

引入语义部位约束的行人再识别

陈莹* 陈巧媛

(江南大学轻工过程先进控制教育部重点实验室 无锡 214122)

摘要: 为减轻行人图片中的背景干扰,使网络着重于行人前景并且提高前景中人体部位的利用率,该文提出引入语义部位约束(SPC)的行人再识别网络。在训练阶段,首先将行人图片同时输入主干网络和语义部位分割网络,分别得到行人特征图和部位分割图;然后,将部位分割图与行人特征图融合,得到语义部位特征;接着,对行人特征图进行池化得到全局特征;最后,同时使用身份约束和语义部位约束训练网络。在测试阶段,由于语义部位约束使得全局特征拥有部位信息,因此测试时仅使用主干网络提取行人的全局信息即可。在大规模公开数据集上的实验结果表明,语义部位约束能有效使得网络提高辨别行人身份的能力并且缩减推断网络的计算花费。与现有方法比较,该文网络能更好地抵抗背景干扰,提高行人再识别性能。

关键词: 行人再识别; 人体语义分割; 语义部位约束

中图分类号: TN911.73; TP391

文献标识码: A

文章编号: 1009-5896(2020)12-3037-08

DOI: [10.11999/JEIT190954](https://doi.org/10.11999/JEIT190954)

Semantic Part Constraint for Person Re-identification

CHEN Ying CHEN Qiaoyuan

(Key Laboratory of Advanced Control Education in Light Industry Process,
Jiangnan University, Wuxi 214122, China)

Abstract: In order to alleviate the background clutter in pedestrian images, and make the network focus on pedestrian foreground to improve the utilization of human body parts in the foreground. In this paper, a person re-identification network is proposed that introduces Semantic Part Constraint(SPC). Firstly, the pedestrian image is input into the backbone network and the semantic part segmentation network at the same time, and the pedestrian feature map and the part segmentation label are obtained respectively. Secondly, the part segmentation label and the pedestrian feature maps are merged to obtain the semantic part feature. Thirdly, the pedestrian feature map is obtained and the global average pooling is used to gain global features. Finally, the network is trained using both identity constraint and semantic part constraint. Since the semantic part constraint makes the global features obtain the part information, only the backbone network can be used to extract the features of the pedestrian during the test. Experiments on large-scale datasets show that semantic part constraints can effectively make the network improve the ability to identify pedestrians and reduce the computational cost of inferring networks. Compared with the state of art, the proposed network can better resist background clutter and improve person re-identification performance.

Key words: Person re-identification; Human semantic segmentation; Semantic Part Constraint (SPC)

1 引言

行人再识别研究的是行人在监控系统内不同摄像头下的检索问题,如何使同一行人在差异条件下

的不同图像的特征一致是研究的重点。客观的差异条件有背景干扰、场景光线不同、相机视角不同以及遮挡物等,主观的差异条件有行人姿态变化、增减携带物以及行人间遮挡等。任意一个差异条件将直接导致行人的外观产生变化,而行人外观仍然是行人再识别的主要辨别手段。

目前行人再识别方法分为传统非深度方法和基于深度学习的方法,侧重于特征提取及度量学习两方面的研究。在特征提取方面,二者皆设法使得同一行人的不同图像获得相近特征,不同的是传统方

收稿日期: 2019-11-27; 改回日期: 2020-06-04; 网络出版: 2020-07-28

*通信作者: 陈莹 chenying@jiangnan.edu.cn

基金项目: 国家自然科学基金(61573168), 江苏省六大人才高峰资助项目(2015-WLW-004)

Foundation Items: The National Natural Science Foundation of China (61573168), The Six Talent Summit Project Talents of Jiangsu Province (2015-WLW-004)

法通过经验设计手工特征作为行人特征,而深度方法旨在通过卷积神经网络自动地学习鲁棒性强的行人特征。从匹配时的距离度量入手,研究人员还提出不同的度量方式,如XQDA^[1]和BRM²L^[2]。相比于传统方法,深度学习的方法能够根据行人再识别中的难点,有针对性地设计网络结构或网络损失,使得卷积神经网络提取的行人特征具有强判别力。

近年来,深度行人再识别考虑将行人的多模态特征进行融合,即除能够直接从行人图片提取的外观特征外,将其他有助于增强特征判别力的模态信息引入网络。Lin等人^[3]以多任务的方式整合了行人再识别和属性识别问题,提出行人-属性网络同时学习行人特征和属性特征。Suh等人^[4]则提出一个双流网络,通过双线性池化的方式将行人的姿态信息与行人的外观信息相融合,使得网络提取姿态对齐的行人特征。Kalayeh等人^[5]第1次将人体语义解析和行人再识别结合起来,提出基于人体语义解析的行人再识别(human Semantic parsing for Person Re-Identification, SPReID)网络,SPReID网络将语义部位图和行人外观特征图进行加权融合,最后串联全局特征、前景特征和语义部位融合特征作为行人特征进行再识别。

SPReID网络融合人体语义解析结果,即语义部位分割标签,从而使最终的行人特征包含了语义部位信息,其与仅使用全局信息的行人再识别网络相比效果更佳。但是,SPReID网络存在两个缺点,一是其将全局特征和语义部位特征直接串联起来,这样得到的行人特征中全局特征和语义部位特征独立性较强,不能兼顾特征多样性和准确率;二是其在测试阶段仍需要对测试图片进行语义分割,大大拖慢了网络的运行速度,增加了网络的运算成本。因此,本文提出引入语义部位约束的行人再识

别(Semantic Part Constraint for person re-identification, SPC)网络,旨在通过人体语义部位约束使得行人特征中的全局特征和语义部位特征互补,在兼顾特征多样性的同时提高准确率,并且在测试阶段不再需要对输入图片进行语义分割,仅需通过主干网络部分,有效加快了网络的预测速度。

2 引入语义部位约束的行人再识别网络

本节将分为4个部分介绍引入语义部位约束的行人再识别网络。首先介绍整体网络结构,然后介绍语义部位标签,接着介绍网络的2个特征提取分支,最后介绍网络的约束。

2.1 网络结构

本文网络结构如图1所示,主要由全局特征提取分支、语义特征提取分支和网络约束组成。全局特征 f_g 由主干网络Resnet50^[6]直接提取输入图片的特征图 A 再池化得到;语义特征提取分支包含采样模块和语义部位分割网络,上采样模块的作用是将特征图 A 的空间尺度上采样至与特征图 U 一致,语义部位分割网络推断输入图片的部位分割图,并将其作为语义部位标签 S ,通道部位图 T 由语义部位标签 S 并变换得到,语义部位特征 F 由上采样后的特征图 U 和语义部位通道部位图 T 按照通道相乘得到,最终的语义特征 f_p 为语义部位特征 F 的最大值融合而得。

本文网络约束由身份约束和语义部位约束组成,身份约束的作用是使全局特征具有判别身份的能力,语义部位约束通过拉近全局特征和语义部位特征的距离,使得全局特征包含语义部位信息,从而兼顾全局和局部信息使二者优势互补。

在训练阶段,训练样本需要分别提取全局特征和语义部位特征,再计算网络的身份约束和语义部

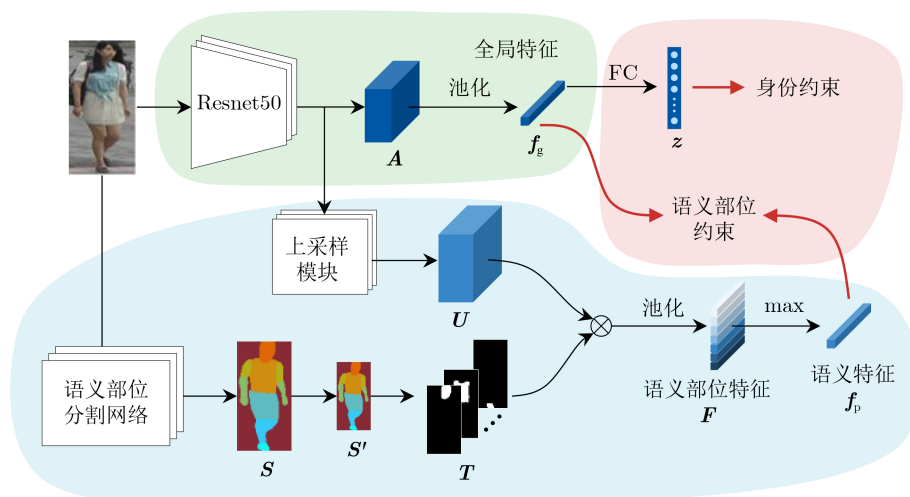


图1 本文网络结构图

位约束的损失值, 进而优化网络权重; 在测试阶段, 因为全局特征经过语义部位约束已包含语义部位信息, 所以仅使用全局特征作为行人特征, 即测试样本仅需要通过绿色虚线框中的主干网络, 无需进行语义部位分割和融合操作, 从而避免由于分割网络在进行推断时的计算成本。提取行人特征后, 再计算检索目标与候选人之间的欧式距离, 距离越近相似度越高。

2.2 语义部位标签

人体语义分割是一个细粒度的语义分割任务, 也称为人体解析, 其指对行人图像上的像素点按照人体的语义部位进行分类。行人外观信息集中于行人图像的前景部分, 即行人主体部分, 而背景和遮挡物属于无用干扰信息。实际上行人图像是一个整体, 通过人眼观察时大脑能够分辨出前景和背景、人体和遮挡, 而对于网络来说它们是无差别的像素点。人体语义分割拥有像素级的精度和轮廓自由的特性, 因此语义部位信息非常适合用于辅助行人再识别问题。

由于常用的行人再识别数据集不含行人语义部位标签, 本文使用Fu等人^[7]提出的场景分割网络(Dual Attention Network, DANet)作为语义部位分割网络并使用COCO Densepose数据集^[8]训练。由于行人再识别不需要例如上臂和下臂这样细致的部位分割, 因此将数据集原本划分较细的24个部位合并为8个。由DANet得到的分割结果实际上属于一种伪标签, 并非数据集的真值, 为理解方便后文仍称其为标签。行人再识别数据集中所有图片对应的语义部位标签, 均在行人再识别网络训练前离线求得。

图2展示了部分行人图片及其通过DANet^[7]提取到的语义标签, 奇数列是行人图片, 偶数列是行



图2 语义部位标签示例

人的语义分割结果, 即本文网络中的语义部位标签为2维图像 $\mathbf{S}(x, y) = 0, 1, \dots, 7$, $\mathbf{S} \in \mathbb{R}^{w_S \times h_S}$, 其宽和长分别为 w_S 和 h_S , 用从0到7共8个值分别代表包括背景、头、胸腹等不同的8个部位, 图2中用不同的颜色表示。

2.3 特征提取

2.3.1 全局特征

行人图像以批次的形式输入网络, 为了说明方便, 在特征提取和网络约束部分都假设批次大小为1。如图1所示, 首先将行人图像 \mathbf{I} 输入Resnet50^[6]网络中, 得到相应的特征图 $\mathbf{A} = G(\mathbf{I})$, $\mathbf{A} \in \mathbb{R}^{w_A \times h_A \times c}$, 其中 G 表示Resnet50网络, w_A , h_A 和 c 分别表示特征图 \mathbf{A} 的宽、高和通道数。然后, 对特征图 \mathbf{A} 进行平均池化, 求得行人的全局特征 $\mathbf{f}_g \in \mathbb{R}^{1 \times 1 \times c}$

$$\mathbf{f}_g = \text{pooling}(\mathbf{A}) = \frac{1}{w \times h} \sum_{x=1}^w \sum_{y=1}^h \mathbf{A}(x, y) \quad (1)$$

2.3.2 语义部位特征

语义部位特征提取主要由上采样模块和语义部位分割网络组成。上采样模块由反卷积层、批归一化层和激活层组成, 其中反卷积层的卷积核为 3×3 , 步长为2。上采样模块的作用是使上采样后的特征图的空间尺度与语义部位标签的空间尺度一致, 通道数不变。对特征图 \mathbf{A} 进行上采样, 得到特征图 $\mathbf{U} \in \mathbb{R}^{w_U \times h_U \times c}$

$$\mathbf{U} = \text{ReLU}(\text{BN}(\text{deconv}(\mathbf{A}))) \quad (2)$$

其中deconv, BN和ReLU分别表示反卷积、批归一化和ReLU激活函数, w_U 和 h_U 分别表示特征图 \mathbf{U} 的宽和高, 且 $w_U = 2w_A$, $h_U = 2h_A$ 。

为了将语义部位标签与特征图 \mathbf{U} 融合以得到语义部位特征, 先将语义部位标签 \mathbf{S} 缩放至与特征图 \mathbf{U} 相同的空间尺度得到 \mathbf{S}' , 再将语义部位标签分类值变为通道值, 得到通道部位图 $\{\mathbf{T}_i \in \mathbb{R}^{w_U \times h_U}\}_{i=1,2,\dots,7}$

$$\mathbf{T}_i(x, y) = \begin{cases} 1, & i = \mathbf{S}'(x, y) \\ 0, & i \neq \mathbf{S}'(x, y) \end{cases} \quad (3)$$

接着, 将除背景($i = 0$)外的通道部位图 \mathbf{T} 与特征图 \mathbf{U} 分别相乘并进行平均池化, 得到语义部位特征 $\mathbf{F}_i = \text{pooling}(\mathbf{U} \times \mathbf{T}_i)$, $i = 1, 2, \dots, 7$, $\mathbf{F} \in \mathbb{R}^{1 \times 1 \times c \times 7}$ 。最后, 对语义部位特征 \mathbf{F} 进行最大值融合, 得到最终的行人语义特征 $\mathbf{f}_p = \max_{i=1}^7 (\mathbf{F}_i)$ 。

2.4 网络约束

本文网络约束由身份约束和语义部位约束组成。身份约束为Softmax交叉熵损失, 设FC层的输出有 K 类, 表示为 $\mathbf{z} = [z_1 z_2 \dots z_K]$, 则身份约束为

$$L_{id}(z, l_{id}) = - \sum_{k=1}^K \lg \left(\frac{e^{z_k}}{\sum_{i=1}^K e^{z_i}} \right) \cdot v_{id} \quad (4)$$

其中 $v_{id} = \begin{cases} 1, l_{id} = k \\ 0, l_{id} \neq k \end{cases}$ 表示当前预测类别 k 与标签 l_{id} 是否一致。

为使训练后的全局特征包含语义信息, 论文网络中采用语义特征来监督全局特征, 即将全局特征和语义部位特征的最小二乘误差作为语义部位约束

$$L_{sp}(f_g, f_p) = \|f_g - f_p\|_2^2 = \frac{1}{c} \sum_{i=1}^c (f_g^{(i)} - f_p^{(i)})^2 \quad (5)$$

最终的网路约束为 $L = L_{id} + \lambda L_{sp}$ 。

通过最小化网络损失, 在全局特征具有身份判别能力的同时, 全局特征和语义特征的距离被拉近, 语义部位约束在网络训练过程中反向作用于主干网络, 从而使得全局特征具有一定的语义信息。语义部位特征是通过语义部位标签与特征图相乘得到的, 相当于是以某个部位作为蒙版取得其特征值, 再将多个语义部位特征通过最大值融合成一个完整的目标特征描述。语义特征不带有背景信息且拥有语义解析信息, 通过语义部件约束, 拉近其与全局特征的距离, 并使得训练后的全局特征包含了语义信息, 进而更具有鉴别力。与串联全局特征和语义特征相比, 引入语义部位约束能够有效提高行人再识别精度, 将在下一节通过实验论证。

3 实验

3.1 数据集与评估标准

本文共在两个大规模数据集上进行实验。分别是Market-1501^[9]和DukeMTMC-reID^[10]。其中Market-1501包含1501个行人共32668张行人图片, 平均每个行人有22张图片。图片由5个高分辨率摄像头和1个低分辨率摄像头采集, 给出的行人保证至少在2个摄像头中出现过, 行人框是由计算机通过检测算法自动生成的, 因此存在误检和误裁。根据官方的评估标准, 训练集包含751个行人共12936张图片, 测试集包含750个不出现在训练集中的行人并划分为查询对象(query)和查询域(gallery), query共有3368张, gallery共有19734张。

DukeMTMC-reID是跟踪数据集DukeMTMC的一个子集, DukeMTMC数据集有1812个行人, 但是有408个行人只曾出现在1个摄像头中, 不符合跨镜头识别的评判标准而被舍弃。因此DukeMTMC-reID包含1404个行人共36441张行人图片, 图片由8个高分辨率摄像头采集, 行人框由人工手工标注。根据官方标准, 训练集包含702个行人共16522张图片, 测试集包含剩余702个人, query和gallery分别有2228和16522张。

行人再识别中常用的评价指标是累积匹配特征曲线(Cumulative Matching Characteristic curve, CMC)和平均准确率的均值(mean Average Precision, mAP)。每个query的平均准确率是从其准确率-召回曲线计算的; 而mAP是所有query的平均准确率的均值, 即CMC反映的是检索精度, 而mAP反映了召回率, CMC中的Rank- n 指的是在匹配结果的前 n 张图片的正确率。本文给出的实验结果和对比均使用单query检索配置。

3.2 实验配置及参数

实验平台的操作系统为Ubuntu16.04, CPU为Intel i7-6850K, 带有2张NVIDIA 1080TI GPU, 显存为12G。实验语言为Python, 使用深度学习框架Pytorch。使用在ImageNet数据集上预训练的Resnet50参数初始化主干网络。训练时, 语义部位标签随着训练集的图片一同被输入到网络中, 输入图像的大小为 256×128 , 批次大小为64, 取值为1, 总共训练180个epoch; 测试时, 批次大小为256, 仅需要输入测试图片即可。

3.3 实验分析

为了验证本文网络及语义约束的有效性, 本节在Market-1501数据集进行实验, 表1展示了所进行的3个对比实验。

3.3.1 语义特征的有效性

实验1是常见的行人再识别基准(Baseline), 使用主干网络提取行人图像的全局特征, 并用身份约束训练网络, 最后用全局特征作为行人特征进行测试。实验2中的 $C_f = \text{cat}(f_g, f_d, f_p)$ 表示按照SPReID^[5]的思想, 串联全局特征、前景特征和语义特征作为行人特征, $\text{cat}(\cdot, \cdot)$ 表示串联操作, 实验2也使用身份约束训练网络, 最后用串联特征作为行人特征进

表 1 在Market-1501数据集上的对比实验(%)

实验编号	行人特征	网络约束	Rank-1	Rank-5	Rank-10	mAP
1	f_g	L_{id}	92.0	96.9	98.2	80.4
2	C_f	L_{id}	92.7	97.5	98.6	80.6
3	f_g	$L_{id} + L_{sp}$	93.6	97.6	98.7	83.6

行测试。前景特征为 $f_a = \text{pooling}(U \times (E - T_0))$, 其中 $E \in \mathbb{R}^{w_U \times h_U}$ 且元素全为1, 作用是使背景通道取反从而得到行人图像的前景蒙版。

实验1和实验2的区别仅在于使用的行人特征不同, 从表1中可以看出, 串联特征的表现要优于全局特征, 这说明语义部位信息从一定程度上补充了全局特征的不足, 原因是前景特征能够获得行人的主体部分, 而语义特征是由语义部位特征得来的, 相当于局部特征的融合, 能够与全局特征互补, 使得最终的行人特征具有不同尺度的信息。

3.3.2 语义部位约束的有效性

表1中的实验3使用全局特征作为行人特征, 与实验1不同之处为实验3使用身份约束和语义部位约束训练网络。对比实验1和实验3可得, 本文提出的语义约束是有效的。与实验1相比, 实验3在Rank-1和mAP分别增长了1.4%和2.6%。

对比实验2和实验3可得, 使用语义约束来优化网络比简单串联行人的全局特征和语义特征更有优势, 因为串联行人全局特征和语义特征无法使二者融合, 实际上等价于分别用全局特征和语义特征进行度量。本文使用欧式距离的平方作为度量函数, 设全局特征的维度为 c_g , 则2个全局特征 $f_{g,a}$ 和 $f_{g,b}$ 之间的距离为

$$d(f_{g,a}, f_{g,b}) = \sum_{i=1}^{c_g} (f_{g,a}^{(i)} - f_{g,b}^{(i)})^2 \quad (6)$$

串联特征为 C_f , 前景特征和语义特征的维度分别为 c_d 和 c_p , 则2个串联特征 $C_{f,a}$ 和 $C_{f,b}$ 之间的距离为

$$\begin{aligned} d(C_{f,a}, C_{f,b}) &= \sum_{i=1}^{c_g+c_d+c_p} (C_{f,a}^{(i)} - C_{f,b}^{(i)})^2 \\ &= \sum_{i=1}^{c_{gl}} (C_{f,a}^{(i)} - C_{f,b}^{(i)})^2 \\ &\quad + \sum_{i=c_{gl}}^{c_{gl}+c_{fg}} (C_{f,a}^{(i)} - C_{f,b}^{(i)})^2 \\ &\quad + \sum_{i=c_g+c_d+c_p} (C_{f,a}^{(i)} - C_{f,b}^{(i)})^2 \\ &= d(f_{g,a}, f_{g,b}) + d(f_{d,a}, f_{d,b}) \\ &\quad + d(f_{p,a}, f_{p,b}) \end{aligned} \quad (7)$$

即串联特征的距离度量实为单个特征距离的和。而在身份约束的基础上增加语义约束, 在网络收敛的同时, 全局特征和语义特征更加靠近, 相当于全局特征拥有了一定的语义含义, 并且行人特征维度较低, 是一种更为有效科学的软约束。

3.3.3 测试时间损耗对比

本文网络以引入语义部位约束的方式, 避免了进行行人再识别预测时, 运算花费较大的部位分割网络预测。表2展示了不同网络测试时长对比, 可以看出本文网络提取一个批次图像时长约为复现SPReID网络时长的1/9。不仅如此, 由于串联特征的维度是全局特征的3倍, 在进行距离度量时, SPReID耗时也显然大于本文网络。引入语义部位约束不仅能提高行人再识别精度, 还能大大缩减计算花费和时间成本。

3.3.4 λ 的取值

图3展示了不同 λ 取值下本文网络的实验结果, 其中横轴为训练的epoch数量, 纵轴为相应模型的Rank-1精度。当 $\lambda=0$ 时, 网络不引入语义部位约束, 其结构相当于表1中的实验1。从图3中可以看出, 在 λ 分别为1.0和1.5并训练至180个epoch时, 本文网络表现最佳, 但 $\lambda=1.0$ 的曲线收敛更为平缓, 故最终 λ 取值为1.0; 随着 λ 的增加, 曲线更早向饱和处移动, 即网络收敛地越快, 例如当 $\lambda=2.5$ 时, 网络在训练到80个epoch时Rank-1就达到了93%以上的精度, 但继续训练就出现了过拟合的现象, 原因是引入语义部位约束后, 网络约束由1个变为2个, 因此更快收敛, 但语义部位约束是一个软约束, 过分加大其权重将适得其反, 故通过实验选取合适值, 在本文实验中, $\lambda=1.0$ 。

3.3.5 检索排序图

图4展示了本文网络SPC和其他网络在测试集上的部分检索结果排序图, 给出前10个检索结果。检索结果(a)和(b)为本文网络和SPReID^[5]的对比,

表2 不同网络测试时长对比(ms)

方法	批次特征提取耗时
复现SPReID	82.87
本文网络	9.45

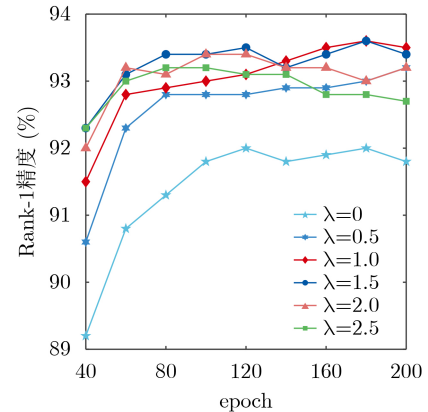


图3 λ 的取值对应Rank-1精度

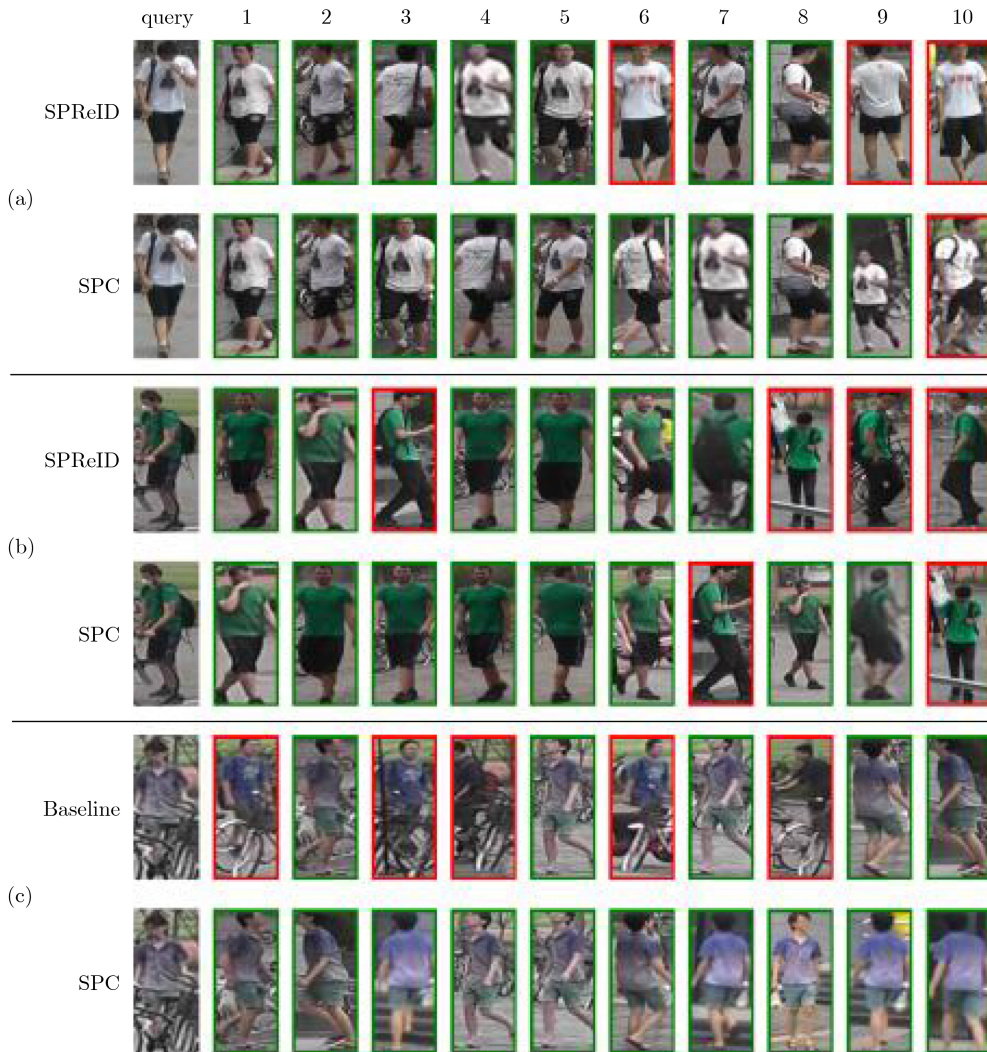


图4 行人检索结果排序图

从中可以看出，本文网络在行人图像完整时优于SPReID。从误检结果来看，两个网络的错误结果与检索对象都十分相似，但本文网络提取到更精细的局部特征。例如检索结果(a)中，错误结果与检索对象都拥有白色短袖上衣和黑色短裤，SPReID匹配到的错误结果胸前的图案是红色的，本文网络则关注到了这个局部特征，并未将其纳入排序，且错误结果的行人图像胸前也拥有深色图案。

检索结果(c)为本文网络和Baseline的对比，Baseline的结构相当于实验1。可以看出，本文网络通过语义部位约束，获得了对局部特征的判别力。Baseline中的错误检索结果都带有自行车的遮挡，说明网络无法正确判别图像中的行人前景及行人部位，而本文网络对带遮挡的行人图像具有较好鲁棒性，虽然在推断时网络结构与Baseline完全一致，但是通过语义部位约束使得主干网络获得部分解析局部特征的能力，此举的优势之处在于不牺牲精度的情况下能减少计算成本，大大加快特征提取的速度。

3.4 方法对比

表3为本文方法在Market-1501和DukeMTMC-reID两个数据集上与不同方法的性能比较。可以看出，本文通过在行人再识别网络中引入语义约束，有效提高了网络的精度，且明显优于其他现有方法。值得一提的是，表3中SPReID方法是多行人再识别数据集预训练后，再在单个数据集上微调的结果，本文网络则是直接在单数据集上训练并测试。根据实验，在Market-1501数据集上，本文方法的Rank-1/mAP比SPReID高1.1%/2.3%；在DukeMTMC-reID上，本文方法的Rank-1/mAP比SPReID高1.0%/0.3%。

4 结束语

为加入轮廓级语义部位信息以提高行人再识别精度，本文提出引入语义部位约束的行人再识别网络。与其他在行人再识别中加入语义部位信息的方法相比，本文在常用行人再识别基准的基础上引入

表3 不同方法在两个数据集上的性能比较(%)

方法	Market-1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
VIM ^[11]	79.5	59.9	68.9	49.3
SVDNet ^[12]	82.3	62.1	76.7	56.8
APR ^[9]	84.3	64.7	70.7	51.2
FMN ^[13]	86.0	67.1	74.5	56.9
PSE ^[14]	87.7	69.0	79.8	62.0
PN-GAN ^[15]	89.4	72.6	73.6	53.2
CamStyle ^[16]	89.5	71.6	78.3	57.6
HA-CNN ^[17]	91.2	75.7	80.5	63.8
Part-Aligned ^[4]	91.7	79.6	84.4	69.3
SPReID ^[5]	92.5	81.3	84.4	71.0
AHR ^[18]	93.1	76.2	81.7	65.9
本文方法	93.6	83.6	85.4	71.3

语义部位约束, 其在缩减推断网络的计算花销的同时, 进一步提高了再识别的精度, 在实际网络预测时仅需要使用基准部分的网络便于部署。随着人体语义分割的发展, 今后提取行人再识别图片的部位标签将会更加精确, 后续工作可以朝着如何充分利用语义部位标签进行。行人再识别在智能监控领域有着极高的应用价值, 现今的精度虽然突破人眼, 但仍没有达到商用级别, 因此今后的研究趋势将会兼顾速度和精度。

参考文献

- [1] LIAO Shengcai, HU Yang, ZHU Xiangyu, *et al.* Person re-identification by local maximal occurrence representation and metric learning[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, America, 2015: 2197–2206. doi: [10.1109/CVPR.2015.7298832](https://doi.org/10.1109/CVPR.2015.7298832).
- [2] 陈莹, 许满月. 基于双向参考集矩阵度量学习的行人再识别[J]. 电子与信息学报, 2020, 42(2): 394–402. doi: [10.11999/JEIT190159](https://doi.org/10.11999/JEIT190159).
CHEN Ying and XU Xiaoyue. Matrix metric learning for person re-identification based on bidirectional reference set[J]. *Journal of Electronics & Information Technology*, 2020, 42(2): 394–402. doi: [10.11999/JEIT190159](https://doi.org/10.11999/JEIT190159).
- [3] LIN Yutian, ZHENG Liang, ZHENG Zhedong, *et al.* Improving person re-identification by attribute and identity Learning[J]. *Pattern Recognition*, 2019, 95: 151–161. doi: [10.1016/j.patcog.2019.06.006](https://doi.org/10.1016/j.patcog.2019.06.006).
- [4] SUH Y, WANG Jingdong, TANG Siyu, *et al.* Part-aligned bilinear representations for person re-identification[C]. The 15th European Conference on Computer Vision, Munich, Germany, 2018: 402–419. doi: [10.1007/978-3-030-01264-9_25](https://doi.org/10.1007/978-3-030-01264-9_25).
- [5] KALAYEH M M, BASARAN E, GOKMEN M, *et al.* Human semantic parsing for person re-identification[C]. The IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, America, 2018: 1062–1071. doi: [10.1109/CVPR.2018.00117](https://doi.org/10.1109/CVPR.2018.00117).
- [6] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, America, 2016: 770–778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [7] FU Jun, LIU Jing, TIAN Haijie, *et al.* Dual attention network for scene segmentation[C]. The 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, America, 2019: 3146–3154.
- [8] GÜLER R A, NEVEROVA N, and KOKKINOS I. DensePose: Dense human pose estimation in the wild[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, America, 2018: 7297–7306. doi: [10.1109/CVPR.2018.00762](https://doi.org/10.1109/CVPR.2018.00762).
- [9] ZHENG Liang, SHEN Liyue, TIAN Lu, *et al.* Scalable person re-identification: A benchmark[C]. 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 1116–1124. doi: [10.1109/ICCV.2015.133](https://doi.org/10.1109/ICCV.2015.133).
- [10] ZHENG Zhedong, ZHENG Liang, and YANG Yi. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro[C]. 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 3754–3762. doi: [10.1109/ICCV.2017.405](https://doi.org/10.1109/ICCV.2017.405).
- [11] ZHENG Zhedong, ZHENG Liang, and YANG Yi. A discriminatively learned CNN embedding for person re-identification[J]. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2018, 14(1): 13. doi: [10.1145/3159171](https://doi.org/10.1145/3159171).
- [12] SUN Yifan, ZHENG Liang, DENG Weijian, *et al.* SVDNet for pedestrian retrieval[C]. 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 3800–3808. doi: [10.1109/ICCV.2017.410](https://doi.org/10.1109/ICCV.2017.410).
- [13] DING Guodong, KHAN S, TANG Zhenmin, *et al.* Feature mask network for person re-identification[J]. *Pattern Recognition Letters*, 2020, 137: 91–98. doi: [10.1016/j.patrec.2019.02.015](https://doi.org/10.1016/j.patrec.2019.02.015).
- [14] SARFRAZ M S, SCHUMANN A, EBERLE A, *et al.* A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking[C]. The IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, America, 2018: 420–429. doi: [10.1109/CVPR.2018.00051](https://doi.org/10.1109/CVPR.2018.00051).
- [15] QIAN Xuelin, FU Yanwei, XIANG Tao, *et al.* Pose-

- normalized image generation for person re-identification[C]. The 15th European Conference on Computer Vision, Munich, Germany, 2018: 650–667. doi: [10.1007/978-3-030-01240-3_40](https://doi.org/10.1007/978-3-030-01240-3_40).
- [16] ZHONG Zhun, ZHENG Liang, ZHENG Zhedong, *et al.* CamStyle: A novel data augmentation method for person re-identification[J]. *IEEE Transactions on Image Processing*, 2019, 28(3): 1176–1190. doi: [10.1109/TIP.2018.2874313](https://doi.org/10.1109/TIP.2018.2874313).
- [17] LI Wei, ZHU Xi Tian, and GONG Shaogang. Harmonious attention network for person re-identification[C]. The 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, America, 2018: 2285–2294. doi: [10.1109/CVPR.2018.00243](https://doi.org/10.1109/CVPR.2018.00243).
- [18] 陈鸿昶, 吴彦丞, 李邵梅, 等. 基于行人属性分级识别的行人再识别[J]. 电子与信息学报, 2019, 41(9): 2239–2246. doi: [10.11999/JEIT180740](https://doi.org/10.11999/JEIT180740).
- CHEN Hongchang, WU Yancheng, LI Shaomei, *et al.* Person re-identification based on attribute hierarchy recognition[J]. *Journal of Electronics & Information Technology*, 2019, 41(9): 2239–2246. doi: [10.11999/JEIT180740](https://doi.org/10.11999/JEIT180740).
- 陈莹: 女, 1976年生, 教授, 博士生导师, 主要研究方向为信息融合、模式识别等.
- 陈巧媛: 女, 1995年生, 硕士生, 研究方向为行人再识别.

责任编辑: 余蓉