ISSN 1003-501X CN 51-1346/O4

光电工程

Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊
Scopus CSCD

伪标签细化引导的相机感知无监督行人重识别方法

程思雨，陈莹

引用本文:

程思雨，陈莹. 伪标签细化引导的相机感知无监督行人重识别方法[J]. 光电工程，2023， **50**(12): 230239.

Cheng S Y, Chen Y. Camera-aware unsupervised person re-identification method guided by pseudo-label refinement[J]. *Opto-Electron Eng*, 2023, **50**(12): 230239.

https://doi.org/10.12086/oee.2023.230239

收稿日期: 2023-09-24; 修改日期: 2023-12-27; 录用日期: 2023-12-27

相关论文

多特征聚合的红外–可见光行人重识别

郑海君，葛斌，夏晨星，邬成

光电工程 2023, **50**(7): 230136 doi: 10.12086/oee.2023.230136

基于多任务学习框架的红外行人检测算法

苟于涛，马梁，宋怡萱，靳雷，雷涛

光电工程 2021, **48**(12): 210358 doi: 10.12086/oee.2021.210358

深度双重注意力的生成与判别联合学习的行人重识别

张晓艳，张宝华，吕晓琪，谷宇，王月明，刘新，任彦，李建军

光电工程 2021, **48**(5): 200388 doi: 10.12086/oee.2021.200388

更多相关论文见光电期刊集群网站

OEE 光电工程 Opto-Electronic Engineering

http://cn.oejournal.org/oee

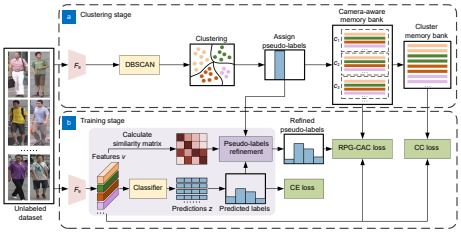
OE_Journal Website

Opto-Electronic Engineering
光 电 工 程

Article
2023 年, 第 50 卷, 第 12 期

DOI: 10.12086/oee.2023.230239

伪标签细化引导的相机感知无监督行人重识别方法



程思雨，陈　莹*

江南大学轻工过程先进控制教育部重点实验室 物联网工程学院，江苏 无锡 214122

摘要：无监督行人重识别因其广泛的实际应用前景而受到越来越多的关注。大多数基于聚类的对比学习方法将每个集群视为一个伪身份类，忽略了由相机风格差异造成的类内差异。一些方法引入了相机感知对比学习，根据相机视角将单一集群划分为多个子集群，但它们容易受到噪声伪标签的误导。为解决这一问题，本文首先基于实例在特征空间中的相似性，采用最近邻的预测标签和原始聚类结果的加权组合细化伪标签。然后，采用细化伪标签动态地关联实例可能属于的类别中心，同时剔除可能存在的假阴性样本。这一方法改进了相机感知对比学习中正负样本的选择机制，有效地减轻了噪声伪标签对对比学习任务的误导。在 Market-1501、MSMT17、Personx 数据集上 mAP/Rank-1 分别达到了 85.2%/94.4%、44.3%/74.1%、88.7%/95.9%。

关键词：行人重识别；无监督；相机感知对比学习；细化伪标签

中图分类号：TP391　　**文献标志码：**A

程思雨，陈莹. 伪标签细化引导的相机感知无监督行人重识别方法 [J]. 光电工程，2023，**50**(12): 230239

Cheng S Y, Chen Y. Camera-aware unsupervised person re-identification method guided by pseudo-label refinement[J]. *Opto-Electron Eng*, 2023, **50**(12): 230239

Camera-aware unsupervised person re-identification method guided by pseudo-label refinement

Cheng Siyu, Chen Ying*

Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China

Abstract: Unsupervised person re-identification has attracted more and more attention due to its extensive practical application prospects. Most clustering-based contrastive learning methods treat each cluster as a pseudo-identity class, overlooking intra-class variances caused by differences in camera styles. While some methods have introduced camera-aware contrastive learning by partitioning a single cluster into multiple sub-clusters based on camera views, they are susceptible to misguidance from noisy pseudo-labels. To address this issue, we first refine pseudo-labels by leveraging the similarity between instances in the feature space, using a weighted combination of the nearest neighboring predicted labels and the original clustering results. Subsequently, it dynamically associates

收稿日期：2023-09-24；修回日期：2023-12-27；录用日期：2023-12-27

基金项目：国家自然科学基金资助项目 (62173160)

*通信作者：陈莹 (1976-)，chenying@jiangnan.edu.cn。

instances with possible category centers based on refined pseudo-labels while eliminating potential false negative samples. This method enhances the selection mechanism for positive and negative samples in camera-aware contrastive learning, effectively mitigating the influence of noisy pseudo-labels on the contrastive learning task. On Market-1501, MSMT17 and Personx datasets, mAP/Rank-1 reached 85.2%/94.4%, 44.3%/74.1% and 88.7%/95.9%.

Keywords: person re-identification; unsupervised; camera-aware contrastive learning; refined pseudo-labels

1 引 言

行人重识别 (Re-identification, Re-ID) 旨在通过不重叠的相机检索特定行人的图像，在智能监控和公共安全领域拥有重要的应用价值[1]。近年来，随着深度学习技术的蓬勃发展，有监督的行人重识别方法取得了显著的性能提升。然而，这些方法依赖于大量的数据标注，标注工作耗时耗力，使得它们难以在实际场景中部署和应用。因此，无监督的行人重识别方法逐渐受到研究者们的关注。

无监督行人 Re-ID 可大致分为两大类：无监督领域自适应 (unsupervised domain adaptation，UDA) 方法和完全无监督学习 (unsupervised learning，USL) 方法。USL 方法需要的未标记行人图像很容易通过目标检测技术从监控系统中获得，相比于 UDA 方法更加灵活和易于部署。

近年来，对比学习方法在无监督视觉表征学习领域表现出色。其核心思想是在特征空间中缩小正样本对之间的距离，同时扩大负样本对之间的距离。这一思想也被成功地应用于 USL 行人重识别任务。根据内存字典设置的不同，对比学习方法在 USL 行人重识别中可分为两类：实例级和聚类级方法。实例级方法将每个图像视为一个单独的类，主要利用每个样本本身的自监督信息，而不太考虑样本间的结构或者相关性。例如，Zhong 等人[2] 提出将目标域中每张图片视为单独的类别，通过构建一个样本存储器，存储所有目标图像的最新特征，进而在目标域中实施样本不变性，相机不变性和邻域不变性三种约束。聚类级方法则通过聚类算法生成伪标签，根据输入图像与集群中心特征之间的相似性来构建对比损失。这类方法将同一集群内的样本视为正样本，而将其他集群的样本视为负样本。例如，Ge 等人[3] 提出了一种自步对比学习框架，将每个集群和离群值视为单一类，并计算集群级别的对比损失。然而，由于聚类大小和采样的随机性，导致每个集群的更新过程不一致。为解决这一问题，Dai 等人[4] 提出了聚类级对比学习机制，并在聚类级别更新内存字典。

这些方法将每个集群视为一个伪身份类，侧重改善类间差异，而忽略了由相机间风格差异引起的类内差异。最近一些方法探索了这一问题。Tian 等人[5] 提出采用 StarGAN[6] 将样本迁移到每个相机。其他一些方法采用相机内和相机间的联合训练方式，在相机内训练阶段，为每个相机设置一个分类器，所有的分类器共享一个主干网络。在相机间训练阶段，现有的方法设计了不同的策略统一从不同相机中学到的知识：Yang 等人[7] 提出了相机感知的元学习算法，将训练数据根据相机标签拆分为元训练和元测试部分，并进行梯度交互，以迫使模型学习相机不变的特征表示；Xuan 等人[8] 提出使用实例归一化 (instance normalization, IN) 和批处理归一化 (batch normalization, BN) 的组合提高分类器的泛化能力；Wang 等人[9] 提出将单一的聚类集群根据相机视角分成多个子集群，并设计了相机内和相机间的对比学习模块；Li 等人[10] 提出了一种相机风格分离模块，以明确地从特征图中分离出特定的相机信息从而减少类内方差；Lee 等人[11] 提出一种基于摄像头的课程学习 (CaCL) 方法，利用相机标签逐渐将从有标签的源域训练的模型推广到无标签的目标域，目标域数据集根据相机标签分为多个子集，模型最初使用单个摄像头捕捉的图像进行训练，然后根据调度规则逐渐利用更多的子集进行训练，调度规则考虑每个子集与源域数据集之间的最大均值差异 (MMD)，以确保在课程中较早地利用更接近源域的子集。

不同相机拍摄的图像受到视角、光照、背景等环境因素的影响，导致来自不同相机的同一身份标签之间的图像具有较大的特征差异。这使得聚类算法也很难将同一身份的样本准确地聚类到同一集群中，生成的伪标签也不可避免地含有噪声。伪标签中的噪声严重降低了无监督行人重识别方法的性能，为解决这一问题，Ge 等人[3] 采用了自步对比学习方案逐步获得更可靠的簇，用于对伪标签进行细化。Ge 等人[12] 和 Zhai 等人[13] 提出采用相互教学的方法，以相互监督

的方式生成更加可信且鲁棒的伪标签。Zhang 等人[14]引入了一种基于聚类共识矩阵的标签传播方案，以降低标签噪声，并鼓励连续两次迭代的聚类结果之间保持一致。Cho 等人[15]使用人体全局特征和部分可靠特征之间的互补关系来进行伪标签细化。Wu 等人[16]提出一种多质心存储器，自适应地捕获集群内的不同身份信息，并通过选择适当的正负质心来有效缓解标签噪声问题。Chen 等人[17]提出使用成对相似度得分作为软伪标签，增强实例之间的一致性，从而减轻标签噪声。Lan 等人[18]提出一种噪声标签净化模块，旨在利用教师模型的知识缓解聚类过程中引入的标签噪声。Chen 等人[19]提出一种双聚类协同教学网络，利用两个网络提取的特征，通过不同参数的聚类分别生成两组伪标签，每个网络都采用其对等网络生成的伪标签进行训练，通过增加两个网络的互补性从而减少噪声的影响。

虽然上述方法在缓解伪标签噪声方面已经取得了显著的性能，但主要关注图像对间的相似性或具有复杂的网络架构，未充分考虑不同相机之间的域转移，也忽略了相机风格影响和噪声伪标签这两个问题之间的内在联系。为探索这一问题，Pang 等人[20]提出了一种新的聚类方法 (DBSCAN-NN) 来缓解类内相机多样性不足的问题，并对特征进行聚类，以提高伪标签的精度。Wang 等人[21]设计了一种动态相机自适应聚类模块来对目标域的全局特征进行分组，通过自适应补偿相机间隙来缩小相机内样本对和相机间样本对之间的特征分布间隙，从而提高伪标签的质量。为了充分利用相机之间的相似性，Li 等人[22]提出一种伪标签细化框架，采用相机内局部聚类结果细化相机间的聚类结果，并采用细化后的可靠伪标签训练模型。与上述方法不同，本文提出了一种伪标签细化引导的相机感知无监督行人重识别方法，重点研究伪标签噪声对相机感知对比学习方法中正负样本选择的干扰，以及如何通过细化伪标签来减轻这种影响，以指导模型的学习过程。本文的主要贡献如下：

1) 通过计算特征空间中训练实例之间的相似性，为每个实例确定邻域集合。然后通过将模型对邻域内样本的预测标签与实例原始聚类结果进行加权组合来细化传统的 one-hot 伪标签。这种方法不仅鼓励模型将实例靠近其所属的集群中心，还将其与可能包含其身份信息的最近邻样本建立联系。该方法能够有效地提高模型对噪声标签的鲁棒性，同时减轻过拟合的风险。

2) 提出采用细化伪标签指导的相机感知对比学习方法。采用细化伪标签动态地关联实例可能属于的集群中心，而不再依赖于单一的集群中心，同时剔除可能存在的假阴性样本，从而降低噪声伪标签对相机感知的误导。

3) 将本文方法在三个大规模公开数据集上进行实验验证，结果表明所提方法较基准方法提升明显且优于目前同类先进方法。在 Market-1501[23]、MSMT17[24]、Personx[25] 数据集上 mAP/Rank-1 分别达到了 85.2%/94.4%、44.3%/74.1%、88.7%/95.9%。

2 本文方法

2.1 网络整体框架

给定一个未标记数据集$D=\{\boldsymbol{x}_i\}_{i=1}^{N}$，其中$x_i$表示第$i$幅图像，$N$表示图像总数。本文方法的整体框架如图 1 所示。在聚类阶段，首先使用主干网络F_θ提取图像特征，然后采用 DBSCAN[26] 算法对这些特征进行聚类。DBSCAN 聚类算法是基于密度的空间聚类算法，通过识别高密度区域来发现数据中的聚类结构，其核心思想是通过确定每个数据点周围的邻域内是否存在足够数量的其他数据点来判断是否为核心点、边界点或噪声点。相较于 K-means 等与距离或几何形状相关的聚类算法，具有不受簇形状限制、处理噪声点、自适应簇大小和区域密度、不需要设置初始化簇中心等优点。在无监督行人重识别任务中，行人的姿态、外观等因素变化较大，DBSCAN 算法的自适应性和对噪声的鲁棒性能够有效适应行人的多样性姿势、形状和密度。聚类的结果用于为每张图像分配伪标签$Y=\{y_i\}_{i=1}^{N}$，其中$y_i\in\mathbb{R}^K$，K表示集群数量。同时，初始化聚类内存字典 (cluster memory bank，$M_{\rm clu}$) 和相机感知内存字典 (camera-aware memory bank，$M_{\rm cam}$)。在训练阶段，对于每个训练实例x_i，首先通过主干网络F_θ和分类器获得其特征向量$v_i\in\mathbb{R}^d$和预测向量$z_i\in\mathbb{R}^K$。然后根据特征空间中实例之间的相似性，寻找每个实例的邻域，并使用邻域内样本的预测标签对伪标签进行细化。细化伪标签用于指导相机感知对比损失的学习。在损失优化阶段，采用聚类对比 (cluster contrastive，CC) 损失[4]、细化伪标签引导的相机感知 (refined pseudo-labels guided camera-aware contrastive，RPG-CAC) 损失以及改进后的交叉熵 (cross entropy, CE) 损失对模型进行联合优化。

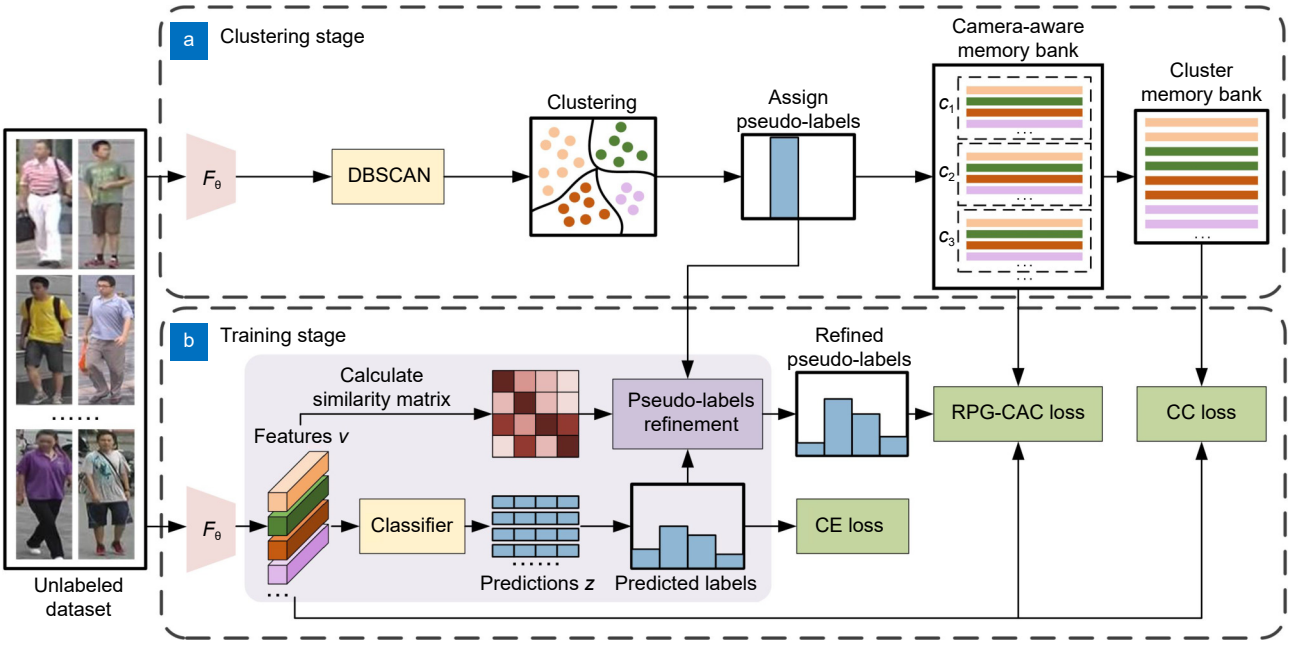


图 1　网络整体框架

Fig. 1　The overall framework of our method

2.2 最近邻伪标签细化

无监督行人重识别模型的性能高度依赖于聚类结果生成的伪标签，但是在聚类过程中不可避免地出现错误，比如将多个行人的图像合并到一类中，或将一个行人的图像分配到多个类中。使用这些被错误分配的伪标签进行训练，网络很容易对训练数据过拟合并过度适应噪声标签。传统的 one-hot 标签更加重了这一问题，因为 one-hot 标签仅用元素 1 表示所属类别，其他元素都为 0，每个类别之间是相互独立的。而无监督本身聚类不准确，one-hot 标签会导致信息的严重丢失，因为它忽略了类别之间的相似性和关联性。但在实际场景中，行人图像可能存在跨类别的共享特征，例如不同行人可能穿着相似的衣服或处于相似的环境中，因此传统的 one-hot 标签会导致模型无法充分捕捉这些重要的共享信息，限制了模型的性能。

受到标签传播思想[27]的启发，本文提出了一种基于最近邻细化的伪标签指导网络训练的方法。标签传播通常采用每个数据点与其相邻节点的相似性来分配标签。这种相似性通常是基于特征的相似性计算的，比如欧氏距离、余弦相似度等。不同于传统的 one-hot 标签，标签传播生成的标签通常为软标签，表示每个数据点属于每个类别的置信度或概率。在本文中，存储在内存字典$M_{\rm clu}$中的中心特征与每个类别的语义信息密切相关，它们代表了不同类别的关键特征。将这些中心特征作为分类层的权重矩阵有助于提高模型分类的性能，并降低模型对噪声数据过拟合的风险。为了进行有效的训练，本文使用了交叉熵损失，如式 (1)：

$$L_{\rm ce}=-\frac{1}{b}\sum_{i=1}^{b}\ell(z_i,y_i)\,, \tag{1}$$

其中：z_i表示分类器对图像x_i的特征向量的预测标签，分类器由一个全连接层和一个 softmax 函数组成。b表示 mini-batch 的大小，$\ell(\cdot)$表示交叉熵损失。

为了确定每个样本的邻域，使用余弦相似度来计算样本在特征空间中的两两相似性，如式 (2)：

$$\cos(v_i,v_j)=\frac{v_i^{\rm T}v_j}{\|v_i\|\bullet\|v_j\|}\,, \tag{2}$$

其中：v_i表示图像x_i的特征向量，$\|\cdot\|$表示 L2 范数。余弦相似度的值越大，表示两个实例在特征空间中的相似性越高。

采用余弦相似度最高的前n个样本作为x_i的最近邻。然后，利用这些最近邻样本进行伪标签细化，如式 (3)：

$$\tilde{y}_i=\alpha y_i+(1-\alpha)\sum_{z_j\in N(x_i,n)}\frac{1}{n}z_j\,, \tag{3}$$

其中：$z_j\in N(x_i,n)$表示实例x_i邻域中的第 j 个样本的预测标签，$1/n$表示每个预测标签z_j的权重，α是控制原始伪标签y_i和其预测标签之间的插值程度。

如图 2 所示，X_1，X_2和X_3三张图像为同一个身份，外观相似的图像 Y 实际为另一身份。由于外观及视角的相似性，聚类算法错误地将X_1和Y分为了一类。这些噪声伪标签造成的错误会在训练过程中不断被放大，阻碍特征的学习。为了有效减轻噪声伪标签的影响，本文在特征空间中采用余弦相似度计算样本之间的相似性，找到X_1的邻域，包括 Y，X_2，X_3三张图像。根据模型给出的邻域内样本的预测标签，利用式 (3) 对原始的 one-hot 标签进行细化，细化伪标签以概率分布的形式呈现了样本属于不同类别的可能性，能够在训练过程中为模型提供更加丰富的类别信息，并指导相机感知对比学习中正负样本的选择。

采用细化伪标签进行交叉熵计算，修改后的交叉熵损失公式为

$$\tilde{L}_{\mathrm{ce}}=\frac{1}{b}\sum_{i=1}^{b}\ell(z_i,\tilde{y}_i)\,. \tag{4}$$

2.3 细化伪标签引导的相机感知对比学习

为降低不同相机之间视角、光照等因素造成的类内差异，Wang 等人[9] 提出相机感知对比学习方法，进一步将聚类后的集群按照相机标签细分为多个子集群，使模型能够更好地捕捉不同相机下的微小特征差异。对于每个图像x_i，分别计算相机内和相机间对比损失。相机内对比损失的目标是将x_i向同一相机中的正类别中心靠近，同时将其与同一相机中其他类别中心推远。相机间对比损失则考虑了x_i与所有相机中的类别中心之间的关系，其目标是将x_i向所有相机中的正类别中心靠近，同时将其与在所有相机中挖掘出的困难负类别中心推远。

然而，这种方法易受到伪标签噪声的干扰。由于聚类结果的不准确性，可能会出现以下情况：推开假阴性样本对或者拉近假阳性样本对。如图 3(a) 所示，橙色标记的圆形实例在聚类阶段被错误地分配到了三角形类别中，原始的相机内对比损失将其靠近类别中心B_1、远离类别中心A_1，但实际应将其靠近类别中心A_1、远离类别中心B_1。这种情况会使模型受到严重的误导，混淆相机内不同类别之间的差异性信息。在相机间对比学习中，这一问题会更加严重，如图 3(c) 所示，橙色标记的圆形实例会被错误地拉近或者推远到多个不正确的类别中心。

因此，本文提出了一种细化伪标签引导的相机感知对比损失。具体做法为，首先使用相机标签和伪标签构建相机感知内存字典$M_{\mathrm{cam}}\in\mathbb{R}^{C\times d}$，其中$C$表示所有相机中心的总数，$d$表示特征维度。以第 i 个集群为例，具有相机标签 j 的中心c_{ij}计算如式 (5)：

$$c_{ij}=\frac{1}{|H_{ij}|}\sum_{v_l\in H_{ij}}v_l\,, \tag{5}$$

其中：H_{ij}表示集群 i 中相机标签为 j 的实例集合，$|H_{ij}|$表示该集合中的实例总数。

然后采用细化伪标签动态地选取实例的正负样本集合，而不再依赖传统的 one-hot 噪声标签选取。与实例x_i具有相同伪标签的正类别中心$\tilde{c}_i^{+}$计算如式 (6)：

$$\tilde{c}_i^{+}=\sum_{j\in S_i}w_j c_j\,, \tag{6}$$

其中：S_i表示根据细化伪标签中的类别概率选取的x_i的前 m 个相似类的类别中心的集合，c_j表示x_i的第 j 个相似类的类别中心。$w_j=\mathrm{softmax}(\mathrm{top}_m(\tilde{y}_i))$表示$x_i$的第 j 个相似类类别中心的权重，其中$\mathrm{top}_m(\tilde{y}_i)$表示根据$x_i$的细化伪标签得到的前 m 个相似类的概率，softmax()表示归一化函数。

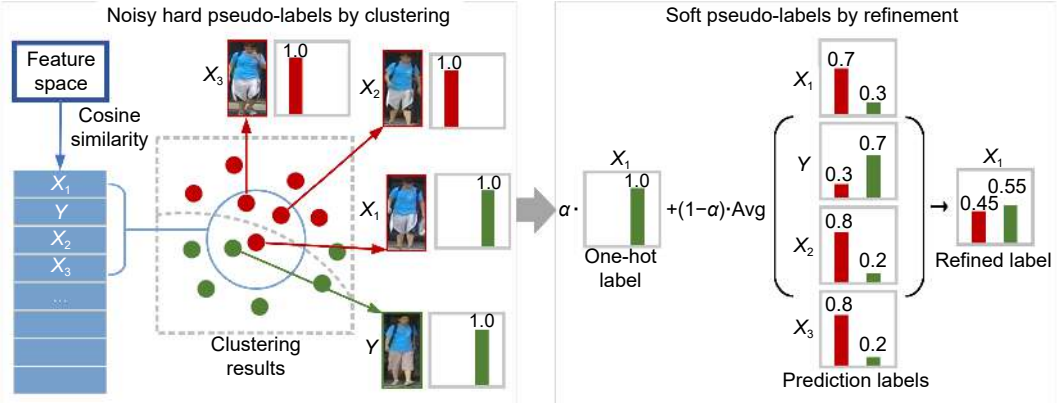


图 2 邻域伪标签细化模块

Fig. 2 Neighborhood pseudo label refinement module

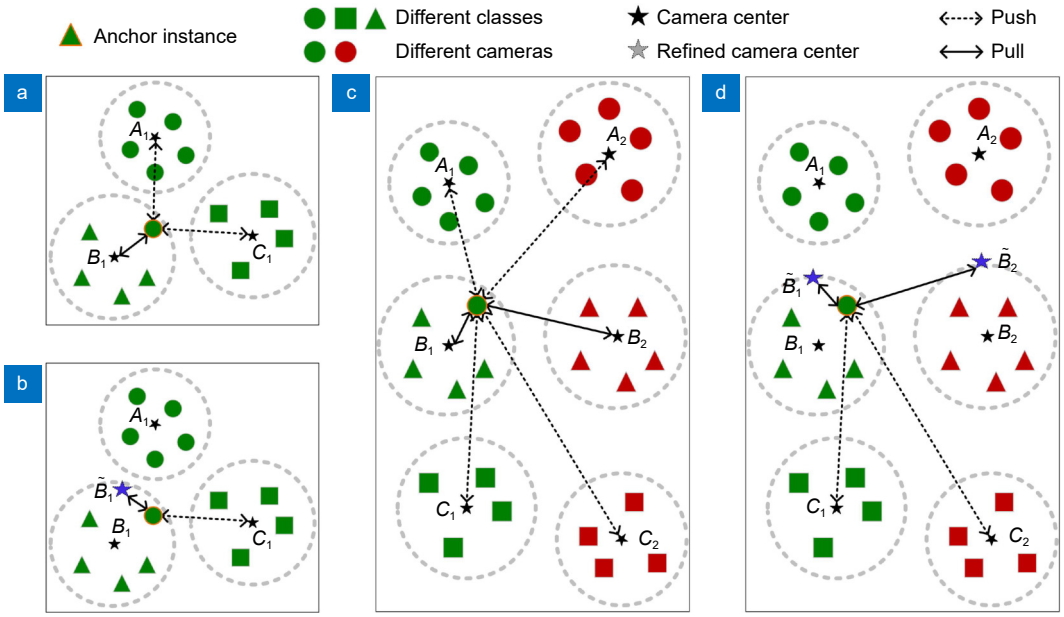


图 3 细化伪标签引导相机感知示意图。(a) 原始相机内对比; (b) 校正后相机内对比; (c) 原始相机间对比; (d) 校正后相机间对比

Fig. 3 Schematic diagram of camera-aware guided by refined pseudo-labels. (a) Original intra-camera contrast; (b) Corrected intra-camera contrast; (c) Original inter-camera contrast; (d) Corrected inter-camera contrast

如图 3(b) 所示，在相机内对比学习中，橙色标记的圆形实例不再被错误地靠近类别中心B_1，而是向平衡后的类别中心$\tilde{B}_1$靠近，同时不再远离类别中心A_1。计算公式如式 (7)：

$$\tilde{L}_{\text{intra}}=-\sum_{i=1}^{N}\log\frac{\exp((\tilde{c}^{+})^{\top}\cdot v_i/\tau_{\text{intra}})}{\sum\limits_{l\in\tilde{P}_i^{\text{intra}}\cup\tilde{Q}_i^{\text{intra}}}\exp(c_l^{\top}\cdot v_i/\tau_{\text{intra}})}, \tag{7}$$

其中：$\tilde{P}_i^{\text{intra}}=\{\tilde{c}^{+}\}$表示相机内与$x_i$具有相同伪标签的正类别中心集合，$\tilde{Q}_i^{\text{intra}}$表示相机内与$x_i$伪标签不同的负类别中心集合。$\tau_{\text{intra}}$表示相机内温度系数。

相机间对比学习采用类似的操作，如图 3(d) 所示，橙色标记的圆形实例不再被错误地靠近类别中心B_1和B_2，而是向平衡后的类别中心$\tilde{B}_1$和$\tilde{B}_2$靠近，同时也不再远离类别中心A_1和A_2。计算公式如式 (8)：

$$\tilde{L}_{\text{inter}}=-\sum_{i=1}^{N}\frac{1}{|\tilde{P}_i^{\text{inter}}|}\sum_{j\in\tilde{P}_i^{\text{inter}}}\log\frac{\exp((\tilde{c}_j^{+})^{\top}\cdot v_i/\tau_{\text{inter}})}{\sum\limits_{l\in\tilde{P}_i^{\text{inter}}\cup\tilde{Q}_i^{\text{inter}}}\exp(c_l^{\top}\cdot v_i/\tau_{\text{inter}})}, \tag{8}$$

其中：$\tilde{P}_i^{\text{inter}}=\{\tilde{c}_j^{+}\}_{j=1}^{N_{\text{pos}}}$表示所有相机中与$x_i$具有相同伪标签的$N_{\text{pos}}$个正类别中心集合，$\tilde{c}_j^{+}$表示$\tilde{P}_i^{\text{inter}}$中第$j$个类别中心，$\tilde{Q}_i^{\text{inter}}$表示所有相机中与$x_i$伪标签不同，但在特征空间中与$x_i$相似度最高的$N_{\text{neg}}$个负困难类别中心集合。$\tau_{\text{inter}}$表示相机间温度系数。

平衡后的相机中心对噪声伪标签没有那么敏感，修正了正负样本的选择，有助于提高模型的稳健性。细化伪标签引导的相机感知对比损失公式为

$$L_{\text{RPG-CAC}}=\tilde{L}_{\text{inter}}+\lambda\tilde{L}_{\text{intra}}, \tag{9}$$

其中，λ是控制相机内和相机间损失平衡的权重参数。

2.4 损失函数

本文方法的整体损失函数是对聚类对比 (CC) 损失、细化伪标签引导的相机感知 (RPG-CAC) 损失以及改进后的交叉熵 (CE) 损失的加权和：

$$L=L_{\text{CC}}+\beta L_{\text{RPG-CAC}}+\tilde{L}_{\text{ce}}, \tag{10}$$

其中，β是控制相机感知对比损失重要性的权重参数。

3 实验结果与分析

3.1 数据集与评价指标

为了评估模型效果，本文使用三个规模行人重识别公开数据集进行了一系列实验，分别是 Market-1501、MSMT17、Personx。Market-1501 数据集是在清华大学校园中采集的，包含了来自 6 台相机、1501 名行人的 32668 张图像。MSMT17 数据集包含了来自 15 台相机 (3 台室内相机、12 台室外相机)、4101 个行人的 126411 张图片。Personx 数据集包含了来自 6 个相机、1266 个行人的 45792 张图像。

本文采用累计匹配特性 (cumulative matching characteristics，CMC) 以及平均准确率均值 (mean average precision，mAP) 作为评价指标。CMC 用于衡量前 K 幅图像匹配成功的概率，本文采用的是前 1 幅、5 幅、10 幅图像匹配成功的概率，分别记为 Rank-1、Rank-5、Rank-10。每个查询图像的平均准确率是通过准确率-召回率曲线计算得出的，mAP 则表示所有查询图像平均准确率的均值。

3.2 实验设置

本文进行实验的硬件环境如下：操作系统为 Ubuntu16.04，使用 2 张 NVIDIA 1080TI GPU 显卡，每张显卡拥有 12 GB 显存。使用 Pytorch 框架搭建整个网络，以在 ImageNet[28] 上预训练的 ResNet50[29] 网络作为特征提取的主干网络，但删除了第 4 层之后的所有层，然后添加了广义平均池化 (generalized mean pooling，GeM) 层和使用 BNNeck 的全连接层。

在训练过程中，将数据集的图像大小调整为 256×128，然后执行了随机水平翻转、像素填充、随机裁剪和随机擦除多项数据增强操作。Batch size 设置为 128，每个批次包含 16 个伪身份，每个身份包含 8 个实例。采用权重衰减为5×10^{-4}的 Adam 优化器更新模型梯度。总共进行 50 个 epoch 的训练，初始学习率设置为3.5×10^{-4}，然后每 20 个 epoch 将学习率缩减为之前的 1/10。每个 epoch 开始时，使用 DBSCAN 算法和基于 k-相互近邻的杰卡德距离[30] 进行聚类以生成伪标签，并使用与文献 [3] 相同的参数设置。相机感知 InfoNCE 损失中温度系数τ_{intra}、τ_{inter}分别设置为 0.05、0.07，困难负相机中心数N_{neg}设置为 50。改进后交叉熵损失中邻域大小n设置为 7，控制伪标签细化程度的参数α设置为 0.3。损失平衡参数λ、β分别设置为 0.6、0.5。以上参数的实验设置分析见 3.3.4 节。测试时，仅对图像大小进行调整，并使用 GeM 池化层的特征来计算距离。

3.3 实验结果

3.3.1 与最新方法的比较

为了验证本文方法的有效性，本文将实验结果与近几年最新的方法进行了比较，包括 UDA 方法和 USL 方法。Market-1501、MSMT17 和 Personx 三个数据集上的比较结果如表 1 所示，其中最优结果加粗表示，次优结果加下划线表示，“-”表示原论文中没有该项结果，其中 SPCL[3] 和 MMT[12] 方法在 Personx 数据集上的结果是基线方法 CC[4] 复现的结果。

如表 1 上半部分所示，即使没有使用额外的标记源域数据集，本文方法也在 3 个数据集上取得了优越的性能。相较于次优模型 CaCL[11]，对于 mAP 和 Rank-1 指标，本文方法在 Market-1501 上分别提升了 0.5%、0.6%，在 MSMT17 上分别提升了 4.0%、4.1%。表 1 下半部分提到的 CC[4] 是本文的基线论文，本文使用的参数设置与基线论文存在一些差异，主要是 Batch size 的大小、显卡的数量和型号。尽管存在这些差异，本文方法在三个数据集上的性能表现都超过了大部分 USL 方法，但在 Market-1501 数据集上略低于 LP[18] 和 DCCT[19] 方法。Market-1501 数据集中每个身份出现较少相机中且数据集总体规模相对较小，降低了相机差异的挑战性，尽管本文方法着重处理相机间的差异问题，但也因此提升较小。而 MSMT17 数据集涵盖了更多的相机视角和环境，其身份更加容易在多个相机中重叠，即同一身份出现在更多的相机中。由于本文着重于处理相机间的差异问题，因此在 MSMT17 数据集上取得了更为显著的性能提升，这表明本文方法在处理更具挑战性的场景，尤其是涉及相机视角和环境差异的情况下，具有更强的泛化能力。

与同样专注于解决相机风格影响的 MetaCam[7]、IICS[8]、CAP[9]、CA-UReID[10]、CaCL[11] 方法相比，本文方法具有明显的优势。与 MetaCam[7] 相比，本文模型在 Market-1501 上 mAP 和 Rank-1 分别提升了 23.5%、10.5%，MSMT17 上 mAP 和 Rank-1 分别提升了 28.8%、38.9%。与模型 IICS[8] 相比，本文模型在 Market-1501 上 mAP 和 Rank-1 分别提升了 12.3%、4.9%，MSMT17 上 mAP 和 Rank-1 分别提升了 17.4%、17.7%。与模型 CAP[9] 相比，本文模型在 Market-1501 上 mAP 和 Rank-1 分别提升了 6.0%、3.0%，MSMT17 上 mAP 和 Rank-1 分别提升了 7.4%、6.9%。与模型 CA-UReID[10] 相比，本文模型在 Market-1501 上 mAP 和 Rank-1 分别提升了 0.7%、0.3%。与模型 CaCL[11] 相比，本文模型在 Market-1501 上 mAP 和 Rank-1 分别提升了 0.5%、0.6%，MSMT17 上 mAP 和 Rank-1 分别提升了 4.0%、4.1%。总体而言，本文所提方法在处理场景复杂的大规模数据集 MSMT17 上表现出较明显的性能提升，体现了以细化伪标签引导相机感知对比学习在处理相机差异方面的优越性。

表 1 本文方法与最新方法的比较

Table 1 The comparison between the our method and the latest methods

Methods		Market-1501		MSMT17		Personx	
		mAP/%	Rank-1/%	mAP/%	Rank-1/%	mAP/%	Rank-1/%
Unsupervised domain adaptation (UDA)							
ECN[2]	CVPR'19	43.0	75.1	10.2	30.2	-	-
SPCL[3]	NeurIPS'20	77.5	89.7	26.8	53.7	78.5	91.1
MEB-Net[13]	ECCV'20	76.0	89.9	-	-	-	-
MMT[12]	CVPR'21	71.2	87.7	23.5	50.0	78.9	90.6
GLT[31]	CVPR'21	79.5	92.2	27.7	59.5	-	-
MCL[32]	JBUAA'22	80.6	93.2	28.5	58.5	-	-
CACHE[33]	TCSVT'22	83.1	93.4	31.3	58.0	-	-
CIFL[20]	TMM'22	83.3	93.9	39.0	70.5	-	-
MCRN[16]	AAAI'22	83.8	93.8	35.7	67.5	-	-
IICM[34]	JCRD'23	74.9	89.0	27.2	52.3	-	-
NPSS[21]	TIFS'23	84.6	94.1	38.9	69.4	-	-
CaCL[11]	ICCV'23	84.7	93.8	40.3	70.0	-	-
Unsupervised learning (USL)							
SPCL[3]	NeurIPS'20	73.1	88.1	19.1	42.3	72.3	88.1
MetaCam[7]	CVPR'21	61.7	83.9	15.5	35.2	-	-
IICS[8]	CVPR'21	72.9	89.5	26.9	56.4	-	-
RLCC[14]	CVPR'21	77.7	90.8	27.9	56.5	-	-
CAP[9]	AAAI'21	79.2	91.4	36.9	67.4	-	-
ICE[17]	ICCV'21	82.3	93.8	38.9	70.2	-	-
CACHE[33]	TCSVT'22	81.0	92.0	31.8	58.2	-	-
CIFL[20]	TMM'22	82.4	93.9	38.8	70.1	-	-
GRACL[35]	TCSVT'22	83.7	93.2	34.6	64.0	87.9	95.3
PPLR[15]	CVPR'22	84.4	94.3	42.2	73.3	-	-
CA-UReID[10]	ICME'22	84.5	94.1	-	-	-	
NPSS[21]	TIFS'23	82.3	94.0	36.7	68.8	-	-
LRMGFS[36]	JEMI'23	83.3	93.3	27.4	58.4	-	-
PLRIS[22]	ICIP'23	83.2	93.1	43.3	71.5	-	-
AdaMG[37]	TCSVT'23	84.6	93.9	38.0	66.3	87.6	95.0
LP[18]	TIP'23	85.8	**94.5**	39.5	67.9	-	-
DCCT[19]	TCSVT"23	**86.3**	94.4	41.8	68.7	87.6	95.0
CC[4]	CoRR'21	82.1	92.3	27.6	56.0	84.7	94.4
Ours	-	85.2	94.4	**44.3**	**74.1**	**88.7**	**95.9**

与同样进行伪标签细化工作的 SPCL[3]、MMT[12]、MEB-Net[13]、RLCC[14]、PPLR[15]、MCRN[16]、ICE[17]、LP[18]、DCCT[19] 相比，本文方法性能显著提升。与其中综合泛化性能较好的模型 PPLR[15] 相比，本文方法在 Market-1501 上 mAP 和 Rank-1 分别提升了 0.8%、0.1%，MSMT17 上 mAP 和 Rank-1 分别提升了 2.1%、0.8%。与 DCCT[19] 相比，本文方法在 MSMT17 上 mAP 和 Rank-1 分别提升了 2.5%、5.4%，Personx 上 mAP 和 Rank-1 分别提升了 1.1%、0.9%。

与同时考虑相机影响和伪标签噪声两个问题的方法 CIFL[20]、NPSS[21]、PLRIS[22] 相比，本文也具有显著的优越性。与 PLRIS[22] 相比，在 Market-1501 上 mAP 和 Rank-1 分别提升了 3.0%、1.3%，MSMT17 上 mAP 和 Rank-1 分别提升了 1.0%、2.6%。

3.3.2 消融实验

本文在 Market-1501 和 Personx 数据集上进行了一系列消融实验，来证明本文网络中各个模块的有效性，实验结果如表 2 所示。其中，“M1”表示基线模型 CC[4] 的结果，“M2”、“M3”、“M4”、“M5”分别表示在基线模型 CC[4] 上添加本文所提方法中各个模块后的消融实验结果，“M6”表示本文完整方法，即在基线模型 CC[4] 上同时添加三个损失的结果。将“M6”与“M1”进行对比，可以看出，本文方法相较于基线模型 CC[4] 表现出显著的性能提升。具体来说，最近邻伪标签细化模块在 Market-1501 数据集上将 mAP 和 Rank-1 分别提升了 2.6%、1.9%。这表明最近邻伪标签细化方法成功地减轻了噪声标签的不利影响，有效地提升了模型性能。在 Personx 数据集上也将 mAP 和 Rank-1 分别提升了 2.8%、1.1%，证明了该模块的有效性。

相机内对比学习模块在 Market-1501 数据集上将 mAP 和 Rank-1 分别提升了 0.8%、0.9%，在 Personx 数据集上将 mAP 和 Rank-1 分别提升了 3.2%、1.3%，这证明了相机内对比学习相较于全局对比学习对噪声标签的敏感性较低，因此表现出更好的性能。相机间对比学习模块在 Market-1501 数据集上将 mAP 和 Rank-1 分别提升了 1.8%、1.3%，在 Personx 数据集上将 mAP 和 Rank-1 分别提升了 2.6%、1.5%，证明了它能够通过充分利用相机间的相关性来改善行人重识别模型，不仅可以使正相机中心更加聚集，从而降低类内方差，还能够使困难负相机中心更加分散，解决了类间行人图像的相似性问题。整体相机感知对比学习模块在 Market-1501 数据集上将 mAP 和 Rank-1 分别提升了 2.0%、1.3%，在 Personx 数据集上将 mAP 和 Rank-1 分别提升了 3.8%、1.4%。

3.3.3 可视化分析

为了更直观地分析本文模型的检索效果，在 Market-1501 数据集上进行了 Rank-10 可视化排序实验。如图 4 所示，随机选择了 6 张查询图像，分别在基线模型 CC[4]、最相关方法 CAP[9]、最先进方法 PPLR[15] 以及本文模型上进行了可视化实验，其中没有边框的图像代表查询图像，带有绿色边框的图像表示正确匹配的图像，而带有红色边框的图像则表示错误匹配的图像。从图 4(a-c) 中可以观察到，当不是同一身份的行人具有相似的外观时，基线模型容易发生错误检索，特别是第 3、4、6 个查询实例，其模型检索结果绝大部分都是错误的，而本文提出的模型显著地改善了这一情况。如图 4(d) 所示，本文模型更能有效地区分在视觉上相似的行人图像，从而提升了检索性能，这证明了本文方法的有效性。

为体现本文方法的有效性，采用 T-SNE 方法可视化了基线模型 CC[4]、最相关方法 CAP[9]、最先进方法 PPLR[15] 以及本文模型学习到的特征表示。图 5 展示了在 Market-1501 数据集中随机抽取的 10 个行人的图像特征分布图，其中不同颜色表示不同的行人，不同形状表示不同的相机。通过观察图 5 可以看出，基线模型 CC[4]、CAP[9] 方法和 PPLR[15] 方法均不能很好地区分身份标签为 43、62、64、67 的行人，尤其是 CAP[9] 方法，而本文模型能更好地区分这些身份。此外，对于身份标签为 15 和 17 的行人，基线模型 CC[4] 和 PPLR[15] 很明显地将身份标签为 17 号相机标签为 5 号的图像错分为 15 号，而本文模型则显著地改善了这一情况。这表明本文模型显著提高了类内样本的紧凑性和类间样本的可分性，从而减少了由相机风格差异引起的类内差异。表 1 中的实验结果也验证了这一点，完整的本文模型相较于基线模型在 Market-1501 数据集上将 mAP 和 Rank-1 分别提升了 3.1%、

表 2 Market-1501 数据集上消融实验结果

Table 2 Results of ablation studies on Market-1501

	$\widetilde{L}_{ce}$	$\widetilde{L}_{intra}$	$\widetilde{L}_{inter}$	Market-1501				Personx			
				mAP/%	Rank-1/%	Rank-5/%	Rank-10/%	mAP/%	Rank-1/%	Rank-5/%	Rank-10/%
M1(CC[4])	-	-	-	82.1	92.3	96.7	97.9	84.7	94.4	98.3	99.3
M2	-	√	-	82.9	93.2	97.3	98.2	87.9	95.7	98.8	99.5
M3	-	-	√	83.9	93.6	97.5	98.3	87.3	95.9	98.8	99.4
M4	-	√	√	84.1	93.6	97.6	98.5	88.5	95.8	98.9	99.6
M5	√	-	-	84.7	94.2	97.9	98.7	87.5	95.5	98.9	99.5
M6(Ours)	√	√	√	85.2	94.4	98.1	98.6	88.7	95.9	99.2	99.7

图 4 不同方法在 Market-1501 数据集上 Top-10 排序列表的比较。(a) Baseline 方法; (b) CAP[9] 方法; (c) PPLR[15] 方法; (d) 本文方法

Fig. 4 Comparison of Top-10 ranking lists between on Market-1501 dataset among different methods. (a) Baseline method; (b) CAP[9] method; (c) PPLR[15] method; (d) Our method

2.1%，在 MSMT17 数据集上将 mAP 和 Rank-1 分别提升了 16.7%、18.1%，在 Personx 数据集上将 mAP 和 Rank-1 分别提升了 4.0%、1.5%。针对 MSMT17 数据集这个更加复杂的场景，本文模型提升尤为明显，该数据集涵盖了更多的场景和行人外观变化，表明了本文模型在处理严重类内差异和标签噪声方面的优越性。

3.3.4 超参数分析

为了深入研究超参数对模型性能的影响，本文在 Market-1501 数据集上进行了一系列实验。在仅添加伪标签细化模块的实验设置下，本文观察了在α的不同取值下，最终标签的来源对模型性能的影响，实验结果如图 6(a) 所示。当$\alpha = 0.3$时，模型取得较好结果，mAP 为 84.7%，Rank-1 为 94.2%。这表明合理地结合原始聚类结果得到的伪标签和最近邻的平均预测标签有助于提高模型的性能，伪标签细化的必要性和有效性。此外，本文还研究了邻域大小 n 对模型性能的影响。实验结果如图 6(b) 所示，当$n = 7$时模型取得最好的性能。若n过小，模型难以捕捉相邻样本的类别信息；若n过大，则可能会从标签信息中引入更多的干扰。

在实验中，本文还重点分析了相机感知 InfoNCE 损失中的温度系数对模型性能的影响。先前的研究 Wang 等人[38] 提出对比损失函数具有困难负样本自发现的性质，即对于那些已经远离的样本，不再继续将它们推离，而更注重如何使那些靠近但是被错误匹配

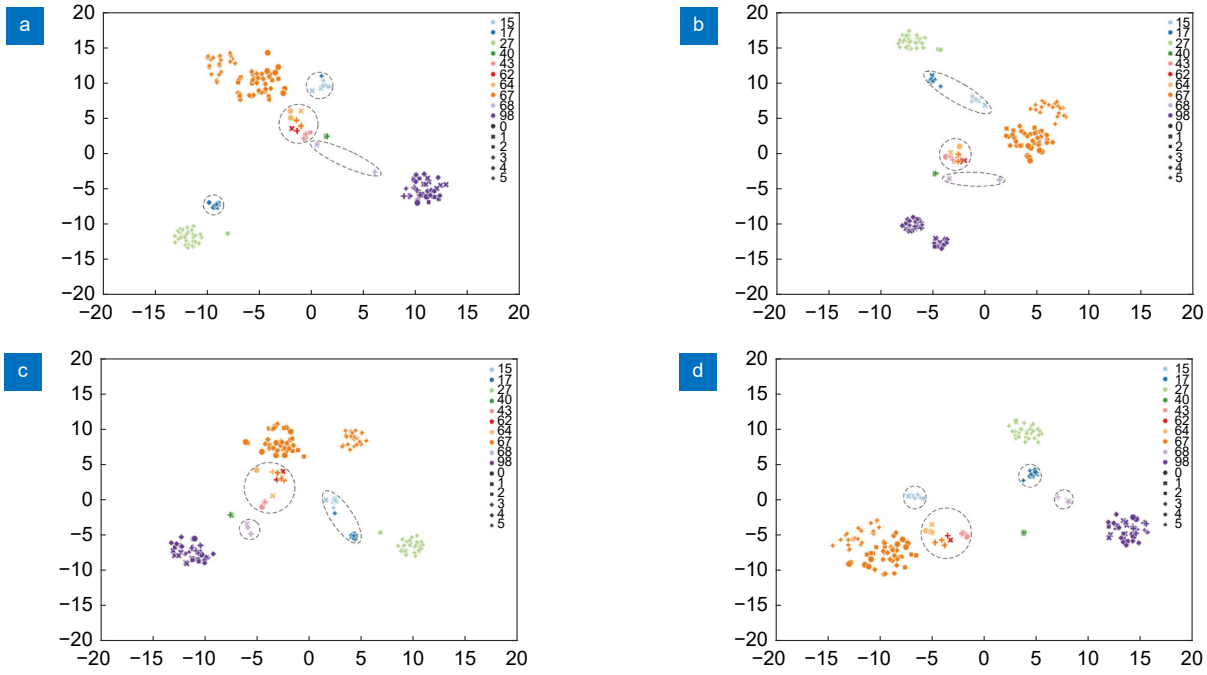


图 5 不同方法在 Market-1501 数据集上特征 T-SNE 可视化结果。(a) Baseline 方法; (b) CAP[9] 方法; (c) PPLR[15] 方法; (d) 本文方法

Fig. 5 Feature T-SNE visualization results of different methods on Market-1501 dataset. (a) Baseline method; (b) CAP[9] method; (c) PPLR[15] method; (d) Our method

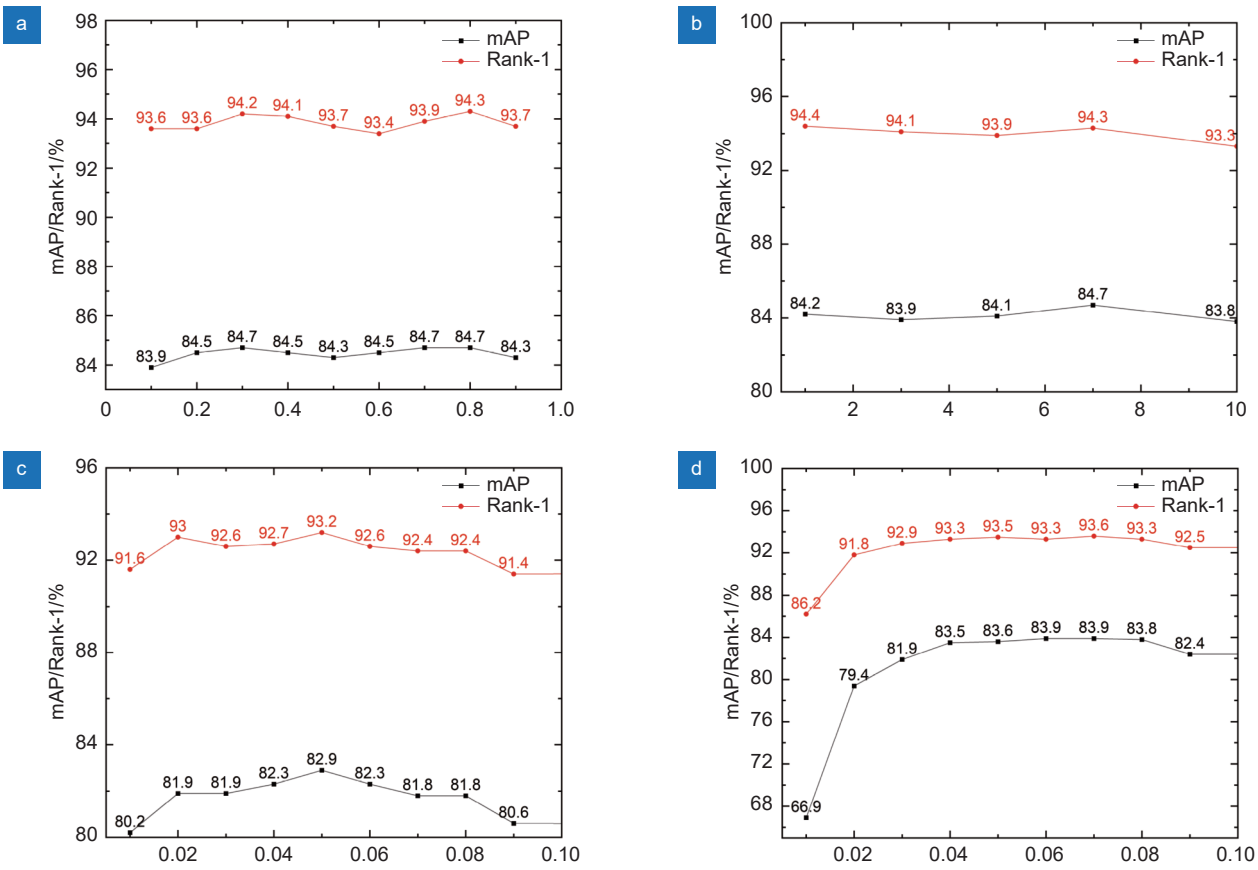


图 6 各超参数在 Market-1501 数据集上对本文模型的影响。(a) α; (b) k; (c) τ_{intra}; (d) τ_{inter}

Fig. 6 The impact of each hyperparameter to our model on Market-1501. (a) α; (b) k; (c) τ_{intra}; (d) τ_{inter}

的困难负样本推离正样本，以使表示空间更加均匀。在这个背景下，温度系数起到了关键的作用，它决定了对比损失对困难负样本的关注程度。温度系数越小，损失越关注与正样本相似但是被错误匹配的困难负样本，从而给予这些困难负样本更大的梯度，将它们与正样本分离。相机域特征空间的偏移导致相机内与相机间匹配的平均成对相似度不一致。具体而言，相机内负样本对比相机间正样本对更容易聚集在一起。在这种情况下，相机内的困难负样本应该受到更多的关注，这也与本文的实验结果相符。实验结果如图 6(c) 所示，当相机内对比损失的温度系数τ_{intra}取 0.05 时，模型表现最佳；如图 6(d) 所示，当相机间对比损失的温度系数τ_{inter}取 0.07 时，模型性能最佳。

4 结　论

本文提出了一种伪标签细化引导的相机感知无监督行人重识别方法。该方法首先根据训练实例在特征空间中的相似性，为每个实例确定邻域集合。然后将模型对邻域内样本的预测标签与实例原始聚类结果进行加权组合，以细化传统的 one-hot 伪标签。最后采用这些细化后的伪标签指导相机感知对比学习，动态地关联实例可能属于的多个正类别中心，而不再依赖于单一的集群中心，同时过滤可能存在的假阴性样本，从而减少噪声伪标签对相机感知的误导。通过在三个大规模数据集上进行与现有方法的对比实验，证明了本文方法的优越性。

利益冲突: 所有作者声明无利益冲突

参考文献

[1] Zhang X Y, Zhang B H, Lv X Q, et al. The joint discriminative and generative learning for person re-identification of deep dual attention[J]. *Opto-Electron Eng*, 2021, **48**(5): 200388.
张晓艳, 张宝华, 吕晓琪, 等. 深度双重注意力的生成与判别联合学习的行人重识别[J]. 光电工程, 2021, **48**(5): 200388.

[2] Zhong Z, Zheng L, Luo Z M, 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*, 2019: 598–607. https://doi.org/10.1109/CVPR.2019.00069.

[3] Ge Y X, Zhu F, Chen D P, et al. Self-paced contrastive learning with hybrid memory for domain adaptive object re-ID[C]//*Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020: 949. https://doi.org/10.5555/3495724.3496673.

[4] Dai Z Z, Wang G Y, Yuan W H, et al. Cluster contrast for unsupervised person re-identification[C]//*Proceedings of the 16th Asian Conference on Computer Vision*, 2023: 319–337. https://doi.org/10.1007/978-3-031-26351-4_20.

[5] Tian J J, Tang Q H, Li R, et al. A camera identity-guided distribution consistency method for unsupervised multi-target domain person re-identification[J]. *ACM Trans Intell Syst Technol*, 2021, **12**(4): 38.

[6] 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/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 8789–8797. https://doi.org/10.1109/CVPR.2018.00916.

[7] Yang F X, Zhong Z, Luo Z M, et al. Joint noise-tolerant learning and meta camera shift adaptation for unsupervised person re-identification[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 4853–4862. https://doi.org/10.1109/CVPR46437.2021.00482.

[8] Xuan S Y, Zhang S L. Intra-inter camera similarity for unsupervised person re-identification[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 11921–11930. https://doi.org/10.1109/CVPR46437.2021.01175.

[9] Wang M L, Lai B S, Huang J Q, et al. Camera-aware proxies for unsupervised person re-identification[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*, 2021: 2764–2772. https://doi.org/10.1609/aaai.v35i4.16381.

[10] Li X, Liang T F, Jin Y, et al. Camera-aware style separation and contrastive learning for unsupervised person re-identification[C]//*2022 IEEE International Conference on Multimedia and Expo*, 2022: 1–6. https://doi.org/10.1109/ICME52920.2022.9859842.

[11] Lee G, Lee S, Kim D, et al. Camera-driven representation learning for unsupervised domain adaptive person re-identification[Z]. arXiv: 2308.11901, 2023. https://doi.org/10.48550/arXiv.2308.11901.

[12] Ge Y X, Chen D P, Li H S. Mutual mean-teaching: pseudo label refinery for unsupervised domain adaptation on person re-identification[C]//*8th International Conference on Learning Representations*, 2020.

[13] Zhai Y P, Ye Q X, Lu S J, et al. Multiple expert brainstorming for domain adaptive person re-identification[C]//*16th European Conference on Computer Vision*, 2020: 594–611. https://doi.org/10.1007/978-3-030-58571-6_35.

[14] Zhang X, Ge Y X, Qiao Y, et al. Refining pseudo labels with clustering consensus over generations for unsupervised object re-identification[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 3435–3444. https://doi.org/10.1109/CVPR46437.2021.00344.

[15] Cho Y, Kim W J, Hong S, et al. Part-based pseudo label refinement for unsupervised person re-identification[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2022: 7298–7308. https://doi.org/10.1109/CVPR52688.2022.00716.

[16] Wu Y H, Huang T T, Yao H T, et al. Multi-centroid representation network for domain adaptive person re-ID[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*, 2022: 2750–2758. https://doi.org/10.1609/aaai.v36i3.20178.

[17] Chen H, Lagadec B, Bremond F. ICE: inter-instance contrastive encoding for unsupervised person re-identification[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 14940–14949. https://doi.org/10.1109/ICCV48922.2021.01469.

[18] Lan L, Teng X, Zhang J, et al. Learning to purification for unsupervised person re-identification[J]. *IEEE Trans Image Process*, 2023, **33**: 3338−3353.

[19] Chen Z Q, Cui Z C, Zhang C, et al. Dual clustering co-teaching with consistent sample mining for unsupervised person re-identification[J]. *IEEE Trans Circuits Syst Video Technol*, 2023, **33**(10): 5908−5920.

[20] Pang Z Q, Zhao L L, Liu Q Y, et al. Camera invariant feature learning for unsupervised person re-identification[J]. *IEEE Trans Multimed*, 2023, **25**: 6171−6182.

[21] Wang H J, Yang M, Liu J L, et al. Pseudo-label noise prevention, suppression and softening for unsupervised person re-identification[J]. *IEEE Trans Inf Forensics Secur*, 2023, **18**: 3222−3237.

[22] Li P N, Wu K Y, Zhou S P, et al. Pseudo labels refinement with intra-camera similarity for unsupervised person re-identification[C]//*2023 IEEE International Conference on Image Processing*, 2023: 366–370. https://doi.org/10.1109/ICIP49359.2023.10222317.

[23] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification: a benchmark[C]//*Proceedings of the 2015 IEEE International Conference on Computer Vision*, 2015: 1116–1124. https://doi.org/10.1109/ICCV.2015.133.

[24] Wei L H, Zhang S L, 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*, 2018: 79–88. https://doi.org/10.1109/CVPR.2018.00016.

[25] Sun X X, Zheng L. Dissecting person re-identification from the viewpoint of viewpoint[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 608–617. https://doi.org/10.1109/CVPR.2019.00070.

[26] Ester M, Kriegel H P, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]//*Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 1996: 226–231. https://doi.org/10.5555/3001460.3001507.

[27] Zhou D Y, Bousquet O, Lal T N, et al. Learning with local and global consistency[C]//*Proceedings of the 16th International Conference on Neural Information Processing Systems*, 2003: 321–328. https://doi.org/10.5555/2981345.2981386.

[28] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009: 248–255. https://doi.org/10.1109/CVPR.2009.5206848.

[29] 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*, 2016: 770–778. https://doi.org/10.1109/CVPR.2016.90.

[30] Zhong Z, Zheng L, Cao D L, et al. Re-ranking person re-identification with k-reciprocal encoding[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 3652–3661. https://doi.org/10.1109/CVPR.2017.389.

[31] 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*, 2021: 5306–5315. https://doi.org/10.1109/CVPR46437.2021.00527.

[32] Li H, Zhang X W, Zhao X P, et al. Multi-label cooperative learning for cross domain person re-identification[J]. *J Beijing Univ Aeronaut Astronaut*, 2022, **48**(8): 1534−1542.
李慧, 张晓伟, 赵新鹏, 等. 基于多标签协同学习的跨域行人重识别[J]. 北京航空航天大学学报, 2022, **48**(8): 1534−1542.

[33] Liu Y X, Ge H W, Sun L, et al. Complementary attention-driven contrastive learning with hard-sample exploring for unsupervised domain adaptive person re-ID[J]. *IEEE Trans Circuits Syst Video Technol*, 2023, **33**(1): 326−341.

[34] Chen L W, Ye F, Huang T Q, et al. An unsupervised person re-Identification method based on intra-/inter-camera merger[J]. *J Comput Res Dev*, 2023, **60**(2): 415−425.
陈利文, 叶锋, 黄添强, 等. 基于摄像头域内域间合并的无监督行人重识别方法[J]. 计算机研究与发展, 2023, **60**(2): 415−425.

[35] Zhang H W, Zhang G Q, Chen Y H, et al. Global relation-aware contrast learning for unsupervised person re-identification[J]. *IEEE Trans Circuits Syst Video Technol*, 2022, **32**(12): 8599−8610.

[36] Qian Y P, Wang F S, Xiong L. Unsupervised person re-identification method based on local refinement multi-branch and global feature sharing[J]. *J Electron Meas Instrum*, 2023, **37**(1): 106−115.
钱亚萍, 王凤随, 熊磊. 基于局部细化多分支与全局特征共享的无监督行人重识别方法[J]. 电子测量与仪器学报, 2023, **37**(1): 106−115.

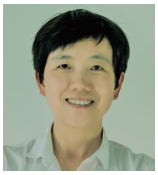
[37] Peng J J, Jiang G Q, Wang H B. Adaptive memorization with group labels for unsupervised person re-identification[J]. *IEEE Trans Circuits Syst Video Technol*, 2023, **33**(10): 5802−5813.

[38] Wang F, Liu H P. Understanding the behaviour of contrastive loss[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 2495–2504. https://doi.org/10.1109/CVPR46437.2021.00252.

作者简介

程思雨 (1999-)，女，硕士研究生，主要从事深度学习、行人重识别方面的研究。

E-mail: 2446297319@qq.com



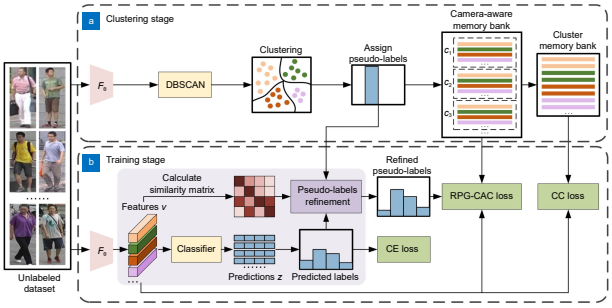
【通信作者】陈莹 (1976-)，女，博士，教授，博士生导师，主要从事机器视觉、信息融合、模式识别方面的研究。

E-mail: chenying@jiangnan.edu.cn

扫描二维码，获取PDF全文

Camera-aware unsupervised person re-identification method guided by pseudo-label refinement

Cheng Siyu, Chen Ying*



The overall framework of our method

Overview: Unsupervised person re-identification has received increasing attention due to its wide practical application prospects. Most clustering-based contrastive learning methods treat each cluster as a pseudo-identity class, focusing on improving inter-class differences while ignoring intra-class differences caused by factors such as perspective, lighting, and background between different cameras. This makes it difficult for clustering algorithms to accurately cluster samples with the same identity into the same cluster, inevitably leading to noisy pseudo-labels. Some methods have introduced camera-aware contrastive learning, which divide a single cluster into multiple sub-clusters based on the camera's perspective, and calculate the intra-camera and inter-camera contrastive loss separately. However, the noise in pseudo-labels may interfere with the selection of positive and negative samples in camera-aware contrastive learning, thereby misleading the model's learning process. To address this issue, this paper proposes a camera-aware unsupervised person re-identification method guided by refined pseudo-labels. By calculating the similarity between training instances in feature space, a neighborhood set is determined for each instance. Subsequently, the model refines one-hot pseudo-labels by combining the predicted labels for samples within the neighborhood with the original clustering results using weighted aggregation. The core idea behind this approach is to encourage the model to not only bring samples closer to their respective cluster centers but also establish associations with other nearby samples that may contain identity information. This strategy effectively enhances the model's robustness against noisy labels while reducing the risk of over-fitting. Building upon this, this paper further proposes camera-aware contrastive learning guided by refined pseudo-labels. By leveraging the probability distribution of each class in the refined pseudo-labels for instances, the model dynamically associates instances with potential class centers, no longer relying on a single class center as the positive sample. Additionally, potential false positive and false negative samples are filtered out. This method enhances the selection mechanism of positive and negative samples in camera-aware contrastive learning, effectively mitigating the influence of noisy pseudo-labels on the contrastive learning task. The method proposed in this article was validated on three large-scale public datasets, and the results showed that this method has significantly improved compared to the baseline method and is superior to current advanced methods in the same field. This method achieved mAP/Rank-1 of 85.2%/94.4%, 44.3%/74.1%, and 88.7%/95.9% on the Market-1501, MSMT17, and Personx datasets, respectively, demonstrating superiority. Specifically, on the Market-1501, MSMT17, and Personx datasets, this paper's method achieves mAP/Rank-1 scores of 85.2%/94.4%, 44.3%/74.1%, and 88.7%/95.9%, respectively, showcasing its superiority.

Cheng S Y, Chen Y. Camera-aware unsupervised person re-identification method guided by pseudo-label refinement[J]. *Opto-Electron Eng*, 2023, **50**(12): 230239; DOI: 10.12086/oee.2023.230239

Foundation item: Project supported by National Natural Science Foundation of China (62173160)

Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China

* E-mail: chenying@jiangnan.edu.cn