

基于相机类对比学习的完全无监督行人重识别

田青^{1,2,3}, 周子巢¹

(1. 南京信息工程大学软件学院, 南京 210044; 2. 南京信息工程大学无锡研究院, 无锡 214101; 3. 南京大学计算机软件新技术国家重点实验室, 南京 210023)

摘要: 最近的无监督行人重识别研究使用聚类和记忆字典中的伪标签来训练模型。但是, 这些研究忽略了行人重识别的数据集是通过不同相机采集的, 即相机之间的分布差异较大, 较大的相机方差会导致模型精度降低。因此, 提出了相机类对比学习, 包括类对比损失和相机对比损失, 其中类对比损失可以实现对内存字典的一致性更新, 并减少噪声标签对模型的影响; 而相机对比损失通过为每个相机中的每个类构建相机类中心, 拉近同属一个类的相机类中心距离, 并使不同类的相机类中心距离相距更远, 从而减少相机方差。通过相机类对比学习, 减少了相机方差和噪声标签对模型的影响, 从而提高了行人重识别的性能。在4个公开数据集上, 相机类对比学习都表现出优异的结果, 有效地缓解了相机方差对模型的影响。

关键词: 行人重识别; 类对比损失; 相机对比损失; 内存字典; 相机方差

中图分类号: TP391 **文献标志码:** A

Completely Unsupervised Person Re-identification Based on Camera Cluster Contrast Learning

TIAN Qing^{1,2,3}, ZHOU Zixiao¹

(1. School of Software, Nanjing University of Information Science and Technology, Nanjing 210044, China; 2. Wuxi Institute of Technology, Nanjing University of Information Science and Technology, Wuxi 214101, China; 3. State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China)

Abstract: Recent unsupervised person re-identification studies have used clustering and memory dictionaries for pseudo labels to train models. However, these studies ignore that the datasets of person re-identification are collected by different cameras, that is, the distribution difference between cameras is large, and a larger camera variance will lead to decrease in model accuracy. Therefore, camera cluster contrast learning is proposed, which includes cluster contrast loss and camera contrast loss. The cluster contrast loss can realize the consistent update of memory dictionary and reduce the influence of noise labels on the model. Camera contrast loss reduces camera variance by building camera cluster center for each cluster in each camera, narrowing the camera cluster center distance of the same cluster, and making different camera cluster centers farther apart. By camera cluster contrast learning, the impact of camera

基金项目: 国家自然科学基金(62176128); 江苏省自然科学基金(BK20231143); 南京大学计算机软件新技术国家重点实验室开放课题(KFKT2022B06); 中央高校基本科研基金(NJ2022028); 江苏省“青蓝工程”人才计划项目。

收稿日期: 2024-02-08; **修订日期:** 2024-03-23

variance and noise labels on the model is reduced, and the performance of person re-identification is improved. On the four public datasets, camera cluster contrast learning has shown excellent results, effectively alleviating the impact of camera variance on the model.

Key words: person re-identification; cluster contrast loss; camera contrast loss; memory dictionary; camera variance

引 言

行人重识别(Person re-identification, ReID)^[1-4]的目的是通过比较不同地点、时间和相机下外观特征的相似性来识别同一个人。它在现实世界中有着广泛的应用,如寻找失踪人员、追踪罪犯等。有监督行人ReID^[5-8]是使用人工标注的真实标签来构建判别性外观特征。然而,将训练好的行人ReID模型直接应用到现实场景上,性能会严重下降,这是因为训练数据集和现实数据之间存在领域差距。此外,由于人工标注标签的成本较高,有监督方法在实际应用中的可扩展性较差。无监督行人ReID无需标注数据就能训练模型,具有更好的可扩展性。因此,无监督行人ReID吸引了越来越多研究人员的关注。

现有的无监督行人ReID根据是否使用带有真实标签的源域数据集分为两类。第一类方法是学习迁移学习中领域自适应^[9-10]的思想,被称为无监督域适应学习(Unsupervised domain adaptation, UDA)^[11-19]。无监督域适应方法训练过程分为两个阶段,首先利用源域数据集中人工标注的标签预训练行人ReID模型,然后在无标签的目标域数据集中对模型进行微调。第二类方法则是直接在标记数据集中训练的完全无监督学习(Completely unsupervised learning, USL)^[20-24]。

目前,无监督域适应学习主要分为3类。第一类是基于语义属性的方法。这种方法假设源域数据和目标域数据共享一个共同的中层特征空间,并利用共同的中层特征空间来推断目标域中行人的身份,如文献[11]和文献[12]。第二类是基于对抗生成网络(Generative adversarial networks, GAN),从样式迁移的源域图像中学习域不变特征。例如文献[13]和文献[14]通过利用循环生成对抗网络(Cycle-consistent adversarial networks, CycleGAN)^[25]训练相机风格迁移模型。而文献[15]和文献[16]考虑了相机差异对模型的影响,使用星形对抗网络^[26]训练相机风格迁移模型。利用训练好的相机迁移模型,生成不同相机风格的行人图像。第三类方法是基于伪标签,首先使用带有真实标签的源域数据集对模型进行预训练,然后使用目标域数据来微调预训练过的行人模型。文献[17]提出了相互学习模型(Mutual mean-teaching, MMT),利用离线硬标签和在线软标签联合训练网络,对伪标签进行提炼。文献[18]提出了迈向判别表征学习(Towards discriminative representation learning, TDRL),通过逐渐减少源域的训练权重和增加源域的权重来实现领域迁移。文献[19]提出了深度可信度量学习(Deep credible metric learning, DCML),通过选择置信度高的伪标签来训练模型,减少噪声标签对行人ReID模型的污染。

一般来说,无监督域适应的方法比完全无监督表现更好,因为它们有源域数据的真实标签支持。然而,无监督域适应的性能受到源域数据集质量和规模的影响。此外,在现实世界中,具有人工注释的大规模数据并不总是可用的。因此,完全无监督行人ReID具有更大的灵活性,可以更好地应用于现实场景。近年来,一些研究人员开始关注如何直接利用无标注数据训练行人ReID模型,即在完全无监督的场景下训练行人ReID模型。现有的完全无监督行人ReID主要面临着两个问题。首先,在完全无监督场景中,行人ReID数据集是没有人工标注的行人标签。因此,研究人员大多使用无监督聚类方法为

行人 ReID 数据分配伪标签,从而对行人 ReID 模型进行监督训练。但通过聚类方法获得的伪标签存在噪声标签,会将具有相似特征但不属于一类的行人图片归为同一类,从而降低模型精度。此外,行人 ReID 数据集是由不同的相机采集得到的,不同相机拍摄的行人外观会因为相机位置和参数的不同而发生改变,使得模型难以识别出相同身份的行人图片。而现有的完全无监督行人 ReID 方法通常关注于噪声标签问题,如文献[20]为了缓解噪声标签对模型的影响,提出了一种自下而上的聚类(Bottom-up clustering, BUC)方法,将每个行人图像作为一个单独的类,并逐步将相似的行人图像合并到更大的聚类中。文献[21]则将内存字典和对比损失相结合,提出了基于内存的多标签分类损失(Memory based multi-classification loss, MMCL),使用相似性计算和循环一致性来预测伪标签。文献[22]提出了一种自步对比学习框架,通过逐步创建更可靠的类特征来改进内存字典中的实例特征。文献[23]提出了类对比学习,利用类对比损失一致地更新类级内存字典,从而减少噪声标签对模型的影响。而对于相机差异问题,鲜有研究人员考虑。如文献[24]通过生成相机内和相机间的伪标签,迭代优化相机间的相似性,从而减少相机分布差异对行人 ReID 模型的影响。

因此,针对上述相机差异问题,提出了相机对比损失的方法,即为每个类中的每个相机创建相机类中心。通过相机类中心减少相机之间的分布差异。此外,还使用类对比损失函数来改善噪声标签对模型的影响。将类对比损失和相机对比损失联合训练模型,并将其称为相机类对比学习(Camera cluster contrast learning, CACCL)。概括而言,本文的主要贡献如下:

(1)提出了相机对比损失来有效解决行人 ReID 数据中相机分布差异问题。

(2)联合使用相机对比损失和类对比损失训练行人 ReID 模型,减少相机方差和噪声标签对模型的影响。

1 基于相机类对比学习的完全无监督行人 ReID 模型

1.1 完全无监督行人 ReID

在本文中,一个未标记的行人数据集表示为 $X = \{x_1, x_2, \dots, x_N\}$,其中 x_i 是指第 i 张行人图片, N 为行人图片的总数,此外该数据集是由 V_{cam} 个相机拍摄构成的。在完全无监督行人 ReID 任务中,其目的是训练一个深度神经网络 $f_\theta(\cdot)$,将 X 中的行人图像投影到特征空间 F 中,从而使查询图像特征 f 与 F 中属于同一身份的图像具有相似的特征。在完全无监督行人 ReID 的背景下,行人 ReID 的难度主要体现在两个方面。首先,由于没有源域数据集中真实标签的指导,完全无监督行人 ReID 一般使用如基于密度的噪声空间聚类算法(Density based spatial clustering of application with noise, DBSCAN)或 K 均值聚类(K-means clustering)算法对行人图像特征进行聚类,为无标签行人数据集生成对应伪标签 $Y = \{y_1, y_2, \dots, y_N\}$ 。然而,生成的伪标签通常含有噪声标签,从而影响模型的优化。此外,行人 ReID 是一个跨相机检索过程,每个行人的图像都可能由不同参数的多个相机拍摄,这些因素会极大地改变行人身份的外观。换言之,各个相机之间的领域差距会增加模型优化类内特征相似性的难度。为了解决这两个问题,相机类对比学习联合类对比损失和相机对比损失来训练模型。

所提算法的总体框架如图 1 所示,首先利用特征提取器获取未标记数据的特征,并使用无监督聚类方法为未标记数据分配伪标签。然后,考虑到噪声标签对内存字典的污染,使用带有伪标签的特征构建类级内存字典,并在计算类对比损失同时对内存字典进行更新。此外,为了缓解相机因素对行人 ReID 模型的影响,对各相机中每个类构建相机类中心特征计算相机对比损失。最后,联合使用类对比损失和相机对比损失共同训练模型。

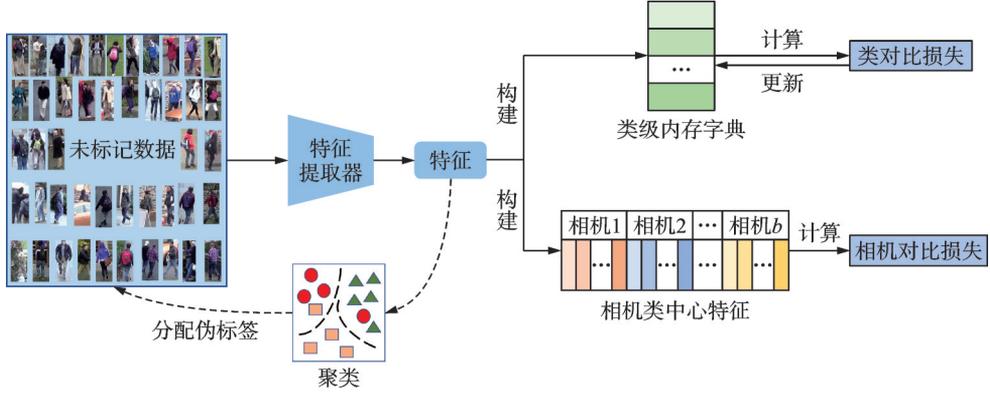


图1 整体框架图

Fig.1 Overall framework diagram

1.2 类对比损失

在完全无监督行人 ReID 中,有一些方法使用信息噪声对比估计损失(Info noise contrastive estimation, InfoNCE)进行模型训练。尽管不同的方法对 InfoNCE 有着不同的改进,但本文将将其归纳为一个通用公式

$$L_{\text{Info}} = -\log \frac{\left(\exp(\text{sim}(\mathbf{f} \cdot \mathbf{v}_+) / \tau) \right)}{\sum_{b=1}^B \exp(\text{sim}(\mathbf{f} \cdot \mathbf{v}_b) / \tau)} \quad (1)$$

式中: \mathbf{f} 为查询样本特征, \mathbf{v}_+ 为与查询特征 \mathbf{f} 属于同一类的特征, B 为训练批次的大小, τ 为控制相似性空间的温度参数, $\text{sim}(\cdot)$ 表示两个特征向量的相似性。

有些方法为未标记的行人数据集的特征构建内存字典 $\mathbf{M} \in \mathbf{R}^{(d \times N)}$,并通过查询特征 \mathbf{f} 对内存字典 \mathbf{M} 进行动量更新。为了改善内存字典的不一致更新和噪声标签对内存字典的污染,本文使用类对比损失

$$L_{\text{Cluster-contrast}} = -\log \frac{\left(\exp(\text{sim}(\mathbf{f} \cdot \mathbf{c}_+) / \tau) \right)}{\sum_{k=1}^K \exp(\text{sim}(\mathbf{f} \cdot \mathbf{c}_k) / \tau)} \quad (2)$$

式中: $\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_k\}$ 为类中心特征, K 为类中心特征的数量。类对比损失通过计算查询实例与所有类中心特征之间的距离来训练模型。计算查询样本特征 \mathbf{f} 与类中心特征 \mathbf{c}_k 之间的距离来训练模型。类中心特征是由每个类的平均特征计算得出的,即

$$\mathbf{c}_k = \frac{1}{|H_k|} \sum_{f_i \in H_k} \mathbf{f}_i \quad (3)$$

式中: H_k 表示属于第 K 个类的特征集合, $|H_k|$ 表示每个类的实例数。使用查询特征 \mathbf{f} 更新类级别的内存字典,即

$$\mathbf{c}_k = m\mathbf{c}_k + (1 - m)\mathbf{f} \quad (4)$$

式中 m 为动量更新因子。

1.3 相机对比损失

虽然类对比损失解决了内存字典更新不一致的问题,并减轻了噪声标签对模型准确性的影响,但

忽略了相机方差对行人ReID模型的影响。不同相机下的行人外观可能会受到视角和光照等环境因素的影响,从而显著改变行人的外观特征。如果不考虑相机变化引起的类内变化,不同相机间的领域差异使模型很难识别具有相同身份的行人图片。因此,提出了相机对比损失来缓解行人数据集中相机方差问题。

首先,相机对比损失通过式(5)为每个相机构建一个唯一的相机类中心特征表示。

$$c_{ab} = \frac{1}{N_{ab}} \sum_{f_i \in y_a \cap f_j \in v_b} f_i \quad (5)$$

式中 N_{ab} 为相机标记为 v_b 的属于类别 a 相机类中心特征的总数。

通过相机对比损失,来自一台相机的实例会被拉近到其他相机类中心特征,从而减少类内相机的风格差异。

$$L_{\text{Cross-camera}} = -\frac{1}{|V|} \sum_{i \neq b \cap i \in V_{\text{cam}}} \log \frac{\exp(\text{sim}(\mathbf{f} \cdot \mathbf{c}_{ai}) / \tau_b)}{\sum_{j=0}^{V_{\text{neg}}} \exp(\text{sim}(\mathbf{f} \cdot \mathbf{c}_j) / \tau_b)} \quad (6)$$

式中 $|V|$ 为相机类中心特征的数量, V_{neg} 为与查询特征 \mathbf{f} 距离较近但不属于同一相机标签的相机类中心特征的数量。

1.4 算法概述

通过上述问题分析和考虑,提出的相机类对比学习方法的最终损失如下

$$L_{\text{total}} = L_{\text{Cluster-contrast}} + \lambda L_{\text{Cross-camera}} \quad (7)$$

式中 λ 为用于平衡两个损失的超参数并设置为0.8。使用相机对比损失,从而减少相机差异,减轻相机因素对特征提取的干扰,并为类对比损失构建更高质量的类中心。而内存字典的类中心特征通过类对比损失不断更新,从而间接更新相机类中心特征。这两个损失函数相互促进相互优化,从而提高了模型的精确度。相机类对比学习的具体细节如算法1所示。

算法1 总体框架

输入:一个未标记的行人数据集 X ,一个神经网络 $f_\theta(\cdot)$,训练批次数量 B ,迭代次数 iters 和最大训练次数 epochs

输出:训练好的神经网络

- (1) While $\text{epoch} < \text{epochs}$ do
- (2) 通过 $f_\theta(\cdot)$ 从行人数据集 X 中提取特征向量 \mathbf{f} ;
- (3) 进行全局聚类并去除异常值;
- (4) 根据式(3)初始化内存字典 \mathbf{M} ;
- (5) 通过式(5),为每个相机中的每个类建立唯一的相机中心特征。
- (6) While $\text{iter} < \text{iters}$ do
- (7) 从训练集 X 中抽取 B 张查询图像;
- (8) 根据式(7)计算总损失;
- (9) 反向更新模型 $f_\theta(\cdot)$;
- (10) 更新内存字典 \mathbf{M} 。
- (11) End while

2 实验结果与分析

2.1 数据集介绍

本文在4个大规模行人ReID基准上评估了所提方法: Market-1501、DukeMTMC、MSMT17和PersonX。Market-1501包含1501个身份的32668张图像,由6个相机拍摄得到,用于训练的有751个身份的12936张图像,用于测试的有750个身份的19732张图像。DukeMTMC包含由8台相机拍摄的1404个身份的16522张图像,训练集为16522张图像,包含702个身份,测试集为17661张图像,包含702个身份。MSMT17包含15台相机拍摄的4101个身份的126441张图像,训练集包含1041个身份的32621张图像,测试集包含3060个身份的93820张图像。PersonX包含1266个身份的9840张图像,有6个相机视图,训练集有410个身份的9840张图像,测试集有856个身份的30816张图像。Market-1501、DukeMTMC和MSMT17是广泛使用的真实世界数据集,而PersonX是基于Unity的合成数据集,其中包含人工设计的障碍,如随机遮挡、分辨率和光照差异。

2.2 评估指标

对于行人ReID任务,有两种常见的评估标准,一种是累积匹配特征(Cumulative match characteristic, CMC),另一种是平均精度(Mean average precision, mAP)。CMC用于反映分类器的准确度,通常用Top- k 准确度表示,即查询图像在列表前 k 个匹配项中正确匹配的概率。在行人ReID任务中,一般考虑 $k=\{1,5,10\}$ 的精度。AP是指查询样本的平均精度,表示模型对某个样本的效果,mAP是指所有查询样本的平均AP,表示模型对所有查询样本的总体效果。

2.3 实验环境及参数设置

行人ReID模型的骨干网络是在ImageNet上预先训练好的ResNet-50。输入图像大小调整为 $256*128$,批次 B 的大小为64。在每个epoch开始时使用DBSCAN聚类技术生成伪标签。DBSCAN中两个样本之间的最大距离为0.6,核心点的最小邻居数为4。本文使用聚类中心来初始化内存库,并在每个epoch开始时重新初始化内存库。使用Adam优化器,权重衰减为 $5e-4$ 。初始学习率设置为 $3.5e-4$,每20个epoch减小到之前的 $1/10$,总共100个epoch。损失权重设为0.8。

2.4 与相关方法的比较实验

将所提算法与最近在Market-1501、DukeMTMC、MSMT17和PersonX中使用的无监督行人ReID方法进行比较,结果分别如表1~4所示。

在Market-1501数据集中,获得了84.3%的mAP和93.4%的Top-1。与Cluster-contrast相比,分别提高了1.3%和0.5%。在DukeMTMC数据集中,获得了74.1%的mAP和85.5%的Top-1。与P2LR相比,分别提高了3.3%和2.9%。在MSMT17数据集中,获得36.1%的mAP和67.0%的Top-1。与Cluster-contrast相比,分别提高了3.1%和5.0%。在PersonX数据集中,获得了89.4%的mAP和95.7%的Top-1。与Cluster-contrast相比分别提高了4.7%和1.3%。这表明所提

表1 在Market-1501数据集上的算法比较结果

方法	mAP	Top-1	Top-5	Top-10	%
BUC ^[20]	38.3	66.2	79.6	84.5	
SSL ^[27]	37.8	71.7	83.8	87.4	
MMCL ^[21]	45.5	80.3	89.4	92.3	
HCT ^[28]	56.4	80.0	91.6	95.2	
CycAs ^[29]	64.8	84.8			
AD-cluster ^[30]	68.3	86.7	94.4	96.5	
JGCL ^[31]	66.8	87.3	93.5	95.5	
UGA ^[32]	70.3	87.2			
IICS ^[23]	72.9	89.5	95.2	97.0	
SPCL ^[22]	73.1	88.1	95.1	97.0	
Cluster-contrast ^[23]	83.0	92.9	97.2	98.0	
所提方法	84.3	93.4	97.4	98.3	

表2 在DukeMTMC数据集上的算法比较结果

Table 2 Results of algorithm comparison on DukeMTMC dataset

方法	mAP	Top-1	Top-5	Top-10
BUC ^[20]	27.5	47.4	62.6	68.4
SSL ^[27]	28.6	52.5	63.5	68.9
MMCL ^[21]	51.4	72.4	82.9	85.0
HCT ^[28]	50.7	69.6	83.4	87.4
UGA ^[32]	53.3	75.0		
JGCL ^[31]	62.8	82.9	87.1	88.5
IICS ^[23]	64.4	80.0	89.0	91.6
SPCL ^[22]	65.3	81.2	90.3	92.2
P2LR ^[33]	70.8	82.6	90.8	93.7
所提方法	74.1	85.5	92.5	94.4

表3 在MSMT17数据集上的算法比较结果

Table 3 Results of algorithm comparison on MSMT17 dataset

方法	mAP	Top-1	Top-5	Top-10
ECN ^[16]	10.2	30.2	41.5	46.8
TAUDL ^[34]	12.5	28.4		
MMCL ^[21]	11.2	35.4	44.8	49.8
UTAL ^[35]	13.1	31.4		
CycAs ^[29]	26.7	50.1		
MMT ^[17]	24.0	50.1	63.5	69.3
SPCL ^[22]	19.1	42.3	55.6	61.2
JGCL ^[31]	21.3	45.7	58.6	64.5
IICS ^[24]	26.9	56.4	68.8	73.4
Cluster-contrast ^[23]	33.0	62.0	71.8	76.7
所提方法	36.1	67.0	77.0	80.6

方法充分利用了相机信息,减少了相机方差对模型的影响,在完全无监督行人ReID场景中有较好的效果。

2.5 参数敏感性分析

在本节中,将验证式(7)中的超参数λ对模型性能的影响。在Market-1501和DukeMTMC数据集进行参数敏感性分析,结果分别如图2、3所示。由图2、3可知,随着λ的增加,相机对比损失逐步减少

表4 在PersonX数据集上的算法比较结果

Table 4 Results of algorithm comparison on PersonX dataset

方法	mAP	Top-1	Top-5	Top-10
SPCL ^[22]	72.3	88.1	96.6	98.3
Cluster-contrast ^[23]	84.7	94.4	98.3	99.3
所提方法	89.4	95.7	99.0	99.6

相机方差对模型的影响,而模型的精确度也随之增加,当λ=0.8时模型精度最高。而当λ=1.0时,模型精度出现降低,这是因为相机对比损失是辅助类对比损失训练行人ReID模型。当相机对比损失的比重高于类对比损失,将会影响模型获取判别性行人特征,从而降低了模型的性能。

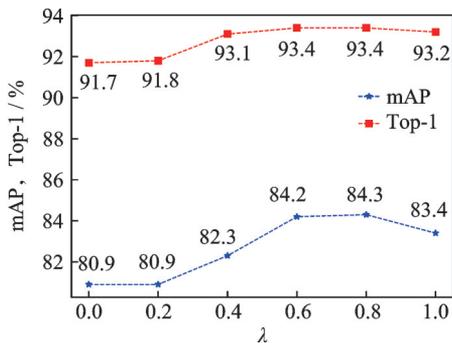


图2 不同λ值Market-1501实验

Fig.2 Experiment of different values of λ on the Market-1501

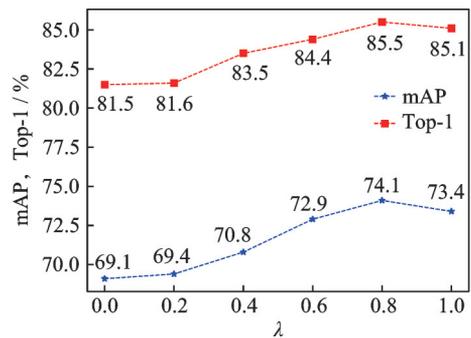


图3 不同λ值DukeMTMC实验

Fig.3 Experiment of different values of λ on the DukeMTMC

2.6 消融实验

通过消融实验分析类对比损失和相机对比损失对模型的影响。通过比较自步对比损失和类别对比损失的效果,展现出类对比损失有效地降低噪声标签对模型的影响。自步对比损失采用的是实例级内存字典,内存字典存储所有样本的特征,因此较依赖伪标签的准确度,使得模型容易受到噪声标签的影响。而类对比损失采用的是类级内存字典,内存字典存储各类的类中心特征,相较于自步对比损失中的实例级内存字典,减少了模型对伪标签的依赖,从而缓解了噪声标签对模型精度的影响。具体的消融结果分别如表 5~8 所示。由于批次从原文的 256 减少到 64,在 Market1501、DukeMTMC、MSMT17 和 PersonX 四个数据集中,自步对比损失和类对比损失相较于原文提供的精度有明显降低。

表 5 在 Market-1501 数据集上的消融实验

方法	mAP	Top-1	Top-5	Top-10	%
自步对比损失	70.8	86.5	94.1	96.3	
类对比损失	80.9	91.7	95.3	97.1	
类对比损失+相机损失	84.3	93.4	97.4	98.3	

表 6 在 DukeMTMC 数据集上的消融实验

方法	mAP	Top-1	Top-5	Top-10	%
自步对比损失	62.0	79.3	87.1	90.1	
类对比损失	69.1	81.5	90.7	93.4	
类对比损失+相机损失	74.1	85.5	92.5	94.4	

表 7 在 MSMT17 数据集上的消融实验

方法	mAP	Top-1	Top-5	Top-10	%
自步对比损失	17.4	40.2	53.4	60.1	
类对比损失	22.8	48.6	59.7	66.5	
类对比损失+相机损失	36.1	67.0	77.0	80.6	

表 8 在 PersonX 数据集上的消融实验

方法	mAP	Top-1	Top-5	Top-10	%
自步对比损失	70.3	85.9	94.2	96.1	
类对比损失	82.6	92.4	96.9	98.0	
类对比损失+相机损失	89.4	95.7	99.0	99.6	

在 Market-1501 数据集上,类对比损失与自步对比损失相比,mAP 提高了 10.1%,Top-1 提高了 5.2%;在 DukeMTMC 数据集上,mAP 提高了 7.1%,Top-1 提高了 2.2%;在 MSMT17 数据集上,mAP 提高了 5.4%,Top-1 提高了 8.4%;在 PersonX 数据集上 mAP 提高了 12.3%,Top-1 提高了 6.5%。由此可看出,类对比损失有效地缓解了噪声标签对模型的影响。此外,在增加了相机对比损失之后,mAP 在 Market-1501 数据集上增加了 4.1%,在 DukeMTMC 数据集上增加了 5.0%,在 MSMT17 数据集上增加了 13.3%,在 PersonX 数据集上增加了 6.8%。由此可以得出结论,通过相机对比损失,可以有效地缓解相机方差对模型的影响。

3 结束语

联合使用了相机对比损失和类对比损失来训练模型。行人 ReID 任务一般存在两个问题:相机方差和噪声标签污染,这两个因素严重影响了行人 ReID 模型的准确性。因此,本文为每个相机的每个类生成相机类中心特征,通过缩小属于同一类的相机类中心特征距离来减少相机差异,并促进类对比损失以形成更好的唯一类中心特征。利用类对比损失不断更新唯一类中心特征,减少噪声标签对模型的影响,并间接更新相机类中心特征,从而生成更高质量的相机类中心特征。通过相机对比损失和类对比损失的相互促进,很好地解决了行人 ReID 中相机方差和噪声标签这两个问题。实验结果也证明了

所提方法的优越性。

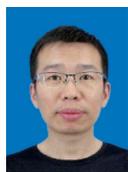
本文不足之处在于,使用 ResNet50 提取行人特征会丢失细粒度信息,而视觉转换器(Vision transformer, ViT)目前在计算机图像处理领域取得了卓越的进展。因此,下一步计划用 ViT 模型取代传统的 ResNet 网络作为特征提取器,以更好地提取行人特征。

参考文献:

- [1] ZHENG L, YANG Y, HAUPTMANN A G. Person re-identification: Past, present and future[EB/OL]. (2016-10-10). <https://arxiv.org/abs/1610.02984>.
- [2] YE M, SHEN J, LIN G, et al. Deep learning for person re-identification: A survey and outlook[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 44(6): 2872-2893.
- [3] MING Z, ZHU M, WANG X, et al. Deep learning-based person re-identification methods: A survey and outlook of recent works[J]. *Image and Vision Computing*, 2022, 119: 104394.
- [4] 杨永胜,邓森磊,李磊,等.基于深度学习的行人重识别综述[J].*计算机工程与应用*, 2022, 58(9): 51-66.
YANG Yongsheng, DENG Miaolei, LI Lei, et al. Overview of pedestrian re-identification based on deep learning[J]. *Computer Engineering and Applications*, 2022, 58(9): 51-66.
- [5] SUN Y, ZHENG L, YANG Y, et al. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)[C]//*Proceedings of the European Conference on Computer Vision (ECCV)*. [S.l.]: ECCV, 2018: 480-496.
- [6] CHEN H, LAGADEC B, BREMOND F. Learning discriminative and generalizable representations by spatial-channel partition for person re-identification[C]//*Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. [S.l.]: IEEE, 2020: 2483-2492.
- [7] LUO H, GU Y, LIAO X, 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*. [S.l.]: IEEE, 2019: 1-9.
- [8] 孙明浩,王洪元,吴琳钰,等.基于特征金字塔分支和非局部关注的行人重识别[J].*数据采集与处理*, 2023, 38(1): 121-131.
SUN Minghao, WANG Hongyuan, WU Linyu, et al. Person re-identification based on feature pyramid branch and non-local attention[J]. *Journal of Data Acquisition and Processing*, 2023, 38(1): 121-131.
- [9] TIAN Q, SUN H, MA C, et al. Heterogeneous domain adaptation with structure and classification space alignment[J]. *IEEE Transactions on Cybernetics*, 2021, 52(10): 10328-10338.
- [10] TIAN Q, MA C, CAO M, et al. A convex discriminant semantic correlation analysis for cross-view recognition[J]. *IEEE Transactions on Cybernetics*, 2020, 52(2): 849-861.
- [11] WANG J, ZHU X, GONG S, et al. Transferable joint attribute-identity deep learning for unsupervised person re-identification [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE, 2018: 2275-2284.
- [12] LIN S, LI H, LI C T, et al. Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification[EB/OL]. (2018-07-04). <https://doi.org/10.48550/arxiv.1807.01440>.
- [13] WEI L, ZHANG S, GAO W, et al. Person transfer gan to bridge domain gap for person re-identification[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE, 2018: 79-88.
- [14] DENG W, ZHENG L, YE Q, et al. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE, 2018: 994-1003.
- [15] ZHONG Z, ZHENG L, LI S, et al. Generalizing a person retrieval model hetero- and homogeneously[C]//*Proceedings of the European Conference on Computer Vision (ECCV)*. [S.l.]: ECCV, 2018: 172-188.
- [16] ZHONG Z, ZHENG L, LUO Z, et al. Invariance matters: Exemplar memory for domain adaptive person re-identification[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.]: IEEE, 2019: 598-607.
- [17] GE Y, CHEN D, LI H. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification[EB/OL]. (2020-01-06). <https://doi.org/10.48550/arxiv.2001.01526>.

- [18] ISOBE T, LI D, TIAN L, et al. Towards discriminative representation learning for unsupervised person re-identification[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2021: 8526-8536.
- [19] CHEN G, LU Y, LU J, et al. Deep credible metric learning for unsupervised domain adaptation person re-identification[C]// Proceedings of Computer Vision—ECCV 2020: 16th European Conference. Glasgow: Springer International Publishing, 2020: 643-659.
- [20] LIN Y, DONG X, ZHENG L, et al. A bottom-up clustering approach to unsupervised person re-identification[C]// Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2019: 8738-8745.
- [21] WANG D, ZHANG S. Unsupervised person re-identification via multi-label classification[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2020: 10981-10990.
- [22] 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.
- [23] DAI Z, WANG G, YUAN W, et al. Cluster contrast for unsupervised person re-identification[C]// Proceedings of the Asian Conference on Computer Vision. [S.l.]: [s.n.], 2022: 1142-1160.
- [24] XUAN S, ZHANG S. Intra-inter camera similarity for unsupervised person re-identification[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2021: 11926-11935.
- [25] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]// Proceedings of the IEEE International Conference on Computer Vision. [S.l.]: IEEE, 2017: 2223-2232.
- [26] CHOI Y, CHOI M, KIM M, et al. StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2018: 8789-8797.
- [27] LIN Y, XIE L, WU Y, et al. Unsupervised person re-identification via softened similarity learning[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2020: 3390-3399.
- [28] ZENG K, NING M, WANG Y, et al. Hierarchical clustering with hard-batch triplet loss for person re-identification[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2020: 13657-13665.
- [29] WANG Z, ZHANG J, ZHENG L, et al. CycAs: Self-supervised cycle association for learning re-identifiable descriptions[C]// Proceedings of Computer Vision—ECCV 2020: 16th European Conference. Glasgow: Springer International Publishing, 2020: 72-88.
- [30] ZHAI Y, LU S, YE Q, et al. AD-cluster: Augmented discriminative clustering for domain adaptive person re-identification [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2020: 9021-9030.
- [31] CHEN H, WANG Y, LAGADEC B, et al. Joint generative and contrastive learning for unsupervised person re-identification [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2021: 2004-2013.
- [32] WU J, YANG Y, LIU H, et al. Unsupervised graph association for person re-identification[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2019: 8321-8330.
- [33] HAN J, LI Y L, WANG S. Delving into probabilistic uncertainty for unsupervised domain adaptive person re-identification [C]// Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2022: 790-798.
- [34] LI M, ZHU X, GONG S. Unsupervised person re-identification by deep learning tracklet association[C]// Proceedings of the European Conference on Computer Vision (ECCV). [S.l.]: ECCV, 2018: 737-753.
- [35] LI M, ZHU X, GONG S. Unsupervised tracklet person re-identification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 42(7): 1770-1782.

作者简介:



田青(1983-),通信作者,男,博士,教授,硕士生导师,研究方向:机器学习、模式识别, E-mail: tianqing@nuist.edu.cn。



周子泉(1999-),男,硕士研究生,研究方向:机器学习、行人重识别, E-mail: 20211249364@nuist.edu.cn。