

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2023)05-1372-12

论文引用格式: Wu Y H and Sang N. 2023. Consistency constraints and label optimization-relevant domain-unsupervised adaptive pedestrians' re-identification. Journal of Image and Graphics, 28(05):1372-1383(吴禹航, 桑农. 2023. 基于一致性约束和标签优化的无监督域适应行人重识别. 中国图象图形学报, 28(05):1372-1383)[DOI:10.11834/jig.220618]

# 基于一致性约束和标签优化的无监督域适应行人重识别

吴禹航, 桑农\*

华中科技大学人工智能与自动化学院图像信息处理与智能控制教育部重点实验室, 武汉 430074

**摘要:** 目的 无监督域适应行人重识别(unsupervised domain adaptive pedestrians' re-identification, UDA Re-ID)旨在通过已有标记的应用场景(即源域)数据和新的无标记应用场景(即目标域)数据, 训练一个可以在目标域泛化性能好的行人重识别模型。现有方法没有考虑实例特征在训练过程中的不稳定问题, 也没有显式考虑由于相机变化所导致的行人类内距离变大、类间距离变小的问题, 以及无标注目标域数据聚类误差带来的伪标签噪声问题。针对这些问题, 提出了一种具有一致性约束和标签优化的方法。方法 首先提出了实例一致性以约束同一实例在不同增广下的特征距离, 提升行人实例特征稳定性; 然后提出相机一致性以约束跨相机正实例特征对之间的距离, 提升对相机变化的鲁棒性; 最后提出了基于标签集成的标签优化, 将one-hot编码的伪标签转换为更可靠的软标签, 提升了监督信号的鲁棒性。结果 本文方法在 Duke→Market, Market→Duke, Duke→MSMT, Market→MSMT 等常用的 UDA Re-ID 任务上的平均精度均值(mean average precision, mAP)分别为 85.0%, 73.5%, 41.3%, 39.3%; Rank-1 分别为 94.0%, 85.6%, 71.6%, 69.5%。通过消融实验验证了本文提出的3个模块的有效性。结论 提出的实例一致性约束和相机一致性约束可以使模型学习到更鲁棒的行人特征表达, 提升行人特征的稳定性, 提出的基于标签集成的标签优化可以减少伪标签噪声的过拟合风险。

**关键词:** 计算机视觉; 行人重识别(Re-ID); 无监督域适应(UDA); 一致性约束; 标签优化

## Consistency constraints and label optimization-relevant domain-unsupervised adaptive pedestrians' re-identification

Wu Yuhang, Sang Nong\*

Key Laboratory of Ministry of Education for Image Processing and Intelligent Control, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China

**Abstract:** **Objective** Pedestrians' re-identification (Re-ID) can be as one of key techniques for multi-camera pedestrians' retrieval in the video surveillance system. To achieve good performance, current Re-ID model is often trained on a scene based on a large number of annotations (source domain), but the performance will be dropped significantly when a new scene (target domain) is applied straightforward. However, re-labeling is time-consuming and labor-intensive for the

收稿日期: 2022-06-24; 修回日期: 2022-11-26; 预印本日期: 2022-12-03

\* 通信作者: 桑农 nsang@hust.edu.cn

**基金项目:** 国家自然科学基金项目(61876210); 中央高校基本科研业务费专项资金资助(2019kfyXKJC024); 高等学校学科创新引智计划项目(B18024)

**Supported by:** National Natural Science Foundation of China (61876210); Fundamental Research Funds for the Central Universities (2019kfyXKJC024); 111 Project on Computational Intelligence and Intelligent Control (B18024)

new scene, which is beneficial for Re-ID-related optimization. The unsupervised domain adaptive pedestrians' re-identification (UDA Re-ID) method is focused on a model training, which can be generalized on the target domain well using the existing source domain data-labeled and target domain data-unlabeled. But, these methods are still challenged for the instability of instance features and image heterogeneity of the intra-class distance-wider and the inter-class distance-narrowed. Furthermore, current cluster unlabeled target domain data can be melted into multiple clusters, and the encoded pseudo labels can be assigned for each cluster. However, due to the limited representation ability of the model, the clustering results are incredible, especially in the early stage of training. One pedestrian-related image is grouped into different clusters while some images of different pedestrian are merged into a cluster, called pseudo label noises. However, it is still challenged for the problem of pseudo label noises-over-fitted and the performance of the model-suppressed although pseudo labels are recognized as a supervision signal for the feature learning process (e. g., contrastive learning). To resolve these problems, we develop a multi-centroid representation network with consistency constraints method (MCRNCC) in terms of the popular multi-centroid representation network method (MCRN). **Method** The MCRNCC is designed on the basis of three MCRN-related modules to improve the stability of instance features and the robustness of pedestrian features, and the overfitting risk of the pseudo-label noise can be reduced. First, to optimize the instance feature stability and semantic information, an instance-consistent is demonstrated to suppress the feature distance of the same instance under different augmentation. The exponential moving average model is illustrated to output additional features based on recent self-supervised learning works. For each image of the training batch, it can be augmented twice in random, the features can be extracted in relevance to original model and exponential moving average model, and cosine distance is used to constrain the feature pairs. Second, to improve robustness of multiple variations-captured, its homogeneity is concerned for suppressing the distance between feature pairs of positive instances. Specifically, two instances are opted to construct a positive pair in the context of same labels-without identity label, and two instances is opted as well to construct a negative pair in related to same label-within multiple identity labels, and a triplet is built up to optimize the network as well. Finally, the label-ensemble-based optimization is carried out to convert one-hot encoded pseudo labels into more reliable soft labels, which improves the robustness of supervision signals. In detail, we add a target domain classifier to generate additional label predictions, followed by linearly weighting the predictions and one-hot encoded pseudo labels into refined soft labels. **Result** To verify the effectiveness of our method, adequate experiments are carried out on 4 popular UDA Re-ID tasks like 1) Duke→Market, 2) Market→Duke, 3) Duke→MSMT, and 4) Market→MSMT tasks. At the beginning, the ablation studies are carried out about the modules in MCRNCC. The four tasks-derived instance consistency constraints can be reached to 0.6%, 0.2%, 0.7% and 0.8% of each mean average precision (mAP), which demonstrates the effectiveness of the instance consistency constraint. The camera consistency constraint yields a general improvement for all of 4 tasks. For example, mAP/Rank-1 is increased by 3.5%/3.2% and 5.3%/4.7% on Duke→MSMT and Market→MSMT. In addition, we visualize the feature space of adding camera consistency constraint before and after. Furthermore, we compare the feature space of some pedestrian-focused close to the camera consistency constraint, and the visualization results show that the camera consistency can make the feature space more compactible. The label-ensemble-optimized can be improved to 0.6%, 0.6%, 1.4% and 0.4% of the mAP for each 4 tasks. Second, our proposed MCRNCC is compared to the existing methods. The comparative analysis shows that the MCRNCC can be reached to 85.0%/94.0%, 73.5%/85.6%, 41.3%/71.6% and 39.3%/69.5% for the optimization of mAP/Rank-1 performance, and the MCRN is surpassed by 1.2%/0.2%, 2.0%/1.1%, 5.6%/4.1%, and 6.5%/5.1% as well. **Conclusion** we develop a method MCRNCC to resolve the UDA Re-ID problem further. The instance consistency constraint and camera consistency constraint proposed in MCRNCC can enable the model to learn more robust pedestrian-related feature representations, while the proposed label ensemble-based optimization can reduce the overfitting risks of pseudo label noises. Experiments show that the effectiveness of three-module based MCRNCC has its potentials for future works.

**Key words:** computer vision; pedestrians' re-identification(Re-ID); unsupervised domain adaption(UDA); consistency constraint; label refinery

## 0 引言

随着经济和社会的发展,我国的视频监控和海量数据检索等计算机技术得到了飞速发展。通过对监控视频数据进行分析处理,可以检索出特定的行人,提升城市的治理水平。目前,由于安全城市和智慧城市的建设需要,视频监控的设备数量和数据规模迅速增加,使得依靠人力手动进行数据处理和行人检索的方法的成本和难度大幅提高,已经无法适应视频监控系统的更大规模推广。行人重识别(pedestrians' re-identification, Re-ID)是一项在视频监控系统中通过计算机视觉技术在监控视频数据中进行跨摄像头行人检索的技术。给定一幅由摄像头采集的待查询行人图像,行人重识别从其他摄像头采集的行人图像库里检索出属于该行人的所有图像。由于行人重识别主要依靠计算机算法程序自动检索,相比人工检索的方式,可以很大程度上减少视频监控系统中的人工成本和检索难度,具有重要的实际应用价值。

目前,在一个具有大量标注数据的应用场景(源域)下,基于深度学习的行人重识别算法(史维东等,2020;郑鑫等,2020)可以使用有监督的方式训练一个模型,使该模型在该场景下能达到应用水平。然而,将这个模型直接应用目标域(即新的应用场景)上时,由于场景之间存在着领域偏移(domain shift),其在目标域上的性能会急剧下降。在实际应用中,仍存在着大量的目标域需要应用行人重识别以解决实际需求,然而为每个目标域人工标注训练数据费时费力,使得行人重识别的大规模推广受到限制。全面推广行人重识别,需要解决算法对目标域标注数据的依赖问题。使用目标检测算法(Ren等,2015;Tian等,2019)对目标域场景图像进行行人检测可以容易地得到大量无标注的行人图像。如何利用这些无标注数据来帮助模型在目标域取得良好性能成为了行人重识别领域在实际应用中亟待解决的问题。本文研究的无监督域适应行人重识别(unsupervised domain adaptive Re-ID, UDA Re-ID)算法旨在利用已有的有标记源域行人图像和无标记目标域行人图像训练模型,使模型在目标域上取得较好的泛化性能。相比于有监督训练的方法,无监督域适应的方法不需要额外的人工标注成本,具有更

重要的实际意义。

现有无监督域适应行人重识别方法主要分为两类,即基于域转换的方法(Deng等,2018;Wei等,2018)和基于聚类的方法(Ge等,2020b;Wu等,2022)。基于域转换的方法通常借助生成对抗网络(generative adversarial network, GAN)(Goodfellow等,2014)转换源域图像的风格到目标域,然后在源域标签的监督下,使用原始图像特征和风格转换后的图像特征计算损失函数并优化模型。为了提升风格转换的图像质量,保留相似性的生成对抗网络(similarity preserving GAN, SPGAN)引入语义分割来约束风格迁移过程中人体区域的一致性,行人转换生成对抗网络(person transfer GAN, PTGAN)(Wei等,2018)考虑了图像风格转换前后的自相似性,也考虑了目标域图像与转换后的源域图像之间的差异性。然而,这类方法生成的图像质量仍不够高。此外,真实场景的目标域风格复杂而多变,通过GAN获得的生成图像并不能真正服从目标域的真实数据分布,对模型的学习帮助有限,使得这类方法难以取得较好的性能表现,所以目前学术界主要研究基于聚类的方法。基于聚类的方法,如自步学习网络(self-paced contrastive learning, SpCL)(Ge等,2020b)和多中心表征网络(multi-centroid representation network, MCRN)(Wu等,2022),采用准备阶段和优化阶段的两阶段交替训练。在准备阶段(每个epoch的初始时刻),这类方法使用聚类算法将无标记目标域训练样本聚类为若干簇,为每个簇分配一个one-hot编码的伪标签。在优化阶段(每个epoch的剩余时刻),这类方法将伪标签作为监督信号,计算损失函数并优化模型。这类方法取得了不错的效果,但存在一些问题有待改进。1)由于模型的表征能力有限,实例特征在训练过程中并不稳定(特别在网络训练的早期),现有方法常常使用数据增强策略学习一个鲁棒的特征,但很少从实例自身挖掘有用信息;2)同一行人在不同相机下拍摄的图像的特征存在较大差异,若不对其进行约束,模型对相机的变化不鲁棒,难以学习到一个类内紧凑性较好的特征空间;3)one-hot编码的伪标签噪声过拟合问题。由于聚类的结果并不可靠,伪标签中存在噪声,即属于相同行人的实例可能被聚到不同簇中而被分配不同的伪标签,或者属于不同行人的实例被聚到同一个簇里而被分配相同的伪标签。在模型的训练过程中,使用

这种 one-hot 编码的伪标签作为特征学习(如对比学习)的监督信号容易使模型对噪声过拟合。

为此,本文提出了实例一致性约束、相机一致性约束和基于标签集成的标签优化这3个模块,可以分别解决行人的实例特征不稳定问题、行人特征对相机变化的不鲁棒问题以及伪标签噪声的过拟合问题,并将这3个模块用于基于聚类的方法 MCRN (Wu 等, 2022)中,得到具有一致性约束的多中心表征网络 (multi-centroid representation network with consistency constraints, MCRNCC)。本文的主要贡献如下: 1) 针对行人的实例特征不稳定问题,在现有方法 MCRN 的基础上,引入实例一致性约束。实例一致性约束通过约束同一实例在不同增广下的特征相近,从实例内部挖掘更丰富的信息,学习到更鲁棒的特征表达。2) 针对行人特征对相机变化的不鲁棒问题,进一步引入相机一致性约束。相机一致性约束以跨相机的正样本对和同相机的负样本对构建难例三元组,提升了行人在不同相机下的特征稳定性。3) 针对伪标签噪声的过拟合问题,引入了基于标签集成的标签优化。基于标签集成的标签优化引入了额外的标签估计,通过将多组额外的标签与聚类生成的 one-hot 编码伪标签集成,减少了伪标签噪声的过拟合风险。

## 1 多中心表征网络

本文的基准网络 MCRN (Wu 等, 2022) 的整个训练过程采用两阶段(准备阶段和优化阶段)交替进行的策略。

### 1.1 准备阶段

在每个准备阶段(每个 epoch 的初期), MCRN 生成目标域无标记图像的伪标签,并初始化多中心记忆体 (multi-centroid memory, MCM)。首先,对于源域数据集  $D_s = \{(x_s^i, y_s^i) | i=1, \dots, N_s\}$  和目标域数据集  $D_t = \{(x_t^i) | i=1, \dots, N_t\}$ , 使用主干网络分别提取所有源域训练集和目标域训练集上的图像的特征  $F_s = \{f_s^i | i=1, \dots, N_s\}$  和  $F_t = \{f_t^i | i=1, \dots, N_t\}$ 。然后,使用 DBSCAN (density-based spatial clustering of applications with noise) 算法 (Ester 等, 1996) 对  $F_t$  进行聚类,并为每个簇分配一个伪标签。对于没有聚类上的离群点,因为这些离群点不能用多个中心表示,将未聚上类的离群点丢弃。

MCM 中每个类别的中心数设置为  $K$ , 若源域的真实类别数为  $n_s$ , 当前目标域类别(即簇)数为  $n_t$ , 则构建的多中心记忆体为  $K(n_s + n_t) \times C$ , 其中  $C$  为主干网络输出的特征维度。最后,根据源域真值标签与目标域伪标签,求得每个类别的所有实例的平均特征,并用其初始化其类别对应的  $K$  个中心。

### 1.2 优化阶段

MCRN 在优化阶段的网络框架如图 1 所示。在每个优化阶段(每个 epoch 的剩余时间)的每次迭代(iteration)中, MCRN 首先分别从源域和目标域各采样  $B = P \times K$  个查询样本构成训练批, 其中  $P$  代表类别数,  $K$  代表每个类别的样本数, 设置其等于每个类别的中心数。然后, 对于每个训练批中的查询样本, MCRN 从多中心记忆体中的  $K$  个正类中心中, 选择与查询样本中相似度排在中间的中心作为后续对比学习中的正样本, 选择每个负类的  $K$  个负类中心的平均特征作为后续对比学习中的负样本。此外, MCRN 使用二阶近邻插值为目标域的查询样本  $q_i^j$ , 生成  $\lambda$  个额外的难负样本  $\{s^{i,l} | l=1, \dots, \lambda\}$ 。最后, MCRN 使用域分离对比学习 (domain-specific contrastive learning, DSCL) 优化网络。具体为

$$L_{\text{DSCL}, D_t} = -\frac{1}{B} \sum_{i=1}^B \log \frac{\exp\left(q_i^i \cdot \frac{c^+}{\tau}\right)}{\sum_{j=1}^{n_t} \exp\left(q_i^i \cdot \frac{c_s^j}{\tau}\right)} \quad (1)$$

$$L_{\text{DSCL}, D_t} = -\frac{1}{B} \sum_{i=1}^B \log \bar{p}^{ij} \quad (2)$$

$$\bar{p}^{ij} = \frac{\exp\left(q_i^i \cdot \frac{c_t^j}{\tau}\right)}{\sum_{k=1}^{n_t} \exp\left(q_i^i \cdot \frac{c_t^k}{\tau}\right) + \sum_{l=1}^{\lambda} \exp\left(q_i^i \cdot \frac{s^{i,l}}{\tau}\right)} \quad (3)$$

式中,  $n_s$  和  $n_t$  分别为源域类别数和目标域簇数,  $q_s^i$  和  $q_t^i$  分别为源域训练批和目标域训练批的第  $i$  个查询样本,  $\tau$  为温度系数。

在每次迭代结束后, MCRN 计算训练批中每个采样类别的  $K$  个查询样本和多中心记忆体中对应的  $K$  个正类中心之间的相似度, 根据整体相似度最高的查询样本—正类中心对的排列  $\hat{\sigma}$  更新中心。具体为

$$c^{\hat{\sigma}(i)} \leftarrow m c^{\hat{\sigma}(i)} + (1 - m) q^i \quad (4)$$

式中,  $m$  为动量系数,  $c^{\hat{\sigma}(i)}$  为处于排列  $\hat{\sigma}$  的第  $i$  个正类中心。



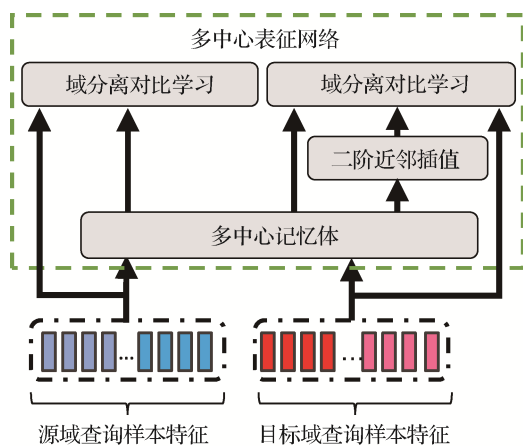


图1 MCRN框架(优化阶段)

Fig. 1 The framework of MCRN in optimization stage

## 2 具有一致性约束的多中心表征网络

本文在基准网络的基础上,提出了具有一致性约束的多中心表征网络 MCRNCC,其在优化阶段的网络框图如图2所示,包含实例一致性约束、相机一致性约束和基于标签集成的标签优化等3个模块。图2中sg代表中断梯度反传(stop gradient)。

### 2.1 实例一致性约束

由于网络本身的表达能力有限,图像的实例特征在训练过程中并不稳定,尤其在网络训练的前期,目标在图像上的一点偏移也会在特征上产生波动。行人重识别任务中通常使用了随机擦除(Zhong等, 2020)、随机翻转和随机裁剪等数据增广策略提高特征的鲁棒性,但未从实例自身挖掘有用的信息。

基于自监督的表征学习研究(Chen等, 2020; He等, 2020; Grill等, 2020)通过挖掘实例内部信息

以学习容易泛化到下游任务的特征表达。这些方法通常会将同一实例在不同增广下的多幅图像的特征构成实例级别的正样本对,然后通过损失函数约束正样本间的距离相近,以从实例内部挖掘语义信息。由于在自监督学习中没有额外的学习目标,仅使用正样本对的约束会导致网络塌陷。为了避免这一问题,SimCLR(simple framework for contrastive learning of visual representation)(Chen等, 2020)和 MoCO(momentum contrast)(He等, 2020)等考虑了将不同实例的样本作为负样本,即同时考虑实例级别的正样本对和负样本对的约束。BYOL(bootstrap your own latent)(Grill等, 2020)则引入了一个非对称映射头(多层感知机)以引入非对称性避免网络塌陷问题,这样就只需考虑正样本对的约束。为了从实例正样本对内部挖掘语义信息,提升实例特征的稳定性,本文借鉴自监督方法,提出了实例一致性约束,并将其应用于MCRN网络中。与已有自监督学习方法不同的是,由于MCRN的损失函数部分已经包含了区分类别级别的负样本对的部分,所以本文在不引入非对称映射头的条件下仅考虑实例级别正样本对的约束,也不会有网络坍塌的问题。

#### 2.1.1 指数滑动平均模型

与 MoCO 和 BYOL 相同,引入了一个指数滑动平均(exponential moving average, EMA)模型  $M_{ema}$ , 该 EMA 模型  $M_{ema}$  的参数  $\theta_{ema}$  不根据梯度优化,而是由原模型  $M_{ema}$  的参数  $\theta$  通过指数滑动平均得到,即

$$\theta_{ema} \leftarrow \varphi \theta_{ema} + (1 - \varphi) \theta \quad (5)$$

式中,  $\varphi$  是控制参数移动的超参数,通常设置为 0.99。通过引入 EMA 模型,可以得到两个参数不一

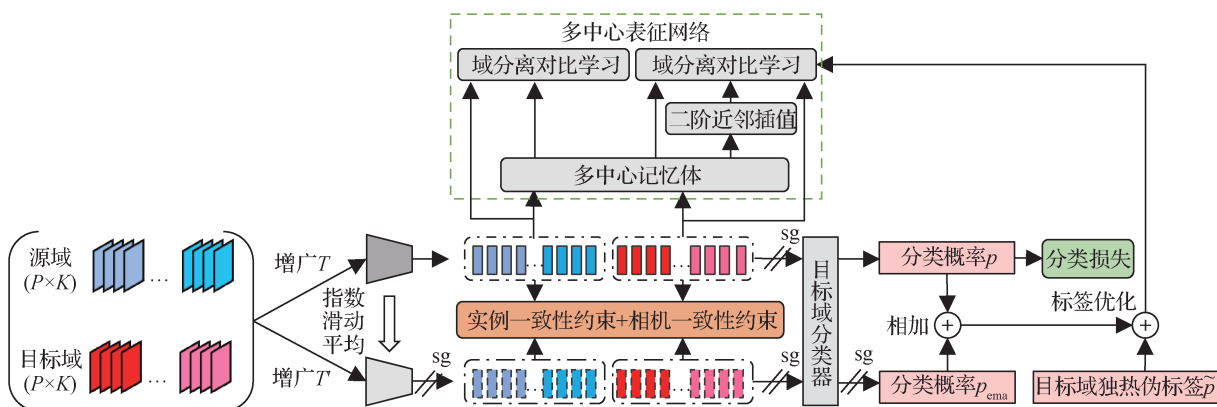


图2 MCRNCC框架(优化阶段)

Fig. 2 The framework of MCRNCC (optimization stage)

样的网络,使得输出的特征具有多样性,同时EMA模型由于在时间上集成了不同时期的模型参数,相对于普通的模型,参数过渡得更平滑,性能通常更好,因此可以在测试阶段使用EMA模型进行测试。

### 2.1.2 实例一致性约束损失函数

实例一致性(instance consistency, IC)约束如图3所示,将训练批中的实例(即查询样本)在不同增广下的特征距离进行约束,即实例 $I$ 经增广 $T$ 和 $T'$ 后得到 $T(I)$ 和 $T'(I)$ ,再分别经模型 $M$ 和EMA模型 $M_{\text{ema}}$ 后得到的特征 $q$ 和 $q_{\text{ema}}$ 的距离要小。具体的损失函数为

$$L_{IC,D_i} = \frac{1}{B} \sum_{i=1}^B (1 - q_s^i \cdot q_{\text{ema},s}^i) \quad (6)$$

$$L_{IC,D_i} = \frac{1}{B} \sum_{i=1}^B (1 - q_t^i \cdot q_{\text{ema},t}^i) \quad (7)$$

通过实例一致性约束,模型可以挖掘更丰富的语义信息,提升模型的性能,后续实验验证了实例一致性约束的有效性。

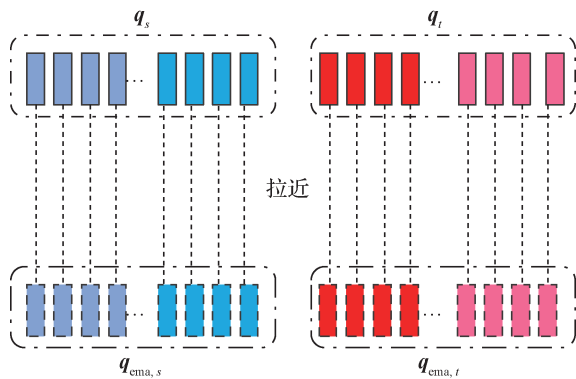


图3 实例一致性约束

Fig. 3 Instance consistency constraint

### 2.2 相机一致性约束

由于行人在不同相机下拍摄的图像往往存在较大的变化,这些差异由光照、遮挡、分辨率和视角等场景因素决定。此外,由于模型的表征能力有限,同一行人在不同相机下的实例特征可能相差较大,而不同行人在同一相机下的实例特征可能相差较小。为了提升行人特征对相机变化的鲁棒性,本文提出了相机一致性(camera consistency, CC)约束,即约束同一行人在不同相机下的实例特征距离近,而不同行人在同一相机下的实例特征距离远。

相机一致性约束依次选取当前训练批中模型 $M$ 提取的实例特征作为锚点 $a$ ,然后根据伪标签和相

机标签,构建 $a$ 的难样本三元组损失来优化网络。由于在实例一致性约束中,EMA模型已经额外产生了训练批大小的特征,为了充分利用多样的特征,在构建三元组时,每个锚点 $a$ 会按2批次,分别从模型 $M$ 的训练批同域特征集合 $\{q|_{i=1}^B\}$ 中和EMA模型 $M_{\text{ema}}$ 的训练批同域特征集合 $\{q_{\text{ema}}^i|_{i=1}^B\}$ 中构建一次三元组。图4展示了一个锚点 $a$ 构建难样本三元组的示例。锚点 $a$ 的正样本选取为当前批次的特征集合中与锚点 $a$ 伪标签相同、但相机标签不同的样本中距离最远的样本,而负样本选取为当前批次的特征集合中与锚点 $a$ 标签不同、但相机标签相同的样本中距离最近的样本。若无法找到满足条件的正样本或负样本,则认为在这一批次中的这个锚点无效,放弃该三元组的构建。相机一致性约束的具体损失函数为

$$L_{CC,D_i} = \frac{1}{B_1} \sum_{i=1}^{B_1} \max(d(\alpha_s^i, p^i) - d(\alpha_s^i, n^i) + \gamma, 0) + \frac{1}{B_2} \sum_{j=1}^{B_2} \max(d(\alpha_s^j, p_{\text{ema}}^j) - d(\alpha_s^j, n_{\text{ema}}^j) + \gamma, 0) \quad (8)$$

$$L_{CC,D_i} = \frac{1}{B_3} \sum_{i=1}^{B_3} \max(d(\alpha_t^i, p^i) - d(\alpha_t^i, n^i) + \gamma, 0) + \frac{1}{B_4} \sum_{j=1}^{B_4} \max(d(\alpha_t^j, p_{\text{ema}}^j) - d(\alpha_t^j, n_{\text{ema}}^j) + \gamma, 0) \quad (9)$$

式中, $\gamma$ 代表三元组的阈值, $d(\alpha_s^i, p^i)$ 和 $d(\alpha_s^i, n^i)$ 分别代表第1批次中,源域第 $i$ 个有效锚点 $\alpha_s^i$ 与在模型 $M$ 的训练批同域特征集合中挑选的难正样本的余弦距离和难负样本的余弦距离,而 $d(\alpha_s^j, p_{\text{ema}}^j)$ 和 $d(\alpha_s^j, n_{\text{ema}}^j)$ 分别代表第2批次中,源域第 $j$ 个有效锚点 $\alpha_s^j$ 与EMA模型 $M_{\text{ema}}$ 的训练批同域特征集合中挑选

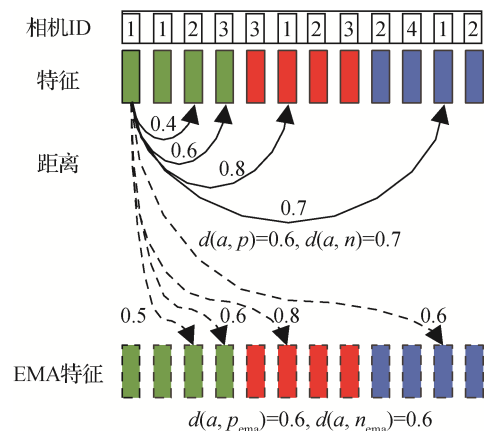


图4 相机一致性中的三元组示例

Fig. 4 An example of triplet in camera consistency

的难正样本的余弦距离和难负样本的余弦距离。目标域的符号含义同理可知,不再赘述。 $B_1, B_2, B_3, B_4$ 分别代表各域各批次中有效的锚点数。

通过相机一致性约束同相机和跨相机下的行人的特征关系,可提升行人特征对相机变化的鲁棒性,并学习到一个具有良好类内紧凑性和类间可分性的特征空间。后续实验验证了相机一致性约束的有效性。

### 2.3 基于标签集成的标签优化

SpCL(Ge 等, 2020b)和MCRN(Wu 等, 2022)等方法使用了在目标域上聚类得到的 one-hot 编码伪标签作为监督信号,指导模型通过对比学习(式(2))优化特征空间。正如前文所提及,由于模型的表现能力有限以及聚类算法的不完美,聚类生成的 one-hot 编码伪标签中存在噪声,即一些属于不同行人的图像被赋予了相同的伪标签,而一些属于相同行人的图像被赋予了不同的伪标签。在网络训练过程中,模型的表现能力会不断提升,本可以自我矫正一部分聚类错误,但由于每个训练周期的监督信号是这个训练周期初期聚类得到的伪标签,模型仍会根据有噪声的监督信号继续拟合,从而阻止了模型的性能提升。为了减少模型对伪标签噪声的过度拟合,本文提出了一种基于标签集成(label ensemble, LE)的标签优化(label refinery, LR)方法,通过将多组预测标签与 one-hot 编码的伪标签集成,使硬的 one-hot 编码伪标签转化为鲁棒的软标签,进一步改善特征学习过程。

#### 2.3.1 额外的预测标签

为了获得预测标签,一个自然的想法是引入一个新的分类器,用分类器的分类概率作为预测标签。这里分类器用一个去掉偏置向量的全连接(fully connected, FC)层实现,其权重可以看做是一组可学习的类中心。

由于分类器需要固定的类别数,而目标域的类别数是由每个训练周期的聚类结果决定,其在不同的训练周期数量不一样。为了解决这个问题,在每个训练周期聚类完成后重新生成一个分类器(分类数为簇的数量),然后用每个簇的均值初始化分类器的权重。

对于训练批中第  $i$  个目标域查询样本特征  $q_i^i$ , 其估计的标签为  $p^i$ , 具体为

$$p^i = \text{softmax}\left(\frac{FC(q_i^i)}{\tau}\right) \quad (10)$$

式中,  $\tau$  为温度系数,  $FC(q_i^i)$  为分类器输出。设置与对比学习中的温度系数一致。为了使分类器的权重能跟随网络训练的变化,分类器会通过一个交叉熵(cross entropy, CE)损失训练,其分类的监督信号为聚类得到的伪标签。具体为

$$L_{CE} = -\frac{1}{B} \sum_{i=1}^B \tilde{p}^i \cdot \log(FC(q_i^i)) \quad (11)$$

式中,  $\tilde{p}^i$  为第  $i$  个目标域查询样本  $q_i^i$  的伪标签。由于分类器中使用了伪标签作为监督信号,所以这里也会有伪标签噪声问题,但是这里的目的是得到一个随网络变化的分类器权重,以用于后续估计额外标签,并且这里设置了梯度不会反传至主干网络(如图2所示),所以伪标签噪声的影响对模型性能的影响较小。此外,为了充分利用已产生的特征,这里也用实例一致性中 EMA 模型  $M_{ema}$  额外产生的特征  $q_{ema,i}^i$  再额外产生一个预测标签  $p_{ema}^i$ , 具体为

$$p_{ema}^i = \text{softmax}\left(\frac{FC(q_{ema,i}^i)}{\tau}\right) \quad (12)$$

#### 2.3.2 标签优化

在得到将估计的两个预测标签后,求得它们的平均,再与查询样本的 one-hot 编码伪标签  $\tilde{p}^i$  相加后,求平均得到最终的标签。具体为

$$\hat{p}^i = \frac{\frac{p^i + p_{ema}^i}{2} + \tilde{p}^i}{2} \quad (13)$$

将优化过的伪标签  $\hat{p}^i$  作为对比学习的监督信号,目标域的对比损失(式(2))改写为

$$L_{LR,D_i} = -\frac{1}{B} \sum_{i=1}^B \sum_{j=1}^{n_i} \hat{p}^{ij} \log \bar{p}^{ij} \quad (14)$$

通过上述基于标签集成的标签优化方法,得到了鲁棒的软标签,减少了伪标签噪声的过拟合风险。在后续实验中,验证了该方法的有效性。

### 2.4 优化目标

网络最终的优化目标为

$$L_{total} = L_{DSCL,D_i} + L_{LR,D_i} + L_{IC,D_i} + L_{IC,D_i} + L_{CC,D_i} + L_{CC,D_i} + L_{CE} \quad (15)$$

式中,  $L_{DSCL,D_i}$  为源域域分离对比损失,  $L_{LR,D_i}$  为标签优化过的目标域域分离对比损失,  $L_{IC,D_i}$  和  $L_{IC,D_i}$  为源域和目标域的实例一致性损失,  $L_{CC,D_i}$  和  $L_{CC,D_i}$  分别为源域和目标域的相机一致性损失,  $L_{CE}$  为目标域分类器



损失。

### 3 实验结果与分析

#### 3.1 实验数据集与评估标准

实验所采用的数据集为 Market-1501 (Zheng 等, 2015)、DukeMTMC-reID (Duke multi-tracking multi-camera reIDentification) (Ristani 等, 2016) 和 MSMT17 (multi-scene multi-time) (Wei 等, 2018), 并以一个数据集为源域 (如 Market), 另一个数据集作为目标域 (如 Duke), 构建域适应任务 (如 Market→Duke)。在目标域上测试的平均精度均值 (mean average precision, mAP) 和 CMC (cumulative match characteristic) 曲线的 R1 (rank-1) 指标用来评价无监督域适应模型的性能。

#### 3.2 实验设置

为了验证所提算法的有效性, 在 4 个常用的无监督域适应行人重识别 (UDA Re-ID) 任务, 即 Duke→Market、Market→Duke、Duke→MSMT 和 Market→MSMT 上进行实验, 包括实例一致性约束、相机一致性约束和基于标签集成的标签优化的相关消融实验。除 MCRNCC 中提出的模块外, 实验的网络结构、训练过程和测试过程与 MCRN 保持一致。

##### 3.2.1 网络结构

主干网络为在 ImageNet (Deng 等, 2009) 上预训练过的 ResNet-50 (residual neural network) (He 等, 2016), 并使用 DSBN (domain-specific batch normalization) (Chang 等, 2019) 缩小域之间的差异。主干网络将最后一层输出的特征图依次通过全局平均池化、批归一化和 L2 归一化后得到的 2 048 维特征作为行人的特征。

##### 3.2.2 训练过程

在每个训练周期的准备阶段, 采用 DBSCAN 算法对目标域无标注图像聚类, 并使用了自步学习 (Ge 等, 2020b) 改善聚类过程。

在每个训练周期的优化阶段, 每个训练批由 64 幅源域图像 (16 个源域行人, 每个行人 4 幅图像) 和 64 幅目标域图像 (16 个目标域簇, 每个簇 4 幅图像) 构成。所有的训练图像会调整至  $256 \times 128$  像素, 并会通过多种数据增强策略对图像进行增广, 包括随机裁剪、随机翻转和随机擦除 (Zhong 等, 2020)。总训练周期 (epoch) 数设置为 50, 每个训练周期的迭

代次数在目标域为 Market 和 Duke 时设置为 400, 而在目标域为 MSMT 时设置为 800。采用 Adam 优化器优化网络, 权重衰减设置为 0.000 5。初始学习率设置为 0.000 35, 每隔 20 个训练周期衰减为原来的 1/10。记忆体更新时的动量系数  $m$  设置为 0.2, 对比损失中的温度系数  $\tau$  设置为 0.05,  $\lambda$  设置为 0.003 与目标域簇数的乘积。MCRNCC 的 EMA 模型的控制参数移动的超参数  $\varphi$  设置为 0.99, 相机一致性约束的阈值  $\gamma$  设置为 0.3。网络使用 Pytorch 实现, 并用 4 个 NVIDIA RTX-2080TI GPU 并行训练。

##### 3.2.3 测试过程

在测试阶段, 使用训练好的 EMA 模型在目标域的测试集上测试, 采用的评判标准为 mAP 和 R1。在训练过程中, 每隔 5 个训练周期测试一次, 以 mAP 最高的模型的结果作为模型的性能表现。

#### 3.3 消融实验

本文开展了关于实例一致性约束、相机一致性约束以及基于标签集成的标签优化方法的有效性的多组消融实验, 其中不同消融实验方法中涉及的相同的损失函数权重保持一致。实验结果如表 1 所示。从表 1 可以看出, 1) 在实例一致性约束的有效性方面, 在多个无监督域适应行人重识别任务上探究了行人的实例一致性 (IC) 约束的有效性。在 MCRN 网络上, 加入实例一致性约束后, 性能有了普遍提升。在 4 个任务上, 实例一致性约束的 mAP 性能分别提升了 0.6%, 0.2%, 0.7%, 0.8%, 这是由于实例一致性约束通过拉近同一实例在不同增广下的特征, 从实例内部挖掘了更丰富的语义信息, 使得性能得到了提升。2) 在相机一致性约束的有效性方面, 探究了相机一致性 (CC) 的有效性。在加入实例一致性后, 进一步加入相机一致性约束增广实例中跨相机的正样本的距离小于同相机的负样本距离, 网络在 4 个任务上性能有了进一步提升。其中, 在 Duke→MSMT 和 Market→MSMT 上, mAP 性能分别提升了 3.5% 和 5.3%, R1 分别提升了 3.2% 和 4.7%, 这是因为 MSMT 数据集有 15 个相机, 相机数目比较多, 相机拍摄的图像间差异较大, 通过相机一致性约束, 可以很好地约束类内不同相机的距离, 从而学习到一个更加紧凑的类内空间。3) 在基于标签集成的标签优化的有效性方面, 探究了基于标签集成的标签优化方法的有效性。在一致性约束的基础上进一



步加入标签优化后,得到本文提出的完整网络 MCRNCC,其在4个任务上,mAP分别提升了0.6%,0.6%,1.4%,0.4%,这是因为通过多组标签集成得

到的软标签会在训练过程中反映模型的变化,比原来在每个训练周期中固定不变的 one-hot 编码伪标签更鲁棒,可以减少模型对噪声的过度拟合。

表1 MCRNCC中的消融实验  
Table 1 Ablation study on MCRNCC

| 方法             | Duke→Market |             | Market→Duke |             | Duke→MSMT   |             | Market→MSMT |             |
|----------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
|                |             |             |             |             |             |             |             |             |
|                | mAP         | R1          | mAP         | R1          | mAP         | R1          | mAP         | R1          |
| MCRN           | 83.8        | 93.8        | 71.5        | 84.5        | 35.7        | 67.5        | 32.8        | 64.4        |
| MCRN +IC       | 84.4        | 93.6        | 71.7        | 84.6        | 36.4        | 67.2        | 33.6        | 64.5        |
| MCRN +IC+CC    | 84.4        | 93.9        | 72.9        | <b>85.7</b> | 39.9        | 70.4        | 38.9        | 69.2        |
| MCRN +IC+CC+LR | <b>85.0</b> | <b>94.0</b> | <b>73.5</b> | 85.6        | <b>41.3</b> | <b>71.6</b> | <b>39.3</b> | <b>69.5</b> |

注:加粗字体表示各列最优结果。

本文将使用相机一致性约束前后的所有行人构成的特征空间进行了可视化,并在图5中展示了多个不同行人的特征可视化对比示例,其中数字代表真实ID,不同颜色代表不同相机。由图5(a)(b)可以观察到,相机一致性约束使得相同行人在不同相机下的距离变近,而由图5(c)可以看到,相机一致性约束使得不同行人在相同相机下的特征距离变远。以上实验说明了相机一致性约束可以使网络的特征空间类内更加紧凑,类间更具可分性。

表2为标签集成(LE)的标签优化方式与传统的

标签软化(label smooth, LS)的标签优化方法的比较结果。其中,LR(LE)代表基于标签集成的标签优化,LR(LS)代表基于标签软化的标签优化方法。这里的标签软化是将 $N$ 维的 one-hot 标签中数值为1的维度的数值设置为0.8,将其他 $N-1$ 维度的数值设置为 $0.8/(N-1)$ 。由实验结果可见,若用简单的标签软化进行伪标签的优化,性能反而变差,在 Duke→Market 和 Market→Duke 上的 mAP 分别下降了0.6%和1.7%,R1分别下降了0.3%和1.6%,进一步说明标签集成的标签优化方式的有效性。

表2 不同标签优化方式的对比  
Table 2 The comparisons between different label refinery methods

| 方法       | Duke→Market |             | Market→Duke |             |
|----------|-------------|-------------|-------------|-------------|
|          |             |             |             |             |
|          | mAP         | R1          | mAP         | R1          |
| wo/LR    | 84.4        | 93.9        | 72.9        | <b>85.7</b> |
| w/LR(LE) | <b>85.0</b> | <b>94.0</b> | <b>73.5</b> | 85.6        |
| w/LR(LS) | 83.8        | 93.6        | 71.2        | 84.1        |

注:w代表with,表示使用, wo代表without,表示不使用;加粗字体表示各列最优结果。

3.4 与现有方法的对比

表3为基于MCRN改进的MCRNCC与现有的其他无监督域适应行人重识别方法在4个任务上的性能比较结果。比较的方法包括MMT(mutual mean-

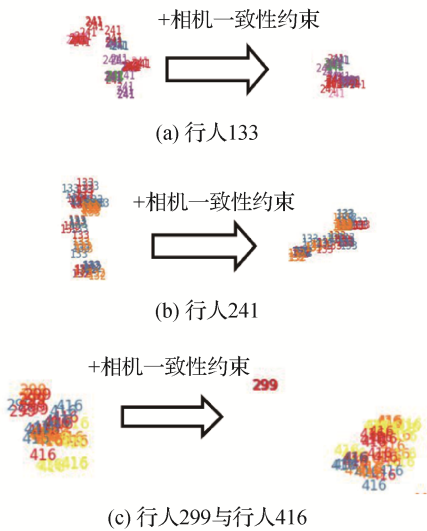


图5 相机一致性的有效性的可视化示例

Fig. 5 Visual examples of the effectiveness of camera consistency((a)pedestrian ID 133; (b)pedestrian ID 241; (c) pedestrian ID 299 and ID 416)

teaching)(Ge 等, 2020a)、MEB-Net (multiple expert brainstorming)(Zhai 等, 2020)、SpCL(Ge 等, 2020b)、UNRN (uncertainty-guided noise resilient network)(Zheng 等, 2021a)、GCL (generative and contrastive learning)(Chen 等, 2021)、GLT (group-aware label transfer)(Zheng 等, 2021b)、OPLG (online pseudo label generation)(Zheng 等, 2021c)、IDM(intermediate domain module)(Dai 等, 2021)、TDRL(Isobe 等, 2021)和MCRN(Wu 等, 2022)。可以看出, MCRNCC 在 4 个任务上都取得了最佳的性能, 显著优于其他方法,

mAP 分别为 85. 0%, 73. 5%, 41. 3%, 39. 3%; R1 分别为 94. 0%, 85. 6%, 71. 6%, 69. 5%。特别地, 在 Duke→MSMT 和 Market→MSMT 任务上, 与 MCRN 相比, MCRNCC 的 mAP 分别超越了 5. 6% 和 6. 5%, R1 分别超越了 4. 1% 和 5. 1%。为了使学习过程更鲁棒, MMT 使用了 4 个网络进行标签软化, MEB-Net 使用了 6 个网络进行集成学习, 而 MCRNCC 只使用了 2 个网络, 并通过充分挖掘 2 个网络的特征进行实例一致性约束、相机一致性约束和基于标签集成的标签优化, 取得了更好的性能。

表 3 与先进的 UDA Re-ID 方法的比较  
Table 3 Comparison with state-of-the-art UDA Re-ID methods

| 方法      | 引用出处   | /%          |             |             |             |             |             |             |             |
|---------|--------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
|         |        | Duke→Market |             | Market→Duke |             | Duke→MSMT   |             | Market→MSMT |             |
|         |        | mAP         | R1          | mAP         | R1          | mAP         | R1          | mAP         | R1          |
| MMT     | ICLR20 | 71.2        | 87.7        | 65.1        | 78.0        | 23.3        | 50.1        | 22.9        | 49.2        |
| MEB-Net | ECCV20 | 76.0        | 89.9        | 66.1        | 79.6        | —           | —           | —           | —           |
| SPCL    | NIPS20 | 76.7        | 90.3        | 68.8        | 82.9        | 26.5        | 53.1        | 26.8        | 53.7        |
| UNRN    | AAAI21 | 78.1        | 91.9        | 69.1        | 82.0        | 26.2        | 54.9        | 25.3        | 52.4        |
| GCL     | CVPR21 | 75.4        | 90.5        | 67.6        | 81.9        | 29.7        | 54.4        | 27.0        | 51.1        |
| GLT     | CVPR21 | 79.5        | 92.2        | 69.2        | 82.0        | 27.7        | 59.5        | 26.5        | 56.6        |
| OPLG    | ICCV21 | 80.0        | 91.5        | 70.1        | 82.2        | 29.3        | 56.1        | 28.4        | 25.3        |
| IDM     | ICCV21 | 82.8        | 93.2        | 70.5        | 83.6        | 35.4        | 61.3        | 33.5        | 61.3        |
| TDRL    | ICCV21 | 83.4        | 94.2        | 70.8        | 83.5        | 36.3        | 66.6        | 35.8        | 65.8        |
| MCRN    | AAAI22 | 83.8        | 93.8        | 71.5        | 84.5        | 35.7        | 67.5        | 32.8        | 64.4        |
| MCRNCC  | 本文     | <b>85.0</b> | <b>94.0</b> | <b>73.5</b> | <b>85.6</b> | <b>41.3</b> | <b>71.6</b> | <b>39.3</b> | <b>69.5</b> |

注: 加粗字体表示各列最优结果, “—”表示无相应的实验数据。

4 结 论

针对无监督域适应行人重识别问题, 本文提出一种具有一致性约束的方法 MCRNCC。其中, 设计了实例一致性约束、相机一致性约束和基于标签集成的标签优化方法等 3 个模块, 以分别解决现有方法中存在的行人实例特征不稳定问题、行人特征对相机变化的不鲁棒问题以及伪标签噪声的过拟合问题。实例一致性约束通过约束同一实例在不同增广下的特征相近, 从实例内部挖掘丰富的信息, 学习到更鲁棒的特征表达; 相机一致性约束对跨相机的正样本对和同相机的负样本对构建难例三元组, 提升

了行人在不同相机下的特征稳定性; 基于标签集成的标签优化引入了额外的伪标签估计, 通过将标签估计与原来聚类生成的 one-hot 编码伪标签集成, 减少了伪标签噪声的过拟合风险。本文通过大量消融实验验证了 MCRNCC 中各模块设计的有效性。与其他现有方法的比较表明 MCRNCC 取得了先进的性能。

本文方法主要针对的是单源域的域适应问题, 对于多个源域的域适应问题尚未探究。通常, 由于源域之间存在域差异, 将多个源域简单地一起训练并不能有效提升性能。而在实际应用中, 已有很多标注好的源域数据集, 这些丰富的数据集已经使用了大量的人力物力, 如何充分利用这些数据集, 将各

个源域通用的知识有效地传递到目标域上,进一步提升模型的性能,具有非常高的实用价值。

## 参考文献(References)

- Chang W G, You T, Seo S, Kwak S and Han B. 2019. Domain-specific batch normalization for unsupervised domain adaptation//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 7346-7354 [DOI: 10.1109/CVPR.2019.00753]
- Chen H, Wang Y H, Lagadec B, Dantcheva A and Bremond F. 2021. Joint generative and contrastive learning for unsupervised person re-identification//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual, USA: IEEE: 2004-2013 [DOI: 10.1109/CVPR46437.2021.00204]
- Chen T, Kornblith S, Norouzi M and Hinton G. 2020. A simple framework for contrastive learning of visual representations//Proceedings of the 37th International Conference on Machine Learning. Virtual: PMLR.org: 1597-1607
- Dai Y X, Liu J, Sun Y F, Tong Z K, Zhang C and Duan L Y. 2021. IDM: an intermediate domain module for domain adaptive person Re-ID//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 11844-11854 [DOI: 10.1109/ICCV48922.2021.01165]
- Deng J, Dong W, Socher R, Li L J, Li K and Fei-Fei L. 2009. ImageNet: a large-scale hierarchical image database//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE: 248-255 [DOI: 10.1109/CVPR.2009.5206848]
- Deng W J, Zheng L, Ye Q X, Kang G L, Yang Y and Jiao J B. 2018. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 994-1003 [DOI: 10.1109/CVPR.2018.00110]
- Ester M, Kriegel H P, Sander J and Xu X W. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise//Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining. Portland, USA: AAAI Press: 226-231
- Ge Y X, Chen D P and Li H S. 2020a. Mutual mean-teaching: pseudo label refinery for unsupervised domain adaptation on person re-identification [EB/OL]. [2022-06-24]. <https://arxiv.org/pdf/2001.01526/pdf>
- Ge Y X, Zhu F, Chen D P, Zhao R and Li H S. 2020b. Self-paced contrastive learning with hybrid memory for domain adaptive object Re-ID //Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: 11309-11321
- Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y. 2014. Generative adversarial nets//Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press: 2672-2680
- Grill J B, Strub F, Althé F, Tallec C, Richemond P H, Buchatskaya E, Doersch C, Pires B A, Guo Z D, Azar M G, Piot B, Kavukcuoglu K, Munos R and Valko M. 2020. Bootstrap your own latent a new approach to self-supervised learning//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc.: 21271-21284
- He K M, Fan H Q, Wu Y X, Xie S N and Girshick R. 2020. Momentum contrast for unsupervised visual representation learning//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 9726-9735 [DOI: 10.1109/CVPR42600.2020.00975]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Isobe T, Li D, Tian L, Chen W H, Shan Y and Wang S J. 2021. Towards discriminative representation learning for unsupervised person re-identification//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 8506-8516 [DOI: 10.1109/ICCV48922.2021.00841]
- Ren S Q, He K M, Girshick R and Sun J. 2015. Faster R-CNN: towards real-time object detection with region proposal networks//Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press: 91-99
- Ristani E, Solera F, Zou R, Cucchiara R and Tomasi C. 2016. Performance measures and a data set for multi-target, multi-camera tracking//Proceedings of 2016 European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 17-35 [DOI: 10.1007/978-3-319-48881-3\_2]
- Shi W D, Zhang Y Z, Liu S W, Zhu S D and Bao J N. 2020. Person re-identification based on deformation and occlusion mechanisms. *Journal of Image and Graphics*, 25(12): 2530-2540 (史维东, 张云洲, 刘双伟, 朱尚栋, 暴吉宁. 2020. 针对形变与遮挡问题的行人再识别. *中国图象图形学报*, 25(12): 2530-2540) [DOI: 10.11834/jig.200016]
- Tian Z, Shen C H, Chen H and He T. 2019. FCOS: fully convolutional one-stage object detection//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 9626-9635 [DOI: 10.1109/ICCV.2019.00972]
- Wei L H, Zhang S L, Gao W and Tian Q. 2018. Person transfer GAN to bridge domain gap for person re-identification//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 79-88 [DOI: 10.1109/CVPR.



- 2018.00016]
- Wu Y H, Huang T T, Yao H T, Zhang C, Shao Y J, Han C C, Gao C X and Sang N. 2022. Multi-centroid representation network for domain adaptive person Re-ID. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(3): 2750-2758 [DOI: 10.1609/aaai.v36i3.20178]
- Zhai Y P, Ye Q X, Lu S J, Jia M X, Ji R R and Tian Y H. 2020. Multiple expert brainstorming for domain adaptive person re-identification//*Proceedings of the 16th European Conference on Computer Vision*. Glasgow, UK: Springer: 594-611 [DOI: 10.1007/978-3-030-58571-6\_35]
- Zheng K C, Lan C L, Zeng W J, Zhang Z Z and Zha Z J. 2021a. Exploiting sample uncertainty for domain adaptive person re-identification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(4): 3538-3546 [DOI: 10.1609/aaai.v35i4.16468]
- Zheng K C, Liu W, He L X, Mei T, Luo J B and Zha Z J. 2021b. Group-aware label transfer for domain adaptive person re-identification//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, USA: IEEE: 5306-5315 [DOI: 10.1109/CVPR46437.2021.00527]
- Zheng L, Shen L Y, Tian L, Wang S J, Wang J D and Tian Q. 2015. Scalable person re-identification: a benchmark//*Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago, Chile: IEEE: 1116-1124 [DOI: 10.1109/ICCV.2015.133]
- Zheng X, Lin L, Ye M, Wang L and He C L. 2020. Improving person re-identification by attention and multi-attributes. *Journal of Image and Graphics*, 25(5): 936-945 (郑鑫, 林兰, 叶茂, 王丽, 贺春林. 2020. 结合注意力机制和多属性分类的行人再识别. *中国图象图形学报*, 25(5): 936-945) [DOI: 10.11834/jig.190185]
- Zheng Y, Tang S X, Teng G L, Ge Y X, Liu K J, Qin J, Qi D L and Chen D P. 2021c. Online pseudo label generation by hierarchical cluster dynamics for adaptive person re-identification//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*. Montreal, Canada: IEEE: 8351-8361 [DOI: 10.1109/ICCV48922.2021.00826]
- Zhong Z, Zheng L, Kang G L, Li S Z and Yang Y. 2020. Random erasing data augmentation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(7): 13001-13008 [DOI: 10.1609/aaai.v34i07.7000]

## 作者简介

吴禹航,男,硕士研究生,主要研究方向为行人重识别、无监督域适应。E-mail:wuyyuhang@gmail.com

桑农,通信作者,男,教授,博士生导师,主要研究方向为计算机视觉、图像检索、行人重识别。E-mail:nsang@hust.edu.cn