那么底层的链路是否出现释放均不会对 IP 地址造成影响,从而实现永久在线的功能。

3 结束语

5G(5th-Generation,第五代移动通信技术)承载着 LTE 网络的发展,但 LTE 代表着长期演进。在现代化信息技术发展中,LTE 无线通信网络亦有着多方面的变化,逐渐朝着高效率、精准化、高速化以及现代化的方向发展。将 LTE 无线通信与物联网两种技术有效结合,则可以增快宽带信息化发展速度,不但可以实现更好的物联网服务,还可以给广大使用群体提供快速、稳定的无线承载。

参考文献

[1] 李珊. 从国外发展经验看中国 4G 发展[J]. 移动通信, 2014 (11):49-50.

[2] 李洪,吴泽龙. 浅析 LTE 无线通信技术与物联网技术的结合与发展[J]. 数字通信世界,2018 (2):122.

[3] 胡玉佩. LTE 技术及应用前景浅析[J]. 长沙通信职业技术学院学报, 2012.

[4] 郎为民, 焦巧, 刘建中. LTE 系统架构研究[J]. 数据通信, 2009 (5):6-9.

[5] 蔡剑. 物联网技术与 LTE 无线通信技术探究[J]. 通信电源技术,2020,37(3):191-192.

[6] 吴端兴. LTE 无线通信技术与物联网技术的结合研究[J]. 现代信息科技,2019,3(11):186-187.