

文章编号: 2095-2163(2020)07-0306-05

中图分类号: TS 941

文献标志码: A

结合 Faster R-CNN 与 Grabcut 算法的服装图像自动分割

杨 思, 徐增波, 陈 冲

(上海工程技术大学 服装学院, 上海 201620)

摘要: 为了实现服装图像自动分割处理, 本文提出基于 Faster R-CNN 模型结合 GrabCut 的图像分割方法。利用快速区域卷积神经网络的基本框架, 将街拍图的待检测任务分为上衣、裙子、包等六个类别, 对原有的基本框架模型的全连接层参数进行调整, 得到前景目标框作为 GrabCut 分割算法的初始框, 再使用 GrabCut 算法进行服装区域提取, 从复杂背景的图片中定位服装位置, 去除复杂背景, 实现服装区域分割。实验结果显示, 本文方法能够很好的实现服装自然轮廓检测和提取, 适用于图像局部弱轮廓边缘的检测及大批量服装图像分割处理, 并且可供大批量处理图片时选择性自动款式类别提取, 提高了服装图像分割处理的效率。

关键词: 服装图像分割; 快速区域卷积神经网络; GrabCut 分割算法; 轮廓检测

Automatic segmentation of clothing images based on Faster R-CNN and Grabcut algorithm

YANG Si, XU Zengbo, CHEN Chong

(Shanghai University Of Engineering Science Costume Institute, Shanghai 201620, China)

[Abstract] In order to realize the automatic segmentation of clothing images, an image segmentation method based on Faster R-CNN combined with GrabCut is proposed. Firstly, the basic framework of the Fast R-CNN is used to subdivide the to-be-detected tasks of street photographs into six categories: tops, skirts, bags, etc. Next, after adjusting the model full connection layer parameters based on the original basic framework, the foreground object box is obtained as the initial frame of the GrabCut segmentation algorithm. Then we use GrabCut algorithm to extract garment area. The method locates the clothing position from the picture of the complex background, removes the complex background, and realizes the segmentation of the clothing area. The experimental results show that the proposed method has good natural contour detection and extraction ability, and is suitable for the detection of local weak contour edges of images and the processing of large-scale clothing image segmentation. Beyond that, it can be used for selective automatic style category extraction in large batch processing of images. It improves the efficiency of the garment image segmentation process.

[Key words] clothing image segmentation; Faster R-CNN; GrabCut; image extraction

0 引言

服装图像分割是指在某张带有服装的自然图像上对服装部分的形状进行检测, 识别并提取出来的操作。服装图像分割与边缘提取在数字化服装的研究中占据了极其重要的地位: 一方面快速提取款式图, 通过提取到的款式图匹配出最为相近的样板, 可以提高制版师的工作效率, 节省版型制作和设计开发时间^[1]; 另一方面在基于图像序列的三维重建中, 目标提取的准确性和速度对重建的质量和效率起决定性作用^[2]。

以往的计算机视觉处理中, 图片提取大多以灰度图像为主。而随着日益发展的数码拍照技术, 人们不再满足于灰度图像单一的颜色, 因此对于彩色图像处理的研究也被重视起来。常见的支持彩色图

像分割的算法有基于边缘检测的算法、卷积神经网络、Graph cuts、Grabcut 等。其中, 边缘检测算法不适用于噪声多、图像质量差的图像, 卷积神经网络在服装图像分割上存在缺少样本数据集的问题, Grabcut 是在 Graph cuts 的基础上改进的算法, 结合了图像处理技术中常用的高斯混合模型以及图割理论, 能获得准确的分割结果, 该算法一经提出便获得了业界广泛的关注。但是大多数对于 GrabCut 的优化研究重点都在算法性能方面, 使这种方法长期处于半自动状态, 虽然其在一些领域有一定的应用, 但不适合应用于大量图像的提取和识别。直到 2014 年 Dina 等人提出一种以聚类算法与 GrabCut 算法的自动图像提取方法^[3]。本文在此基础上提出了结合 Faster R-CNN 与 Grabcut 算法的服装图像自动

基金项目: 上海市科学技术委员会科技创新行动计划资助项目(18030501400); 研究生创新项目(E3-0903-18-01185)。

作者简介: 杨 思(1996-), 女, 硕士研究生, 主要研究方向: 服装数字化。

通讯作者: 徐增波 Email: 1299860310@qq.com

收稿日期: 2019-12-20

分割方法。以在多场景拍摄,具有不同背景的图片作为研究对象,针对服装款式图识别与提取进行研究。以服装区域的提取作为切入点,通过深度神经网络训练,自动进行图像检测,并标注出服装的种类,使用GrabCut算法对图片进行提取。本文将深度学习技术与服装行业相结合,为服装设计服装制造提供了技术支持,对从复杂背景中自动提取服装款式具有现实意义。

1 Fast-RCNN 服装区域检测模型

1.1 Fast-RCNN 概述

快速区域卷积神经网络(Fast Region CNN)是Ross Girshick在区域卷积神经网络和SPP-Net的基础上增加了Multi task training,即利用DeepCNN进行目标的快速检测^[4]。该方法在很大程度上节省了提取特征和分类的时间,并实现了End-to-End的Joint Training,直接框选出要检测的区域。故本文采用Faster R-CNN算法,实现图片中服装区域的有效识别。图1为快速区域卷积神经网络的架构图。

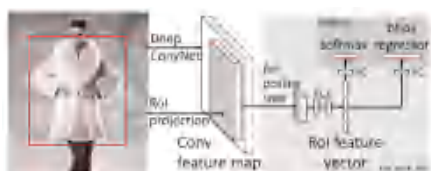


图1 快速区域卷积神经网络架构图

Fig. 1 Fast regional convolutional neural network architecture

1.2 基于Fast-RCNN确定初始提取区域

服装风格款式多样,服饰搭配多元,使服装检测的困难加大,快速定位服装的区域成为当前研究的核心内容。如图2所示,本文需要快速且准确检测出服装图像所在区域的位置,也会检测出其他配件的位置,如鞋,包等。

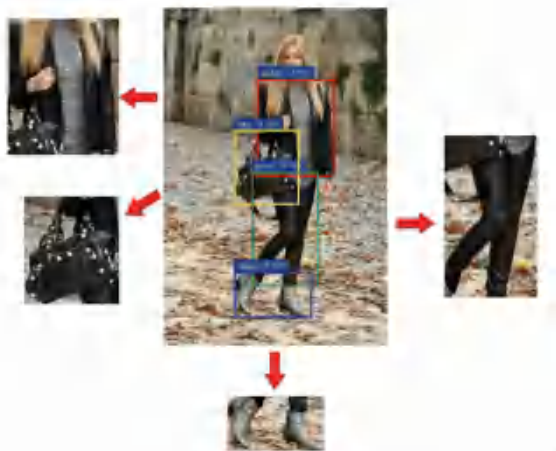


图2 服装图像所在区域的位置检测

Fig. 2 Position detection of the area where the clothing image is located

本文将常见的服饰标签分为上衣、裤子、连衣裙、半身裙、鞋子、包六类,采用快速区域卷积神经网络对这六类目标进行检测。首先,以全部训练样本作为输入对象,通过对图片进行选择性搜索,得到约2k个RoI;其次,对图像进行放缩,得到conv5的特征金字塔;再利用单层SPP layer(RoI pooling layer)将其统一为相同尺寸;最后,将这些特征分别共享到两个新的全连接,并对连接上的两个目标进行优化。这两个优化目标分别使用柔性最大值传输函数进行分类,使用平滑的L1损失函数进行边界框回归。

由于对目标检测对象进行了分类识别处理,可供大批量处理图片时选择性自动款式类别提取,如只需提取上衣,就可设定快速区域卷积神经网络检测识别时只留下上衣区域的初始框,再进行后续分割处理。本文提出的方法区别于对服装区域分割后再识别操作的一般服装图像分割方法,能有效提高批量服装图像分割处理的工作效率。

1.3 实验结果

本文将常见的三个数据库:时尚服饰数据集Clothing Co-Parsing (CCP) Dataset(包含2 098张已标注数据)、Colorful Fashion Parsing Data(包含2 600张已标注图片)、DCSA(包含1 856张图片)合并,合并后样本数据集共包含5 554张图片,使用3 894张用作训练样本,1660张作为测试样本。基于快速区域卷积神经网络框架,采用VGG-16网络训练架构训练,并与传统区域卷积神经网络进行比较。实验基于Ubuntu17.04,64位操作系统,采用开源深度学习框架Tensorflow,硬件配置为NVIDIA GeForce GTX1080,8GB显存。实验结果显示,区域卷积神经网络的图像平均检测时间约为74s,而快速区域卷积神经网络的图像平均检测时间约为0.51s。从mAP值来看,RCNN和Fast R-CNN的检测准确率差别并不明显,但是快速区域卷积神经网络的检测速度比区域卷积神经网络的检测速度高出100多倍。

为了提高快速区域卷积神经网络检测的准确性,本实验在Deep Fashion数据集中选取了4 000张图片,逐个手工标记,并加入先前合并的数据集中。从新的数据集中选取2 318张图片水平反转,新的数据集共包含12 172个样本,8 521个训练样本,3 651个测试样本,重新训练模型,并再次测试新的模型。模型测试准确率结果详见表1。其中S表示小样本训练集,L表示大样本训练集。从表1可以看出,随着样本集数量的增加,各类别的准确率和

mAP 都会有一定程度的提高,如果再增加样本数量,可能会得到更好的结果。

表 1 各类别检测准确率平均值

Tab. 1 Average accuracy of each category %

方法	RCNN(S)	Fast R-CNN(S)	RCNN(L)	Fast R-CNN(L)
上衣	82.7	84.2	86.9	87.2
裤子	57.8	57.1	64.1	65.3
连衣裙	65.5	65.7	71.1	72.8
半身裙	78.3	80.6	84.3	84.9
鞋子	74.1	75.3	78.9	79.5
包包	58.7	59.1	61.9	62.3
mAP	69.52	70.33	74.63	75.33

快速区域卷积神经网络的实际检测效果如图 3 所示,(a)为样本实际标签位置(GT),(b)为区域卷积神经网络的检测结果,(c)为快速区域卷积神经网络的检测结果。实验结果表明,本文所采用的快速检测方法效果良好。除了一些角度问题和遮挡问题导致误检,上衣,鞋子和裙子的检测都比较准确。两种区域识别算法的识别精度是可以接受的,但是对袖子、裙摆等部分在检测框外的部分识别还不够完整。



图 3 实际检测效果对比

Fig. 3 Comparison of actual detection results

2 改进的 GrabCut 算法实现

2.1 GrabCut 算法介绍

GrabCut 算法是一种需要用户交互的图像分割方法,是基于 Graph Cut 算法的改进^[5]。该算法利用图像中纹理(颜色)信息和边界(反差)信息,基于

被分割对象的指定边界框开始,使用高斯混合模型估计被分割对象和背景的颜色分布,只要少量的用户交互操作就可得到较好的分割效果,得到的结果也比较精确^[6]。

2.2 实验图像选取

本文选取的实验图像一部分来自于前述的数据库,另一部分来源于自行拍摄,以保证实验的普遍性和有效性。所选图像分为三类:目标简单背景简单、目标复杂背景简单和目标复杂背景复杂。图 4 为部分待分割图像,第一行是目标简单背景简单类的图像,第二行是目标复杂背景简单类的图像,第三行是背景复杂且有其他人物干扰的图像。



图 4 实验选取的部分图片

Fig. 4 Some images selected for this experiment

2.3 实验步骤

首先利用快速区域卷积神经网络确定服装区域,将获取到的区域坐标提取出来,使用 GrabCut 进行分割。主要步骤如下:

Step 1 采用快速区域卷积神经网络训练模型。输入目标图像位置,得到识别出的矩形服装图像的区域坐标;

Step 2 将 Step1 识别的框外像素作为背景像素,框内像素作为目标像素,利用这两组像素训练背景高斯混合模型和目标高斯混合模型;

Step 3 利用这两组训练好的 GMM 模型以及能量函数 E 中的数据项,计算出每个像素属于目标和属于背景的概率。能量函数中的平滑项的计算与 GraphCut 的计算大致相同;

Step 4 通过能量函数的优化得到最终的分割结果;

Step 5 用 Step3 得到的结果中的背景像素和前景像素训练背景高斯混合模型和目标 GMM 模型;

Step 6 重复 Step3、4、5 直到目标结果收敛(两次迭代的结果变化小于一定程度)。

3 实验结果与分析

3.1 分割效果评价方法

本文通过有监督学习的训练集的分类准确性(ground truth)来计算所得结果图像的过分割率 E_p 、分割的准确度 T_p 和欠分割率 D_p 等指标, 以此对分割效果进行评价^[7]。为了评价分割算法的准确性, 需要人工手动对图片的目标区域即所穿着的服装区域进行标注, 将提取到的区域与算法提取的结果进行比较。服装分割准确度 T_p 的计算公式(1)为:

$$T_p = \frac{\frac{\sum |A_s - P_s|}{A_s}}{\frac{\sum |A_s - P_s|}{A_s}} \times 100\% \quad (1)$$

其中, TP 表示分割准确的面积占有监督学习的训练集的分类准确性图像中真实区域的百分比; A_s 表示人工标注出待分割图像的参考区域; P_s 表示该算法获得结果图像的实际区域; $|A_s - P_s|$ 表示像素点被错误分割的数量。 T_p 为该算法取得的结果, T_p 的值越接近 100%, 表示算法得到的目标区域准确度越高, 分割效果越好; T_p 的值越接近 0, 表示该算法得到的目标区域准确率越低, 分割效果越差; 当 $T_p = 0$ 时, 分割效果最差, 即完全错误的分割结果, 没有获取正确的目标。

图像的过分割率 E_p (即提取 GT 图像的参考区域与区域外的像素点比率)的计算公式(2)为

$$E_p = \frac{E_s}{A_s + E_s} \times 100\% \quad (2)$$

其中, E_s 表示在实际的最终结果中错误出现的像素数量, 这些像素本不应该存在于目标中, 但在实际分割后却包括在目标中。即在实际获得的图像中的像素点出现在 E_s 中, 但在理论上该像素在目标图像 A_s 中不会出现。

欠分割率 D_s (即最终获得的图像结果在 GT 图像中缺少的像素点占参考区域像素点的比率)的计算公式(3)为

$$D_p = \frac{D_s}{A_s + E_s} \times 100\% \quad (3)$$

其中, D_s 表示最终结果中不存在但是本应该在结果中包含的像素点。即 D_s 中的像素没有出现。

即在实验中, 像素没有出现在目标图像中, 但是在理论上应该出现。

3.2 分割结果对比

图 5 展示了不同分割方法得到的结果, (a) 列展示的是原图; (b) 列是手动获取的目标图像; (c) 列是通过 GrabCut 检测得到图像区域后分割得到的结果; (d) 列是通过改进的 GrabCut 自动分割得到的结果。从图 5 可以看出, 对于背景简单的服装图像, 几种分割方法的提取效果都比较好, 对于背景复杂的或有干扰的图片, 有人工交互的方法的准确率比自动识别的准确率高, 因为自动识别会识别到属于背景的一些服装, 造成过度分割等情况。



图 5 不同分割方法比较

Fig. 5 Comparison of different segmentation methods

3.3 结果分析

本文通过有监督学习的训练集的分类准确性计算提取服装图像的 T_p 、 E_p 、 D_p 等指标。评估算法的最终结果, 见表 2。

表 2 三种方法平均分割准确率

Tab. 2 Average splitting accuracy of three methods

方法	T_p	E_p	D_p
传统 GrabCut	0.978	0.035	0.048
自动 GrabCut	0.959	0.064	0.036

根据表2的目标分割结果可以看出,有人工交互的方法分割的准确率较高,但是其处理每一幅图像都需要进行人工标注,时间成本较高。自动分割方法实现了全自动分割目标图像,很大程度上节省提取时间,传统方法分割的准确率与自动分割准确率相差不大。

4 结束语

本文针对服装图像分割处理这一问题,提出结合 Faster R-CNN 模型的 GrabCut 图像分割方法。采用快速区域卷积神经网络自动检测目标服装区域作为 GrabCut 分割算法的初始框,再使用 GrabCut 算法进行提取,实现了从复杂背景的图片中提取指定类别的服装区域的功能。结果表明,本文方法拥有良好的自然轮廓检测和提取能力,并可以进行大量的图片处理,推动了图像分割研究从半自动分割到自动分割的进程。

(上接第305页)

5 系统演示

系统设计完成后,应用在其默认域名下所展示的网页效果如图5和图6所示。



图5 水参数监测报警状态

Fig. 5 Monitoring and alarm status of water parameters



图6 水参数监测正常状态

Fig. 6 Normal state of water parameter monitoring

该 web 界面主要包含水温参数监测显示、水温阈值的设定、报警信息显示以及水的 PH 值参数监测的仪表盘显示等。图5所示水参数监测报警状态,系统监测到的实时水温为 35 °C,超过水温阈值设定的 25 °C,故报警灯为红色,表示此时水参数监

参考文献

- [1] Zhou Wencan. The future of pattern maker in clothing [J]. Melliland China, 2015, 43(1):66-68.
- [2] 栾广宇. 基于多幅图像的建筑物三维重建关键技术研究[D]. 哈尔滨:哈尔滨工业大学,2009.
- [3] KHATTAB D, EBIED H M, HUSSIEN A S, et al. Automatic GrabCut based on unsupervised clustering for image segmentation [J]. Journal of Computer Science and Technology, 2015.
- [4] ROSS B. Girshick. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision, Santiago, Chile: ICCV, 2015: 1440-1448.
- [5] ROTHER C, KOLMOGOROV V, BLAKE A. " GrabCut" interactive foreground extraction using iterated graph cuts [J]. ACM transactions on graphics (TOG), 2004, 23(3): 309-314.
- [6] BOYKOV Y Y, JOLLY M P. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images[C]// Proceedings of the Eighth International Conference on Computer Vision. Canada: ICCV. 2001:105.
- [7] HAN Zhe. The research on identification technologies of benign and malignant breast tumor based on convolutional neural network[D]. Taiyuan University of Technology, 2018.

测超过阈值,处于报警状态,此时的 PH 值监测数据为 6.8。图6所示水参数监测正常状态中,系统监测到的实时水温为 35 °C,未超过水温阈值设定的 50 °C,此时报警灯为蓝色,表示当前水参数监测未超过阈值,处于正常状态,此时的 PH 值监测数据为 7.2。

6 结束语

面向人们在生产、生活中对水质要求越来越高的现状,针对传统水环境参数监控方式的不足,设计实现了以 STM32 单片机为采集节点的管理中心,以 WIFI 为通信形式,以阿里云平台为网络媒介的远程水参数监测管理系统。系统运行结果表明,该设计硬件模块结构简单;web 交互界面友好、便捷;水质参数监测实时、有效。对开发者来说,系统的后期功能扩展也比较方便,如可以通过增加传感器,开展其他水质参数的检测,也可以通过更改 ESP8266 模块的工作模式,完成多点监控组网工作,还可以通过云端和手机端的连接,开发移动终端的 APP 监测软件。

参考文献

- [1] 余志剑. 基于 B/S 架构的远程水质监测系统的设计与实现[D]. 广州:华南理工大学,2017.
- [2] 陈亮. 基于 Zigbee 的鱼塘水质监测系统的设计[J]. 电子技术与软件工程, 2019(17): 63-65.
- [3] 连丽红,江伟,洪伟捷. 基于阿里云 IoT 平台的智能停车系统研究[J]. 中国集成电路, 2019, 28(12): 76-79, 86.
- [4] 杨飞,谢涛,伍英,等. 基于 WIFI 的农业物联网温室大棚环境监测系统的设计[J]. 计算机测量与控制, 2017, 25(2): 50-53.
- [5] 丘源,经本钦,李精华. 基于 ESP8266WiFi 模块和 MQTT 协议的物联网传感节点设计[J]. 物联网技术, 2019, 9(6): 24-26, 29.