精细化局部语义与属性学习的行人重识别

肖进胜[1)] 吴婧逸[1)] 郭浩文[1)] 郭 圆[2)] 赵持恒[1)] 王 银[1)]

[1)](武汉大学电子信息学院 武汉 430072)

[2)](武汉理工大学资源与环境工程学院 武汉 430070)

摘 要 行人的随身物品信息与属性描述是提高行人重识别任务性能的有效途径.本文提出了一种精细化局部语义与属性学习的行人重识别网络,来提取行人的随身物品信息,同时从语义区域中获得行人的属性描述.首先,将特征聚类方法生成的随身物品区域融合到额外语义模型生成的语义解析结果中,解决目前较多额外语义解析模型遗漏行人随身物品信息的问题.其次,利用生成的语义区域作为身体标签,网络由全局特征构建这些区域的语义特征映射,然后从语义特征中预测与之相关的属性信息,增强行人的描述.最后,考虑到行人某些属性之间包含强相关性,重新构建加权模型来提高某些属性的置信分数,优化属性的预测准确率.将属性预测结果和行人的全局特征连接在一起,形成行人的鲁棒特征表示.在 Market-1501 和 DukeMTMC-reID 属性数据集上的实验表明,所提算法较基线网络分别得到了 3.6%和 6.4%的 *mAP* 指标增益,可以提高行人重识别任务的性能.

关键词 语义分析;属性预测;相关性;行人重识别

中图法分类号 TP18 **DOI 号** 10.11897/SP.J.1016.2024.02387

Person Re-Identification Network with Fine-Grained Local Semantics and Attribute Learning

XIAO Jin-Sheng[1)] WU Jing-Yi[1)] GUO Hao-Wen[1)] GUO Yuan[2)] ZHAO Chi-Heng[1)] WANG Yin[1)]

[1)](*School of Electronic Information*, *Wuhan University*, *Wuhan* 430072)

[2)](*School of Resources and Environmental Engineering*, *Wuhan University of Technology*, *Wuhan* 430070)

Abstract Pedestrian Re-IDentification (Person Re-ID) aims to search for the same pedestrian across multiple different camera views. Due to its importance in various practical applications such as video surveillance and content-based image retrieval, it has garnered extensive attention in recent years. However, the task still faces numerous challenges, including significant variations in pedestrian poses, lighting, and backgrounds in different camera views. Additionally, the similar appearance of clothing among different pedestrians and inaccurate pedestrian detection bounding boxes further complicate its practical application. Personal belongings information (e.g., backpacks, handbags) is often overlooked by semantic models because these items are not person body parts, but personal belongings information provide crucial contextual information for re-identification. On the other hand, attribute descriptions, such as gender, type of upper body clothes, color of upper body clothes, are also discriminative information in person re-identification and can effectively enhance Person Re-ID task performance. Addressing the issues that current semantic methods cannot effectively extract potential personal belongings information and clustering methods are too coarse, failing to fully utilize the attribute information of local semantic features,

收稿日期:2023-11-13;在线发布日期:2024-06-25.本课题得到国家自然科学基金(42201480)、国家重点研发计划(2021YFB2501104)资助.**肖进胜**,博士,副教授,中国计算机学会(CCF)高级会员,主要研究方向为计算机视觉、图像处理与分析.E-mail:xiaojs@whu.edu.cn.**吴婧逸**,硕士研究生,主要研究方向为图像处理与计算机视觉.**郭浩文**,硕士,主要研究方向为图像处理、视频分析与计算机视觉.**郭 圆**(通信作者),博士,讲师,主要研究方向为高清地图、众包数据和计算机视觉.E-mail:yuang@whut.edu.cn.**赵持恒**,硕士研究生,主要研究方向为图像处理与计算机视觉.**王 银**,硕士研究生,主要研究方向为计算机视觉、图像分割.

this paper proposed a pedestrian re-identification algorithm based on fine-grained local semantics and attribute learning. This algorithm extracts information about personal belongings and obtains attribute descriptions from semantic regions. The proposed method involves several key modules. Firstly, the Fine-grained Local Semantics (FLS) Module integrates the personal belonging regions generated by the feature clustering method into the semantic parsing results generated by an additional semantic model, addressing the problem of many additional semantic parsing models missing personal belongings information and resulting in smooth and more comprehensive semantic regions. Secondly, Attribute Learning Module (ALM) uses the generated semantic regions as body labels, allowing the network to construct semantic feature mappings of these regions from global features, and then predicts the associated attribute information from the semantic features to capture detailed and contextually relevant information about the pedestrian. Lastly, considering the strong correlations between certain pedestrian attributes, such as female and long hair, Attribute Weighted Module (AWM) is constructed to improve the confidence scores of certain attributes and optimize the prediction accuracy of attributes. Then the model combines the attribute prediction results with the global features of the pedestrian to form robust feature representations. In addition, the high confidence attribute information is used to filter out the irrelevant pedestrian images from the image gallery to be retrieved to improve the speed of similarity sorting. To evaluate the performance of the proposed model, experiments were conducted on two public datasets widely used for pedestrian re-identification tasks (Market-1501 and DukeMTMC-reID) and their attribute datasets Experiments on the Market-1501 and DukeMTMC-reID attribute datasets show that the proposed algorithm achieves 3.6% and 6.4% *mAP* index gains, and 1.1% and 5.3% *mAP* index gains compared with the baseline network respectively, indicating that the proposed model can improve the performance of person re-identification task. Visual analyses of person attribute prediction results and person similarity ranking were also performed to verify the effectiveness of the proposed model in accurately predicting pedestrian attributes and utilizing them to improve matching accuracy and efficiency.

Keywords semantic analysis; attribute prediction; correlation; person re-identification

1 引 言

行人重识别(Person Re-IDentification,Re-ID)是计算机视觉领域一项重要的任务,旨在匹配从不重叠摄像机中获得的行人[1],具有广泛的应用场景,如智能视频监控[2]、智能零售和人员关联等,近年来引起了学术界和工业界越来越多的关注.该任务的一个主要问题在于,受到背景、遮挡、低分辨率、光照和摄像机视角等因素的影响[3],行人在不同画面中的外观会发生巨大的变化,给模型匹配相同的行人增加了困难.

现有的研究方法主要致力于设计一个鲁棒性的行人特征,来提高模型的识别性能.通常的做法是从行人的部位中提取局部特征,与全局特征进行互补,利用身份标签进行监督,来丰富行人的表示能力.这些局部特征可以分为局部区域的特征[4-6]、注意力机制引导的局部特征[7-8]和语义模型获得的局部特征[9-10].此外,跨模态行人重识别任务也应用语义级特征进行对齐融合[11-12].前两种方法直观简便,但相对粗糙,获得的特征常常包含有背景干扰和局部区域之间的特征重叠,并且获得的局部特征存在特征对齐和特征一致性的问题.

通过语义模型,尤其是语义解析模型获得人体部位,然后提取局部特征,通过其像素级的精度和任意轮廓建模的能力,具有可观的性能增益.然而,语义解析模型通常针对行人的身体部位进行解析,行人的随身物品(如背包、手提包和雨伞等)作为重要的语义判别信息而发生漏检[13-14].

行人的随身物品信息可以提高行人表示能力,也属于行人的一个属性.行人的重识别任务和属性识别目标相同,只是粒度有区别.已有的一些研究通

过增加局部描述的属性信息提高了模型的区分能力[15-16]，证明了行人属性信息的价值. 然而，现有的方法未能充分利用精确的语义特征上的属性信息，或者使用信息量大的全局特征和带有背景杂波的局部特征来预测属性，降低了属性预测的置信度. 另外，行人不同属性之间存在一定的相关性[15]，如可以利用头发长度、衣服与性别的强关联性提高其判别置信度.

为了解决上述问题，本文将语义特征和属性特征集成到一个统一的分类框架中，实现了对行人更强、鲁棒性的特征提取. 本文通过特征聚类获取行人语义解析伪标签的方法得到行人随身物品区域，并利用精确的语义特征来预测相应的多个属性. 此外，本文充分利用属性之间的相关性增强预测的可靠性，提高行人的特征表示能力. 所提算法的主要贡献如下：

(1) 本文提出了精细化局部语义模块. 该模块将特征聚类产生的随身物品语义区域融入精准的额外语义模型产生的语义区域，从而获得了包含随身物品信息的语义解析结果.

(2) 本文提出了属性学习模块. 在精细的语义特征上，基于人类先验知识，从各个语义特征获得行人属性信息，提高了属性预测的准确率和行人的表示能力.

(3) 本文提出了一个属性加权模块，根据属性之间的相关性对预测属性的置信分数进行了重新赋权. 然后，将最终的属性预测结果联合行人的全局特征，构造了一个辨别性的行人特征表示.

2　相关工作

现阶段，许多的研究工作集中于开发一个鲁棒性的行人特征来应对各种场景因素[17]. 这些方法主要以卷积神经网络[18]作为特征提取器，通过构造分类损失来辨别不同的行人. 根据特征的不同，这些方法可以分为：局部特征方法、语义特征方法和属性特征方法. 本节回顾了与所提算法相关的一些研究工作.

2.1　局部特征方法

早期的一些方法将行人整幅图像和行人的 ID 身份信息作为输入，存在背景干扰等外界因素的影响，不足以学习出一个泛化性足够强的网络. 局部特征方法是对全局特征方法之后的进一步改进，部分方法以水平分割来生成这些局部区域[4-6]，另外一些方法引入注意力机制增强感兴趣区域的特征表示[7-8].

Sun 等人[4]将图像水平均匀分割为 6 个部分，提出基于部分的卷积基线网络. 与均匀划分方法不同，Zheng 等人[5]提出从粗到细的金字塔模型，局部区域的大小遵循金字塔模型给出的比例. 在红外-可见光行人重识别任务中，Liu 等人[6]提出对齐灰度模态解决交叉模态匹配问题，利用头肩特征来辅助全局特征，形成更强的特征描述符.

基于注意力机制的方法旨在增强行人图像的重要特征以及抑制干扰信息. Li 等人[7]提出了一种协调注意力模型，联合了粗糙的空间注意力和通道注意力，最大限度地利用了不同层次视觉注意的互补信息，优化了特征表示. 近年来，Transformer 架构也被应用于重识别领域. He 等人[8]提出了一个行人重识别框架 TransReID，解决卷积神经网络无法处理多个局部区域已经存在细节信息损失的问题.

综上所述，局部区域特征方法可以进一步增强行人的特征表示，关注局部的细微差异，注意力机制可以融入局部特征方法，实现性能的提高. 但局部区域方法存在特征对齐问题，需要有效的对齐策略提高匹配精度；基于注意力机制的方法，对于数据量的要求较高，在信息量较大的场景才能充分发挥获取重要信息的能力.

2.2　语义特征方法

基于语义特征的方法通过预先训练的语义模型来获得人体区域，去除背景等因素的干扰. 在语义区域获取方式上，分为姿态关键点模型和语义解析模型两种方式，或通过无监督的聚类方法获得行人的语义区域.

在 2017 年，Zhao 等人[9]提出了以人体区域为导向的特征分解和融合的行人重识模型 Spindle Net，通过卷积姿态信息检测人体 14 个身体关节和 7 个身体部分区域. 一些方法还将语义特征与属性特征进行结合，Nguyen 等人[10]提出了一种基于图的行人表示(Graph-based Person Signature，GPS)，将详细的属性标签和语义特征聚合成图网络，学习行人属性和肢体语义特征的关系，进行对行人更强有力的表示.

在可见光-红外行人重识别(VI-ReID)任务中，Li 等人[11]提出了一种以形状为中心的表征学习框架，旨在学习形状特征和与形状相关的外观特征. Chen 等人[12]从语义结构信息建模和位置部分相互作用挖掘出发，提出了一种基于结构感知的位置变压器网络，用于提取语义级可共享的跨模态表示.

对于使用无监督方法定位人体语义区域，聚类

是一种经典的无监督学习方法，它对相似的特征进行分组. Zhu 等人[13]在行人重识别中提出了一个基于特征映射的级联聚类算法，生成人体语义区域的伪标签，并利用可见的语义特征进行相似度计算，提高了对遮挡场景的鲁棒性.

上述的语义特征方法可以解决背景干扰和遮挡的问题，能实现精准的语义特征提取，在性能上优于基于局部特征的方法. 然而缺点在于，该方法依赖额外语义模型的精度，同时不能够识别与行人相关的随身物品的上下文信息. 与上述工作相比，本文方法学习包含行人随身物品在内的语义区域特征，以及精确的语义区域中包含的细粒度行人属性信息.

2.3 属性特征方法

属性特征方法是利用额外的属性信息来增强行人表示的一种新方法. 常见的全局特征方法在学习过程中会遗漏对小尺度的上下文线索，通过数据集事先标注的属性，网络可以增加多个属性预测分支，通过这些属性标签计算属性损失，发现上下文线索.

为了解决公开重识别数据集的属性标注问题，Lin 等人[15]为两个大规模的重识别数据集手工标注属性标签，然后提出了一个属性识别（Attribute-Person Recognition，APR）网络，同时学习重识别的嵌入特征和预测行人属性，并对预测的属性进行重新赋权. 基于上述数据集，Chen 等人[16]提出了一种属性引导的度量蒸馏方法，从语义角度上解释行人相似的原因.

综上，属性特征的方法可以通过增加属性预测分支来预测行人图像包含的属性，从而为行人提供更具辨别力的特征描述. 另外，属性信息不受环境变化影响，可以提高行人表示的鲁棒性. 在检索行人图像时，属性会加快过滤不正确的行人结果，提高识别速度.

然而，相较于上述方法，现有的一些方法还未考虑到利用精确的语义特征来针对性地预测属性信息. 常规利用包含背景干扰的全局特征和局部特征来获得属性预测的方法，在性能上未达到最优.

3 本文方法

本文提出的精细化局部语义与属性信息的行人重识别算法的网络结构如图 1 所示.

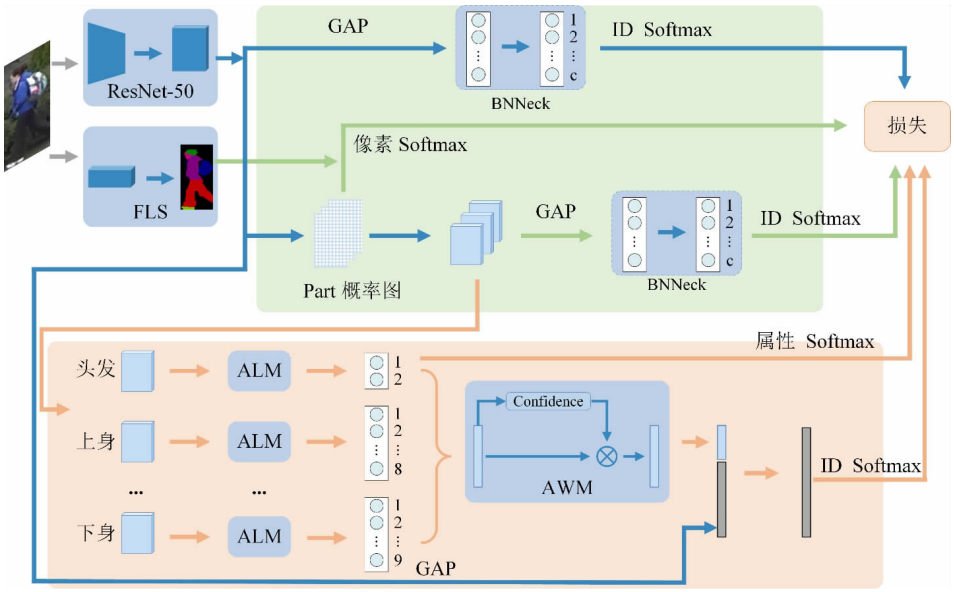

图 1 精细化局部语义与属性信息的行人重识别网络结构

网络包括三个部分，分别是包含随身物品信息的精细化局部语义模块（Fine-grained Local Semantics Module，FLS）、基于语义特征的属性学习模块（Attribute Learning Module，ALM）和属性置信分数的加权修正模块（Attribute Weighted Module，AWM）. 首先，主干网络对输入的图片提取特征，作为行人的全局表示. 然后通过卷积生成多个语义特征的概率图，与原特征相乘得到多个语义特征. 基于这些特征和组合成的前景特征，来预测各自的属性信息. 此外，进行属性预测分数的加权修正，以充分利用属性之间的相关性信息，提高属性的识别率. 最后，联合生成的精细的语义特征以及预测的多个属性，增强了整体行人的表示能力.

3.1 精细化局部语义模块

额外语义模型的解析结果中通常不包含行人的随身物品信息，而聚类方法生成的语义区域较为粗

糙却包含随身物品信息. 本节提出了精细化局部语义模块，融合额外语义模型获得的语义区域和使用聚类产生的语义区域，该模块示意图如图 2.

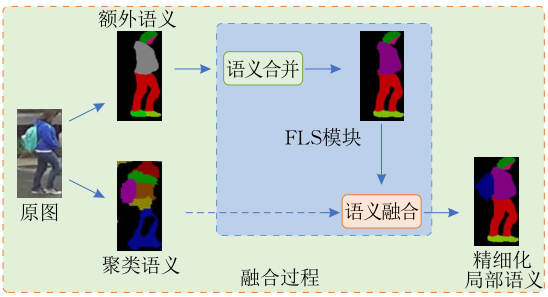

图 2　精细化局部语义模块的网络结构

为了获得行人正确的语义标签，本节基于 SCHP 语义解析模型[14]来收集行人的 20 个语义区域标签. 语义合并阶段将这些局部语义区域合并为 9 个集合：背景、头发、帽子、头部、上身、手部、下身、腿部、鞋子. 将算法生成的属于同一集合的区域分配同一标签，并从上到下依次赋值为{0,1,…,8}.

为了获取行人的随身物品区域，本文算法借鉴了 ISP 特征聚类算法[13]聚类产生的伪标签作为最终的聚类语义结果. 一共包含了 7 个区域，即背景、头部、肩部、随身物品、下身、腿部和鞋子. 根据伪标签的分布情况，随身物品的标签值为 3 或 4. 语义融合阶段将他们融合到上述的 SCHP 语义解析结果中，并保留较大的聚类区域(筛选像素点数高于 100 的同一聚类区域). 从而增加为 10 个语义区域，如表 1 所示.

表 1　额外模型与聚类算法的语义融合结果

新的语义区域	组成
背景	背景
帽子	帽子
头发	头发
头部	面部、太阳镜
上身	上衣、外套、围巾
手部	左手、右手、手套
下身	围裙、裙子、裤子、连衣裤
腿部	左腿、右腿、袜子
鞋子	左鞋、右鞋
随身物品	聚类产生的第 3、4 部分

值得注意的是，聚类产生的语义解析结果中不一定包含随身物品等属性. 但因为聚类产生的语义结果具有一致性，可以使用上述方法来增加一个新的语义区域.

得到所有的 10 个局部语义区域的像素标签后，假设 $\boldsymbol{F}_{global}$ 代表主干网络生成的全局特征，网络自身需要进一步映射出各个局部语义区域的特征，并使用语义标签作为约束. 以 f_{conv} 来表示卷积过程，全局特征 $\boldsymbol{F}_{global}=\boldsymbol{F}_{\theta}^{c\times h\times w}$，生成 $K=10$ 维通道的中间特征为

$$\boldsymbol{F}_1^{k\times h\times w}=f_{conv}(\boldsymbol{F}_{\theta}^{c\times h\times w}) \tag{1}$$

其中，θ 是主干网络学习的参数. 然后，将该特征进行 Softmax 运算处理，分为 K 类，得到了各个局部语义局域的概率图：

$$P_{part\text{-}k}^{h\times w}(x,y)=\mathrm{Softmax}(\boldsymbol{F}_1^{k\times h\times w}(x,y))=\frac{\exp(\boldsymbol{W}_k^{\mathrm{T}}\boldsymbol{F}_1^{k\times h\times w}(x,y))}{\sum_{i=0}^{K-1}\exp(\boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{F}_1^{i\times h\times w}(x,y))},$$
$$k=0,1,\cdots,K-1,\ (x,y)\in(w,h) \tag{2}$$

这 K 个概率图分别代表了背景和其它局部语义区域. $\boldsymbol{W}$ 为线性层的参数. 值得注意的是，个人的随行物品对应最后一个语义区域. 理想情况下，若局部语义区域 k 在图像中的真实语义标签中不存在，网络对不可见区域不产生表示，有 $\boldsymbol{P}_{part\text{-}k}^{h\times w}(x,y)=0$. 最终每个局部语义区域都是一组概率加权的像素级表示的聚合. 各个局部语义区域的特征如下：

$$\boldsymbol{F}_{part\text{-}k}=\boldsymbol{F}_{global}\cdot\boldsymbol{P}_{part\text{-}k}^{h\times w},\ k=1,2,\cdots,K-1 \tag{3}$$

然后，将除去背景后的 $K-1$ 个概率图按照像素对齐进行相加，得到了前景概率图：

$$\boldsymbol{P}_{foreground}^{h\times w}=\sum_{k=1}^{K-1}\boldsymbol{P}_{part\text{-}k}^{h\times w},\ k=1,2,\cdots,K-1 \tag{4}$$

该概率图代表了除背景 $k=0$ 以外所有局部语义区域的组合，相比于全局特征只包含了前景区域. 与上述类似，得到前景特征为

$$\boldsymbol{F}_{foreground}=\boldsymbol{F}_{global}\cdot\boldsymbol{P}_{foreground}^{h\times w} \tag{5}$$

至此，便得到了行人的全局特征 $\boldsymbol{F}_{global}$、前景特征 $\boldsymbol{F}_{foreground}$ 和局部语义特征 $\boldsymbol{F}_{part\text{-}k}$.

3.2　属性学习模块

行人的属性信息是除 ID 标签信息以外能对行人进一步描述的特征. 增加更多的属性，就会更加有利于网络区分相似的行人. 同时可以从待查询的图像库中过滤掉不相关的图像，提高相似度排序的速度. 本节提出了属性学习模块来从行人的各个语义特征中预测与之相关的属性. 从精确的语义区域提取语义特征并预测与之相关的属性信息，相比于全局图像预测不易出现背景杂波和不相关区域的干扰，且计算量小，网络容易收敛.

目前，基于 Market-1501 数据集与 DukeMTMC 数据集的属性数据集[15]标签分别有 27 和 23 个行人属性. 基于这些属性标签作为监督信息，对于 Market-1501 数据集的 17 种不同服装颜色属性

合并为 2 种(上身服装颜色和下身服装颜色),得到 12 个可用的属性信息. 同样对 DukeMTMC-reID 数据集合并颜色属性,得到了 10 个可用的属性信息. 根据之前的精细化局部语义模块获得的 10 个语义区域和前景区域,选取了帽子、头发、上身、下身、鞋子、随身物品和前景区域. 这些区域对应的属性如表 2 所示.

表 2 7 个语义区域及相应预测属性的对应关系

语义区域	相关的属性
前景	年龄、性别
帽子	戴帽子
头发	头发长度
上身	袖子长度、上身服装颜色、上身服装长度
下身	下身服装颜色、下身服装长度
鞋子	鞋型、鞋子颜色
随身物品	背包、手提包、其它类型的包

具体实现上,假设一系列输入图像是 $x_1, x_2, \cdots, x_i$,其中 x_i 包含的属性集合为 $a_i=(a_i^1, a_i^2, \cdots, a_i^m)$. 那么对于其中的一个属性 a_i^j,代表了行人图像 x_i 的第 j 个属性. 根据表 2,图像 x_i 包含了 6 个局部语义特征以及一个前景特征用来预测相应的属性,分别表示为 $\boldsymbol{F}_{foreground}$ 和 $\boldsymbol{F}_{part}=(\boldsymbol{F}_{part\text{-}1}, \boldsymbol{F}_{part\text{-}2}, \cdots, \boldsymbol{F}_{part\text{-}k})$. 将所有的属性 a_i 根据上表,相应的分配给每个特征进行预测. 其中第 k 个语义特征需预测的 n 个属性集合表示为 $a_{i,k}=(a_i^1, a_i^2, \cdots, a_i^n)$. 本节的 ALM 模块对于每个属性的预测过程如图 3 所示.

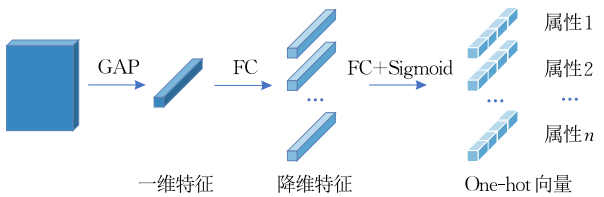

图 3 ALM 模块的属性预测过程

首先,对于第 k 个局部语义特征 $\boldsymbol{F}_{part\text{-}k}^{c\times h\times w}$,通过全局平均池化(Global Average Pooling,GAP)将其压缩为大小为 $1\times c$ 的一维特征. 然后,通过一个大小为 $n\times c_1$ 的全连接 FC 层,将上述一维特征变成 n 个大小为 $1\times c_1$ 的一维特征,如下:

$$\boldsymbol{F}_{k,2}^{n\times c_1}=\mathrm{FC}(\mathrm{GAP}(\boldsymbol{F}_{part\text{-}k}^{c\times h\times w})) \tag{6}$$

上述的过程主要有两个目的,一方面,将大小为 $1\times c$ 的一维特征减小为 $1\times c_1$,是为了防止过拟合. 另一方面,由于包含 n 个属性的预测,所以相应的由一维特征经过不同的全连接 FC 层扩展为 n 个一维特征. 全连接层的作用实际上就是对一维向量进行不同方式的加权组合,然后映射到每个属性标签上,这些参数通过训练可以获得.

最后,对于获得的 n 个大小为 $1\times c_1$ 的向量,通过后续的全连接 FC 层和 Sigmoid 函数进行激活分类. 由于每个属性包含的类别数不一致,因此需要对于每个向量设计相应大小的全连接层. 如年龄为 4 个类别,那么全连接层的输出维度是 4. 对于其中的一个向量 $\boldsymbol{F}_{k,2}^{1\times c_1}$,经过全连接层和 Sigmoid 函数激活后,就得到了与类别相关的 One-hot 向量,过程如下:

$$\boldsymbol{F}_{k,\text{one-hot}}^{1\times c_2}=\mathrm{Sigmoid}(\mathrm{FC}(\boldsymbol{F}_{k,2}^{1\times c_1})) \tag{7}$$

其中,c_2 代表该属性的类别数量,根据实际情况进行改变. 将输出的 One-hot 向量通过属性交叉熵分类损失进行约束,使属性预测的结果接近真实值.

由于每个行人的属性特征都不同,那么这些属性的组合可以对行人的全局特征进行补充,形成对行人更详细的描述. 为了提高分类的准确性,将所有的属性向量 $\boldsymbol{F}_{k,\text{one-hot}}^{1\times c_2}$ 组合在一起,与全局特征进行拼接,作为行人全局特征的额外线索. 这样就得到了一个新的联合属性的全局特征,如下:

$$\boldsymbol{F}_{global\text{-}attr}=\boldsymbol{F}_{global}+\sum_{j=1}^{m}\boldsymbol{F}_{j,\text{one-hot}}^{1\times c_2} \tag{8}$$

上式中,m 代表属性的总个数,$\boldsymbol{F}_{global}$ 是主干网络后经过 GAP 处理生成的全局特征,$\boldsymbol{F}_{j,\text{one-hot}}^{1\times c_2}$ 是降维后的第 j 个属性分类后的置信度向量. 将联合属性的全局特征也经过 ID 分类损失进行约束,这样即就增强了原始全局特征的表示.

3.3 属性加权模块

尽管可以通过特征映射来预测行人的各个属性,然而属性信息之间的相关性是另一个可以利用的信息. 对于同一个行人,很多属性是同时出现的,属性之间的相关性可以提高某些属性的预测置信分数. 因此,为了提高属性模型的泛化性,本节提出了属性加权模块 AWM. 每个属性的加权值来自于对前面生成的属性的预测得分. 这个属性的预测得分被连接并输入至 AWM 模块中. 然后将 AWM 模块的输出与全局特征连接起来,使用 ID 分类损失进行约束,从而更新上述的式(8),该过程如图 4 所示.

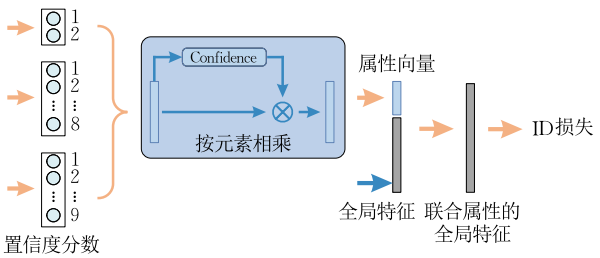

图 4 AWM 模块对属性预测分数的加权修正过程

假设图像 x_i 经过属性分类器得到的各个属性预测得分集合为 $\tilde{a}_i=(\tilde{a}_i^1,\tilde{a}_i^2,\cdots,\tilde{a}_i^m)$，其中，$\tilde{a}_i^j$ 是第 j 个属性的预测分数，$\tilde{a}_i^j\in[0,1]$. 每个属性的预测分数可以经过下式来生成：

$$\tilde{a}_i^j=\max(\mathrm{Softmax}(\boldsymbol{F}_{j,\text{one-hot}}^{1\times c_2})) \tag{9}$$

然后，将 $\tilde{a}_i$ 连接为一个向量 $\tilde{\boldsymbol{a}}_i\in R^{1\times m}$，输入至 AWM 模块. 对于 AWM 模块预测的所有属性的置信度得分向量 $\boldsymbol{c}_i$ 为

$$\boldsymbol{c}_i=\mathrm{Sigmoid}(\boldsymbol{w}\cdot\tilde{\boldsymbol{a}}_i^{\mathrm{T}}+\boldsymbol{b}) \tag{10}$$

其中，$\boldsymbol{w}\in R^{m\times m}$ 且 $\boldsymbol{b}\in R^{m\times 1}$，是可训练的参数. 置信度得分向量 $\boldsymbol{c}_i\in R^{m\times 1}$，是学习到的每个属性的权重. 于是，AWM 模块将原来的预测得分向量 $\tilde{\boldsymbol{a}}_i$ 变换为

$$a_i'=\boldsymbol{c}_i\circ\tilde{\boldsymbol{a}}_i^{\mathrm{T}} \tag{11}$$

上述的“∘”代表向量按照对应元素进行相乘. 因此联合新的加权分数，式(8)的联合属性的全局特征被更新为

$$\boldsymbol{F}_{global\text{-}attr}=\boldsymbol{F}_{global}+\sum_{j=1}^{m}a_i'\cdot\boldsymbol{F}_{j,\text{one-hot}}^{1\times c_2} \tag{12}$$

当间隔较长时间，两次捕捉到的同一行人随身物品发生变化时，经过属性加权模块，针对关联性高的属性提高行人身份判断的置信度，随身物品变化带来的影响也会降低.

总的来说，通过利用可训练的参数进行组合，重新校准每个属性不同的激活程度，这样就考虑到了各个属性之间的相关性. 例如，当“红色上衣”和“长发”的预测得分非常高时，AWM 模块就会倾向于对性别属性“女性”的预测得分的加权，提高该属性的响应程度.

3.4　相似性度量

在测试阶段，对于每个待测试的图像，一共可以获得 4 类特征：全局特征 $\boldsymbol{F}_{global}$、前景特征 $\boldsymbol{F}_{foreground}$、局部语义特征 $\boldsymbol{F}_{part\text{-}k}=\{\boldsymbol{F}_{part\text{-}1},\boldsymbol{F}_{part\text{-}2},\cdots,\boldsymbol{F}_{part\text{-}K-1}\}$ 和联合属性的全局特征 $\boldsymbol{F}_{global\text{-}attr}$. 依据 BoT[19] 基线网络的测试方法，本文算法使用特征之间的余弦距离来衡量查询图像和待测试图像的相似性. 由于局部语义特征的数量是自适应变化的，在遮挡时某些局部语义区域不可见，所以只选取可见的局部语义特征进行距离计算.

总的特征距离分为两大部分：固定不变的特征距离（全局特征、前景特征、联合属性的全局特征），和数量可变的局部语义特征距离.

以 $D(\cdot,\cdot)$ 来表示特征之间的余弦距离，全局特征、前景特征和联合属性的全局特征的距离分别为

$$\begin{aligned}d_{global}&=D(\boldsymbol{F}_{global}^q,\boldsymbol{F}_{global}^g),\\ d_{foreground}&=D(\boldsymbol{F}_{foreground}^q,\boldsymbol{F}_{foreground}^g),\\ d_{global\text{-}attr}&=D(\boldsymbol{F}_{global\text{-}attr}^q,\boldsymbol{F}_{global\text{-}attr}^g)\end{aligned} \tag{13}$$

上标 q 和 g 分别代表待查询图像和查询库中图像.

然后，为了获得可见的局部语义特征，可以根据精细化局部语义模块中的概率图 $P^{k\times h\times w}(x,y)$ 进行判断. $\mathrm{argmax}\,P^{k\times h\times w}(x,y)$ 代表该点所对应的语义标签，假设所有点的值都不等于 k，表明第 k 个局部语义特征是不存在的 $l_k=0$，否则 $l_k=1$.

若查询图像 q 和待查询图像 g 的第 k 个局部语义特征均存在，则 $l_k^q\cdot l_k^g=1$. 其它情况下，$l_k^q\cdot l_k^g=0$. 以 $d_k=D(F_{part\text{-}k}^q,F_{part\text{-}k}^g)$ 来代表两者之间的距离，于是，最终的局部语义特征的平均相似距离为

$$d=\frac{\sum_{k=1}^{K-1}(l_k^q\cdot l_k^g)d_k+(d_{global}+d_{foreground}+d_{global\text{-}attr})}{\sum_{k=1}^{K-1}(l_k^q\cdot l_k^g)+3} \tag{14}$$

此外，本文根据属性预测筛选库数据，过滤掉一些与查询不具有相同属性的图库图像，从而加快评估过程. 在进行距离计算之前，对高置信分数的属性作为查询和图库图像的可靠属性，然后删除与可靠属性不相关的候选图库图像. 本文依据已发表文献中的经验值[14]，将阈值设置为 0.7，当置信分数高于阈值时，该属性为可靠属性. 然后，检查查询图像和图库图像中该可靠属性的预测值. 若相同则进行后续的距离计算，若不同，则将该图像从候选图像中移除. 通过该方法，能够有效提升相似度排序的过程，快速去除不匹配的行人图像.

3.5　损失函数

本文算法在特征优化时主要遵循 3 个部分的损失指导网络收敛，分别是：基础的特征损失、语义解析损失和属性预测损失.

对于基础的特征损失，依据 BoT[19] 基线网络的 BNNeck 模块对每个特征进行后续处理，其基本的损失组包括了增加了标签平滑的 ID 分类损失、三元组损失[14] 和中心损失，可以表示为

$$L_{\mathrm{BNNeck}}=L_{\mathrm{ID}}+L_{Tri}+\beta L_C \tag{15}$$

上式中的 β 代表中心损失对于损失组的贡献权重. 在 BoT 基线网络中，β 被设置为 0.0005.

由于本文算法包含了 4 类特征，即全局特征 $\boldsymbol{F}_{global}$、前景特征 $\boldsymbol{F}_{foreground}$、局部语义特征 $\boldsymbol{F}_{part\text{-}k}=\{\boldsymbol{F}_{part\text{-}1},\boldsymbol{F}_{part\text{-}2},\cdots,\boldsymbol{F}_{part\text{-}K-1}\}$ 和联合属性的全局特征 $\boldsymbol{F}_{global\text{-}attr}$. 所以每类特征根据述式(15)建基础损失，

也全为损失组的形式，分别表示为 L_{global}，$L_{foreground}$，L_{part} 和 $L_{global\text{-}attr}$. 其中，L_{part} 代表所有局部语义特征的损失总和，其中每个局部语义特征，也均为上述公式的形式.

另外，为了指导精细化局部语义模块中的概率图生成的准确性，使用了以伪标签作为监督的交叉熵损失进行约束. 对于 FLS 模块的概率图 $P^{k\times h\times w}(x,y)$，按照像素级别进行展开. 假设 (x,y) 点处的 $1\times K$ 向量表示为 $(p_0,p_1,p_2,\cdots,p_i,\cdots,p_{K-1})$，其中 p_i 表示该像素点属于语义类别 i 的概率，然后用 k 代表 (x,y) 点代表融合的语义解析图的真实的标签. 当 $i=k$ 时，真值 q_i 为 1，否则为 0，那么像素点 (x,y) 处的语义交叉熵损失为

$$L_{parsing}(x,y)=\sum_{i=0}^{K-1}-q_i\log(p_i) \tag{16}$$

将所有像素点的损失进行相加，总的语义交叉熵损失为

$$L_{parsing}=\sum_{x,y}L_{parsing}(x,y),\ (x,y)\in(w,h) \tag{17}$$

在预测属性信息时，假设 m 个属性的真实标签为 $a=(a_1,a_2,\cdots,a_j,\cdots,a_m)$. 第 j 个属性特征 $F_{attr,j}$ 经过 Softmax 函数运算后的向量为 $(p_0,p_1,\cdots,p_{c-1})$，c 是该属性的类别数. 当 i 等于 a_j 时，真值 q_i 为 1，否则为 0，那么第 j 个属性对应预测的交叉熵损失为

$$L_{attr,j}=\sum_{i=0}^{c-1}-q_i\log(p_i) \tag{18}$$

然后，将 m 个属性的交叉熵损失进行相加，得到属性预测的损失为

$$L_{attr}=\sum_{j=0}^{c-1}L_{attr,j} \tag{19}$$

最后，联合由行人 ID 标签引导的基础特征损失（全局特征、前景特征、包含随身物品的局部语义特征、和联合属性信息的全局特征）以及由语义标签引导的语义解析损失，与属性标签引导的属性预测损失，作为网络的总损失，如式(20)：

$$L=L_{global}+L_{foreground}+L_{part}+L_{global\text{-}attr}+\alpha L_{parsing}+\alpha L_{attr} \tag{20}$$

其中 α 代表语义交叉熵损失与属性预测损失的权值，实验取 0.01. 总的来说，本文网络通过增强语义区域的完善性来提高网络预测的性能. 通过将聚类产生的随身物品区域融合至额外语义模型的语义解析结果，形成了更准确的行人表示. 同时随身物品区域的增加为网络对行人背包属性的预测提供了精确的特征信息，体现了本文精细化局部语义模块相对于通用语义解析模型的优越性. 另外，通过将每个属性的预测分数进行加权，使得结果充分考虑了每个属性之间的相关性，提高了行人关键属性描述的置信分数. 引入属性的全局特征更加详细地描述了行人，从而提高了分类的准确性.

4 实验及分析

本节首先介绍了本文算法使用的实验配置，包括数据集、算法评估指标和实验实施的细节. 然后，对提出的 3 个模块进行消融实验，验证各个模块的有效性. 进一步地，将本文所提算法与经典的行人重识别算法进行性能对比. 最后，展示了本文算法相关的可视化结果，并进行分析.

4.1 实验配置

为了体现所提方法的有效性和普遍性，本文网络在 Lin 等人[15]提出的 Market-1501 和 DukeMTMC-reID 属性数据集上验证性能. 表 3 展示了各数据集的详细信息.

表 3 实验所使用属性数据集的详细信息

属性数据集	ID 个数	图像总数	相机个数	属性数量
Market-1501	1501	32668	6	27
DukeMTMC-reID	1404	36441	8	23

按照 Re-ID 常见的做法，算法使用 *Rank*-1 的累积匹配特征 CMC（Cumulative Matching Characteristic）和平均准确率 *mAP*（mean Average Precision）来评估性能[20]. 对于语义标签的预测性能，使用区域交并比 IoU（Intersection over Union）[21]来衡量. 在评估属性的预测性能时，将 Market-1501 数据集上的干扰图像从测试集中排除. 对于每个类别属性的分类精度，使用准确率（Precision）来衡量. 然后，将所有属性预测的平均准确度作为总体的属性分类准确率.

本文算法基于开源框架 PyTorch 1.2.0 版本来实现网络构建，并使用 NVIDIA-Tesla-v100 作为网络训练的硬件环境. 网络的训练参数与 BoT 基线网络保持一致. 在数据预处理方面，将输入图像的大小调整为 $3\times256\times128$，全局特征图 $\boldsymbol{F}_{global}$ 的大小是输入大小的 1/4，通道数为 256. 然后，采用了概率为 0.5 的随机裁剪和随机擦除变换来进行数据增强. 主干网络使用 ResNet50 结构在整个网络中提取特征，并使用在 ImageNet 上预训练的模型进行参数初始化. 学习率采用 Warmup 预热学习的方式进行先增后减，从 3.5×10^{-5} 到 3.5×10^{-4} 进行了 10 个

Epoch 的预热. 在第 40 个和第 70 个 Epoch，学习率降低了 0.1 倍. 模型在训练了 120 个 Epoch 之后达到了收敛. 网络训练的 Batch-size 设置为 64，并使用 Adam 方法[22] 对模型进行损失优化. 测试时的 Batch-size 设置为 128，以提高计算速度.

4.2　消融实验

本节内容探讨了精细化局部语义模块 FLS、属性学习模块 ALM 和属性加权模块 AWM 对网络性能的影响. 基于 BoT 基线网络，逐渐增加各个模块的组合实验，结果如表 4 所示.

表 4　各个模块消融实验结果　（单位：%）

Baseline	FLS	ALM	AWM	Market-1501		DukeMTMC-reID	
				mAP	*Rank*-1	*mAP*	*Rank*-1
✓				85.9	94.5	76.4	86.4
✓	✓			88.2	95.2	80.9	90.6
✓	✓	✓		89.2	95.5	82.3	91.5
✓	✓	✓	✓	**89.5**	**95.6**	**82.8**	**91.7**

4.2.1　精细化局部语义模块 FLS 的选择

使用精细化局部语义模块 FLS 融合聚类语义和额外语义模块生成的语义解析结果既包含精细的效果，同时还含有行人的随身物品信息，为下一步预测背包的属性提供了基础. 从实验结果可以看出，增加 FLS 模块后，在两个数据集上相比 Baseline 的 *mAP*/*Rank*-1 指标分别提高了 2.3%/0.7% 和 4.5%/4.2%，这验证了 FLS 的有效性. 因为增加 FLS 模块后，网络包含了更具区分度的局部特征，从而为全局特征进行了补充，增强了行人的特征表达. 为了证明 FLS 模块分别融合额外语义和聚类语义的有效性，表 5 显示了单独使用每个语义情况的结果.

表 5　FLS 模块与单个语义的消融实验结果（单位：%）

模块	Market-1501		DukeMTMC-reID	
	mAP	*Rank*-1	*mAP*	*Rank*-1
Baseline	85.9	94.5	76.4	86.4
＋聚类语义	86.8	94.6	78.3	88.4
＋SCHP 语义	87.6	95.0	79.4	89.7
＋FLS	**88.2**	**95.2**	**80.9**	**90.6**

由表 5 结果可以发现，在原始的 Baseline 网络上增加语义信息时，因为聚类语义的标签并不那么精细，仅使用聚类语义的效果是最不好的；使用 SCHP 语义的性能优于聚类语义；使用 FLS 模块得到的语义比单独使用每个语义的结果更优. 显然，在 DukeMTMC-reID 数据集上使用 FLS 模型的性能提升更多. 这是因为该数据集中行人背包的数量较多，SCHP 语义解析结果中大多不包含该结果，所以融合后的语义信息进一步提升了算法的性能. 图 5 显示了 FLS 模块优化后的语义解析结果与聚类语义、SCHP 解析结果的对比.

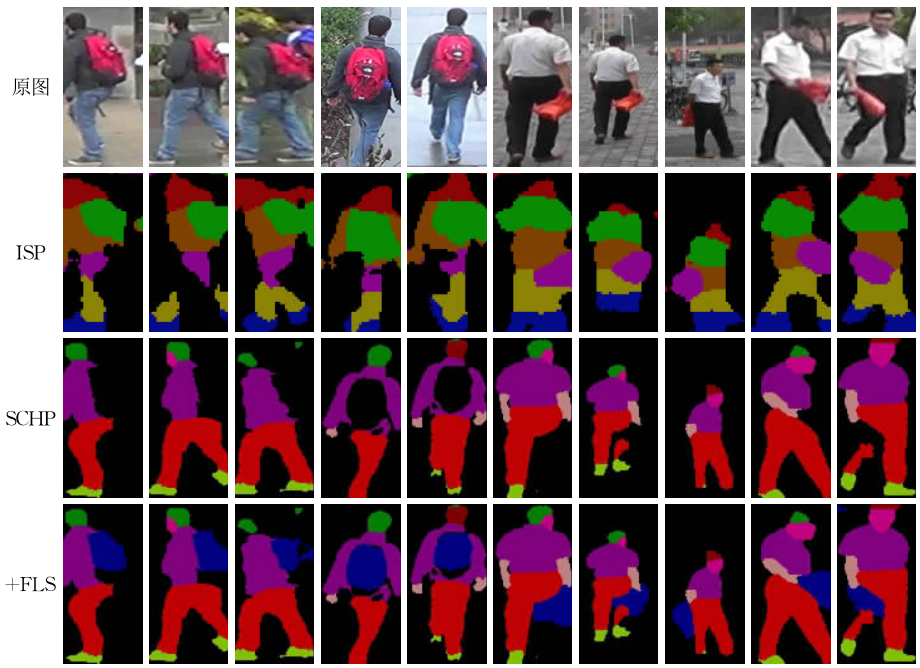

图 5　FLS 模块与其它语义解析结果的对比图

从图 5 可以看出，当行人包含潜在的随身物品时，FLS 模块的效果是最好的. FLS 模块相对于 ISP 聚类语义方法，具备了 SCHP 额外语义模型生成的语义区域精准正确的优点，并且包含随身物品区域，

使得行人的描述信息更加全面. 当 SCHP 模型对行人的随身物品漏检时,FLS 模块可以在背景空缺的地方填补聚类语义产生的随身物品区域,从而完善了行人的语义解析效果. 以上结果证明了本文所提 FLS 模块相比聚类语义和额外语义的优势.

4.2.2 属性学习模块 ALM 的选择

考虑到精细化的语义特征更有利于预测行人的属性信息,本文提出了属性学习模块来生成行人的多个属性. 本文通过局部的语义特征来预测行人的属性,提出的 ALM 模块可以从 FLS 模块生成的各个语义区域中预测相关的属性,避免了大量背景噪声的影响和大量网络的参数量. 从表 4 的实验结果可以看出,增加 ALM 模块后,*mAP*/*Rank*-1 性能在两个数据集相比之前提高了 1.0%/0.3%和 1.4%/0.9%,这验证了增加属性信息对于网络性能的提升. 值得注意的是,属性信息对于网络的性能增益不如全局特征和语义特征,如语义特征为 2.3%/0.7%和 4.5%/4.2%. 这主要是因为,属性信息本身是对于全局特征的一种补充,网络主要还是依赖于信息量更大的全局特征. 但是,属性信息有利于判别相似的行人,同时加速网络的相似度排序过程,对于提升性能起到额外的小幅度加强作用.

为了验证本文算法提出的 ALM 模块通过语义特征来预测属性的优势,本节进行了通过全局特征预测属性的对比实验. 具体来说,对比实验的属性信息完全来自于主干网络后生成的全局特征. 然后,将这些属性信息与全局特征进行连接,经过 ID 分类损失进行网络训练. 在同时包含了语义特征的情况下,得到的属性信息对于网络性能的提升效果如表 6 所示.

表 6 全局特征与 ALM 模块的属性性能对比结果

(单位:%)

特征类别	Market-1501		DukeMTMC-reID	
	mAP	*Rank*-1	*mAP*	*Rank*-1
Baseline	88.2	95.2	80.9	90.6
+全局特征	88.8	95.3	81.7	90.9
+ALM	**89.2**	**95.5**	**82.3**	**91.5**

从表 6 结果可以得到结论,使用全局特征来预测属性是可以一定程度上提高网络的性能的,但不如局部特征预测属性带来的提升多. 第二行的结果显示,使用全局特征预测的属性信息相比之前的基础 Baseline 网络,即只使用全局特征和局部特征做 ID 分类,*mAP*/*Rank*-1 性能分别提高了 0.6%/0.1%和 0.8%/0.3%. 这些性能提升不如 ALM 模块带来的 1.0%/0.3%和 1.4%/0.9%性能增益多. 原因在于全局特征中包含了很多与某个属性无关的信息,这些信息与背景噪声一起成为了干扰因素,从而降低了属性预测的准确性. 总之,对比结果显示了基于局部的语义信息进行属性预测,能带来比全局特征更好的性能提升.

基于数据集的属性标签,在 Market-1501 数据集上,本文统计对比了使用全局特征与使用 ALM 模块的 12 个属性的预测准确率,及二者之间的变化量 1 如表 7 所示.

表 7 Market-1501 数据集基于 ALM 和 AWM 的 12 个属性的预测准确率对比

(单位:%)

方法	性别	年龄	头发长短	袖子长短	下身服装长短	下身服装类型	是否有背包	是否有手提包	其它类型的包	是否戴帽子	上身服装颜色	下身服装颜色	平均准确率
全局特征	90.7	87.5	83.6	93.4	93.2	92.2	87.0	89.3	76.7	95.7	73.1	73.7	86.3
加 ALM	92.4	88.4	86.2	95.7	95.2	94.4	91.3	93.0	81.4	97.2	84.5	81.9	90.1
变化量 1	1.7	0.9	2.6	2.3	2.0	2.2	**4.3**	**3.7**	**4.7**	1.5	**11.4**	**8.2**	3.8
加 AWM	94.2	89.1	87.5	96.4	95.7	95.4	92.2	94.5	82.6	98.9	85.6	82.5	91.2
变化量 2	**1.8**	0.7	**1.3**	0.7	0.5	1.0	0.9	**1.5**	**1.2**	**1.7**	1.1	0.6	1.1

可以发现,增加 ALM 模块后,属性的平均准确率提高了 3.8%,这证明了 ALM 模块对于提升属性预测准确率的作用. 另外,通过单个属性的变化量可以看出,通过局部语义特征来预测属性会显著提升与局部信息有关属性的预测准确率,即表中的上下身服装颜色和背包信息等. 此外,对使用高置信度属性信息筛选查询图库对行人匹配的评估速度提升效果进行验证. 实验表明,在没有加速的情况下,测试过程使用 NVDIA TITAN Xp 每次查询 1.91 s. 在加速时,每张图片的测试只需要 0.27 s.

4.2.3 属性加权模块的选择

本文算法提出属性加权模块来充分利用属性之间的相关性,通过某些属性的高度相关性来提高属性自身的预测置信分数. 从表 4 的实验结果可以得到结论,增加 AWM 模块后,在两个数据集上的 *mAP*/*Rank*-1 指标的性能提升为 0.3%/0.1%和 0.5%/0.2%. 对于改善网络的识别性能有进一步的作用,验证了 AWM 模块的有效性. 这与预期的效果一致,相比于属性信息带来的 1.0%/0.3%和 1.4%/0.9%提升,AWM 模块的性能增益也是

可观的. 通过提升一些相关性高的属性的置信分数，能够优化该属性的预测准确率，从而提升网络的整体性能.

表 7 后半部分为本节针对 AWM 模块的属性预测准确率及准确率变化量 2 的统计结果.

从表 7 可以发现，增加 AWM 模块后，可以获得 1.1%的平均准确率的性能增益. 该模块对于性别、头发长短、手提包、其他类型的包、戴帽子和上身服装颜色的性能提升很高，实验结果验证了通过相关性改善属性预测准确率的有效性.

通过上述对比实验，验证了 ALM 和 AWM 模块对网络性能的提升. 实际上，AWM 模块起到了良好的辅助作用，通过属性之间的相关性信息提高了属性预测的准确性，从而为行人提供了更详细、更精准的信息描述.

4.3　与经典算法对比

为了体现本文算法性能与已有工作的对比，本节选取了基于语义特征的 GPS[10] 和 ISP[13] 算法，以及使用属性信息的 AANet[23] 和 AAB[24] 算法，和近年来发表的性能先进的 AOPS[25]、MHSA-Net[26]、BPBReID[27] 算法. 上述算法的主干网络大部分为 ResNet-50，实验数据取自文献中的性能结果. 本文算法与各个算法的性能对比如表 8 所示.

结果显示，本文算法在两个数据集上均取得了较优和最优的性能. 相比于 BoT 基线算法，本文算法的 *mAP*/*Rank*-1 指标增益为 3.6%/1.1%和 6.4%/5.3%. 对于使用语义特征的算法 GPS 和 ISP，本文算法的 *mAP*/*Rank*-1 指标至少提高了 0.9%/0.3%和 2.8%/2.1%. 原因在于本文算法能生成精细的语义区域，不仅语义区域的解析效果良好，同时包含了行人潜在的随身物品信息和属性信息，提高了行人表示. 对于使用了属性信息的 AAB 和 AANet 算法，本文算法均取得了更优的性能. 主要原因在于本文算法的属性信息是通过精细的语义特征来预测的，而不是依赖包含噪声干扰的全局特征和局部特征. 这种方法能显著提高属性预测的准确率. 最后，本文算法还充分利用了属性之间的关系来辅助增强属性预测的准确性，使得性能更优. 综合来看，本文算法在已有的语义特征方法和属性方法中，识别性能具有一定的竞争力.

表 8　本文算法与经典算法横向对比的结果（单位：%）

算法	Market-1501		DukeMTMC-reID	
	mAP	*Rank*-1	*mAP*	*Rank*-1
AANet[23]（CVPR 19）	83.4	93.3	74.3	87.7
ISP[13]（ECCV 20）	88.6	95.3	80.0	89.6
AAB[24]（TIP 20）	88.6	**96.1**	80.4	89.9
GPS[10]（CVPR 21）	87.8	95.2	78.7	88.2
TransReID[7]（ICCV 21）	88.9	95.2	82.2	90.6
AOPS[25]（TCSVT 22）	85.3	94.6	76.3	87.5
AGW[3]（TPAMI 22）	87.8	95.1	79.6	89.0
MHSA-Net[26]（TNNLS 23）	88.8	95.2	82.3	90.5
BPBReID[27]（WACV 23）	87.0	95.1	78.3	89.6
BoT[19]（CVPR 19）	85.9	94.5	76.4	86.4
本文算法	**89.5**	95.6	**82.8**	**91.7**

对于本文算法的属性预测性能，本文选取了 AANet 和 ARN[28] 算法的结果与所提算法进行比较，Market-1501 数据集的结果如表 9 所示.

表 9　Market-1501 数据集 12 个属性的预测准确率　（单位：%）

方法	性别	年龄	头发长短	袖子长短	下身服装长短	下身服装类型	是否有背包	是否有手提包	其它类型的包	是否戴帽子	上身服装颜色	下身服装颜色	平均准确率
APR[15]	88.9	88.6	84.4	93.6	93.7	92.8	84.9	90.4	76.4	97.1	74.0	73.8	86.6
AANet[23]	92.3	88.2	86.6	94.5	94.2	94.8	87.8	90.0	79.7	98.0	77.1	70.8	87.8
ARN[28]	91.5	87.7	89.1	93.8	94.7	93.6	90.2	90.5	81.9	97.5	76.4	68.6	88.0
本文算法	**94.2**	**89.1**	**87.5**	**96.4**	**95.7**	**95.4**	**92.2**	**94.5**	**82.6**	**98.9**	**85.6**	**82.5**	**91.2**

从表 9 中数据可以得到结论，本文算法的属性预测性能获得了最优. 属性的平均准确率相比 AANet 和 ARN 算法分别提高了 3.4%和 3.2%. AANet 和 ARN 算法的特征中均包含背景噪声，性能不如本文精细化的像素级语义特征，因此本文算法实现了更好的属性预测效果.

4.4　实验效果展示

在测试阶段，本文将网络的特征进行距离计算，然后生成查询图像的相似度排序结果. 图 6 展示了本文算法和 BoT 基线算法的相似度排序结果图.

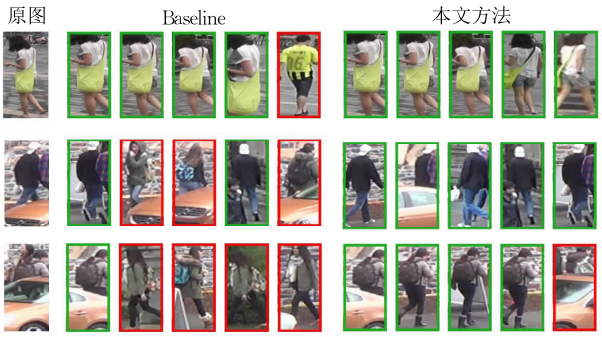

图 6　本文算法与基线算法的相似度排序结果对比

从图 6 结果可以显示本文算法利用语义特征作为行人描述的优点.语义特征能够解决遮挡、背景干扰和行人姿态变换等问题,从而提高了匹配过程的鲁棒性.另外,属性信息在相似度排序时也具有进一步区分行人的能力,图 7 展示了利用属性信息对行人描述进行详细补充的好处.

图 7 属性信息对于相似度排序的作用示例

5 结 论

针对现有的语义解析模型无法识别行人随身物品,以及未在语义特征上充分利用属性信息的不足,本文提出了一种精细化局部语义与属性信息的行人重识别网络.首先,基于额外语义模型形成的语义区域与聚类算法形成的伪标签语义区域,从中捕获行人的随身物品区域信息,然后融合至额外语义模型生成的语义区域.这些新的语义解析结果,区域显现得非常平滑良好,同时也包含了潜在的随身物品信息.因此提高了语义区域的完整性,且行人的描述更加充分.其次,基于这些精细化的语义特征,通过构建属性预测子分支,预测了行人的多个属性.因为这些语义特征不包含背景杂波的影响,所以预测性能更优.最后,基于属性之间的相关性先验知识,可以通过某些属性的联合来增加相关程度高的属性的置信分数.该过程对行人的属性置信分数进行重新赋权,增加了行人描述的准确性.本文算法将精细化的语义信息和详细的属性信息相结合,学习到了更具有鉴别性的行人特征表示,并加快了行人相似度排序的过程.在公开数据集上的大量实验验证了本文算法的有效性,显示了语义信息和属性信息相互受益的互补性.

致 谢 本论文的数值计算得到了武汉大学超级计算中心的计算支持和帮助!

参 考 文 献

[1] Luo Hao, Jiang Wei, Fan Xing, et al. A survey on deep learning based person re-identification. Acta Automatica Sinica, 2019, 45(11): 2032-2049(in Chinese)
(罗浩,姜伟,范星等.基于深度学习的行人重识别研究进展.自动化学报,2019,45(11):2032-2049)

[2] Xiao J, Wu Y, Chen Y, et al. LSTFE-Net: Long short-term feature enhancement network for video small object detection//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Vancouver, Canada, 2023: 14613-14622

[3] Ye M, Shen J, Lin G, et al. Deep learning for person re-identification: A survey and outlook. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(6): 2872-2893

[4] Sun Y, Zheng L, Yang Y, et al. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)//Proceedings of the European Conference on Computer Vision. Munich, Germany, 2018: 480-496

[5] Zheng F, Deng C, Sun X, et al. Pyramidal person re-identification via multi-loss dynamic training//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 8514-8522

[6] Liu H, Xia D, Jiang W. Towards homogeneous modality learning and multi-granularity information exploration for visible-infrared person re-identification. IEEE Journal of Selected Topics in Signal Processing, 2023, 17(3): 545-559

[7] Li W, Zhu X, Gong S. Harmonious attention network for person re-identification//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 2285-2294

[8] He S, Luo H, Wang P, et al. TransReID: Transformer-based object re-identification//Proceedings of the IEEE/CVF International Conference on Computer Vision. Nashville, USA, 2021: 15013-15022

[9] Zhao H, Tian M, Sun S, et al. Spindle net: Person re-identification with human body region guided feature decomposition and fusion//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 1077-1085

[10] Nguyen B X, Nguyen B D, Do T, et al. Graph-based person signature for person re-identifications//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Nashville, USA, 2021: 3492-3501

[11] Li S, Leng J, Gan J, et al. Shape-centered representation learning for visible-infrared person re-identification. arXiv preprint arXiv:2310.17952, 2023

[12] Chen C, Ye M, Qi M, et al. Structure-aware positional transformer for visible-infrared person re-identification. IEEE Transactions on Image Processing, 2022, 31: 2352-2364

[13] Zhu K, Guo H, Liu Z, et al. Identity-guided human semantic parsing for person re-identification//Proceedings of the European Conference on Computer Vision. Glasgow, UK, 2020: 346-363

[14] Li P, Xu Y, Wei Y, et al. Self-correction for human parsing. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(6): 3260-3271

[15] Lin Y, Zheng L, Zheng Z, et al. Improving person re-identification by attribute and identity learning. Pattern Recognition, 2019, 95: 151-161

[16] Chen X, Liu X, Liu W, et al. Explainable person re-identification with attribute-guided metric distillation//Proceedings of the IEEE International Conference on Computer Vision. Montreal, Canada, 2021: 11813-11822

[17] Xiao Jin-Sheng, Guo Hao-Wen, Zhang Shu-Hao, et al. Pedestrian re-identification algorithm based on dual-domain filtering and triple metric learning. Journal of Electronics & Information Technology, 2022, 44(11): 3931-3940(in Chinese)
(肖进胜，郭浩文，张舒豪等. 双域滤波三元组度量学习的行人再识别. 电子与信息学报，2022，44(11)：3931-3940)

[18] Kuang Cheng, Chen Ying. Multi-granularity feature fusion network for person re-identification. Acta Electronica Sinica, 2021, 49(8): 1541-1550(in Chinese)
(匡澄，陈莹. 基于多粒度特征融合网络的行人重识别. 电子学报，2021，49(8)：1541-1550)

[19] Luo H, Gu Y, Liao X, et al. Bag of tricks and a strong baseline for deep person re-identification//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Online, 2019

[20] Zheng L, Shen L, Tian L, et al. Scalable person re-identification: A benchmark//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile, 2015: 1116-1124

[21] Zheng Z, Wang P, Liu W, et al. Distance-IoU loss: Faster and better learning for bounding box regression//Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA, 2020, 34(7): 12993-13000

[22] Zhang Z. Improved adam optimizer for deep neural networks//Proceedings of the 2018 IEEE/AWM 26th International Symposium on Quality of Service. Banff, Canada, 2018: 1-2

[23] Tay C P, Roy S, Yap K H. AANet: Attribute attention network for person re-identifications//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Online, 2019: 7134-7143

[24] Zhang J, Niu L, Zhang L. Person re-identification with reinforced attribute attention selection. IEEE Transactions on Image Processing, 2020, 30: 603-616

[25] Jin H, Lai S, Qian X. Occlusion-sensitive person re-identification via attribute-based shift attention. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(4): 2170-2185

[26] Tan H, Liu X, Yin B, et al. MHSA-Net: Multihead self-attention network for occluded person re-identification. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(11): 8210-8224

[27] Somers V, De Vleeschouwer C, Alahi A. Body part-based representation learning for occluded person re-identification//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, USA, 2023: 1613-1623

[28] Yin J, Fan Z, Chen S, et al. In-depth exploration of attribute information for person re-identification. Applied Intelligence, 2020, 50: 3607-3622

XIAO Jin-Sheng, Ph. D., associate professor. His research interests include computer vision, image processing and analysis.

WU Jing-Yi, M. S. candidate. Her research interests include image processing and computer vision.

GUO Hao-Wen, M. S. His research interests include image processing, video analysis and computer vision.

GUO Yuan, Ph. D., lecturer. Her research interests include high-definition map, crowdsourcing data and computer vision.

ZHAO Chi-Heng, M. S. candidate. His research interests include image processing and computer vision.

WANG Yin, M. S. candidate. Her research interests include computer vision and image segmentation.

Background

In recent years, many advanced image analysis technologies have been introduced into video surveillance systems, including facial recognition, license plate recognition, pedestrian detection, and action recognition. However, facial recognition technology fails when the distance is too great or when faces are obscured. To address these shortcomings, person re-identification (Re-ID) technology has emerged. Person re-identification involves recognizing specific individuals across multiple non-overlapping camera views. With the gradual deepening of research, the performance of person re-identification methods has improved annually. Essentially, person re-identification involves designing a robust feature representation and determining whether individuals are the same by calculating the similarity between the feature vectors of images.

Different camera positions result in varying appearances of individuals in different camera views, complicating the identification process. Additionally, factors affecting person re-identification performance include low resolution, lighting variations, unconstrained pose changes, occlusion, complex camera environments, background clutter, unreliable detection boxes, similar appearances, and changes in attire. These challenges hinder the generalization performance of many re-identification methods in actual scenes, though average accuracy rates on certain datasets reach around 80%. This study primarily focuses on improving the performance of person re-identification methods in various complex environments, which holds significant potential for advancement in the field.

Personal belongings information and attribute description are effective ways to improve the performance of person re-identification tasks. A person re-identification network based on fine-grained local semantics and attribute learning is proposed to extract the information of personal belongings and, at the same time, obtain attribute descriptions of the person from the semantic region. First, the belongings area generated by the feature clustering method is fused into the semantic analysis results generated by an additional semantic model. Second, using the generated semantic regions as body labels, the network constructs semantic feature maps of these regions from global features, and then predicts attribute information related to them from the semantic features. Finally, considering that some attributes of a person contain strong correlations, a reweighting model is constructed to improve the confidence of some attributes and optimize the prediction accuracy of attributes. Concatenating the attribute prediction results and the global features of a person forms a robust feature representation. Experiments on the Market-1501 and DukeMTMC-reID attribute datasets show that the proposed algorithm can improve the performance of person re-identification tasks.

This work is supported by the National Natural Science Foundation of China (No.42201480) and the National Key Research and Development Program of China (No.2021YFB2501104).