



基于深度学习的服装图像语义分析与检索推荐

徐 慧¹, 白美丽¹, 万韬阮², 薛 涛¹, 汤 汶³

(1. 西安工程大学 计算机科学学院/陕西省服装设计智能化重点实验室, 陕西 西安 710048;

2. 布拉德福德大学 工程与信息学院, 英国 布拉德福德 BD7 1DP;

3. 伯恩茅斯大学 科学与工程技术学院, 英国 伯恩茅斯 BH12 5BB)

摘要:针对服装检测推荐方法中的服装图像分割算法分割精度不高且存在多尺度特征融合不佳的问题,提出一种基于注意力机制和多尺度特征融合的深度卷积网络服装特征提取分析检索与推荐方法。该方法以图像语义分割算法为基础,从服装图像中提取高级语义信息,通过哈希函数对提取的特征进行编码,并引入 ASPP 对图像进行多尺度特征提取,然后使用注意力模型对提取到的特征进行融合并构建索引库,最终实现语义相似性服装推荐的功能。实验表明:本方法可以有效提高分割准确度,解决多尺度特征提取的问题,能够成功提取图像中的高层次语义特征,具有较高的分割准确率,在很大程度上提高了服装相似性语义推荐的效率。

关键词:深度卷积神经网络;多尺度特征融合;服装语义分割;服装多尺度特征提取;服装检索;服装推荐

中图分类号:TP 391.41

文献标志码:A

Image semantic analysis and retrieval recommendation for clothing based on deep learning

XU Hui¹, BAI Meili¹, WAN Taoruan², XUE Tao¹, TANG Wen³

(1. School of Computer Science/Shaanxi Key Laboratory of Clothing Intelligence,
Xi'an Polytechnic University, Xi'an 710048, China;

2. Faculty of Engineering and Informatics, University of Bradford, Bradford, BD7 1DP, United Kingdom;

3. Faculty of Science and Technology, Bournemouth University, Poole BH12 5BB, United Kingdom)

Abstract: Aiming at the problems of low segmentation accuracy and poor multi-scale feature fusion of clothing image segmentation algorithm in clothing detection and recommendation method, a deep convolution network clothing feature extraction analysis retrieval and recommendation

收稿日期:2020-04-14

基金项目:陕西省技术创新引导专项计划项目(2020CGXNG-012)

第一作者:徐慧(1996—),女,西安工程大学硕士研究生。

通信作者:汤汶(1963—),女,英国伯恩茅斯大学教授,研究方向为计算机视觉、VR/AR等。

E-mail: wtang@bournemouth.ac.uk

引文格式:徐慧,白美丽,万韬阮,等.基于深度学习的服装图像语义分析与检索推荐[J].纺织高校基础科学学报,2020,33(3):64-72.

XU Hui, BAI Meili, WAN Taoruan, et al. Image semantic analysis and retrieval recommendation for clothing based on deep learning[J]. Basic Sciences Journal of Textile Universities, 2020, 33(3): 64-72.

method based on attention mechanism and multi-scale feature fusion is proposed. Based on the image semantic segmentation algorithm, this method extracts high-level semantic information from clothing images, encodes the extracted features by hash function, and extracts multi-scale features of the image by introducing ASPP, and then merges the extracted features with attention model to build an index database, and finally the function of clothing recommendation with semantic similarity is realized. The experimental results show that this method can effectively improve the segmentation accuracy, solve the problem of multi-scale feature extraction, successfully extract the high-level semantic features in the image, with high segmentation accuracy, and greatly improve the efficiency of semantic recommendation of clothing similarity.

Key words: deep convolutional neural network; multi-scale feature fusion; clothing semantic segmentation; clothing multi-scale feature extraction; clothing retrieval; clothing recommendation

0 引言

早期服装检索与推荐方法主要是基于文本检索与分析^[1],即通过关键词进行图像检索。随着大数据时代的到来,在线图像数量成指数级增长,并且由于其自身属性的复杂性,真实场景中的服装图像很难通过文字清晰表达图像内容。然而,图像具有比文字更具体的内容,所以人们开始考虑研究基于内容的图像分析^[2],也就是“以图搜图”技术。目前,服装检索与推荐方法大多利用图像的全局特征(例如颜色、纹理)进行相似性检索^[3-4],难以获得良好的搜索结果。

服装分割方法主要利用手工设计的特征和预估的人体姿态预测像素级标注。YAMAGUCHI等提出了一种良好的服装分割框架,但该方法需要在模型训练之前手工标注图像的像素级标签,这需要耗费巨大的人力物力以及时间^[5];LIU等提出了使用颜色类别标签的弱监督迭代方法,该算法不需要像素级标注,只使用服装颜色标签和人体姿态估计完成图像分割,省时省力,但过于依赖于人体姿态^[6];JUAN等针对复杂纹理服装提出了一种改进的服装分析方法,该方法在不考虑人体姿态的情况下,从图形模型和图像特征2个方面对传统的基于图像的分割方法进行了改进,避免了复杂纹理服装的过分割和像素标签失败导致的估计不准确,然而该方法对类似肤色的复杂不规则的纹理服装解析效果较差^[7];LEDERER等提出集服装检测、分割、识别及特征识别为一体的可扩展的自动服装推荐方法,虽然实现了自动服装推荐,但该方法图像分割模型粗糙,分割精度普遍较低^[8];黄冬艳等提出了基于HOG和E-SVM的服装图像联合分割算法,该算法不依赖于人体姿态,能够较好地分割服装图像,但不适用于多种服装风格或相似服装颜色和背景的服装图像^[9]。

近年来,研究者们开始利用深度学习技术进行语义分割。CHEN等提出了基于context aware的语义分割方法,可以有效提取真实环境中物体表面的材料信息^[10];YAMAGUCHI等提出了一种预先训练网络提取服装全局与局部特征的网络,该方法虽然可以将服装较好地分割开来,但是实现的过程却很繁琐^[11];郭鑫鹏等提出了一种无监督的分割算法,该算法结合了颜色和纹理特征以及CNN和HOG提取服装的关键部分,但是结果的特征维数不高,分割精度也不高。CNN可以自动学习图像特征,有效减少了手工设计特征和姿态估计造成的损失^[12]。CYCHNERSKI等以衣服的属性为研究点,使用深度神经网络提取衣物特征并分类,最终取得了较好的效果,但是对处理复杂背景图像仍然存在不足^[13]。文献[14]提出的深度卷积神经网络语义分割算法是利用服装语义分析预测图结合自监督学习算法学习人体估计,实现了端到端的语义分割,解决了服装分割过程中依赖人体估计以及需要手工设计特征等问题,其与其他经典的语义分割方法相比,具有较高的分割性能。但是,对于现实场景下的人体街拍图像或者含有复杂背景的图像效果仍然不够理想。

综上所述,为了更好地利用全局语义信息以及进一步优化文献[14]中Deeplabv2-SSL网络的分割性能,本文首先通过引入注意力机制,使模型可以有针对性地学习相应的目标区域,从而获取更加精细的图像特征,以此提高语义分割性能。在此基础上,提取服装高纬度特征,通过在网络中添加哈希函数,对提取到的高维语义特征进行哈希编码,生成对应的二进制编码,然后将数据库中的二进制编码构建索引库,待检索的图像经过网络模型提取语义特征,再通过相似性度量,得出服装推荐排序图。

1 服装图像语义分割模型

在计算机视觉中,注意力机制^[15]被广泛应用于图像分类和目标检测。由于注意力机制具有聚集最相关的特性,因此在网络中加入注意力机制,使网络可以自适应地选择要处理的图像区域,从而提高网络输出质量。对于每一个尺度而言,注意力机制输出一个权值图,权值图以像素为单位进行加权,然后使用 Deeplabv2-SSL 生成所有尺度上的得分图加权进行分类。基于多尺度融合的服装图像语义分割模型整体框架如图 1 所示,主要实现过程为:

1) 模型构建。以注意力机制为基础,结合文献^[14]中提出的 Deeplabv2-SSL 网络,搭建一个新的多尺度语义分割网络,将其命名为 DSAnet 网络,通过训练数据集获得分割模型。

2) 多尺度特征提取。Deeplabv2-SSL 网络本身可以提取多尺度特征,但是由于提取的多尺度特征缺乏上下文之间的潜在信息,因此,引入聚焦机制的注意力模块,然后通过新的网络结构将提取图像中不同对象的底层特征和高层次特征,再将不同尺度的特征进行融合,获得具有高层次多尺度的特征图,以此来提高服装图像分割的准确度。

3) 基于注意力机制的多尺度融合网络框架设计。从图 1 可以看出,首先在 Deeplabv2-SSL 的 FC7 层引入注意力加权,图像经过计算得到多个尺度的权重图,然后通过 Deeplabv2-SSL 获得的多尺度图像进行求和,实现多尺度特征的融合。

4) 通过双线性插值对融合后的特征图进行上采样,以将其恢复到原图大小。即通过插值得到图像中间点的像素值,将低分辨率的特征图扩展到与原始图相同的大小,得到最终的语义分割预测结果图。

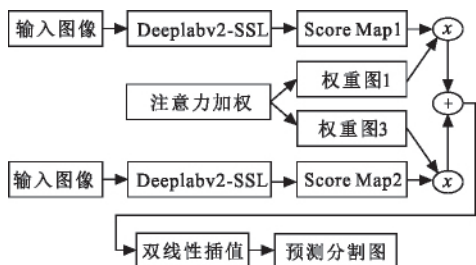


图 1 语义分割模型网络 (DSAnet) 框架

Fig.1 Semantic segmentation network (DSAnet)

1.1 注意力机制的构建

注意力机制构建步骤包括:

1) 网络结构。DSAnet 包含了 2 个卷积层,第

1 个卷积层为主要层,其中含有 1 024 个卷积核,第 2 层为起到补偿作用的附加层。卷积核数目 S 的确定根据第 1 层的卷积结果随机产生。但若只使用 1 个卷积层学习特征,可能会导致特征提取过于单一,最终无法获得有效的权重图。

2) 输入模型。通过多尺度特征提取结构输出的每个分支的 FC7 层尺度特征,被合并后作为统一输入。这是因为与其他网络层相比,FC7 层可以学习更多的高级全局语义功能,并且包含更大范围的信息。

3) 卷积核。为了关注像素本身和分割结果之间的相关性,注意力机制的 2 个卷积层均使用 1×1 卷积核进行卷积运算。卷积的结果主要是为比例特征图的每个位置像素分配权重,而其他尺寸卷积核的焦点是考虑周围像素对位置像素的影响。

4) 尺度计算。注意力机制主要通过尺度 x 中的各个位置像素点 i 进行计算,从而得到多个尺度对应的分数图 p_i^x ,然后使用 softmax 函数对得到的 p_i^x 统一进行归一化处理,得到不同尺度的各个位置像素点 i 的权重。权重计算公式为

$$w_i^x = \frac{\exp p_i^x}{\sum_{t=1}^x \exp p_i^t} \quad (1)$$

式中: w_i^x 表示权重; p_i^x 表示注意力模型在位置 i 处为尺度 x 生成的分数图,也就是 softmax 最后一层输出值。通过计算权重 w_i^x ,得到

$$g_{i,c} = \sum_{i=1}^x w_i^x f_{i,c}^x \quad (2)$$

式中: $g_{i,c}$ 为多尺度特征融合的加权和; $f_{i,c}^x$ 为得到的得分特征映射值。

1.2 数据集

CFPD 数据集由基于超像素注释的 2 682 个服装图像组成,每张图像都是基本可见全身,而且面向前方的女性人物图像。数据集中设置了 22 个服装类别标签以及 1 个表示背景标签 bk。22 个服装标签分别是 T-shirt、bag、belt、blazer、blouse、coat、dress、face、hair、hat、jeans、legging、pants、scarf、shoe、shorts、skin、skirt、socks、stocking、sunglass、sweater 等。将 CFPD 数据集按比例随机拆分为单独的训练集、验证集及测试集,拆分比例为 78%、2%、20%。CFPD 数据集集中的所有图像都是 400×600 像素的彩色图片,实验中不再更改图像的格式。

1.3 评估指标

实验选择总体像素精度 (overall accuracy, a_o), 平均像素精度 (mean accuracy, a_m) 和平均交并比

I_{ou} 等3种指标评估模型性能。假设共有 n_{cl} 个像素级别(其中包括1个背景类),以 n_{ij} 表示像素类别 i 被预测为 j 的像素个数, n_{ii} 则为真正的数量, t_i 表示属于类别 i 的个数,3种指标计算公式分别为

$$a_o = \sum_i n_{ii} / \sum_i t_i \quad (3)$$

$$a_m = \frac{1}{n_{cl}} \sum_i \frac{n_{ii}}{t_i} \quad (4)$$

$$I_{ou} = \frac{P_T}{P_T + P_F + N_F} \quad (5)$$

式中: P_T 为当前类别中正确预测的像素数; P_F 为不在当前类别中并且预测错误的像素数; N_F 为不在当前类别中且预测出来的不是当前类别的像素数。

2 服装语义分析推荐网络结构

2.1 语义特征提取方法与系统设计

现有的服装检索方法^[16]中最重要的步骤的是对图像特征进行提取。不同的图像即使看起来很相似,其包含的语义内容差别很大。如果只考虑图像中诸如颜色、纹理、外观等浅层特征,2种服装相似性就很高。然而从语义层面上来讲,二者之间相似度却很低。

大多数传统的特征提取方法^[17]依赖于提取服装中的颜色和纹理等局部特征,提取到的特征并不能代表图像中的整体内容,因此对图像检索的结果有一定影响。本文提出的DSANet网络基于深度学习的方法对服装图像中所包含的特征进行自动学习,并对学习到的高层语义特征进行提取。由于大部分信息分布在卷积层中排列在前的特征图中,为了直观地分析神经网络自动学习图像特征的过程,并不失一般性,本实验选取了每个卷积层中前25个特征图进行分析,如图2所示。图2中:

1) 浅层特征更多地倾向于对图像边缘的检测,检测到的内容更加全面,同时也会有关键信息提取出来(例如第1个卷积层中明显的人体边缘轮廓);

2) 随着层次的加深,特征图也越来越抽象,同时会忽略了部分局部信息,但却对全局特征可较好提取;

3) 越深的层次,空白区域越多,说明卷积核没有得到它们所需要的特征,因此,深度神经网络并不是网络结构越深效果越好,需要有针对性的对图像进行网络设计。

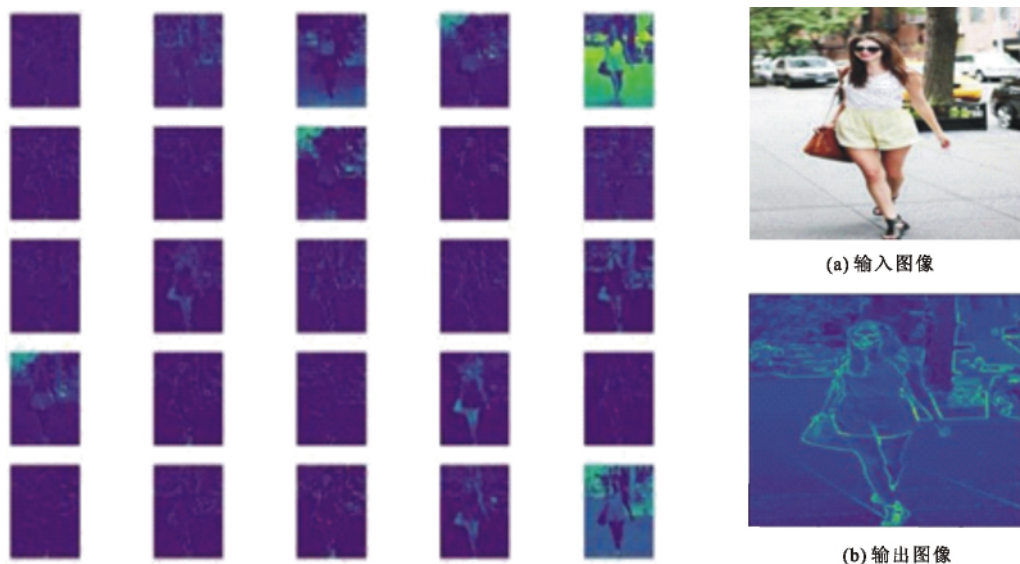


图2 服装图像语义特征提取实验

Fig.2 Experiment of semantic feature extraction of clothing image

传统的图像检索^[18]过程中对提到的图像特征通过相似性度量^[19]匹配进行排序。但是,由于提取到的特征都是图像的浅层特征,不包含语义信息,检索的结果图可能与用户需要的图像差距较大,而且这种基于图像特征的匹配占据了较多的空间维度,可能会造成维度灾难,导致最终检索结果的失败。通过深度卷积神经网络提取到的特征维度较高,计算机的性能可能达不到所需要内存空间。因此,本

文考虑在DCNN中加入哈希函数,将获得的带有语义的特征图信息通过哈希函数映射得到对应二进制的哈希编码;然后,将获得的哈希编码进行存储,构建基于语义特征^[20]的哈希索引库;之后,将待检索的图像通过相似性度量算法计算汉明距离并进行排序;最后,输出检索的结果图。使用的哈希检索结构如图3所示。

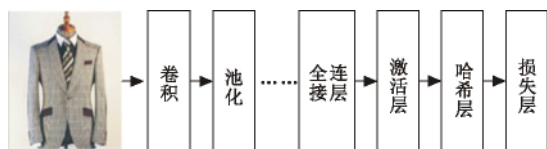


图3 基于CNN的哈希层网络结构

Fig.3 Hash layer network structure based on CNN

由图3可知,基于CNN的通用哈希检索将哈希层置入激活层和损失函数层中间,并在分类前对特征进行哈希编码。哈希层编码图像特征的本质是使用局部敏感哈希算法^[21](locality sensitive Hashing, LSH),基本思想是设计一个特殊的哈希函数,将提取到的特征映射到相应的哈希索引数据库。

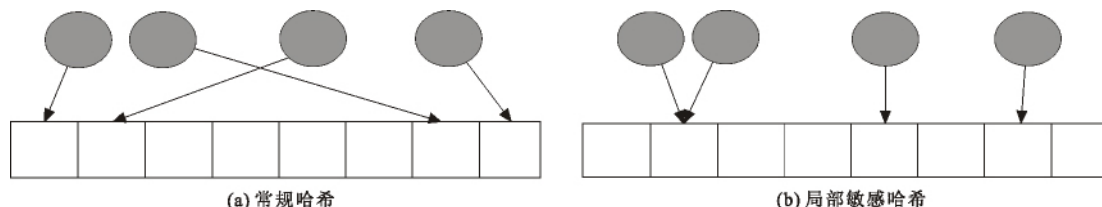


图4 常规哈希与LSH局部敏感哈希

Fig.4 Regular Hash and locality sensitive Hashing

2.2 汉明距离

哈希方法用于将服装图像的相似性度量转换为相应的二进制哈希码度量。汉明距离用于衡量搜索过程中服装图像之间的相似性,对于长度为 u 和 v 的二进制特征向量,汉明距离被定义为具有相同位置但不同二进制代码的两个特征向量的数量。计算汉明距离的公式为

$$D(u, v) = \sum_{i=1}^n \oplus (u_i \neq v_i) \quad (6)$$

式中: \oplus 为异或运算; u_i 、 v_i 分别为 u 和 v 的第 i 维特征向量值。如何计算汉明空间中向量的相似度取决于图像特征,没有通用的相似度度量方法。对于

如果2个数据点之间的相似性很高,它们会大概率地被映射到同一个哈希编码;否则,映射得到的概率很低。然后,通过计算要检索的图像和库中图像之间的汉明距离进行排序,得到最终的搜索结果。

与常规的哈希算法不同,局部敏感哈希算法的目的是将距离相近的点存储在距离较近的哈希表中,并将其分配给距离较远的其他表。当使用LSH进行检索时,通过从网络中提取的特征对要检索的图像进行哈希处理,然后在索引数据库中搜索相似的特征。常规哈希方法和LSH方法对比图如图4所示。

特征提取生成的服装图像的多维特征向量,计算待检索服装图像与检索库中的图像之间的汉明距离,选择相似度高的 $\text{top-}k$ 值对结果进行排序并输出。该方法利用汉明空间进行计算,加快了整体搜索速度。

2.3 算法流程图

将待检索的图像在独立的哈希函数下进行相同的局部敏感哈希特征提取和相似性度量,并根据分类信息在相应类别的服装图像集中,按相似度降序检索图像,实现图像的准确分类和快速检索。为了分类和优化检索,测量测试样本的相似性、模型的性能并调整模型的参数。模型整体结构如图5所示。



图5 服装检索算法流程图

Fig.5 Flow chart of clothing retrieval algorithm

3 实验结果与分析

3.1 实验环境

使用公共可用的,在多尺度特征提取与融合中具有较好的表达能力的注意力机制,作为网络架构的基本架构。基于Deeplabv2-SSL的网络,该网络

采用了空洞卷积(atrous convolution)、多尺度输入和最大池化(max-pooling)以及ASPP合并来自所有尺度的结果。

为了训练基于注意力模型的网络,输入图像被缩放到 321×321 。使用来自Deeplabv2提供的预先训练的模型和网络设置,并使用随机梯度下降

(SGD)对网络权重和偏置进行更新。初始学习率设置为 0.001,每经过 2 000 迭代后学习率会乘以 0.1。实验过程中共迭代 20 000 次,然后将数据输入到构建好的网络中进行模型的训练测试。

3.2 结果与分析

由于 Deeplab 模型可以提取多尺度特征,因此,为了验证注意力模型可以解决服装图像中小部件产品识别精度低的问题,使用本文提出的 Deeplabv2-SSL-Attention(简称为 DSANet)方法与另外 2 种方法进行对比分析,以此来验证多尺度特征加权的注意力模型可以提高语义分割算法的准确性,并可以较好地分割服装中的小部件物体。实验结果如表 1 所示。

表 1 3 种 CFPD 方法性能比较

Tab.1 Performance comparison of three CFPD methods

方法	像素总体 精确度	平均像素 精度	平均 交并比
Deeplabv2-SSL	83.37	52.53	42.46
Deeplabv2-Attention	83.56	54.28	42.97
DSANet(本文算法)	84.53	54.81	44.59

选取 CFPD 的测试集数据对网络模型进行测试,可以得出模型测试的语义分割预测图,如图 6 所示。

以图 6 可以看到:Deeplabv2-SSL 可以较好地分割较大尺度的图像,如 T 恤、连衣裙等,但是对于鞋子、包等小部件物体分割效果不好,缺失了大量的细节信息。这是由于该网络中采用的空洞卷积核 ASPP,虽然增大了感受野,但也丢失了一部分细节特征。Deeplabv2-Attention 中由于引入了注意力机制,可以将目标较好地聚焦于多尺度特征图像中,因此整体分割性能高于 Deeplabv2-SSL。但是,由于 Deeplabv2 网络中没有引入自监督结构敏感损失学习,无法自主学习图像中的人体姿态特征,因此对人体的关键部位的分割效果不好。本文提出的 DSANet 网络,不仅将人体姿态估计整合到语义分割模型中,还针对服装数据中存在的多尺度特征融合问题进行了优化。经过融合的语义分割网络具有良好的分割性能,尤其对于包、太阳镜、腰带等小部件物体具有良好的分割结果,而且可以更好地将背景和前景分割开。

3 种不同情况下,服装类产品基于相似性得分的推荐排序实验结果如图 7、8、9 和表 2、3、4 所示,图 7、8、9 中,除原图像外,其余均为相应的输出图像。

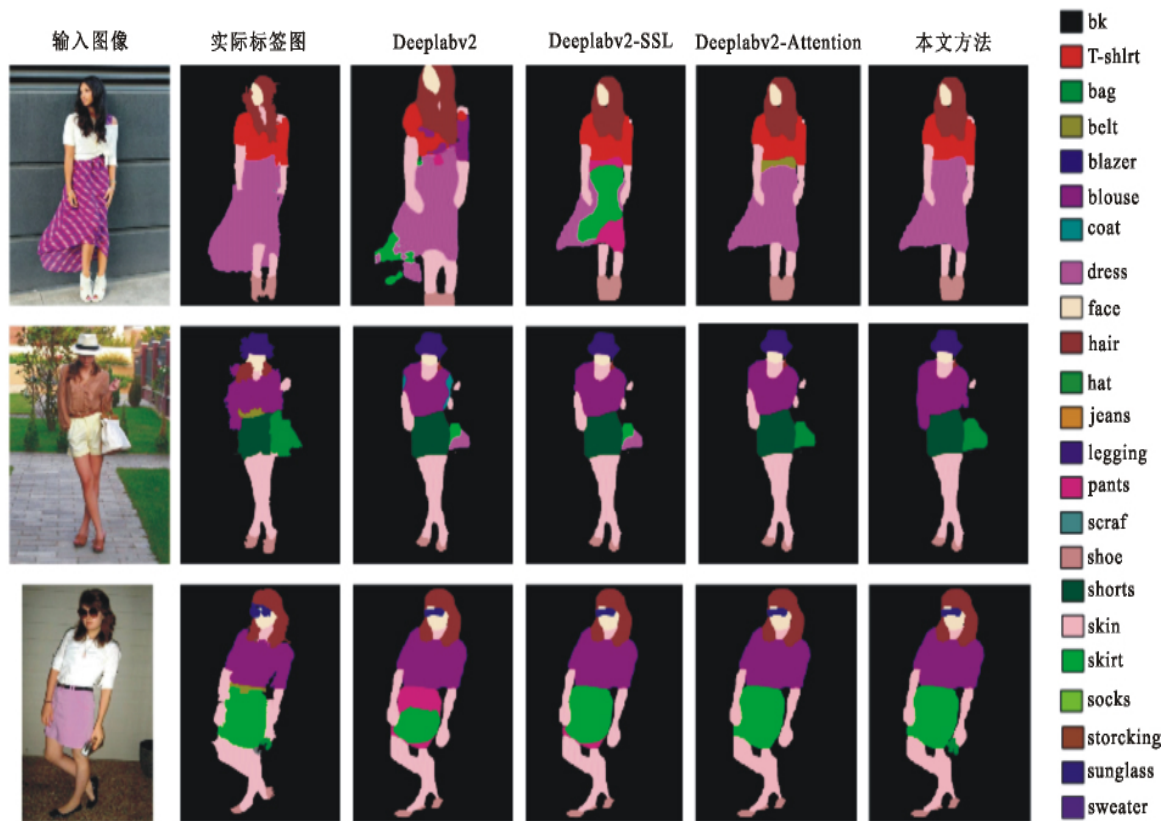


图 6 各种模型性能对比
Fig.6 Comparisons of performance of various model



图7 手提包产品检索及相似度排序实验

Fig.7 Experimental chart of handbag product retrieval and similarity ranking



图8 衣物检索及相似性推荐实验

Fig.8 Experimental chart of clothing retrieval and similarity recommendation



图9 人体图像服装相似推荐实验

Fig.9 Experiment chart of clothing similarity recommendation based on human body image

表 2 手提包产品相似度排序
Tab.2 Ranking of handbag products by similarity

输出图像	相似度	推荐排序
图 7(b)	0.919 454	1
图 7(c)	0.870 477	2
图 7(d)	0.829 969	3
图 7(e)	0.779 671	4
图 7(f)	0.762 682	5
图 7(g)	0.753 429	6
图 7(h)	0.725 592	7
图 7(i)	0.721 642	8
图 7(j)	0.717 422	9

表 3 衣物分析相似性推荐实验排序
Tab.3 Experiment ranking of similarity recommendation based on clothing analysis

输出图像	相似度	推荐排序
图 8(b)	0.843 842	1
图 8(c)	0.8192 89	2
图 8(d)	0.819 289	2
图 8(e)	0.814 488	4
图 8(f)	0.814 488	4
图 8(g)	0.773 488	6
图 8(h)	0.773 045	7
图 8(i)	0.772 494	8
图 8(j)	0.766 414	9

表 4 人体图像服装检索与推荐排序
Tab.4 Ranking of clothing retrieval and recommendation based on human body

输出图像	相似度	推荐排序
图 9(b)	0.820 708	1
图 9(c)	0.808 366	2
图 9(d)	0.808 122	3
图 9(f)	1	1
图 9(g)	0.839 136	2
图 9(h)	0.831 258	3

表 2~4 分别给出了 3 种不情况下的服装类产品基于相似性得分的推荐排序结果。由表 2~4 可知:单独物品检索时像素精度相对较高,最高接近 92%;带人体的服装图像检索稍低,约 82%。另外,对于已经经过训练的图像再次检索时会直接检索到与自己本身语义一致的图像,检索精度为 1。因为深度学习主要是通过对图像特征进行的自主学习,所以对于学习过的图像进行了存储,因此估算精度很高。

4 结 语

通过分析传统服装分割语义模型中存在的问题,提出了一种基于深度卷积神经网络的图像语义分割算法。该算法能利用图像高层信息更精确地分析和理解图像内容,可以直接从像素级别上提取图像数据中的特征和语义信息。为了能进一步提高语义分割准确度以及提升对小部件物体的识别,又在该算法的基础上引入注意力机制优化网络模型,利用深度哈希函数建立服装特征库并进行相似性度量,最终实现语义相似性服装推荐功能。该算法可以较好地对服装图像进行语义分析,能够有效地用于服装检索与推荐。

参考文献(References):

[1] 陈鹏.文本图像检索技术的研究[D].成都:成都理工大学,2017.
CHEN P.The research on retrieval technology of document image[D].Chengdu:Chengdu University of Technology, 2017.(in Chinese)

[2] 李泗兰,郭雅.基于深度学习哈希算法的快速图像检索研究[J].计算机与数字工程,2019,47(12):3187-3192.
LI S L, GUO Y. Fast image retrieval based on hash algorithm in depth learning[J]. Computer & Digital Engineering, 2019, 47(12): 3187-3192. (in Chinese)

[3] 段立娟,高文,林守勋,等.图像检索中的动态相似性度量方法[J].计算机学报,2001,24(11):1256-1262.
DUAN L J, GAO W, LIN S X, et al. A dynamic similarity metric approach for image retrieval [J]. Chinese Journal of Computers, 2001, 24 (11): 1256-1262. (in Chinese)

[4] 万华林.图象检索中高层语义和低层可视特征的提取研究[D].北京:中国科学院研究生院(计算技术研究所),2002.
WAN H L. Research on extraction of high-level semantic and low-level visual feature for image retrieval[D]. Beijing: Institute of Computing Technology, Chinese Academy of Science, 2002. (in Chinese)

[5] YAMAGUCHI K, KIAPOUR M H, ORTIZ L E, et al. Parsing clothing in fashion photographs [C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2012: 3570-3577.

[6] LIU S, FENG J S, DOMOKOS C, et al. Fashion parsing with weak color-category labels[J]. IEEE Transactions on Multimedia, 2014, 16(1): 253-265.

[7] JI J, YANG R Y. An improved clothing parsing method emphasizing the clothing with complex texture[J]. Advances in Multimedia Information Processing, 2018,

- 10735:487-496.
- [8] LEDERER C, ALTSTADT S, KALANTIDIS Y. Getting the look: Clothing recognition and segmentation for automatic product suggestions in everyday photos [C]//Proceedings of the 3rd ACM Conference on International Conference on Multimedia Retrieval. New York: ACM, 2013:105-112.
- [9] 黄冬艳, 刘骊, 付晓东, 等. 基于 HOG 和 E-SVM 的服装图像联合分割算法[J]. 计算机工程与应用, 2017, 53(18):199-203.
- HUANG D Y, LIU L, FU X D, et al. Segmentation algorithm of clothing image based on HOG and E-SVM [J]. Computer Engineering and Applications, 2017, 53(18):199-203. (in Chinese)
- [10] CHEN L, TANG W, JOHN N W, et al. Context-aware mixed reality: A learning-based framework for semantic-level interaction [J]. Computer Graphics Forum, 2020, 39(1):484-496.
- [11] YAMAGUCHI K, KIAPOUR M H, BERG T L. Paper doll parsing: Retrieving similar styles to parse clothing items [C]//2013 IEEE International Conference on Computer Vision. Sydney: IEEE, 2013: 3519-3526.
- [12] 郭鑫鹏, 黄元元, 胡作进. 融合颜色与纹理的复杂场景下的服装图像分割算法[J]. 计算机科学, 2017, 44(S2):228-232.
- GUO X P, HUANG Y Y, HU Z J. Unsupervised complex-scene clothing image segmentation algorithm based on color and texture features [J]. Computer Science, 2017, 44(S2):228-232. (in Chinese)
- [13] CYCHNERSKI J, BRZESKI A, BOGUSZEWSKI A, et al. Clothes detection and classification using convolutional neural networks [C]//2017 22th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). Limassol: IEEE, 2017: 1-8.
- [14] 白美丽, 万韬阮, 汤汶, 等. 一种改进的用于服装解析的自监督网络学习方法[J]. 纺织高校基础科学学报, 2019, 32(4):385-392.
- BAI M L, WAN T R, TANG W, et al. An improved self-supervised neural network learning method for clothing parsing [J]. Basic Sciences Journal of Textile Universities, 2019, 32(4):385-392. (in Chinese)
- [15] CHEN L C, YANG Y, WANG J, et al. Attention to scale: Scale-aware semantic image segmentation [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 3640-3649.
- [16] MIZUOCHI M, KANEZAKI A, HARADA T. Clothing retrieval based on local similarity with multiple images [J]. Transactions on Multimedia, 2014, 16(1): 253-265.
- [17] 刘玉杰, 冯士贺, 李宗民, 等. 用于快速服装搭配的 FMatchNet 算法 [J]. 中国图象图形学报, 2019, 24(6):979-986.
- LIU Y J, FENG S H, LI Z M, et al. FMatchNet algorithm for fast clothing matching [J]. Journal of Image and Graphics, 2019, 24(6):979-986. (in Chinese)
- [18] 齐亚莉. 基于内容的图像检索与图像语义分析 [D]. 天津: 天津大学, 2017.
- QI Y L. Content based image retrieval and image semantics analysis [D]. Tianjin: Tianjin University, 2017. (in Chinese)
- [19] BASELGA A, JIMENEZ-VALVERDE A, NICCOLINI G. A multiple-site similarity measure independent of richness [J]. Biology Letters, 2007, 3(6):642-645.
- [20] FARABET C, COUPRIE C, NAJMAN L, et al. Learning hierarchical features for scene labeling [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8):1915-1929.
- [21] 陈双. 基于深度学习的服装图像分类与检索研究 [D]. 杭州: 浙江理工大学, 2019.
- CHEN S. Research on clothing image classification and retrieval based on deep learning [D]. Hangzhou: Zhejiang University of Science and Technology, 2019. (in Chinese)

责任编辑: 武 晖