



中国科技期刊卓越行动计划项目入选期刊

# 控制与决策

CONTROL AND DECISION



## 基于相关性得分的伪标签优化行人重识别

程德强, 黄绩, 寇旗旗, 张剑英, 李云龙

引用本文:

程德强, 黄绩, 寇旗旗, 张剑英, 李云龙. 基于相关性得分的伪标签优化行人重识别[J]. *控制与决策*, 2024, 39(8): 2579–2587.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2023.0510>

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### 基于多尺度特征表示的行人再识别

Multi-scale feature representation for person re-identification

控制与决策. 2021, 36(12): 3015–3022 <https://doi.org/10.13195/j.kzyjc.2020.0952>

#### Anchor-free的尺度自适应行人检测算法

Anchor-free scale adaptive pedestrian detection algorithm

控制与决策. 2021, 36(2): 295–302 <https://doi.org/10.13195/j.kzyjc.2020.0124>

#### 基于分类特征约束变分伪样本生成器的类增量学习

Class incremental learning based on variational pseudo-sample generator with classification feature constraints

控制与决策. 2021, 36(10): 2475–2482 <https://doi.org/10.13195/j.kzyjc.2020.0228>

#### 行人重识别中度量学习方法研究进展

A survey on metric learning in person re-identification

控制与决策. 2021, 36(7): 1547–1557 <https://doi.org/10.13195/j.kzyjc.2020.0801>

#### 结合注意力机制的循环神经网络复述识别模型

Recurrent neural networks based paraphrase identification model combined with attention mechanism

控制与决策. 2021, 36(1): 152–158 <https://doi.org/10.13195/j.kzyjc.2019.0638>

# 基于相关性得分的伪标签优化行人重识别

程德强<sup>1†</sup>, 黄 绩<sup>1</sup>, 寇旗旗<sup>2</sup>, 张剑英<sup>1</sup>, 李云龙<sup>1</sup>

(1. 中国矿业大学 信息与控制工程学院, 江苏 徐州 221116;

2. 中国矿业大学 计算机科学与技术学院, 江苏 徐州 221116)

**摘要:** 无监督域自适应行人重识别旨在将源域训练的识别能力泛化到目标域上, 以减少对标签的依赖. 目前基于聚类方法的网络, 聚类过程中不可避免地会受到环境噪声的影响, 降低网络原有识别性能. 为了解决这一问题, 提出一种基于相关性得分的伪标签优化行人重识别网络. 首先, 通过计算全局与局部特征间前  $k$  个相似样本集合的相关性得分, 找到两类特征直接可靠的关联性, 从而提取已有伪标签优化方法所忽略的局部细粒度特征; 然后, 利用得分对局部伪标签进行优化处理, 降低网络对与行人无关局部特征的关注; 最后, 依赖于相关性得分, 利用优化后局部伪标签的预测结果对全局伪标签进行细化, 缓解聚类过程中噪声的同时也细化了行人的特征完整表示. 与近年无监督域自适应行人重识别方法相比, 所提出网络在 DukeMTMC-ReID, Market 1 501 和 MSMT 17 三个公开数据集上的实验结果表现优异, 验证了所提出网络的有效性.

**关键词:** 无监督域自适应; 行人重识别; 深度学习; 细粒度特征; 相关性得分; 伪标签优化

中图分类号: TP391

文献标志码: A

DOI: 10.13195/j.kzyjc.2023.0510

引用格式: 程德强, 黄绩, 寇旗旗, 等. 基于相关性得分的伪标签优化行人重识别[J]. 控制与决策, 2024, 39(8): 2579-2587.

## Person re-identification with pseudo label refinement based on correlation score

CHENG De-qiang<sup>1†</sup>, HUANG Ji<sup>1</sup>, KOU Qi-qi<sup>2</sup>, ZHANG Jian-ying<sup>1</sup>, LI Yun-long<sup>1</sup>

(1. School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China; 2. School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China)

**Abstract:** The purpose of unsupervised domain adaptive person re-identification is to generalize the recognition ability of training in the source domain to the target domain to reduce the dependence on labels. At present, the network based on clustering methods will inevitably be affected by environmental noise during the clustering process, which will reduce the original recognition performance of the network. To solve this problem, the person re-identification network with pseudo label refinement based on correlation score is proposed. Firstly, by calculating the correlation scores between the top  $k$  similar sample sets of global and local features, reliable correlations between two types of features are found, so as to extract local fine-grained features that existing pseudo-label optimization methods ignore. Then, the scores are used to optimize the local pseudo-labels, reducing the network's attention to irrelevant local features of the person. In addition, relying on the correlation scores, the prediction results of the optimized local pseudo-labels are used to refine the global pseudo-labels, which alleviates noise during the clustering process and refines the complete representation of person features. Compared with the unsupervised domain adaptive method in recent years, the experimental results of the network on three public data sets, DukeMTMC-ReID, Market 1 501 and MSMT 17, show that the network performance is significantly improved.

**Keywords:** unsupervised domain adaptation; person re-identification; deep learning; fine-grained features; correlation score; local pseudo label refinement

## 0 引言

行人重识别(person re-identification, Re-ID)是一种在多相机域中识别同一行人的任务<sup>[1]</sup>. 由于其在视

频监控和公共安全方面的重要性<sup>[2]</sup>, 此任务已被广泛研究了几十年. 现有较为成熟的识别网络主要是通过有监督方法实现, 这类方法依赖完整的高质量标

收稿日期: 2023-04-19; 录用日期: 2023-07-04.

基金项目: 国家自然科学基金项目(52204177).

责任编委: 张国山.

<sup>†</sup>通讯作者. E-mail: chengdq@cumt.edu.cn.

签,然而,标注标签成本较为昂贵且耗时长.近年来,半监督和无监督行人重识别吸引了越来越多研究者的注意.诸多无监督训练方法的识别精度正在逐步接近有监督训练方法的高识别精度<sup>[3-4]</sup>.此外,无监督域自适应网络训练方式降低了标注成本,也使得行人重识别网络在现实部署中成为可能.

对于无监督域自适应行人重识别方法而言,一般会通过 $k$ -近邻搜索或聚类方法自主生成伪标签,然后利用伪标签构成的新数据集对网络进行训练,将上述两个步骤交替执行即为网络的迭代训练过程.其中:伪标签作为数据集的重要组成部分,与聚类效果直接相关.行人数据集场景复杂,一张图片除包含行人,图片局部还包含诸多无关因素,会作为噪声对聚类精度产生较大影响.为了解决这一问题,研究者提出了采用鲁棒聚类或伪标签优化等方法来去除噪声.已有优秀成果ECN-GPP<sup>[5]</sup>提出了采用复合网络的形式:源域和目标域数据集同时输入网络,基于图的正向预测方法(graph-based positive prediction, GPP)提取可靠的特征,以此来训练网络.然而,多骨干网带来的计算代价也很直观.此外,这类方法对伪标签的优化只考虑到了全局特征,而忽略了重要的细粒度线索.

对于同一图片,全局和局部特征会关注不同的语义信息.全局特征对于行人姿态和相机视点变化较为敏感,而局部特征捕获到不变的细粒度纹理特征,可增加身份判别的权重,从而提高网络性能.通过进一步研究,本文发现:细粒度信息不仅可以为更好的行人表示提供有用的线索,且细粒度特征与全局特征间存在一定的关联性,这种关联性可对减少聚类噪声起到引导作用.

本文为了优化伪标签,减少噪声带来的聚类问题,提出一种基于相关性得分的伪标签优化无监督域自适应网络.本文的主要内容如下.

1) 对于相关性的挖掘,以行人全局和局部特征的 $k$ 个相邻样本集合的相似度为基础,计算相关性得分,后续利用得分引导来减少伪标签噪声.

2) 基于相关性得分,提出两种伪标签优化方案:局部伪标签优化和全局伪标签细化方案.第1种方案主要针对局部特征进行优化,排除与行人无关的特征,减少对聚类过程的影响;第2种方案通过局部细粒度特征丰富行人全局特征,以获取到更加细化的伪标签.

3) 采用热图和空间域可视化的方法,对所提出方案进行直观展示.通过在Market 1501、DukeMTMC-ReID和MSMT 17三个公共数据集上

多次实验,将结果与近年先进无监督域自适应算法相比较,验证所提出伪标签优化方法的有效性.

## 1 相关工作

### 1.1 无监督域自适应行人重识别

作为无监督学习中的重要分支,无监督域自适应方法旨在将一定的行人识别能力从有标签的源域转移至无标签的目标域.在转移过程中,域间差距成为自适应的一大挑战.针对这一挑战,现有解决方法包括图像风格迁移方法、中间特征对齐方法和聚类方法.

图像风格迁移方法的核心是通过引入生成对抗网络<sup>[6]</sup>(generative adversarial network, GAN),将源域数据集转化为目标域风格的新数据集. Wei等<sup>[7]</sup>创建了一个非常接近现实场景的MSMT 17大型数据集,并提出一种图片风格转移生成对抗网络(person transfer GAN, PTGAN)来弥补领域间的差距.对于中间特征对齐方法, Dai等<sup>[8]</sup>提出了中间域模块(an intermediate domain module, IDM),将监督信息从源域传输至目标域,从而使得网络逐渐适应源域和目标域的数据分布.与基于聚类的方法相比,前两种方案由于引入额外计算量增加了网络的部署负担.基于聚类的方法通常利用聚类算法生成的伪标签来训练网络. Fu等<sup>[9]</sup>提出了一种自相似性分组方法,通过将相似性引导的半监督训练策略引入域自适应的网络框架中,联合训练半监督和非监督分支,有效地促进了图像的转化过程.与上述方法不同,所提出方法更关注聚类方法中的噪声优化问题.通过减少噪声的影响,使得网络发挥其应有的识别性能.

### 1.2 伪标签优化

图片特征各异且复杂,导致聚类结果也具有不确定性.因此,为每个未标记的图像分配正确的伪标签是聚类方法的优化方向. Ma等<sup>[10]</sup>提出了一种深度可信度量学习域自适应方法,通过逐步调整采样策略来减少噪声干扰,并通过实例边缘扩展损失函数约束网络来提高识别的能力.抗噪声相互训练网络<sup>[11]</sup>(noise resistible mutual training, NRMT)通过在训练期间维护两个网络来执行协作聚类和相互实例选择.自定节奏对比学习框架<sup>[12]</sup>(self-paced contrastive learning, SpCL)提出了自学习策略来获得更可靠的聚类结果.组标签转换网络<sup>[13]</sup>(group-aware label transfer, GLT)使用了组感知标签转移算法在线细化伪标签.然而,这些工作通常使用平均特征或可学习权重来表示对标签噪声敏感的聚类.所提出方法通过计算相关性得分,不仅可去除不相关局部特征,也可将相关局部

信息充分利用,丰富行人特征.

### 1.3 基于局部特征的方法

全局特征主要对行人图片全局信息进行提取,多聚焦于行人服装款式、颜色、配饰等特征.然而,网络以全局特征中的某些特征作为行人身份的判别条件,往往是网络分类特征相似但身份不同、样本效果较差的主要原因.此外,全局特征内可能包含的背景区域会作为噪声影响网络性能,且行人姿态不固定同样会影响身份的判别.以有监督行人重识别中多粒度网络<sup>[14]</sup>(multiple granularity network, MGN)为起点,行人局部特征、细粒度特征结合全局特征,逐步完善了行人的特征表示.其算法的思路和性能一直影响着后续学者的研究.其中:基于局部补丁的无监督学习网络<sup>[15]</sup>(patch-based unsupervised learning, PAUL)使用了空间变换器网络提取局部特征,用于基于局部的区分学习.通过进一步分析,本文在充分利用局部特征的同时,也考虑到其可能带来的聚类噪声影响,

所提出的两种标签优化方案很大程度上降低了这种影响.

## 2 基本原理

本文提出一种基于相关性得分的伪标签优化无监督域自适应网络,采用在 ImageNet<sup>[16]</sup>上预训练好的 ResNet 50<sup>[17]</sup>作为主干网络.本文充分利用了行人局部特征与全局特征间的相关性,作为去除行人无关特征和优化伪标签的依据.如图1所示,将数据集  $\{x_i\}_{i=1}^{N_D}$  中的  $N_D$  张行人图片  $x_i$  依次输入网络,网络对图片进行水平等分3块提取特征后,得到全局特征  $f_i^g$  和局部特征  $\{f_i^{p1}, f_i^{p2}, f_i^{p3}\}$ ,  $g$  和  $p_n$  分别作为全局和局部特征的标识;然后,利用密度聚类 (density-based spatial clustering of applications with noise, DBSCAN) 的方法<sup>[18]</sup>分配相应伪标签,全局特征和部分特征共享相同的伪标签  $y_i$ ;最后,计算全局和局部特征的前  $k$  个近邻样本集合的相关性得分  $S(g, p_n)$ , 并利用得分实现局部伪标签优化和全局伪标签细化.

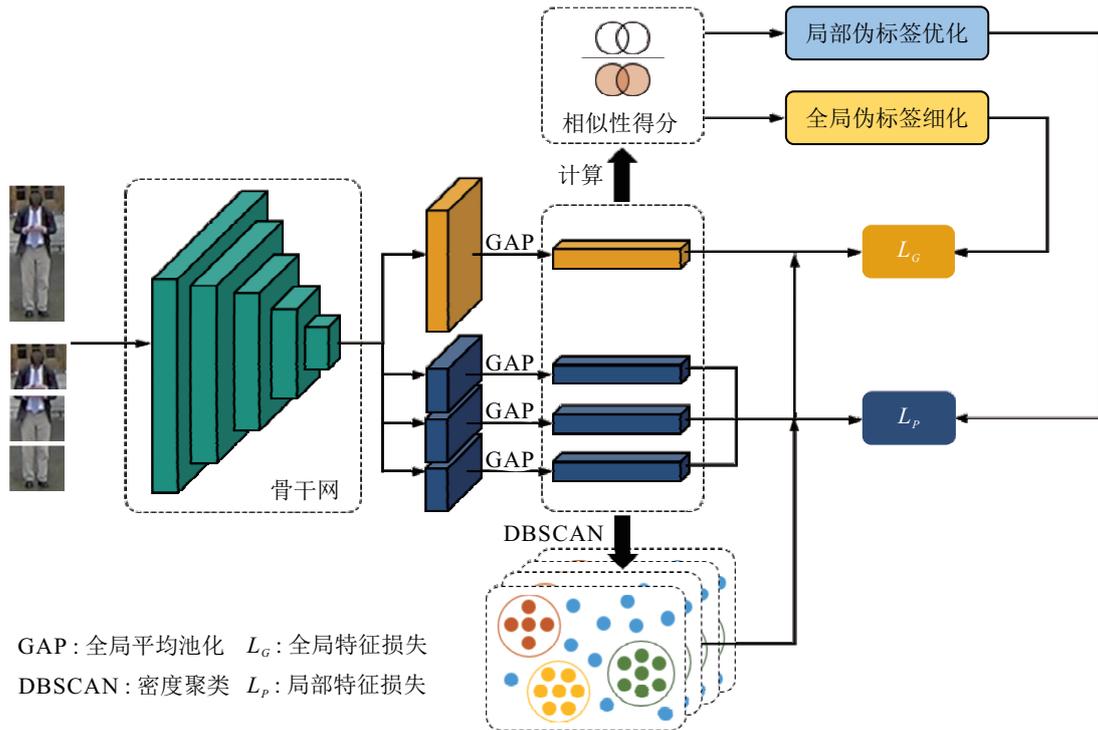


图1 网络框架结构

### 2.1 相关性得分

将行人完整图片输入网络,得到的分块局部特征与全局特征必定存在一定的相关和不相关特征.在网络学习过程中直接引入局部特征,会增加聚类噪声的可能性,这种噪声多由与全局特征不相关特征所导致.图2直观地展示了通过 t-SNE 可视化方法<sup>[19]</sup>生成 DukeMTMC-ReID 数据集<sup>[20]</sup>的全局特征和3个局部特征空间.图2中相同颜色代表同一行人身份,每个

特征空间具有不同的特征分布和不同的语义信息,且可能包含不相关特征.图2中圈出部分包含行人的无关特征,会干扰行人伪标签.

针对这一问题,本文设计了相关性得分.该得分获取了与全局特征最相近  $k$  个局部特征的相关程度,主要计算全局和局部特征的 Jaccard 相似度<sup>[18]</sup>.在全局和每个局部特征空间中独立地执行  $k$  近邻搜索,从而在每个图像上产生4个排序集合,然后通过下式

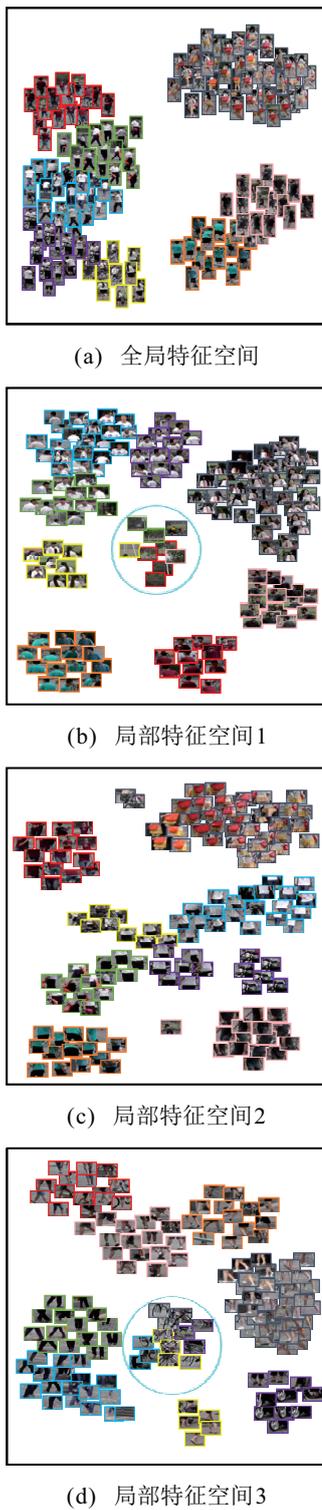


图2 不同特征空间可视化

计算全局与多个局部特征间的相关性得分:

$$S_i(g, p_n) = \frac{|R_i(g, k) \cap R_i(p_n, k)|}{|R_i(g, k) \cup R_i(p_n, k)|} \in [0, 1]. \quad (1)$$

其中:  $|\cdot|$  为集合基数;  $R_i(g, k)$  和  $R_i(p_n, k)$  分别为全局特征  $f^g$  和局部特征  $f^{p_n}$  计算的排名前  $k$  个样本的索引集合,  $n=1, 2, 3$ .

根据 Jaccard 相似度的性质可知, 相似度  $S_i(g, p_n)$  得分越接近 1, 表明  $g$  和  $p_n$  对应的特征空间在数据点  $i$  附近具有越高的相关性, 即包含有用的局部行人特

征; 反之, 相似度  $S_i(g, p_n)$  得分越接近 0,  $g$  和  $p_n$  前  $k$  个样本的索引集合交集越少, 即与行人无关的局部特征越多.

## 2.2 伪标签优化方案

### 2.2.1 局部伪标签优化

使用相同的全局伪标签对模型进行训练, 会使得网络对于同一身份学习到与行人无关的特征, 如背景、遮挡等因素. 因此, 通过去除局部无关特征, 可起到优化全局特征的作用. 对于 3 个局部特征空间, 通过利用相关性得分, 降低对得分较低特征的关注度, 实现局部特征优化的目的. 相应地, 局部伪标签也得到优化. 具体优化公式如下所示:

$$\hat{y}_i^{p_n} = \alpha_i^{p_n} y_i + (1 - \alpha_i^{p_n}) u. \quad (2)$$

其中:  $y_i$  为给定行人图像  $x_i$  相应的伪标签,  $\hat{y}_i^{p_n}$  为优化后的局部伪标签,  $u$  为均匀向量,  $\alpha$  为标签优化的权重. 与使用恒定权重  $\alpha$  的标签优化方法不同, 本文使用相关性得分  $S_i(g, p_n)$  作为标签优化权重  $\alpha$ , 动态调整不同分块集合  $p_n$  的权重. 当某一局部标签优化权重  $\alpha$  较小时, 表示无关局部特征被逐步训练为接近均匀分布的伪标签. 因此, 当 3 个局部优化权重均接近 0 时, 局部伪标签则作为统一无关向量, 对全局伪标签影响较小. 通过这种方式, 可实现减弱无关噪声伪标签对全局伪标签的影响, 从而获得更好的表示学习.

### 2.2.2 全局伪标签细化

细粒度信息从局部出发, 包含更多有用的行人特征线索. 将此类线索与全局特征相结合, 即可实现全局伪标签细化的目的. 在全局特征聚类中, 可利用具有丰富局部上下文的部分特征信息来处理标签噪声. 然而, 由于相关性较低的局部特征会提供误导性信息, 本文根据每个相关性得分, 将具有不同权重的局部特征预测结果进行聚合, 从而用更可靠的信息来细化标签. 具体细化公式如下所示:

$$\hat{y}_i^g = \beta y_i^g + (1 - \beta) \sum_{n=1}^3 w_i^{p_n} q_i^{p_n}, \quad (3)$$

$$w_i^{p_n} = \frac{\exp(S_i(g, p_n))}{\sum_k \exp(S_i(g, p_k))}. \quad (4)$$

其中:  $\hat{y}_i^g$  为细化后的全局伪标签;  $w_i^{p_n}$  和  $q_i^{p_n}$  分别为局部特征  $f_i^{p_n}$  的集合权重和预测向量,  $w_i^{p_n}$  具体计算如式(4)所示; 参数  $\beta \in [0, 1]$  主要控制独热编码(one-hot)伪标签和预测集合的比例.

由式(3)可知, 对于局部特征的预测向量  $q_i^{p_n}$  而言, 集合权重  $w_i^{p_n}$  的大小直接决定局部伪标签对全局伪标签的细化程度. 使得网络在训练的过程中, 能够

在关注全局特征的基础上关注部分与行人信息相关的细粒度特征。

### 2.3 损失函数

基于相关性得分,按照局部伪标签优化和全局伪标签细化方案,分别生成了新的局部伪标签 $\hat{y}_i^g$ 和全局伪标签 $\hat{y}_i^g$ 。为了有效地减少噪声标签的影响,本文利用上述两种精细伪标签对网络进行训练,具体训练方式如下。

对于全局特征,分别计算其交叉熵损失 $L_G$ 和难样本三元组<sup>[21]</sup>损失 $L_{\text{tri}}$ ,具体公式如下所示:

$$L_G = - \sum_{i=1}^{N_D} \hat{y}_i^g \log(q_i^g), \quad (5)$$

$$L_{\text{tri}} = - \sum_{i=1}^{N_D} \log\{\exp(\|f_i^g - f_{i,n}^g\|) - \log\{\exp(\|f_i^g - f_{i,p}^g\|) + \exp(\|f_i^g - f_{i,n}^g\|)\}\}. \quad (6)$$

其中: $N_D$ 为数据集图片数; $q_i^g$ 为全局特征的预测结果; $\|\cdot\|$ 为2阶范数; $f_{i,n}^g$ 表示难样本三元组选择的与全局特征 $f_i^g$ 最不相似的特征,即负样本,反之 $f_{i,p}^g$ 为正样本。

网络对于局部特征同样计算交叉熵损失 $L_P$ ,采用下式进行训练:

$$L_P = - \sum_{i=1}^{N_D} \sum_{n=1}^3 (\alpha_i^{p_n} H(\hat{y}_i^{p_n}, q_i^{p_n}) + (1 - \alpha_i^{p_n}) D_{\text{KL}}(u \| q_i^{p_n})), \quad (7)$$

其中 $H(\cdot)$ 和 $D_{\text{KL}}(\cdot)$ 分别为交叉熵和KL散度。

最终,将网络总体损失函数 $L$ 定义为上述3部分损失函数之和,即

$$L = L_G + L_{\text{tri}} + L_P. \quad (8)$$

## 3 实验和结果分析

### 3.1 实验数据集

本文实验主要在Market 1501<sup>[22]</sup>、MSMT 17<sup>[7]</sup>和DukeMTMC-ReID<sup>[20]</sup>这3个数据集上进行。

Market 1501数据集<sup>[22]</sup>包含1501个身份的行人图像,每个身份均有来自6个不同视点相机的多个图像,包括1个低分辨率相机和5个高分辨率相机,其标签包含27个属性。数据集取景于清华大学内某购物中心前。其中:751个身份用于网络训练,750个身份用于网络识别性能测试。该数据集对于1501个身份行人有32068个带注释的行人边界框。每个带注释身份均至少用两台摄像机拍摄,这样可实现网络的跨摄像机搜索。

DukeMTMC-ReID<sup>[20]</sup>为Duke MTMC数据集的扩展和子集,由1812个身份组成。其中:1404个身份出现在两个以上的摄像机中,408个身份作为干扰项只出现在一个摄像机中。数据集包含训练集702个身份图片共16522张。后文简称DukeMTMC-ReID数据集为Duke。

MSMT 17数据集<sup>[7]</sup>。该数据集是一个接近真实场景的大型数据集,使用12台室外摄像机和3台室内摄像机拍摄。考虑到一个月内的各种天气条件,数据集的取景周期共4天。每天在上午、中午和下午3个不同时间段拍摄1小时的视频。数据集包含126441张图片,共4101个行人身份。其中:训练集图片30248张,身份数为1041个。其取景视点类似于Market 1501数据集,具有更复杂的场景。

### 3.2 实验细节

本文的训练和实验在Ubuntu 18.04操作系统上进行,使用的服务器搭载3块NVIDIA RTX 3090 24 GB GPU、Intel(R) Xeon(R) W-2175 CPU @ 2.50 GHz和64 GB内存。采用ImageNet<sup>[16]</sup>上训练好的ResNet 50<sup>[17]</sup>作为主干网,在网络基础上用平均池化层拼接BNNeck全连接分类器<sup>[23]</sup>替换了layer 3后的所有层。在训练过程中,输入图像尺寸调整为 $384 \times 128$ ,数据集经过常规方式随机翻转、裁剪和擦除等进行了增强。设置mini-batch包含16个身份,每个身份4张图片,共64张图片。网络共训练50个epoch,每个epoch包含400次迭代。对于初始学习率,设置为 $3.5 \times 10^{-4}$ ,且每结束训练20个epoch,学习率缩小10倍。Adam网络优化器权重衰减设置为 $5 \times 10^{-4}$ 。对于聚类方法,本文采用基于Jaccard距离的DBSCAN和 $k$ 近邻编码方法。经实验对比:本文设置参数 $k = 20$ ,如式(1)所示;设置加权参数 $\beta = 0.5$ ,如式(3)所示。

### 3.3 与先进算法进行比较

本文在3个数据集上分别进行域自适应实验,并将实验结果与近年顶级会议相关文章进行对比分析,具体实验结果如表1和表2所示。域自适应实验包括Duke与Market 1501相互自适应,以及Duke和Market 1501分别自适应到MSMT 17。

以行人重识别网络的两个主要评价指标:累计匹配曲线<sup>[37]</sup>(CMC)和平均准确率<sup>[22]</sup>(mAP)作为数据支撑。所有表中R-1表示Rank-1,其他以此类推。

表1为所提出算法与无监督域自适应的15种不同方法识别精度的比较结果。其中:ECN-GPP<sup>[5]</sup>域自适应后识别性能最优,相比于所提出算法,在Duke到Market 1501中,所提出算法的mAP在其基础上提升

表1 不同算法在Duke和Market 1 501的实验结果

算法	Duke → Market 1 501				Market 1 501 → Duke			
	mAP	R-1	R-5	R-10	mAP	R-1	R-15	R-10
UCDA <sup>[24]</sup>	34.5	64.3	—	—	36.7	55.4	—	—
CDS <sup>[25]</sup>	39.9	71.6	81.2	84.7	42.7	67.2	75.9	79.4
ECN <sup>[26]</sup>	43.0	75.1	87.6	91.6	40.4	63.3	75.8	80.4
MAR <sup>[27]</sup>	48.0	67.1	79.8	84.2	40.0	67.7	81.9	87.3
PAST <sup>[28]</sup>	54.6	78.4	—	—	54.3	72.4	—	—
SSG <sup>[9]</sup>	58.3	80.0	90.0	92.4	53.4	73.0	80.6	83.2
UDAP <sup>[29]</sup>	53.7	74.7	86.9	90.3	49.0	68.4	80.1	83.5
NSSA <sup>[30]</sup>	47.9	76.2	88.7	92.4	45.5	65.5	77.9	81.3
SAL <sup>[31]</sup>	38.7	65.3	79.7	84.6	48.5	67.6	80.9	84.7
ADTC <sup>[32]</sup>	59.7	79.3	90.8	94.1	52.5	71.9	84.1	87.5
ACT <sup>[33]</sup>	60.6	80.5	—	—	54.5	72.4	—	—
MPLP <sup>[34]</sup>	60.4	84.4	92.8	95.0	51.4	72.4	82.9	85.0
DCJ <sup>[35]</sup>	51.4	74.5	83.8	87.0	50.9	68.3	79.4	83.6
ECN-GPP <sup>[5]</sup>	63.8	84.1	92.8	95.4	54.4	74.0	83.7	87.4
TAL-MIRN <sup>[36]</sup>	40.0	73.1	86.3	—	41.3	63.5	76.6	—
ours	64.1	85.6	93.7	96.4	56.1	75.0	83.9	89.0

表2 不同算法在MSMT 17的实验结果

算法	Duke → MSMT 17				Market 1501 → MSMT 17			
	mAP	R-1	R-5	R-10	mAP	R-1	R-15	R-10
ECN <sup>[26]</sup>	10.2	30.2	41.5	46.8	8.5	25.3	36.3	42.1
SSG <sup>[9]</sup>	13.3	32.2	—	51.2	13.2	31.6	—	49.6
NSSA <sup>[30]</sup>	12.1	32.8	43.1	48.3	10.6	28.9	38.2	43.3
ECN-GPP <sup>[5]</sup>	16.0	42.5	55.9	61.5	15.2	40.4	53.1	58.7
MPLP <sup>[34]</sup>	16.2	43.6	54.3	58.9	15.1	40.8	51.8	56.7
TAL-MIRN <sup>[36]</sup>	11.2	30.9	43.5	—	14.2	39.0	51.5	—
ours	17.8	44.1	58.0	63.7	16.3	42.6	55.5	62.0

了0.3%,但是R-1提升了1.5%;在Market 1 501到Duke中,所提出算法的mAP在其基础上提升了1.7%,R-1提升了1.0%。ECN-GPP算法<sup>[5]</sup>在原ECN的基础上结合GPP图神经网络来获得可用于网络训练的正样本,网络复杂度和训练复杂度也相应增加。而所提出算法更关注减弱噪声对聚类结果的影响,且未引入额外计算量。

如表2所示,Duke和Market 1 501数据集自适应到大型真实场景数据集MSMT 17上,目前仍然是一项难度较大的任务。由于MSMT 17数据集包含更多的拍摄视角和多种遮挡情况,使得网络学习条件不太理想,识别精度较低。与表2中性能最优的MPLP方法相比,Duke到MSMT 17实验中mAP提升了1.6%,R-1提升了0.5%;Market 1 501到MSMT 17实验中mAP提升了1.2%,R-1提升了1.8%。

### 3.4 消融实验

为了验证所提出方法的有效性,本节通过充足实验进行消融研究。实验具体安排方式如下:将模型在Duke数据集上预训练,然后进一步在Market 1 501数据集上进行实验。

本文主要在相似度得分的基础上,提出两种伪

标签优化方案。通过实验,评估了两种不同方案对骨干网络的提升效果。表3为不同算法在MSMT 17的实验结果。由表3可见,局部伪标签优化方案和全局伪标签细化方案对网络识别性能均有不同程度的提高。两种方案同时作用,对伪标签的优化能力显著提高:与骨干网相比mAP提高了9.8%,Rank-1提高了7.8%。此外,由表3还可以看出,局部伪标签优化方案较全局细化方案对网络性能提升更明显,验证了无关局部特征会对行人特征聚类产生较大影响。

表3 不同算法在MSMT 17的实验结果

方法	Duke → Market 1 501			
	mAP	R-1	R-5	R-10
Baseline	54.3	77.8	88.4	90.9
Baseline + $L_G$	60.7	83.1	91.3	92.8
Baseline + $L_G$	58.9	82.5	89.5	92.6
ours	64.1	85.6	93.7	96.4

对于第3.2节实验安排中涉及的两个参数 $k$ 和 $\beta$ ,分别对应于相关性得分中全局和局部特征最相似前 $k$ 个取值,以及全局伪标签细化公式中的加权参数。本文采用控制变量法,分别对两个参数进行多组实验,实验结果如图3和图4所示。

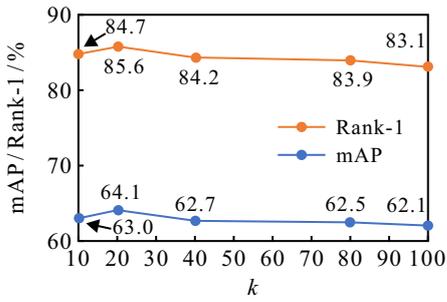


图3 不同k的取值对实验结果的影响

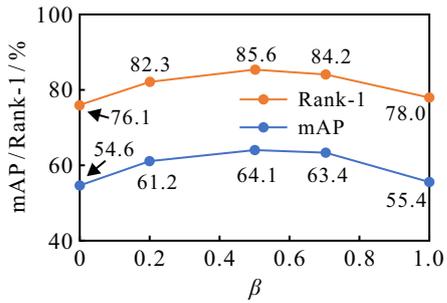


图4 不同beta的取值对实验结果的影响

如图3所示,通过分析可知k取值较大会导致在全局特征与局部特征的相关性匹配过程中,出现包含过多错误匹配的现象,从而降低相关性得分. 这些错误的匹配项会影响网络对全局和局部特征相关性的判断,直接限制了网络性能. 通过多次实验,最终将k的取值确定为20.

通过观察图4,本文将beta在其完整值域内取值:当beta=0时,相当于式(3)中直接忽略了原有全局特征的预测结果,这会导致细化后的伪标签过度关注相关性得分较高的局部特征而完全忽略全局的行人显著性判别特征,属于本末倒置;反之,当beta=1时,伪标签与全局特征预测标签无异. 控制其他变量多次实验后,确定当beta=0.5时网络性能最优.

对于行人图片水平分块的数量n,本文做了相应的消融实验,实验结果如图5所示. 观察可知,不分块的情况网络识别效果最差,网络对行人细粒度特征的捕捉能力不足,仍然依赖于全局特征对行人身份做出判断;反之,过多切分又会导致局部特征被分割,影响对行人身份判别的性能. 通过分析当前公共数据集

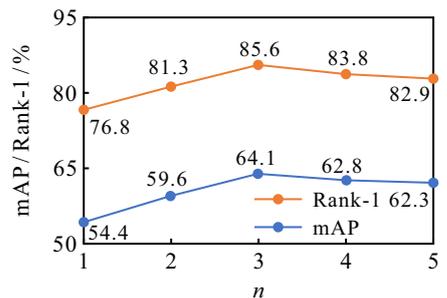


图5 分块数n的取值对实验结果的影响

图片,多数图片通过3等分后可能被分为头部、上半身和下半身3部分,有利于网络对局部特征的学习. 最后,对于输入图片的分块情况,本文将确定其为3.

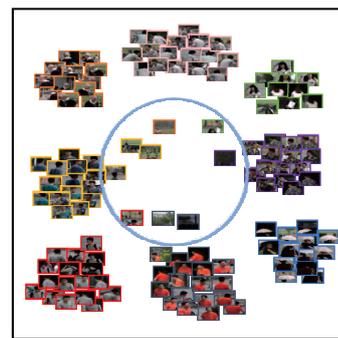
### 3.5 可视化分析

前文对局部伪标签优化和全局伪标签细化方案进行了详细说明和实验论证,为了直观地体现两种方案对伪标签的优化效果,本文采用热图<sup>[38]</sup>(grad-CAM)和空间域可视化<sup>[19]</sup>(t-SNE)方法对实验结果进行展示.

图6用空间域可视化方法<sup>[19]</sup>展示了在主干网基础上有无局部伪标签优化方案的效果. 由图6可见,未使用优化方案(如图6(a)所示)的局部特征出现了与其身份标签过度拟合的情况,与行人无关的特征混杂在局部特征的正样本中. 使用优化方案后(如图6(b)所示),在空间域内,相关性得分使得局部无关特征远离身份簇,降低了网络学习无关特征的可能.



(a) 局部伪标签未优化



(b) 局部伪标签优化后

图6 局部伪标签优化方案的有效性可视化

图7利用热图<sup>[38]</sup>的方法,直观地展现了全局伪标签细化方案对主干网学习特征能力的影响. 图7中包含两组样例图片,每组图片中依次展示了原图、未使用优化方案和使用优化方案的热图. 由图7可见,依靠全局特征训练的网络关注行人显著性特征(如背包、配饰等),且易受到视点变化和遮挡的影响. 而图7中使用优化方案的热图表现了全局伪标签细化方案利用分块内学习到的细粒度特征(如衣服纹理等)对全局特征进行丰富,进一步完善了行人的特征表示.

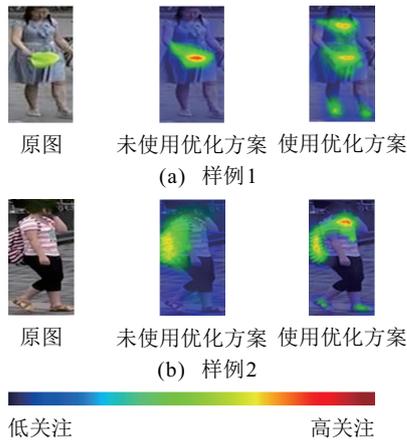


图7 全局伪标签细化方案的有效性可视化

## 4 结论

本文提出了一种基于相关性得分的伪标签优化无监督域自适应网络,网络利用全局与局部特征间的相关性得分,对全局和局部伪标签分别做出了相应的优化处理.减少对局部特征集合中与行人无关特征的关注,可降低其对网络聚类的影响,使得网络发挥其应有的性能;增加了全局特征提取所忽略的局部细粒度特征,有利于对行人的精确分类.实验结果表明:结合两种伪标签优化方案,网络在主干网的基础上mAP提高了9.8%,Rank-1提高了7.8%.与近年域自适应算法相比验证了所提出方法的有效性.在后续的改进中,考虑通过将主干网替换为Transformer来增强网络模型的跨域识别能力和并行运行效率.

## 参考文献(References)

- [1] 刘月, 邹国锋, 陈贵震, 等. 基于分布约束的非对称度量学习无监督行人重识别[J]. 控制与决策, 2023, 38(6): 1703-1711.  
(Liu Y, Zou G F, Chen G Z, et al. Asymmetric metric learning approach based on distribution constraints for unsupervised person re-identification[J]. Control and Decision, 2023, 38(6): 1703-1711.)
- [2] 许文正, 黄天欢, 贲焜焜, 等. 跨视角步态识别综述[J]. 中国图象图形学报, 2023, 28(5): 1265-1286.  
(Xu W Z, Huang T H, Ben X Y, et al. Cross-view gait recognition: A review[J]. Journal of Image and Graphics, 2023, 28(5): 1265-1286.)
- [3] 寇旗旗, 黄绩, 程德强, 等. 基于语义融合的域内相似性分组行人重识别[J]. 通信学报, 2022, 43(7): 153-162.  
(Kou Q Q, Huang J, Cheng D Q, et al. Person re-identification with intra-domain similarity grouping based on semantic fusion[J]. Journal on Communications, 2022, 43(7): 153-162.)
- [4] 张再腾, 张荣芬, 刘宇红. 基于多尺度残差注意网络的轻量级行人属性识别算法[J]. 控制与决策, 2022, 37(10): 2487-2496.  
(Zhang Z T, Zhang R F, Liu Y H. Lightweight pedestrian attribute recognition algorithm based on multi-scale residual attention network[J]. Control and Decision, 2022, 37(10): 2487-2496.)
- [5] Zhong Z, Zheng L, Luo Z M, et al. Learning to adapt invariance in memory for person re-identification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(8): 2723-2738.
- [6] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]. Proceedings of the 27th International Conference on Neural Information Processing Systems. New York: ACM, 2014: 2672-2680.
- [7] Wei L H, Zhang S L, Gao W, et al. Person transfer GAN to bridge domain gap for person re-identification[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 79-88.
- [8] Dai Y X, Liu J, Sun Y F, et al. IDM: An intermediate domain module for domain adaptive person re-ID[C]. IEEE/CVF International Conference on Computer Vision. Montreal, 2022: 11844-11854.
- [9] Fu Y, Wei Y C, Wang G S, et al. Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification[C]. IEEE/CVF International Conference on Computer Vision. Seoul, 2020: 6111-6120.
- [10] Ma A J, Yuen P C, Li J W. Domain transfer support vector ranking for person re-identification without target camera label information[C]. IEEE International Conference on Computer Vision. Sydney, 2014: 3567-3574.
- [11] Zhao F, Liao S C, Xie G S, et al. Unsupervised domain adaptation with noise resistible mutual-training for person re-identification[C]. Computer Vision-ECCV 2020: The 16th European Conference. Cham: Springer International Publishing, 2020: 526-544.
- [12] Ge Y, Zhu F, Chen D, et al. Self-paced contrastive learning with hybrid memory for domain adaptive object re-id[J]. Advances in Neural Information Processing Systems, 2020, 33: 11309-11321.
- [13] Zheng K C, Liu W, He L X, et al. Group-aware label transfer for domain adaptive person re-identification[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, 2021: 5306-5315.
- [14] Wang G S, Yuan Y F, Chen X, et al. Learning discriminative features with multiple granularities for person re-identification[C]. Proceedings of the 26th ACM International Conference on Multimedia. Seoul, 2018: 274-282.
- [15] Yang Q Z, Yu H X, Wu A C, et al. Patch-based discriminative feature learning for unsupervised person re-identification[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, 2020: 3628-3637.
- [16] Deng J, Dong W, Socher R, et al. ImageNet: A large-scale hierarchical image database[C]. IEEE

- Conference on Computer Vision and Pattern Recognition. Miami, 2009: 248-255.
- [17] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 770-778.
- [18] Ester M, Kriegel H, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]. KDD'96: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining. New York: ACM, 1996: 226-231.
- [19] van Der M L, Hinton G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9(11): 2579-2605.
- [20] Zheng Z D, Zheng L, Yang Y. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro[C]. Proceedings of the IEEE International Conference on Computer Vision. Venice, 2017: 3774-3782.
- [21] Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification[J/OL]. 2017, arXiv: 1703.07737.
- [22] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification: A benchmark[C]. Proceedings of the IEEE International Conference on Computer Vision. Santiago, 2016: 1116-1124.
- [23] Qi L, Wang L, Huo J, et al. A novel unsupervised camera-aware domain adaptation framework for person re-identification[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2020: 8079-8088.
- [24] Wu J L, Liao S C, Lei Z, et al. Clustering and dynamic sampling based unsupervised domain adaptation for person re-identification[C]. IEEE International Conference on Multimedia and Expo. Piscataway: IEEE, 2019: 886-891.
- [25] Zhong Z, Zheng L, Luo Z M, et al. Invariance matters: Exemplar memory for domain adaptive person re-identification[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 598-607.
- [26] Yu H X, Zheng W S, Wu A C, et al. Unsupervised person re-identification by soft multilabel learning[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 2143-2152.
- [27] Zhang X Y, Cao J W, Shen C H, et al. Self-training with progressive augmentation for unsupervised cross-domain person re-identification[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2020: 8221-8230.
- [28] Song L C, Wang C, Zhang L F, et al. Unsupervised domain adaptive re-identification: Theory and practice[J]. Pattern Recognition, 2020, 102: 107173.
- [29] Zhao Y R, Lu H T. Neighbor similarity and soft-label adaptation for unsupervised cross-dataset person re-identification[J]. Neurocomputing, 2020, 388: 246-254.
- [30] Jiang K Z, Zhang T Z, Zhang Y D, et al. Self-supervised agent learning for unsupervised cross-domain person re-identification[J]. IEEE Transactions on Image Processing, 2020, 29: 8549-8560.
- [31] Ji Z L, Zou X L, Lin X H, et al. An attention-driven two-stage clustering method for unsupervised person re-identification[C]. Computer Vision-ECCV 2020: The 16th European Conference. Cham: Springer International Publishing, 2020: 20-36.
- [32] Yang F X, Li K, Zhong Z, et al. Asymmetric co-teaching for unsupervised cross-domain person re-identification[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12597-12604.
- [33] Wang D K, Zhang S L. Unsupervised person re-identification via multi-label classification[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 10978-10987.
- [34] Ge Y, Liu L, Zhang H X. A three-stage learning approach to cross-domain person re-identification[J]. Applied Soft Computing, 2021, 112: 107793.
- [35] Li H F, Dong N, Yu Z T, et al. Triple adversarial learning and multi-view imaginative reasoning for unsupervised domain adaptation person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(5): 2814-2830.
- [36] Luo H, Gu Y Z, Liao X Y, et al. Bag of tricks and a strong baseline for deep person re-identification[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2020: 1487-1495.
- [37] Zhao R, Ouyang W L, Wang X G. Person re-identification by saliency matching[C]. IEEE International Conference on Computer Vision. Piscataway: IEEE, 2014: 2528-2535.
- [38] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization[C]. Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 618-626.

## 作者简介

程德强(1979—), 男, 教授, 博士生导师, 从事图像处理、模式识别、图像智能检测与信息处理等研究, E-mail: chengdq@cumt.edu.cn;

黄绩(1995—), 男, 硕士生, 从事图像处理与模式识别的研究, E-mail: huangji@cumt.edu.cn;

寇旗旗(1988—), 男, 讲师, 博士, 从事图像处理与模式识别、计算机视觉等研究, E-mail: kouqiqi@cumt.edu.cn;

张剑英(1964—), 女, 教授, 博士, 从事图像处理与模式识别、电磁场理论及应用等研究, E-mail: zjycumt@126.com;

李云龙(1997—), 男, 硕士生, 从事图像处理与模式识别的研究, E-mail: li330240647@163.com.