



基于 Faster R-CNN 的连衣裙衣领自动识别

张艳清^a, 刘成霞^{a,b}

(浙江理工大学, a.服装学院; b.服装数字化技术浙江省工程实验室 杭州 310018)

摘要: 为探讨数据集标注对 Faster R-CNN 识别连衣裙衣领的影响,以框选范围依次扩大的 3 种标准(即衣领;唇部至肩部;唇部至腰部,后二者中的唇部对应平面服装图中的衣领上端)框选连衣裙图像中的衣领位置,制成 3 个数据集。以 Faster R-CNN 算法训练 3 个数据集,其中预训练网络均选用 AlexNet。结果表明:召回率与平均精度均值随框选范围的扩大均有提高,分别提高了 14% 和 13%,较小的立领、圆领、方领和 V 领提高最明显,而整体准确率随框选范围扩大保持不变。为进一步提升识别效果,以 GoogLeNet 为预训练网络,训练框选范围最大的数据集,得到的召回率、准确率和平均精度均值进一步增至 86%、81% 和 83%。Faster R-CNN 识别连衣裙衣领具有较好的可行性,扩大数据集框选区域可提升 Faster R-CNN 自动识别较小领型的效果。

关键词: Faster R-CNN; 连衣裙; 衣领; 目标检测; 分类

中图分类号: TS941.2

文献标志码: A

文章编号: 1673-3851 (2021) 11-0751-07

Automatic recognition of dress collar based on Faster R-CNN

ZHANG Yanqing^a, LIU Chengxia^{a,b}

(a. School of Fashion Design and Engineering; b. Zhejiang Province Engineering Laboratory of Clothing Digital Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: To explore the impact of dataset annotations on the recognition of dress collars by Faster R-CNN, three datasets were made following three standards to label collar positions in dress images. The regions be labeled of these three standards were expanded in turn (i.e., the collar itself, from the lip to the shoulder, from the lip to the waist. In the last two, the lip corresponds to the upper end of the collar in flat state clothing images). Faster R-CNN algorithm was used to train the three datasets respectively, and AlexNet was chosen as pre-training networks. The results showed that the recall rate and the mean average precision were improved with the enlargement of regions of interest by 14% and 13% respectively. The improvement of recognition of the smaller collars like the stand-up collar, the round collar, the square collar, and the V collar was the most obvious, while the overall accuracy remained unchanged with the enlargement of regions of interest. To further enhance the recognition results, GoogLeNet was used as the pre-training network and the dataset with the largest region in the three standards was selected in the training process. The recall rate, accuracy rate and mean average precision were further increased to 86%, 81%, and 83%. Faster R-CNN has good feasibility in dress collars' identification, and expanding the region of interest in the dataset properly can improve the performance of automatic recognition of smaller collars by Faster R-CNN.

Key words: Faster R-CNN; dress; collar; object detection; classification

收稿日期:2021-02-21 网络出版日期:2021-04-30

基金项目:浙江省自然科学基金(LY20E050017)

作者简介:张艳清(1996—),女,湖南衡阳人,硕士研究生,主要从事数字化服装技术方面的研究。

通信作者:刘成霞, E-mail: gloriol_liu@163.com

0 引言

衣领是最重要的服装部件之一,对服装风格塑造有极大作用。目前电商平台主要依靠人为添加标签来完成其分类^[1],随着服装设计的多样化,领型标注难度也日益加大。传统方法由于不同标注者对领型的认识存在差异,易导致同一衣领标注结果不同,造成消费者检索困难,且浪费大量人力和时间。

基于图像处理技术的服装识别分类方法^[2]在近几年获得了广泛的关注。按识别对象不同,服装图像识别可分为两类:一类是不同品类服装(如连衣裙、衬衫、裤子等)的识别^[3-4];另一类是相同品类不同款式或部件(如领子、袖子、底摆等)的识别^[5-6],后者识别对象更小,特征区别更细微,识别难度也更大。用于服装图像分类检索的方法也可归为两类:一类是通过特征提取算法对仅包含目标的图像提取特征从而分类;第二类是针对包含多个对象的图像,先确定目标位置再分类。由于现实中大部分的图像都包含多个目标,如连衣裙包含衣领和衣袖等部件,第二类能够自动定位目标,因此更符合实际需要。较经典的实现算法有 Faster R-CNN^[7]、R-FCN^[8]等。迄今,小目标及复杂背景服装图像的识别仍具挑战,也是服装检测的关注点,例如:刘正东等^[9]提出了基于尺寸分割和单次多盒检测的西装领识别方法,准确率达 90% 以上;Xiang 等^[10]利用改进 R-CNN 的方法来识别衬衫的领、袖等;Hidayati 等^[6]通过提取风格元素来识别服装款式。上述方法对提

高小目标及复杂背景服装部件的识别率作出了较大贡献,但识别对象种类较少。

连衣裙衣领种类繁多,而 T 恤、衬衫、卫衣、西服等款式的服装衣领种类略显单调,且各有偏重,例如:T 恤偏重圆领和 V 领;衬衫偏重企领和娃娃领;卫衣和西装分别以圆领和翻领为主。相比之下,连衣裙衣领则包含了大部分常见的服装领型,且数量分布更均匀,所以对连衣裙衣领识别的研究结果也适用于其他非连衣裙品类。但由于衣领在连衣裙图像中所占比例较小,尤其是较小的领型,容易产生漏检。若在数据集标注时,适当扩大框选范围,在对目标特征影响较小的情况下能减少漏检,或可提高识别率。为探索能否通过改变数据集标注标准提升 Faster R-CNN 对连衣裙衣领的识别率,本文以不同背景(包含纯色及各种环境背景)下的连衣裙领型为研究对象,针对立领、圆领等体积较小的领型易出现漏检的问题,分析数据集标注方式对 Faster R-CNN 识别连衣裙衣领的影响,以期为解决较小服装部件难以识别的问题提供新思路。

1 领型分类

综合考虑结构、造型及实用性,结合电商平台的相关分类,将领型分为 14 类(示例图见图 1,图片来自淘宝、京东等电商平台):立领、结带领、企领、圆领、V 领、一字领、方领、娃娃领、荷叶领、海军领、翻领、高领、卫衣领、组合 V 领。本文以广受女性欢迎的连衣裙为载体,研究连衣裙领型的自动识别。



图 1 14 类常见连衣裙领型示例

2 数据采集与数据集建立

以选自淘宝、京东等电商平台的 1400 张不同背景、遮挡程度、服装状态(如平面服装或人体着装)的连衣裙图像为样本集,每种领型 100 张,包含单一和复杂背景(由于来自电商平台产品图,背景不会太复杂),复杂背景又分 3 种情况(见图 2),训练集和测试集中复杂与单一背景图像的比例分别约为 9:5 和

4:1;遮挡程度分 3 级(见图 3),主要为头发和手势遮挡,其数量在数据集中占比低于 10%;训练和测试集中人体着装与平面服装的比例分别约为 10:4 和 4:1。在数据集中增加复杂背景、对衣领有遮挡、以及不同状态的连衣裙图像,主要为丰富样本,使模型更符合实际应用的需要。

为提高图像处理效率,裁剪除去文字及空白边框,以 8:2 的比例随机划分为训练集和测试集。

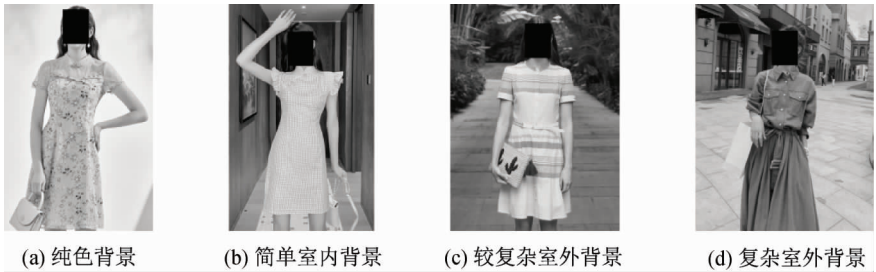


图 2 单一及复杂背景的图片示例



图 3 不同遮挡程度的图像示例

使用 MATLAB 软件中的 Image Labeler App 工具标记衣领信息,形成包含图像读取路径、标记框坐标和标记框标签的 table 文件,构成数据集。为探讨标注方式对识别结果的影响,用 3 种不同的标准 A、B、C 制作数据集,标准 A、B、C 框选的区域分别对应:衣领,唇部(或衣领上端)至肩部,唇部(或衣领上端)至腰部,得到对应的数据集 A' 、 B' 和 C' 。

3 定位与分类方法

Faster R-CNN 由 Ren 等^[7]于 2015 年提出,其在 Fast R-CNN 的基础上,引入了区域建议网络(RPN)代替选择性搜索方法,实现候选框的提取,大大缩短了检测时间,提高了准确率。

3.1 模型框架

Faster R-CNN 的模型框架如图 4 所示,包括特

征提取、目标区域建议和目标分类 3 部分。本文选择 AlexNet 作为初步实验的预训练网络,后续为进一步提高识别率,又增加了以 GoogLeNet 为预训练网络的实验。

3.2 区域建议网络

区域建议网络 (Region proposal networks, RPN)以单张任意大小的图像作为输入,输出一组矩形对象建议框,每个建议框产生一个得分。图 4 中左下虚线框内为 RPN 网络的结构,特征提取网络后面连接一个 3×3 的卷积层,来生成候选框,激活函数为线性整流函数 (Rectified linear unit, ReLU)。候选框内对应的特征将提供给完全连接的回归层(reg)和分类层(cls),分类层用于分类前景和背景,回归层用于计算候选框与真实边界框的偏移量。最后通过区域建议层筛选掉偏移量过大及相互重叠的候选框。

3.3 定位识别

定位识别模块位于图 4 右上角的虚线框中,特征图与候选框共同输入池化层,将候选框内特征转化为固定尺寸的特征图,通过全连接层映射到一个特征向量,再进行回归和多元分类,得到更精确的目标识别框和分类结果。

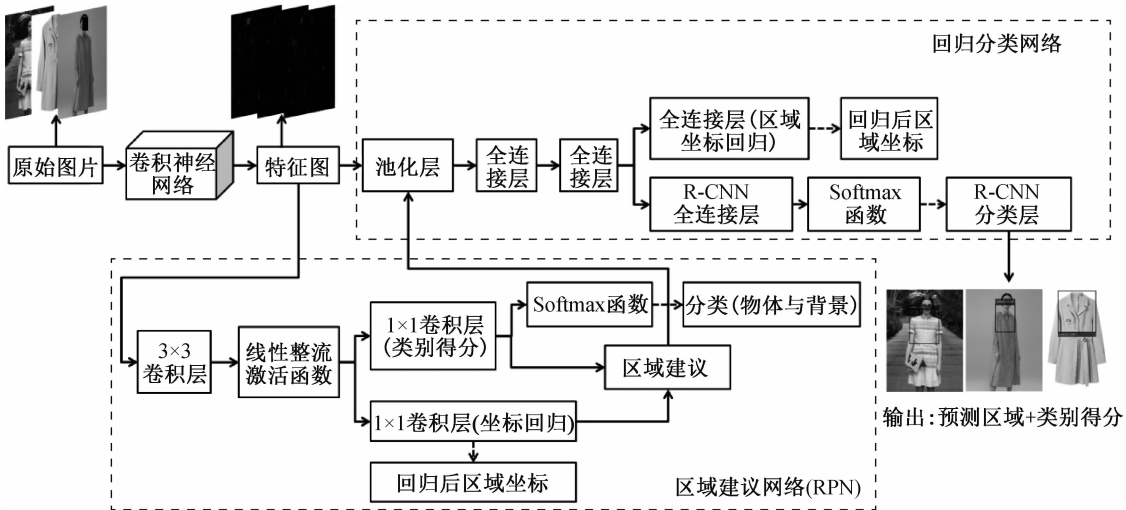


图 4 Faster R-CNN 模型框架

3.4 损失函数

Faster R-CNN 网络的损失包括 RPN 和 Fast R-CNN 的损失,二者均包含回归和分类两阶段的损失,即图 4 中虚线箭头所指阶段。在 RPN 和 Fast R-CNN 中,分类过程采用交叉熵损失函数(Cross-entropy loss),回归采用 smooth L1 函数。RPN 的总损失计算如式(1)所示,分类损失和回归损失的计算如式(2)–(4)所示:

$$L(\{p_i\},\{t_i\})=\frac{1}{N_{\text{cls}}}\sum_iL_{\text{cls}}(p_i,p_i^*)+\lambda\frac{1}{N_{\text{reg}}}\sum_i p_i^*L_{\text{reg}}(t_i,t_i^*)\tag{1}$$

$$L_{\text{cls}}(p_i,p_i^*)=-\log[p_i p_i^*+(1-p_i)(1-p_i^*)]\tag{2}$$

$$L_{\text{reg}}(t_i,t_i^*)=\sum_i\text{smooth L1}(t_i-t_i^*)\tag{3}$$

$$\text{smooth L1}(t_i-t_i^*)=\begin{cases}0.5(t_i-t_i^*), & |t_i-t_i^*|<1; \\ |t_i-t_i^*|-0.5, & \text{其他}\end{cases}\tag{4}$$

式中: i 为一张图片产生候选框的索引; p_i 指候选框内为物体的概率; p_i^* 为候选框的判别标签,若候选框内为物体,则 p_i^* 值为 1,若为背景,则为 0,通过 IoU (Intersection over Union) 值,即候选框与 groundtruth 的交并比值,来判断候选框内是否为物体,本文设置 IoU 值 0.5~1.0 时为物体, IoU 值 0.1~0.5 时为背景; t_i 为预测的候选框坐标向量; t_i^* 为真实边界框; N_{cls} 为分类数量; N_{reg} 为回归数量; λ 为权重参数,文中 λ 值取 10。

3.5 实验过程与评价指标

3.5.1 样本训练

本文中 4 次 Faster R-CNN 网络模型的搭建和训练均通过 MATLAB 2018b 软件实现,系统环境为 Windows10 的 64 位操作系统,12 GB 内存,Intel i5-5200U。参数设置:MaxEpochs 为 30,即最大的迭代次数为 33600 次,初始学习率为 0.0001,学习率的衰减系数为 0.1。在调整设置好参数后,分别训练得到 2 个网络训练数据集的 4 个结果 A'' -AlexNet、 B'' -AlexNet、 C'' -AlexNet(下文简称 A'' 、 B'' 和 C'' ,3 个结果分别与标准 A、B、C 相对应)和 C'' -GoogLeNet。单次训练分 4 步,训练 RPN 和 Fast R-CNN 网络,以得到的权重作为初始值,重复训练过程。由于 4 次训练的损失迭代结果类似,因此以 C'' 训练的损失迭代图为例进行说明(见图 5)。从图 5 可以看出分类损失与回归损失之和(Y 轴)随

迭代次数(X 轴)的变化过程:训练开始阶段,损失值较高;随着迭代次数增加,震荡下降,并趋于平缓,且在训练最后阶段,损失值保持在 0.2 以下,模型收敛较好。

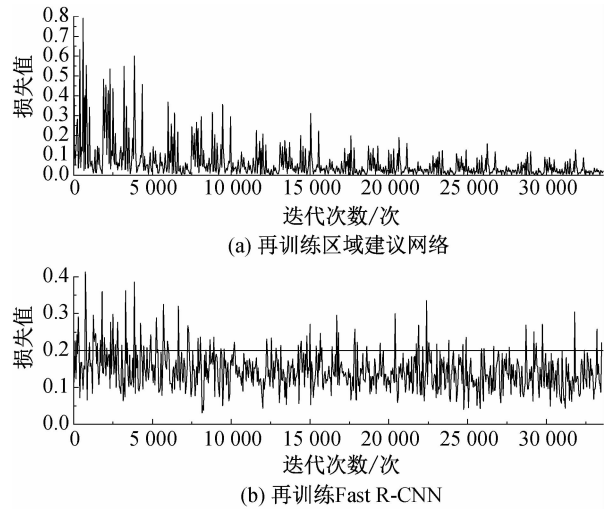


图 5 迭代损失图

3.5.2 评价指标

用于评价区域检测模型好坏的指标有:召回率 R (Recall)、准确率 P (Precision) 和平均精度 AP (Average precision)。其中,对所有类别的平均精度求均值得到平均精度均值 mAP (Mean average precision)。召回率、准确率和平均精度越高,模型分类效果越好。这 3 个指标的计算与正确检测 (TP)、错误检测 (FP) 和漏检 (FN) 密不可分。若预测边界框与对应的真实边界框的交并比 (IoU) 大于某一阈值(本文设为 0.5),且分类结果正确,则为正确检测,否则为错误检测。若某真实边界框缺少对应的预测边界框,则为漏检。其中召回率、准确率和平均精度的计算公式为:

$$R/\%=\frac{TP}{TP+FN}\times100\tag{5}$$

$$P/\%=\frac{TP}{TP+FP}\times100\tag{6}$$

$$AP=\int_0^1PdR\tag{7}$$

4 结果与分析

4.1 数据集标注对小目标识别的影响

将训练好的 3 个网络应用于测试集,得到每类领型的测试回归框数量,如图 6 所示,表 1 为 3 个网络预测的整体召回率、准确率和平均精度均值。

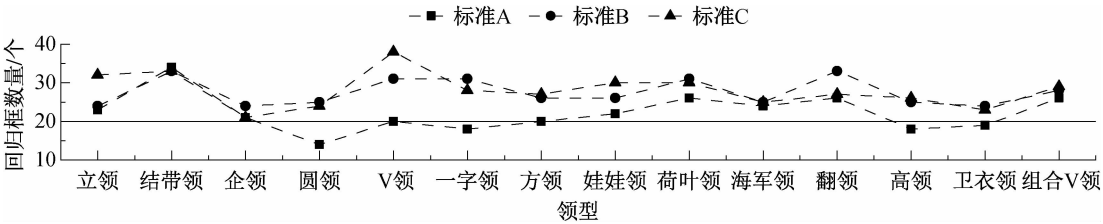


图 6 3 种标注下不同领型测试的回归框数量

图 6 和表 1 表明,随着框选范围的扩大,即由标准 A 到标准 B 及标准 C,召回率从 69% 提高到 83%,平均精度均值从 62% 提高到 75%,大部分领型的回归框数量也随之增加,说明扩大框选范围能够减少漏选;框选范围的增大,使杂乱信息混入,令特征提取及分类的难度增加,从而使一个衣领的周围产生了多个预测类别不同的候选框,漏检减少、正确检测增加,但错误检测也增加,从而出现了表 1 中用 3 种标准标注下的准确率保持不变的现象。

4.2 预训练网络对分类结果的影响

图 7 为不同预训练网络的测试结果。由图 7 可知以 GoogLeNet 为预训练网络的召回率(除企领、

表 1 召回率、准确率和平均精度				%
评价指标	标准			
	A	B	C	
召回率	69	80	83	
准确率	61	61	61	
平均精度均值	62	73	75	

方领和娃娃领)、准确率和平均精度均值均高于 AlexNet,其召回率达到 86%,准确率为 81%,平均精度均值 83%,原因为 GoogLeNet 网络对图像抽象特征提取能力更强,因此网络 C'-GoogLeNet 的误检数量较少,最终分类结果更准确。以上结果表明,增强特征提取网络的性能对提高模型准确率具有较大影响。

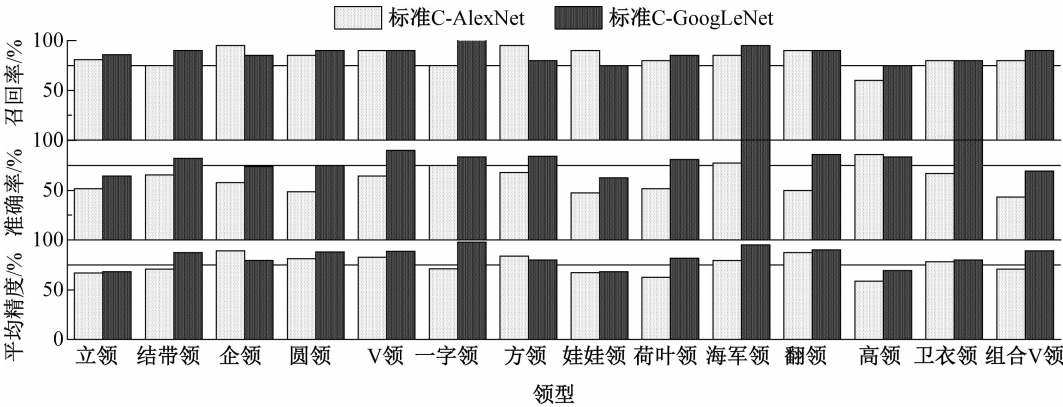


图 7 不同预训练网络下 14 类衣领的召回率、准确率和平均精度

4.3 领型对识别结果的影响

整体而言,扩大数据集标注范围能较好地提高 Faster R-CNN 识别衣领的效果,但对不同领型产生的影响也不尽相同。利用 3 个网络对 14 类领型的识别结果见图 8,各类识别结果对比如图 9—图 11 所示,各图中的文字均为最终识别的种类,文字后对应的数字为这一类别的得分。由图 8 可知,立领、圆领、娃娃领等较小的衣领,受数据标注方式的影响较显著,而一字领、荷叶领、海军领等较大的领型,受标注方式的影响较小。领型特征对结果的影响还存在以下规律:

a) 领型类似但预测结果差异较大的有企领与娃娃领。企领与娃娃领外形相似,但前者的识别效果优于后者(见图 9)。主要原因是娃娃领领面形状

变化较多,且外领周还可能带有装饰,因此易被误识为其他领型,而内领形状又易被误识为无领。企领则较为单一,且线条板正,易于识别。

b) 扩大数据集标注范围反而使组合 V 领的 3 个评价指标呈下降趋势,如图 10 所示。图 10 表明,以标准 C 时,同一衣领产生多个识别结果。主要原因为:组合 V 领本身由立领、翻领或圆领与 V 领组合而成,扩大特征范围易产生歧义。

c) 扩大标注范围,平均精度与召回率上升而准确率下降的情况如圆领,主要原因为:圆领属于典型的小目标,扩大标注范围使平均精度与召回率上升,另一方面,圆领较为简单,但可识别性不如 V 领,其特征识别易受服装花纹的影响,扩大标注范围也扩大了此种影响。图 11 所示为利用网

络 A'' 和 C'' 识别圆领的结果,同为误检时,网络 C'' 对准确率的影响小于误检,使得网络 C'' 的准确率难以提升。
的误检数量多于 A'' ,由于在目标数量相同时,漏检

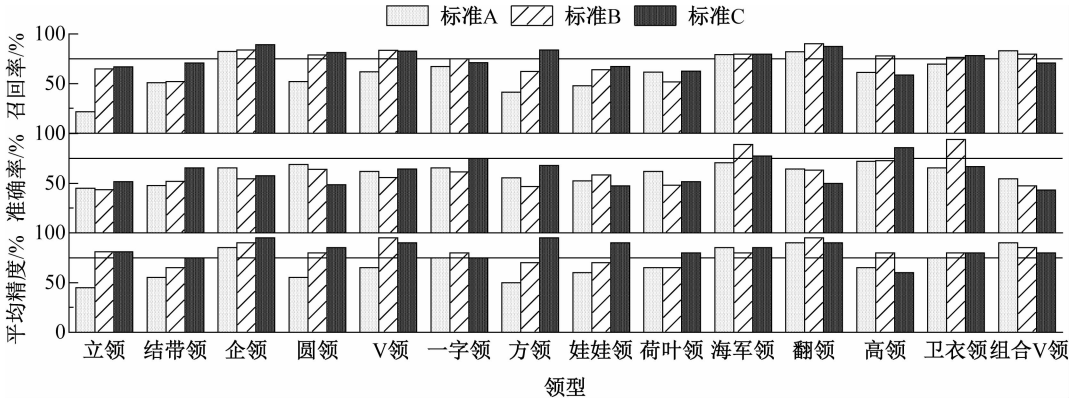


图8 不同标注下14类衣领的召回率、准确率和平均精度



图9 网络 C'' 识别企领、娃娃领结果对比



图10 网络 A'' 与 C'' 识别组合V领结果对比



图11 网络 A'' 与 C'' 识别圆领结果对比

5 结 论

本文以1400张来自淘宝、京东等电商平台的连衣裙图像为研究对象,用3种不同的框选标准制成数据集 A' 、 B' 、 C' ,Faster R-CNN 作为主体算法,首先用 AlexNet 预训练网络分别对3个数据集进行迁移训练,其次以 GoogLeNet 为预训练网络训练数据集 C' ,得到以下结论:

a) 扩大数据集标注范围可减少出现目标未检出的现象,从而提高召回率和平均精度,由标准 A 到标准 C,整体的召回率和平均精度均值分别由

69%和62%提升至83%和75%,且适当扩大数据集标注范围对小目标识别更为适用。

b) 预训练网络为 AlexNet 和为 GoogLeNet 相比,召回率、准确率和平均精度均值前者分别为83%、61%、75%,后者为86%、81%、83%,GoogLeNet 网络对图像抽象特征的提取能力更强。因此,可以通过调整 Faster R-CNN 的预训练网络,来提升识别性能。

通过本文的研究发现:扩大数据集标注范围可在一定程度上解决立领、圆领等体积较小的领型不易识别的问题,从而提高较小服装部件的识别率。

参考文献：

[1] 尹光灿, 罗戎蕾. 基于卷积神经网络的服装领型识别与分类研究[J].现代纺织技术, 2020, 28(3): 48-53.

[2] 李东, 万贤福, 汪军. 采用傅里叶描述子和支持向量机的服装款式识别方法[J]. 纺织学报, 2017, 38(5): 122-127.

[3] Madulid J P A, Mayol P E. Clothing classification using the convolutional neural network inception model [C]// Proceedings of the 2nd International Conference on Information Science and Systems. New York: Assoc Computing Machinery Press, 2019: 3-7.

[4] 吴欢, 丁笑君, 李秦曼, 等. 采用卷积神经网络 CaffeNet 模型的女裤廓形分类[J]. 纺织学报, 2019, 40(4): 117-121.

[5] 吴苗苗, 刘骊, 付晓东, 等. 款式特征描述符的服装图像细粒度分类方法[J]. 计算机辅助设计与图形学学报, 2019, 31(5): 780-791.

[6] Hidayati S C, Hua K L, Tsao Y, et al. Garment detectives: Discovering clothes and its genre in consumer photos [C]//2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). March 28-30, 2019, San Jose, CA, USA. IEEE, 2019: 471-474.

[7] Ren S Q, He K M, 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.

[8] Dai J F, Li Y, He K M, et al. R-FCN: Object detection via region-based fully convolutional networks [C]// Advances in Neural Information Processing Systems. California: Neural Information Processing Systems Press, 2016: 1-9.

[9] 刘正东, 刘以涵, 王首人. 西装识别的深度学习方法[J]. 纺织学报, 2019, 40(4): 158-164.

[10] Xiang J, Dong T T, Pan R R, et al. Clothing attribute recognition based on RCNN framework using L-softmax loss[J]. IEEE Access, 2020, 8: 48299-48313.

(责任编辑:唐志荣)