基于部件检测与检索的行人精细化分割

王 枫,厉 智,刘青山,孙玉宝
(南京信息工程大学信息与控制学院,江苏省大数据分析技术实验室,江苏南京 210044)

摘 要: 针对行人图像外观的多样性以及结构、姿态、场景的复杂性,提出一种有效的精细化行人部件分割方法.该方法实现把一幅行人图像分割成不同的语义区域,主要包含三个阶段,前两个阶段单独训练两个 Fast R-CNN (Fast Region-based Convolutional Neural Network,快速区域卷积神经网络)模型,分别用来检测整个人体以及各个部件以获得各类别部件的大体位置;第三个阶段使用基于检索过分割图像的方法来对检测到的各个部件进行分割,最后把各部件分割结果还原到原图坐标上以得到最终的分割结果.实验表明所提方法在三个公开的数据库上,与其他算法相比,分割准确率更高,边缘效果更好.

关键词: 行人分割; 快速区域卷积神经网络; 过分割; 部件检索
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2019)02-0502-07
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2019.02.035

Fine Pedestrian Segmentation with Parts Detection and Retrieval

WANG Feng, LI Zhi, LIU Qing-shan, SUN Yu-bao
(*B-DAT Lab, Collaborative Innovation Center, School of Information & Control, Nanjing University of Information Science and Technology, Nanjing, Jiangsu* 210044, *China*)

Abstract: Focused on the diversity of appearance and the complexity of configuration, laying, and occasion in human images, a coarse-to-fine method was proposed for effective human parsing. It can decompose a human image into semantic regions which consists of three phases. In the first two phases, two effective models were trained with Fast Region-based Convolutional Network (Fast R-CNN) to respectively detect human body and clothing items. In the third phase, parsing clothing items based on retrieving similar over-segmented images and morphing them into absolute image coordinates. Experiments are conducted on three public databases, and the experimental results show that proposed method has higher accuracy and promising performance.

Key words: pedestrian segmentation; Fast R-CNN; over-segmentation; parts retrieval

1 引言

行人精细化分割主要是将行人图像中的人体部件区域(上衣、背包、脸等)分割出来,由于其在人体图像分析中的广泛应用而备受关注,例如人员识别以及服装分析.随着在线服装销售的迅速发展,这项研究在电子商务中有着巨大的应用前景.许多研究关注于服装属性预测[1],服装推荐[2],以及通过服装的社会身份识别[3].不同于以上这些工作,本文的工作是对人体图像的几类主要部件区域做出精确的分割.

行人部件精细化分割是一个非常具有挑战性的问题,由于行人服装的外观,层次和风格差异很大,目标类别越多,对边界处的效果要求也越严格.对于衣服分割现已提出许多方法.经典的算法有复合与或图模型[4],文献[5]用来模拟和解析衣服结构.文献[6]的工作利用封闭模型对高度遮挡的群图像进行衣服分割,可变形空间先验模型用来提高衣服分割的性能[7].基于形状的人体模型方法[8],姿态估计和监督区域标记[9]等方法都取得了较好的结果.Torr 和 Zisserman 等人提出一种基于 CRF(Conditional Random Field,条件随机场)框架的方法同时实现人体姿态估计和人体部件标记[10],是分割与姿态估计[11]的结合.一些分割方法如 parselet[12] 和 co-parsing[13] 利用基于分割方法[14]的区域假设,由自上而下的信息生成物体区域.但是这些方法假定物体有很大的可能性属于其中至少一种区域,当

收稿日期: 2017-04-12;修回日期:2017-12-12;责任编辑: 郭游
基金项目:国家自然科学基金(No.61532009)

物体外观变化较大时这种假设并不一定成立. 文献[15]结合多模型的融合,通过大量的计算可以得到较好的结果,但是模型较为复杂,效果上也有待提升. 还有一些方法[16,17]检索整张输入图像来进行分割.

之前的方法在人体分割方面已经取得了较好的效果,但是这些算法仍然存在着一些问题. 这些设计的对应关系不能完全找到人体图像的外观与结构之间的相关性. 之前的一些方法对于一些具体的任务通常需要大量的先验信息,例如文献[18]的分割方法依赖于人体的姿态估计[19],姿态估计的结果很大程度影响了分割的效果. 另外,还有一些工作容易混淆不同的部件并且不能分割出衣服的细节[15],并且该工作对相似图像查询是基于整张图像进行的,这样做法比较粗糙,整体效果不够理想,边缘信息融合也比较粗糙,不够平滑.

近年来,随着深度学习在图像分析上的成功运用,不少研究人员也将深度卷积神经网络应用到图像语义分割领域[20,21].

本文提出一种由粗到细的行人部件分割方法. 首先训练一个 Fast R-CNN 模型对原始图像检测整个人体,然后训练另一个 Fast R-CNN 模型在上一步人体检测结果的基础上来检测各个衣服部件. 最后对各个部件进行过分割,利用过分割检索的方法对图像进行进一步分割. 实验在三个公开数据库上进行,实验结果表明本文的方法能够取得较好的分割结果.

2　基于 Fast R-CNN 和过分割检索的部件分割

本文的目标是对衣服人体图像进行精细化的分割. 本文的方法主要包含两个部分:部件检测和分割. 如图 1 所示,给定一幅图像,首先基于 Fast R-CNN[22]框架在原始图像上检测出整个人,然后利用另一个独立的 Fast R-CNN 模型在检测出的人的区域上检测各个部件,得到部件图,并且对每一个部件图像进行过分割. 本文事先建立一个包含各个部件原始图像和对应分割图像的数据库,使用基于结构相似性的方法检索数据库中最相似的部件图. 对过分割块进行组合,产生一系列候选结果,把这些候选结果与最相似的部件图对应的分割图计算相似度,得分最高的组合就认为是分割的结果.

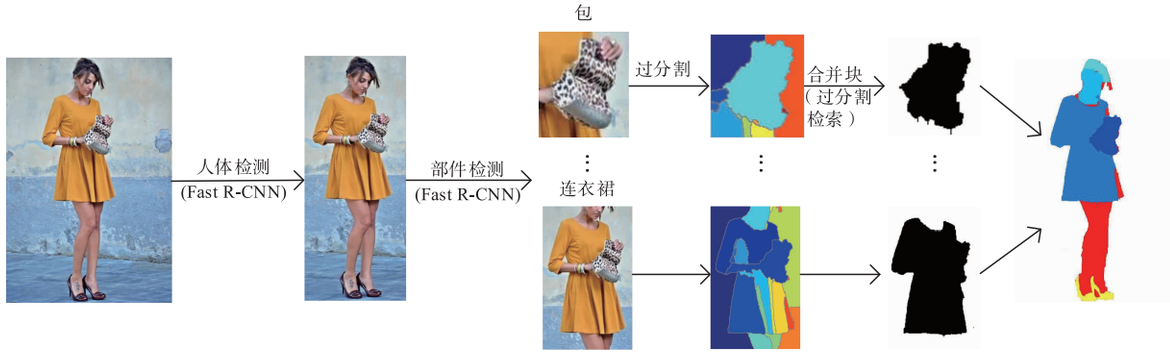


图1　分割整体流程图

2.1　基于 Fast R-CNN 的人体检测和部件分割

本文通过对各个部件进行检测来直接获得部件的位置信息而不是依赖姿态估计. 为了降低误检率,本文采用一种由粗到细的方式,首先对整个人体进行检测,这样,人体检测框里包含了所有的部件. 具体地,本文使用基于 Fast R-CNN[22]的方法检测人体和各个部件,对人体和部件的检测单独训练两个 Fast R-CNN 模型. Fast R-CNN 是一种新的高效的目标检测算法. 它的训练测试过程相对较快,并且准确率较高.

图 2 是 Fast R-CNN 的框架结构(虚线表示人体的检测,实线表示部件的检测). Fast R-CNN 网络把一整幅图像和一系列物体候选框作为输入,候选框由选择性搜索(selective search)得到,图像首先经过一系列卷积以及池化层,获得卷积特征图. 然后,对于每一个物体候选框,通过 RoI 池化层将特征图下采样成为大小固定的特征向量. 这些定长的特征向量经过两个全连接,再分别送到两个新的全连接层. 一个是使用 softmax 对 K 类目标及背景进行分类,另一个任务是对目标的位置进行回归,使用了平滑的 L1-loss.

Fast R-CNN 网络有两个输出层,第一个输出是离散的概率分布 $p=(p_0,\cdots,p_K)$,是一个 $K+1$ 个的 softmax 输出,其中的 K 是类别个数,1 是背景. 第二个输出是框的回归坐标 $t^k=(t_x^k,t_y^k,t_w^k,t_h^k)$,对于 K 个目标类别,索引为 k. 相应的有两个损失函数,分别为分类损失函数和回归损失函数,损失函数 L 可以表示为式(1):

$$L(p,u,t^u,v)=L_{cls}(p,u)+\lambda[u\geq 1]L_{loc}(t^u,v) \quad (1)$$

当 $u\geq 1$ 时,$[u\geq 1]$为 1,否则为 0.

对于分类损失函数,对于类别 u,分类损失函数用式(2)表示:

$$L_{cls}(p,u)=-\log p_u \quad (2)$$

对于回归损失函数,是一个 $4\times K$ 路输出的回归器,对于每个类别都会单独训练一个回归器,回归器的损失函数是一个平滑的 L_1 范数,形式如式(3)所示:

$$L_{loc}(t^u,v)=\sum_{i\in\{x,y,w,h\}}smooth_{L_1}(t_i^u-v_i) \tag{3}$$

$$smooth_{L_1}(x)=\begin{cases}0.5x^2,\\ |x|-0.5,\end{cases} \tag{4}$$

其中,对于类别 u,$v=(v_x,v_y,v_w,v_h)$ 为坐标真实值,$t^u=(t_x^u,t_y^u,t_w^u,t_h^u)$ 为预测值.

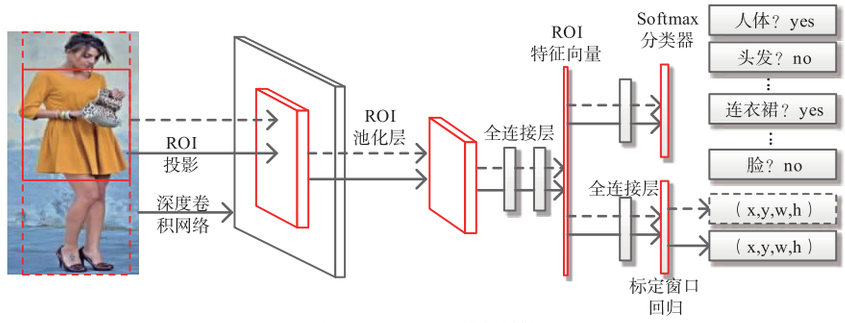


图2 Fast R-CNN框架结构

本文对人的检测以及部件检测分别训练两个模型,在得到的人体的检测框中检测各个部件,并且根据各个部件的检测框把各个部件裁剪出来.

考虑到检测的不准确,人的检测框可能并不是包含了所有的部件,因此本文把检测框扩大 1.2 倍,检测框外的部分都被当作背景.

2.2 过分割检索

本文使用基于相似图像检索的方式分别分割检测到的每一个部件. 在运用 Fast R-CNN 模型检测到每一类部件之后,分别对部件进行过分割,得到若干个区域块,接下来对这些过分割块进行合并,获得每一类部件的分割结果后再把他们转换到原图的绝对坐标上.

本文在边缘检测[23]的基础上进行过分割. 给定一幅图像,首先使用结构边缘检测器[24]检测物体的边缘,然后计算图像中与物体边缘相连的超像素,最终对图像生成几个块. 图 3 是过分割过程图,不同的颜色表示各个不同的块.

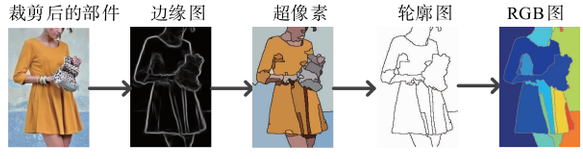


图3 过分割过程图

上一步已经得到了图像的过分割结果,接下来需要对这些块进行合并得到准确的部件区域. 本文采用的方法是基于相似图像检索.

具体地,首先对每一类建立足够数量的真实值数据库,根据分割结果的真实值,裁剪出每一类部件的图像以及每一类的分割结果图像. 在部件检测过程后,把检测到的部件图像都裁剪出来. 给定一幅裁剪出来的图像,本文使用基于 SSIM(Structural Similarity Index Measurement,结构相似性)的方法首先对它在数据库中检索最相似的图片.

结构相似性分别从亮度、对比度、结构三方面度量图像相似性. 给定两幅图像 x 和 y,它们的结构相似性可以通过式(5)求出:

$$SSIM(x,y)=\frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)(\sigma_{xy}+c_3)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)(\sigma_x\sigma_y+c_3)} \tag{5}$$

其中 μ_x 和 μ_y 分别是 x 和 y 的均值,σ_x^2 和 σ_y^2 分别是 x 和 y 的方差,σ_{xy} 是 x 和 y 的协方差. $c_1=(k_1L)^2$,$c_2=(k_2L)^2$,$c_3=(k_3L)^2$ 是用来维持稳定的常数. L 是像素值的动态范围. 一般地,$k_1=0.01$,$k_2=0.03$,$L=255$. $SSIM$ 的取值范围为[0,1],值越大,表示相似度越高.

接下来对过分割得到的块进行组合. 由于当块的数量较大时,随机组合的数量将会成指数级上涨,因此本文对过分割得到的块按照一定的空间关系进行编号,从上到下,从左到右依次编号. 这里本文假设部件是由连续的块组成,因此本文对编号连续的块进行组合并生成一系列候选的结果. 分别对这些候选的结果计算与检索到的图片相对应的分割真实结果计算相似度得分,我们认为得分最高的组合方式就是最终的分割结果. 依次对检测到的每一类部件进行分割,最后将这些结果还原到原图坐标上去得到最终的分割结果. 图 4 是本文分割每一类部件的过程. 特别地,对于皮肤这一类,首先使用相同的方法获得整个人的分割结果,然后再用人的分割结果减去各个部件的分割结果就得到了皮肤区域.

在合并块的过程中,由于在裁剪部件图像的时候把检测框扩大了 1.2 倍,包含了更多的上下文信息,所以事先认为与图像边缘相邻的块是背景. 因为部件一定在检测框的中间部分,当扩大检测框之后,边缘块是

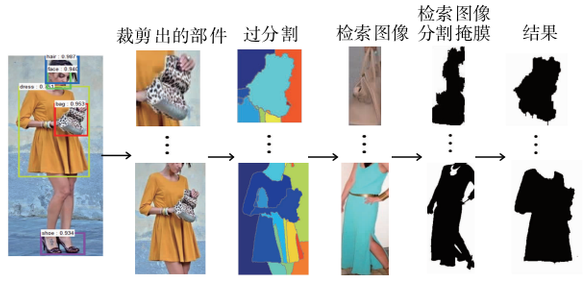


图4　部件分割流程

背景的可能性就更大. 这么做的另一个目的是块的个数减少了,组合产生的结果就会更少,降低计算复杂度的同时也降低了错误率.

2.3　分割结果优化

由于过分割的局限性,与图像边缘相邻的块有可能包含了部件所在的块,而在上一节中提到,我们事先认定与图像边缘相邻的块是背景,因此真正的部件所在的块有可能被丢失. 因此,在一幅图像中如果背景所占比例过大,就认为部件块丢失,此时把最靠近图像中心点的块作为部件区域.

3　实验

3.1　数据库

目前为止,现有的公开数据库相对较小,因此本文整合了三个公开的数据库的作为本文的数据库. 第一个数据库是 CFPD(Colorful Fashion Parsing Data)数据库[18],包含了 2682 张标好的像素级的图像. 第二个数据库是 CCP(Clothing Co-Parsing)数据库[13],包含 1004 张图像. 第三个数据库是 Fashionista[9],包含 685 张图像,最终合并的数据库共有 4371 张图像. 本文中,定义 14 类标签,分别是:背景、包、腰带、连衣裙、脸、头发、帽子、裤子、围巾、鞋、短裙、太阳镜、上衣和皮肤. 为了把这三个数据集运用到本文的分割任务中来,我们把三个数据集的不同的标签整合成本文事先定义好的 14 类. 其中,3671 张图像用来训练,700 张图像用来测试. 具体每种类别的样本个数如表 1 所示. 图 5 是数据库中的部分图片.

表 1　数据库训练和测试分布

	包	腰带	连衣裙	脸	头发	帽子	裤子	围巾	鞋	短裙	太阳镜	上衣	合计
训练	2060	984	967	2663	3616	464	1230	293	3552	1009	717	3094	3641
测试	418	215	201	699	687	93	242	57	684	166	140	604	700
合计	2478	1199	1168	3362	4303	557	1472	350	4236	1175	857	3694	4371

图5　数据样本

每个任务都采用相同的数据集,不同点在于对数据给定的标签. 对于检测任务,根据数据库中图片每个像素的真实值,确定人体以及每一类部件的坐标(x,y,w,h),分别作为人体检测和部件检测的真实标签,构建人体检测与部件检测的训练集及测试集. (x,y)表示左上角的坐标,(w,h)表示宽和高. 对于分割任务,每一张图片给出像素级类别的标签,作为分割的真实值.

3.2　实验结果

本文训练模型采用 GTX Titan X GPU,12GB 显存,训练时间 10 小时. 本文在 700 张测试图像上评测分割结果,使用与 PaperDoll[15] 相同的评测方法,包括像素级的整体正确率(Accuracy),前景正确率(F. g. accuracy),平均准确率(Avg. precision),平均召回率(Avg. recall)和平均 F1 值(Avg. F-1). 具体计算方法如式(6)~(10)所示:

$$A_{all} = \frac{N'}{N_{all}} \times 100\% \tag{6}$$

$$A_{fg} = \frac{N'_{fg}}{N_{fg}} \times 100\% \tag{7}$$

$$P_{avg} = \frac{1}{C}\sum_{i=1}^{c}\frac{N'_i}{N_i^*} \times 100\% \tag{8}$$

$$R_{avg} = \frac{1}{C}\sum_{i=1}^{c}\frac{N'_i}{N_i} \times 100\% \tag{9}$$

$$F1_{avg} = \frac{1}{C}\sum_{i=1}^{c}\frac{2 \times N'_i}{P_i + R_i} \times 100\% \tag{10}$$

其中 N'表示预测正确像素总数,N_{all}表示像素总数,N'_{fg}表示前景像素预测正确总数,N_{fg}表示前景像素总数,C 表示类别数,N'_i 表示第 i 类像素预测正确数,N_i^* 表示

预测为第 i 类的像素数，N_i 表示第 i 类像素数，式(10)中，$P_i=\frac{N_i^{'}}{N_i^{*}}$，$R_i=\frac{N_i^{'}}{N_i}$.

表 2 展示了本文的分割方法的性能. 并且与前期的两种分割方法 Yamaguchi et al[9]. 和 PaperDoll[15] 进行了对比. 通过实验结果可以看出，本文方法的准确率为 90.39%，要胜于 Yamaguchi et al. 83.15% 和 PaperDoll 87.07% 的正确率，本文的方法在前景正确率上达到 68.32%，比 Yamaguchi et al. 56.59% 和 PaperDoll 58.66% 的前景正确率有了很大的提升. 同时，本文的方法能够获得更高的准确率和召回率. 另外，实验还把检测框的大小扩大到 1.1 倍和 1.5 倍，做了实验对比. “Full”是加上优化的结果，而其他实验没有进行优化. “w/o hd”是没有加人体检测的结果，可以看出，通过先检测人体再检测部件的这种由粗到细的方法可以很大程度地提高分割的性能.

同时，实验还给出了每一个前景类的 F1 值如表 3. 对比 PaperDoll 和本文的方法，本文的方法在大部分类别上 F1 值都有了很大的提高. 并且在大部分类别上的准确率都有了一定的提升，特别是在包，裤子，连衣裙，围巾和短裙这几类上，因此前景准确率有了很大的提升.

表 2 实验性能对比

	Accuracy	F. g. accuracy	Avg. precision	Avg. recall	Avg. F-1
Yamaguchi et al.[9]	83.15	56.59	51.34	50.03	45.80
PaperDoll[15]	87.07	58.66	57.63	48.86	48.39
w/ohd	83.67	65.54	65.77	63.45	61.77
Scale-1.1	89.45	67.1	70.12	64.36	65.47
Scale-1.2	89.21	69.52	68.39	66.44	65.6
Scale-1.5	87.44	70.04	60.54	66.67	61.33
Full	90.39	68.32	71.81	64.17	66.01

表 3 前景语义标签的 F1 值在文中的引用

	包	腰带	连衣裙	脸	头发	帽子	裤子	围巾	鞋	短裙	太阳镜	上衣	皮肤
Yamaguchi et al.	28.64	29.89	51.07	74.28	63.77	10.46	45.32	11.42	42.39	45.59	12.27	56.97	55.92
PaperDoll	32.88	31.76	57.21	73.14	66.65	9.56	48.52	25.35	57.69	51.76	1.59	66.85	59.68
w/ohd	70.42	57.31	77.38	61.77	51.24	62.4	77.41	55.16	50.6	82.13	18.27	71.97	40.94
Scale-1.1	75.09	58.36	82.32	67.92	50.59	62.21	72.36	66.95	50.72	88.53	22.34	76	46.74
Scale-1.2	74.55	60.62	81.32	66.07	54.43	66.18	72.11	57.02	55.69	88.72	21.95	76.5	47.04
Scale-1.5	71.3	59.71	74.63	57.84	52.04	56.4	69.52	56.21	49.21	84.58	13.09	72.09	46.78
Full	75.11	59.12	80.8	67.78	50.71	63.85	72.09	67.33	49.69	87.13	26.14	74.6	52.07

但是本文的方法也存在一些缺点. 与一些大部件如短裙和上衣相比较，一些细小的部件准确率相对较低，如腰带，太阳镜. 由于部件较小，能提供的外观信息较少，因此分割小部件就相对比较困难. 另外，当部件颜色与背景比较相似的时候，分割也比较困难.

图 6 展示了本文的实验结果，还包括真实的标签和 PaperDoll 的实验结果. 从效果图中可以看出本文的方法能够基本能够展现出各个部件的轮廓. 尽管有一些部件分割的轮廓不是很精准，但是根据类别跟位置可以看出检测结果还是比较准确的，这样能够更好的区分各个类别. 然而 PaperDoll 方法容易在各个类别之间混淆并且会遗漏一些部件，没有分割出来.

4 结束语

本文提出了一种有效的人体分割的方法，把人体图像分割成不同的语义区域. 本文的方法包括：人体检测，部件检测和部件分割. 使用基于 Fast R-CNN 的方法检测目标并且基于过分割检索分割各个部件. 实验结果表明本文的方法能够取得很好的效果，能够准确的检测各个衣服部件并且获得较高的像素级分割准确率. 在未来的工作中将会致力于提高像腰带、太阳镜这些小部件的准确率. 另外，后续将增加数据量以支持本文的方法.

图6　分割结果

参考文献

[1] Bourdev L, Maji S, Malik J. Describing people: A poselet-based approach to attribute classification[A]. IEEE International Conference on Computer Vision[C]. Barcelona: IEEE Press, 2011. 1543 – 1550.

[2] Liu S, Feng J, Song Z, et al. Hi, magic closet, tell me what to wear! [A]. Proceedings of the 20th ACM international conference on Multimedia [C]. Nara: ACM, 2012. 619 – 628.

[3] Song Z, Wang M, Hua X, et al. Predicting occupation via human clothing and contexts[A]. IEEE International Conference on Computer Vision[C]. Barcelona: IEEE Press, 2011. 1084 – 1091.

[4] Chen H, Xu Z J, Liu Z Q, et al. Composite templates for cloth modeling and sketching[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. New York: IEEE Press, 2006. 943 – 950.

[5] Lin L, Wang X, Yang W, et al. Discriminatively trained and-or graph models for object shape detection[J]. IEEE Transactions on pattern analysis and machine intelligence, 2015, 37(5): 959 – 972.

[6] Wang N, Ai H. Who blocks who: Simultaneous clothing segmentation for grouping images[A]. IEEE International Conference on Computer Vision [C]. Barcelona: IEEE Press, 2011. 1535 – 1542.

[7] Hasan B, Hogg D C. Segmentation using Deformable Spatial Priors with Application to Clothing[A]. British Mahine Vision Conference[C]. Aberystwyth: British Machine Vision Association, 2010. 1 – 11.

[8] Bo Y, Fowlkes C C. Shape-based pedestrian parsing[A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Colorado Springs: IEEE Press, 2011. 2265 – 2272.

[9] Yamaguchi K, Kiapour M H, Ortiz L E, et al. Parsing clothing in fashion photographs[A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Providence: IEEE Press, 2012. 3570 – 3577.

[10] Ladicky L, Torr P H S, Zisserman A. Human pose estimation using a joint pixel-wise and part-wise formulation [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Portland: IEEE Press, 2013. 3578 – 3585.

[11] Kohli P, Rihan J, Bray M, et al. Simultaneous segmentation and pose estimation of humans using dynamic graph cuts[J]. International Journal of Computer Vision, 2008, 79(3): 285 – 298.

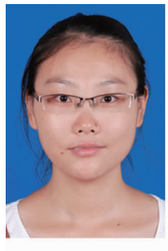
[12] Dong J, Chen Q, Xia W, et al. A deformable mixture parsing model with parselets[A]. IEEE International Conference on Computer Vision[C]. Sydney: IEEE Press, 2013. 3408 – 3415.

[13] Yang W, Luo P, Lin L. Clothing co-parsing by joint image segmentation and labeling[A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Columbus: IEEE Press, 2014. 3182 – 3189.

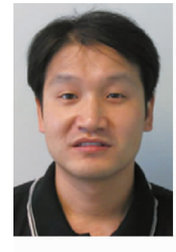
[14] Carreira J, Sminchisescu C. Cpmc: Automatic object segmentation using constrained parametric min-cuts [J].

IEEE Transactions on Pattern Analysis and Machine Intelligence,2012,34(7):1312 – 1328.
[15] Yamaguchi K,Hadi Kiapour M,Berg T L. Paper doll parsing:Retrieving similar styles to parse clothing items[A]. IEEE International Conference on Computer Vision[C]. Sydney:IEEE Press,2013. 3519 – 3526.
[16] Yamaguchi K,Kiapour M H,Ortiz L E,et al. Retrieving similar styles to parse clothing[J]. IEEE transactions on pattern analysis and machine intelligence,2015,37(5):1028 – 1040.
[17] LiuS,Liang X,Liu L,et al. Matching-cnn meets knn:Quasi-parametric human parsing[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Boston:IEEE Press,2015. 1419 – 1427.
[18] LiuS,Feng J,Domokos C,et al. Fashion parsing with weak color-category labels[J]. IEEE Transactions on Multimedia,2014,16(1):253 – 265.
[19] Dantone M,Gall J,Leistner C,et al. Human pose estimation using body parts dependent joint regressors[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Portland:IEEE Press,2013. 3041 – 3048.
[20] Liang X,Liu S,Shen X,et al. Deep human parsing with active template regression[J]. IEEE transactions on pattern analysis and machine intelligence,2015,37(12):2402 – 2414.
[21] Liang X,Xu C,Shen X,et al. Human parsing with contextualized convolutional neural network[A]. IEEE International Conference on Computer Vision[C]. Santiago:IEEE Press,2015. 1386 – 1394.
[22] Girshick R. Fast r-cnn[A]. IEEE International Conference on Computer Vision[C]. Santiago:IEEE Press,2015. 1440 – 1448.
[23] Zitnick C L,Dollár P. Edge boxes:Locating object proposals from edges[A]. European Conference on Computer Vision[C]. Zurich:Springer International Publishing,2014. 391 – 405.
[24] Dollár P,Zitnick C L. Structured forests for fast edge detection[A]. IEEE International Conference on Computer Vision[C]. Sydney:IEEE Press,2013. 1841 – 1848.

作者简介



王 枫 女,1993 年生于江苏扬州. 南京信息工程大学硕士研究生,研究方向为深度学习、计算机视觉、行人属性分析.



刘青山 男,1975 年生于安徽庐江. 南京信息工程大学教授,博士生导师,主要研究方向为人脸图像分析、图像理解、视频分析、模式识别、机器学习等.
E-mail:qsliu@ nuist. edu. cn