

基于相机感知距离矩阵的无监督行人重识别

白梦林, 周非, 舒浩峰

(重庆邮电大学通信与信息工程学院, 重庆 400065)

摘要: 在跨场景、跨设备的行人重识别中虽然增加了可利用的行人数据,但由于行人姿态不同、部分遮挡现象,难以避免引入样本噪声,在聚类过程中易生成错误的伪标签,造成标签噪声,影响模型的优化。为减弱噪声影响,应用相机感知的距离矩阵对抗相机偏移引起的样本噪声问题,利用对噪声鲁棒的动态对称对比损失减少标签噪声,提出基于相机感知距离矩阵的无监督行人重识别算法。在聚类前通过更改度量行人特征相似度的距离矩阵,利用相机感知距离矩阵来增强类内距离度量准确性,减少由于拍摄视角不同对聚类效果造成的负面影响。同时,结合噪声标签学习方法,进行损失设计,提出动态对称对比损失函数,联合损失训练,不断精炼伪标签。在 DukeMTMC-reID 和 Market-1501 两个数据集上进行实验,验证了提出方法的有效性。

关键词: 无监督行人重识别; 聚类; 距离矩阵; 标签噪声; 损失函数

中图分类号: TP391 **文献标志码:** A

Unsupervised Person Re-identification Based on Camera-Aware Distance Matrix

BAI Menglin, ZHOU Fei, SHU Haofeng

(School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: Cross-scene and cross-device shooting greatly increases the data of pedestrians. However, due to the different postures and partial occlusion of pedestrians, it is difficult to avoid the introduction of sample noise. During the clustering process, it is easy to generate false pseudo-labels, resulting in label noise and affecting the optimization of the model. In order to reduce the influence of noise, the camera-aware distance matrix is applied to combat the sample noise problem caused by camera offset, and the noise-robust dynamic symmetric contrast loss is used to reduce label noise. Specifically, the distance matrix that measures the similarity of pedestrian features is changed before clustering, and the camera-aware distance matrix is used to enhance the accuracy of the intra-class distance measurement, reducing the negative impact of different perspectives on the clustering effect. Combined with the noise label learning method, a robust loss is designed, a dynamic symmetric contrast loss function is proposed, and a joint loss training is used to continuously refine the pseudo-labels. Experiments are carried out on DukeMTMC-reID and Market-1501 datasets to verify the effectiveness of the proposed method.

Key words: unsupervised person re-identification; clustering; distance matrix; labels noise; loss function

引言

行人重识别技术^[1-3]主要针对非重叠视域的跨场景下行人的识别和检索,很好地弥补了跨设备、跨区域时特定行人检索困难、人工耗时长等问题,在安防、刑事侦查等领域得到应用。目前有标注数据的重识别(Re-identification, ReID)发展迅速^[4-5],但需要不断地标注新的数据,更新模型使之保持较高的检索精度。不断增长的视频监控需求和昂贵的标注成本极大地限制了其应用。因此,越来越多的研究开始关注不需要大量标注数据的无监督行人重识别^[6]。目前的无监督行人重识别任务主要包含两种形式:第一种形式为域适应方法,在具有标注信息的源域数据集上预训练,进行模型的初始训练,再在具有大量无标注数据的目标域上对模型微调。第二种形式的实验设置更为严格,在没有任何标记数据的情况下,直接用在 ImageNet 上预训练好的模型来训练 ReID 模型,这是基于聚类的完全无监督学习,通过挖掘特征相似性来估计每个实例的标签,在聚类中为每个样本分配伪标签。

本文主要研究学习完全无监督行人重识别任务,目的为利用无标注的数据优化 ReID 模型,提升模型在新场景下的泛化能力。因其实验设置为全程利用无标注的数据,更具有挑战性。而无标注的数据在使用时也面临着噪声问题。行人重识别数据集中的图片拍摄自不同相机,比如 Market-1501 数据集中的图片拍摄自 6 个不同的相机。如图 1 所示,不同相机拍摄的视角、场景不同,行人外观呈现很大差异,甚至存在严重遮挡现象,引起样本噪声,难以对有效特征进行度量,对网络训练造成干扰,因此需要考虑相机偏移给样本造成的影响。同时,这些含有噪声的样本在聚类过程中,由于缺乏对行人特征的准确描述,可能会被分配错误的伪标签,即不能保证将同一行人的样本分配相同的身份标签。错误的伪标签在聚类后的集群中成为标签噪声,影响模型的优化,这对于完全无监督任务来说是一个挑战。



图1 同一行人的跨相机样本

Fig.1 Cross-camera samples of the same pedestrian

针对上述样本噪声和标签噪声的问题,本文提出一种基于相机感知距离矩阵的无监督行人重识别算法。主要工作如下:使用基于聚类的主体框架,将距离度量的结果作为样本相似度,在聚类前的距离度量中考虑减少相机偏差造成的样本噪声影响,应用相机感知距离矩阵,提高对有效特征度量的准确性;结合自监督对比学习和基于伪标签的无监督行人重识别策略,将噪声标签学习应用于无监督行人重识别任务,并进行损失设计来提高模型的抗噪鲁棒性。在不同数据集上的实验结果表明本文方法能够提升无监督行人重识别的模型性能。

1 相关工作

在近期的无监督行人重识别任务中,文献[7]提出混合记忆存储单元,使用对比损失进行训练并不断更新存储单元。分组抽样^[8]的方法将属于同一类的样本分组进行训练,利用类的整体趋势来减弱单个样本的影响。文献[9]通过相机到相机风格转移来对齐目标样本的相机位移。文献[10]将样本相似度计算分为两个阶段,利用伪标签计算相机内的样本相似度和跨相机的样本相似度计算。文献[11]通过考虑相机标签构建代理级内存。这些方法适合具有多个相机的数据集,并考虑了相机对重识别任务的影响,极大地优化了模型的性能。与上述方法不同,本文改进行人相似度度量的距离矩阵,利用相机感知距离矩阵来降低跨相机拍摄的影响。

目前在无监督任务中利用聚类的方法生成伪标签是比较流行的方式,在网络训练过程中,标签噪声中的信息会对网络性能带来干扰。噪声标签学习(Learning with noisy labels, LNL)能够对标签噪声

样本进行合理的训练避免标签噪声的干扰。在无监督任务中应用噪声标签学习,避免噪声对网络模型的训练造成干扰是近年来的一个热门话题。目前关于噪声标签学习的研究主要可以分为样本重新加权方法^[12-13]、标签校正方法^[14-15]和损失设计方法^[16-18]。样本重新加权方法旨在训练过程中为干净的样本分配更高的权重,文献[12]提出了一种元学习的方法来对损失进行重加权,让神经网络“学习”如何给不同的损失进行重加权,元测试集中需要有额外干净数据,这在现实世界的应用程序中可能不可用。标签校正方法尝试在训练期间识别噪声样本并校正它们的标签,需要额外干净数据的支持或昂贵的检测过程来估计噪声模型。文献[16]提出了一种理论来检查损失函数的鲁棒性,文献[17]提出了一种对称交叉熵损失。由于完全无监督行人重识别任务中没有理想状态下干净的数据,因此本文采用损失设计来达到减少嘈杂伪标签的目的。

2 本文算法

目前,无监督行人重识别的经典解决方案是类自步的算法,如图2所示。首先,对无标注信息的图片提取特征并利用聚类算法生成伪标签,即对每个样本生成类标签,再利用生成的伪标签微调原始模型,对微调后的模型继续通过聚类算法生成伪标签,这样反复迭代并持续一

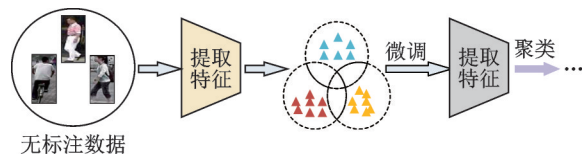


图2 无监督行人重识别系统

Fig.2 Unsupervised person re-identification system

段时间,提升伪标签生成的准确度,ReID模型的检索精度。由于每一次聚类后的簇的数量和伪标签数量都会变化,图2所示的经典无监督行人重识别的方法依然面对标签噪声难以优化的挑战,而目前大多数现有的噪声标签学习方法不能直接应用于无监督行人重识别的场景。本文考虑到无监督任务中噪声对精炼伪标签的影响,提出了基于相机感知距离矩阵的无监督行人重识别算法。

2.1 整体框架

本文算法的整体框架如图3所示。首先,考虑相机位移影响,由相机感知距离矩阵计算实例对之间的特征相似度,通过伪标签生成器进行基于密度的聚类,初始化存放实例特征存储单元,这里采用的聚类方法为具有噪声的基于密度的聚类方法(Density-based spatial clustering of applications with noise, DBSCAN)^[19]。其次,在每次迭代中,利用编码器对小批量样本进行特征编码,分别计算输入和特征存储单元的联合损失。最后,反向传播更新编码器模型并利用编码的特征以动量更新的方式更新特征存储单元。

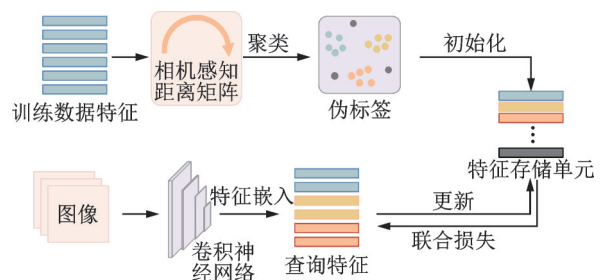


图3 本文方法的整体框架

Fig.3 Overall architecture of the proposed method

2.2 相机感知距离矩阵

由于行人重识别的数据集为跨场景拍摄,不同相机拍摄的图像呈现不同的分布,具体表现为视角或光照差异引起的不同。由于拍摄角度的不同,行人存在外观差距,同时可能遮挡行人的明显特征,对特征相似度的排序以及聚类准确度造成很大干扰。

在无监督行人重识别任务中的每一轮聚类算法前,对需要识别图片进行距离度量,再进行聚类,以生成伪标签。目前最常用的距离度量矩阵为 k -倒排最近邻的检索重排序方法^[20],将杰卡德距离与马氏

距离结合。针对上述相机移位造成干扰的问题,本文在计算距离矩阵和生成伪标签时,通过考虑视角不同造成的样本差异进行改进,尽量缩小相机拍摄视角因素造成的影响,进行具有相机感知的聚类。

首先设置一个反映不同相机域之间(或一个域内)的图像差异的相机域偏移矩阵 $D^c \in N_c \times N_c$,其中 N_c 表示相机的数量。然后,在每次迭代中,矩阵 D^c 由属于相应相机域的实例对的平均相似度计算。最后,通过原始距离矩阵 $D \in N \times N$,其中 N 为实例类别数,和偏移矩阵 D^c 计算相机感知距离矩阵 D' ,尝试从 D 减少相机域偏移 D^c 。在 D' 中,实例之间的距离会更好反映不同行人之间的差异。

原始距离矩阵 $D \in N \times N$ 用于计算实例特征之间的相似度, $D_{i,j}$ 表示第 i 个实例和第 j 个实例之间的相似度。由于行人样本的数量远大于相机域的数量,因此可以预期,每个相机对间的相似度的平均值可以隐式地由相应的域间或域内条件(或受相机标签影响的相似度)引起的样本差异表示。相机域偏移矩阵 D^c 中相机偏移距离计算如下

$$D_{i,j}^c = \frac{1}{N_c^i \times N_c^j} \sum_{i=1}^{N_c^i} \sum_{j=1}^{N_c^j} D_{\{u,v|c_u=i, c_v=j\}} \quad (1)$$

式中: u, v 表示实例; c_u, c_v 表示第 u 个实例和第 v 个实例的相机标签; N_c^i 表示相机 i 捕获的实例数量。 $i=j$ 时,偏移距离 $D_{i,j}^c$ 表示相机内的相似度值,值越大表示域相似度越大,域内(同一相机标签)的值应该最大。相机感知距离矩阵 D' 由原始距离矩阵 D 和偏移矩阵 D^c 计算得到

$$D'_{u,v} = D_{u,v} - \lambda \cdot D_{c_u, c_v}^c \quad (2)$$

式中 λ 为控制相机因素影响的参数。对于相机因素影响显著的数据集,相应地设置得更大。

最终求得实例间的新的距离为由相机感知距离与文献[20]中的杰卡德距离的加权距离和,如图4所示,图中 $x_p, x_1, x_2, \dots, x_N$ 为外观特征向量, $v_p, v_1, v_2, \dots, v_N$ 为 k -倒排特征向量。对新的距离矩阵的每一行应用 softmax 进行归一化,实例之间的距离即为相对距离,表示考虑相机偏移因素后的实例间的相似度。

在无监督行人重识别任务中,通过使用相机感知距离矩阵(Camera-aware distance matrix, CADM)来进行具有相机感知的聚类,考虑了每个相机内和相机间的实例对,减少了来自不同相机域的距离偏差对样本造成的影响,有助于减少拍摄视角不同引起行人样本差异在聚类过程中生成的错误伪标签,且有助于减少标签噪声,更利于模型后续优化。

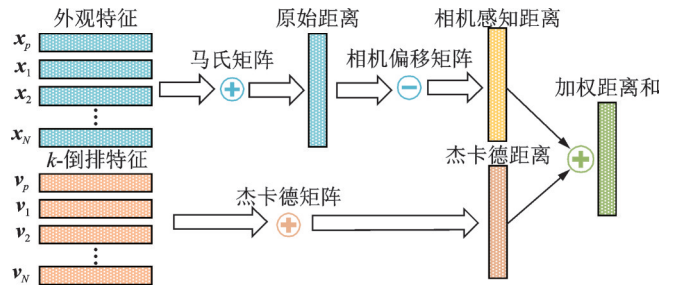


图4 相机感知距离的应用

Fig.4 Application of camera-aware distance

2.3 动态对称对比损失

对于无监督行人重识别任务来说,因其样本非理想状态中干净的数据,本文采用损失设计的方法来提升模型的抗噪鲁棒性。

在无监督任务中,由于每一次聚类后类别数量发生变化,难以利用噪声标签学习中的方法,本文采用文献[4]中提出的利用特征存储单元以及具有对比损失的特征编码器的自步对比学习方法来解决类别数量不固定的问题。给定没有任何真实标签的训练样本集 $X = \{x_1, x_2, \dots, x_{n-1}\}$,目标是通过使用特征嵌入函数 $v = f_\theta(x)$, $x \in X$, 获得特征集 $V = \{v_1, v_2, \dots, v_{n-1}\}$ 。采用伪标签生成器将特征集划分为簇集 C 和离群集 O ,簇集包含 K 个簇,其中每个簇至少包含两个样本,离群集包含单独的样本。为每个实例分配一个伪标签以构建对比损失,即

$$L_{\text{UCL}} = -\ln \frac{\exp(\langle \mathbf{v}, \mathbf{z}^+ \rangle / \tau)}{\sum_{k=1}^{n_c} \exp(\langle \mathbf{v}, \mathbf{c}_k \rangle / \tau) + \sum_{k=1}^{n_o} \exp(\langle \mathbf{v}, \mathbf{o}_k \rangle / \tau)} \quad (3)$$

式中: c_k 表示簇质心; o_k 表示离群点实例特征; \mathbf{z}^+ 表示 \mathbf{v} 对应的正样本类别原型;温度 τ 根据经验设置为0.05; $\langle \cdot, \cdot \rangle$ 表示两个特征向量的内积,衡量它们的相似性; n_c 为聚类集群的数量, n_o 为离群点实例的数量,它们的和为 N 。优化具有对比损失的编码器并动态更新具有编码特征的存储单元,其中更新公式定义为

$$M[\mathbf{x}] \leftarrow m \cdot M[\mathbf{x}] + (1 - m) \cdot f_\theta(\mathbf{x}) \quad (4)$$

式中: M 为特征存储单元; $m \in [0, 1]$ 为存储单元中更新样本特征的动量系数,经验设置为0.2。

在运用了特征存储单元、解决了无监督任务中类别数目不断变化问题的基础上,本文考虑噪声优化问题。文献[17]提出一个理论来判定损失函数是否对噪声鲁棒,即

$$\sum_{k=1}^N L(f, k) = Z \quad (5)$$

式中: f 表示样本; N 为类别总数; Z 为常数; k 为类别数。式(5)表明,对于任何样本 f 和损失函数 L ,如果损失函数 L 是对噪声鲁棒的,则将 f 分类到所有类别(即 $1 \sim N$)的损失之和应该是一个常数。本文根据式(5)的理论,专门为无监督行人重识别任务设计一个对噪声鲁棒的损失,提出动态对称对比损失函数(Dynamic symmetric contrastive loss, DSCL)为

$$L_{\text{DSCL}} = \text{soft max} \left[\ln \frac{\exp(\langle \mathbf{v}, \mathbf{z}^+ \rangle / \tau)}{\sum_{k=1}^{n_c} \exp(\langle \mathbf{v}, \mathbf{c}_k \rangle / \tau) + \sum_{k=1}^{n_o} \exp(\langle \mathbf{v}, \mathbf{o}_k \rangle / \tau)} \right] \ln [\text{soft max}(\hat{y}_i)] \quad (6)$$

式中 $\hat{y}_i \in \mathbf{R}^{N \times 1}$ 表示第 i 个样本伪身份的 one-hot 向量。式(6)的动态对称对比损失函数结合了式(3)的对比损失函数,并满足式(5)的对称性理论。下面证明 DSCL 损失满足式(5)。

定义将 f 分配给 i 类的概率为 p_i 且 $\sum_{i=1}^N p_i = 1$,经过 softmax 归一化后的 \hat{y} 的第 i 个元素为 \tilde{y}_i ,则 \tilde{y}_i 为

$$\tilde{y}_i = \begin{cases} \frac{1}{N-1+e} & \hat{y}_i = 0 \\ \frac{e}{N-1+e} & \hat{y}_i = 1 \end{cases} \quad (7)$$

式(6)可以重新表述为

$$L_{\text{DSCL}} = -\sum_{i=1}^N p_i \ln \tilde{y}_i \quad (8)$$

当类标签为第 k 类时,即 $\hat{y}_k = 1$ 时,式(8)可以表述为

$$L_{\text{DSCL}}(f, k) = -p_k \ln \tilde{y}_k - \sum_{i \neq k} p_i \ln \tilde{y}_i \quad (9)$$

为了方便,定义式(7)中 $\hat{y}_i = 0$ 时, \tilde{y}_i 记为 P ,又由于 $\sum_{i \neq k} p_i = 1 - p_k$,代入式(9),其化简为

$$L_{\text{DSCL}}(f, k) = -p_k \ln(eP) - \ln P \sum_{i=k}^N p_i = -p_k - \ln P \quad (10)$$

对式(9)求和得到

$$\sum_{k=1}^N L_{\text{DSCL}} = (f, k) = -1 - N \ln O \quad (11)$$

由式(11)得到,本文提出的 DSCL 损失的所有类别的损失值之和为常数,满足式(5),因此 DSCL 损失是对噪声鲁棒的。基于以上对比损失与动态对称对比损失,提出优化后的联合损失函数为

$$L = L_{\text{UCL}} + \beta L_{\text{DSCL}} \quad (12)$$

式中 β 是用于控制权重的参数,取值区间为 $[0, 1]$ 。

在联合损失函数中,前者利于模型的收敛,后者用于减少聚类噪声对模型的影响。在如图 3 所示的整体框架中使用联合损失来反向传播更新编码器与实例特征存储单元,将对比损失与噪声鲁棒的损失进行结合,与只使用式(3)的对比损失相比,不仅能够获得良好的类内与类间距离,同时能够提高模型对抗噪声影响的能力。

3 实验与分析

3.1 数据集和实验环境

本文在行人重识别领域常用的两个大规模公开数据集 DukeMTMC-ReID 和 Market-1501 上进行验证,两个数据集采集地不同,有着不同的拍摄条件。表 1 为两个数据集的基本信息,包含训练集与测试集。使用累计匹配特性(Cumulative match characteristic, CMC)和平均精度均值(mean Average precision, mAP)作为评价标准。

表 1 数据集信息

Table 1 Information of datasets

| 数据集 | 训练集图片数/张 | 训练集人数/个 | 测试集图片数/张 | 测试集人数/个 | 相机数量/个 |
|---------------|----------|---------|----------|---------|--------|
| DukeMTMC-ReID | 16 522 | 702 | 17 661 | 702 | 8 |
| Market-1501 | 12 936 | 751 | 19 732 | 750 | 6 |

本文所有实验均在 PyTorch 上进行,使用单个 GTX 2080TI GPU 训练大约需要 3 h。在实验中,采用在 ImageNet 数据集上预训练的 ResNet-50 作为主干来提取特征,移除了 pooling-5 层的后续层,并添加了批量归一化层和 L2 归一化层。同时所有图像在输入网络之前都被调整为 256 像素 \times 128 像素,对于训练图像,执行随机裁剪、随机翻转和随机擦除等策略。本文使用 DBSCAN 与 k 个最近邻进行聚类,其中邻居间最大距离 $d=0.6$, $k=30$ 。训练的 epoch 设置为 80,初始学习率设置为 0.000 35,每 20 个 epoch 后降为先前值的 1/10,批量大小设置为 64。

3.2 方法比较

为了进一步验证方法的有效性,本文将实验结果与目前较为先进的无监督行人重识别方法对比,如表 2 所示,包括 SSL^[21]、JVTC^[22]、HCT^[23]、SpCL^[7]、IICS^[10]、CAP^[11]和 CACL^[24]方法。从表 2 中可以看出,本文方法优于大部分现有方法。根据表 2 所示,本文方法在 DukeMTMC-ReID 和 Market-1501 上分别获得 85.9% 和 92.5% 的 Rank-1 准确率,以及 71.9% 和 81.7% 的 mAP。与 SpCL 方法^[7]相比,本文方法在 DukeMTMC-ReID 数据集上, mAP 和 Rank-1 分别提高 6.6% 和 4.7%;在 Market-1501 数据集上, mAP 和 Rank-1 分别提高 8.6% 和 4.4%。结果表明,本文提出的方法具有一定的有效性。

表2 不同方法对比结果

Table 2 Comparison results of different methods

| 方法 | DukeMTMC-ReID | | | | Market-1501 | | | |
|--------------------------------|---------------|-------------|-------------|-------------|-------------|--------|--------|---------|
| | mAP | Rank-1 | Rank-5 | Rank-10 | mAP | Rank-1 | Rank-5 | Rank-10 |
| SSL(CVPR20) ^[21] | 28.6 | 52.5 | 63.5 | 68.9 | 37.8 | 71.7 | 83.8 | 87.4 |
| JVTC(ECCV20) ^[22] | 42.2 | 67.6 | 78.0 | 81.6 | 41.8 | 72.9 | 84.2 | 88.7 |
| HCT(CVPR20) ^[23] | 50.7 | 69.6 | 83.4 | 87.4 | 56.4 | 80.0 | 91.6 | 95.2 |
| SpCL(NeurIPS20) ^[7] | 65.3 | 81.2 | 90.3 | 92.2 | 73.1 | 88.1 | 95.1 | 97.0 |
| IICS(CVPR21) ^[10] | 64.4 | 80.0 | 89.0 | 91.6 | 72.9 | 89.5 | 95.2 | 97.0 |
| CAP(AAAI21) ^[11] | 67.3 | 81.1 | 89.3 | 91.8 | 79.2 | 91.4 | 96.3 | 97.7 |
| CACL(Arxiv21) ^[24] | 69.9 | 83.3 | 91.5 | 94.1 | 80.9 | 92.7 | 97.4 | 98.5 |
| 本文方法 | 71.9 | 85.9 | 92.1 | 94.1 | 81.7 | 92.5 | 96.7 | 98.0 |

3.3 消融实验

为了验证本文提出模型的有效性,本文在两个大规模的公开数据集 DukeMTMC-ReID 和 Market-1501 上展开实验:(1) Han 等^[8]提出的分组抽样的基准模型 Baseline;(2)加入相机感知距离矩阵模块的模型 Baseline+CADM;(3)加入动态对称对比损失,利用联合损失训练的模型 Baseline+DSCL;(4)将广义平均池化(Generalized mean pooling, GeM)^[25]应用于骨干网 ResNet-50 后的模型 Baseline+GeM;(5)本文提出的结合相机感知距离矩阵和动态对称对比损失,同时利用广义平均池化协同训练的模型 Baseline+DSCL+CADM+GeM,实验结果如表 3 示。通过对比表 3 中不同模块应用到模型中的 mAP 和 Rank-1 的结果,能够看到在实例对相似度度量中考虑相机因素以及损失设计能够提升伪标签生成的准确性,促进伪标签的精炼。同时,结合广义平均池化能够协同增加影响力。在基准模型上,使用本文的方法,mAP 在 DukeMTMC-ReID 和 Market-1501 数据集上分别提升 2.9% 和 2.8%,使伪标签更接近真实标签。通过两个数据集上的 CMC(Cumulative math characteristic)曲线能够更直观地看出不同方法对模型的优化结果,如图 5 所示。

表3 消融实验对比结果

Table 3 Comparison results of ablation experiments %

| 模型 | DukeMTMC-ReID | | Market-1501 | |
|------------------------------|---------------|-------------|-------------|--------|
| | mAP | Rank-1 | mAP | Rank-1 |
| Baseline ^[8] | 69.2 | 83.1 | 78.9 | 91.7 |
| Baseline+CADM | 70.6 | 83.9 | 80.0 | 91.7 |
| Baseline+DSCL | 69.8 | 83.8 | 79.7 | 91.8 |
| Baseline+GeM ^[25] | 69.5 | 83.2 | 80.7 | 92.2 |
| Baseline+DSCL+CADM+GeM | 71.9 | 85.9 | 81.7 | 92.5 |

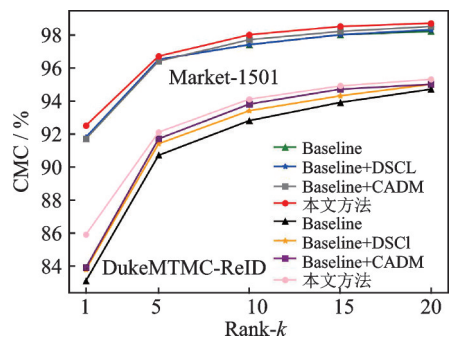


图5 两个基准数据集上的 CMC 曲线

Fig.5 CMC curves on two benchmark datasets

3.4 参数分析

在联合损失训练策略中,式(12)中设置了 β 参数,通过改变此参数来定义动态对称对比损失与对比损失的权重比例。图 6 为 DukeMTMC-ReID 数据集上模型在 $\beta \in [0, 1]$ 时的 mAP 和 Rank-1 准确率,能

够看到当 $\beta = 0.5$ 时, mAP 和 Rank-1 都较小; 当 $\beta = 0.8$ 时, mAP 和 Rank-1 都较大, 且 mAP 和 Rank-1 随 β 参数变化的趋势大致相同, 能够看到当 β 过大或者过小时都会影响识别精度, 因此要设置合理的 β 参数。

3.5 伪标签类别数量分析

在基于聚类的无监督任务中, 随着模型的训练, 目标为生成的伪标签越来越靠近真实标签。因此, 本文分析聚类过程中生成伪标签的数量变化过程。图7为在DukeMTMC-ReID数据集上的实验结果, 曲线为聚类过程中伪标签类别数目的变化曲线, 其中红色和灰色分别为在基准模型和本文提出的模型上的伪标签类别数量变化结果, 绿色为DukeMTMC-ReID数据集的真实行人类别数。从图7中能够看到, 随着迭代次数的增加, 聚类过程中的伪标签类别数越来越接近真实类别数目, 本文所提出的方法相比于基准模型在聚类过程中更接近真实情况。这种结果可能是由于相机感知距离矩阵和动态对称对比损失的应用有助于减弱相机偏移带来的影响并不断纠正错误伪标签, 生成正确的伪标签。

3.6 聚类特征可视化

本文选取了Market-1501数据集中15个行人的跨相机图像作为聚类特征化的测试集, 共395张图像, 不同颜色代表不同行人。图8为用基准模型和本文模型的聚类效果可视化。通过对比特征聚类的t-SNE可视化图, 能够看出在图8(b)中采用相机感知距离矩阵和动态对称对比损失后的模型, 在聚类时的样本类内距离减小, 类间距离增大, 这有助于将同一行人的样本分配同一伪标签, 减少标签噪声, 本文模型的优势得以体现。

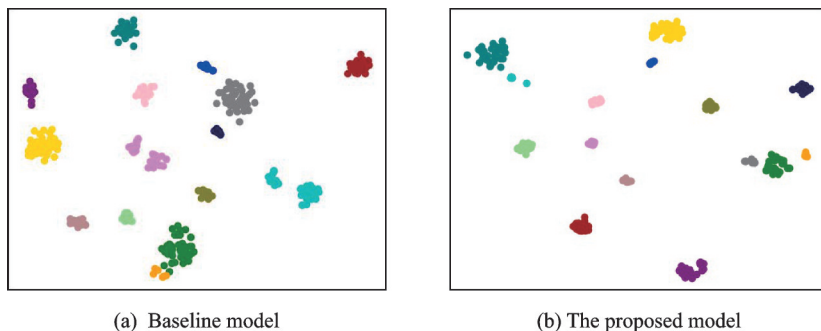


图8 Market-1501子集特征的t-SNE可视化

Fig.8 t-SNE visualization of features from a subset of Market-1501

4 结束语

本文在基于聚类的无监督行人重识别框架基础上对现有的算法进行研究和改进。为了在聚类过

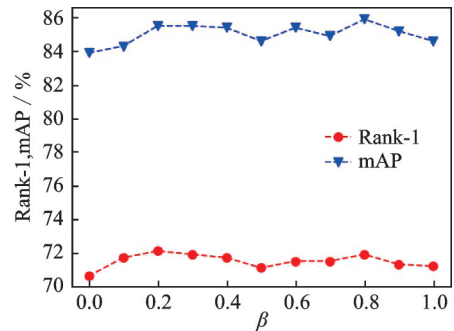


图6 DukeMTMC-ReID上 β 对mAP和Rank-1的影响

Fig.6 Effect of β on Rank-1 and mAP in DukeMTMC-ReID

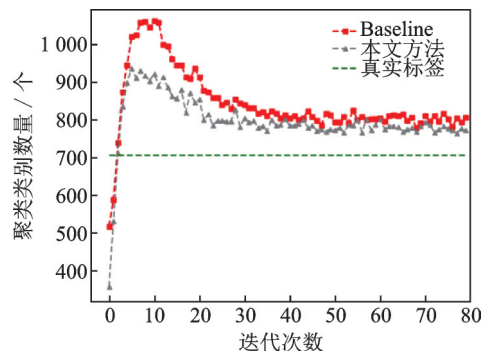


图7 伪标签类别数量变化

Fig.7 Changes in the number of pseudo-label categories

程中减少噪声的影响,本文通过考虑相机内和相机间的样本对,应用相机感知距离矩阵,减弱相机因素带来的偏差。基于噪声标签学习,提出一种对噪声鲁棒的损失函数,有利于减少标签噪声造成的误差积累,减轻负面影响。经实验结果分析,本文提出的基于相机感知距离矩阵的无监督行人重识别模型能够有效地提升识别效果。未来工作将继续关注行人重识别,在现有方法基础上考虑针对特征存储单元的改进,进一步提升重识别模型的性能。

参考文献:

- [1] 李梦静,吉根林,赵斌.基于步行周期聚类的视频行人重识别关键帧提取算法[J].南京航空航天大学学报,2021,53(5): 780-788.
LI Mengjing, JI Genlin, ZHAO Bin. Key frame extraction algorithm for video-based person re-identification based on walking cycle clustering[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2021, 53(5): 780-788.
- [2] 罗浩,姜伟,范星,等.基于深度学习的行人重识别研究进展[J].自动化学报,2019,45(11): 2032-2049.
LUO Hao, JIANG Wei, FAN Xing, et al. A survey on deep learning based person re-identification[J]. Acta Automatica Sinica, 2019, 45(11): 2032-2049.
- [3] 夏道勋,郭方,刘浩杰,等.开放式行人再识别研究进展综述[J].数据采集与处理,2021,36(3): 449-467.
XIA Daoxun, GUO Fang, LIU Haojie, et al. Review on research progress of open-world person re-identification[J]. Journal of Data Acquisition and Processing, 2021, 36(3): 449-467.
- [4] 胡彬,杨斌,邵叶秦,等.基于视频的行人再识别[J].南京航空航天大学学报,2019,51(5): 669-674.
HU Bin, YANG Cheng, SHAO Ye-qin, et al. Video-based person re-identification[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2019, 51(5): 669-674.
- [5] 郝玲,段继忠,庞健.基于难样本混淆增强特征鲁棒性的行人重识别[J].数据采集与处理,2022,37(1): 122-133.
HAO Ling, DUAN Jizhong, PANG Jian. Person re-identification based on hard negative sample confusion to enhance robustness of features[J]. Journal of Data Acquisition and Processing, 2022, 37(1): 122-133.
- [6] 曹亮,王洪元,戴臣超,等.基于多样性约束和离散度分层聚类的无监督视频行人重识别[J].南京航空航天大学学报,2020,52(5): 752-759.
CAO Liang, WANG Hongyuan, DAI Chenchao, et al. Unsupervised video-based person re-identification based on diversity constraint and dispersion hierarchical clustering[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2020, 52(5): 752-759.
- [7] GE Yixiao, CHEN Dapeng, ZHU Feng, 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.
- [8] HAN Xumeng, YU Xuehui, JIANG Nan, et al. Group sampling for unsupervised person re-identification[EB/OL]. (2023-05-16). <https://arxiv.org/abs/2107.03024>.
- [9] ZHONG Zhun, ZHENG Liang, LI Shaozi, et al. Generalizing a person retrieval model hetero-and homogeneously[C]//Proceedings of the European Conference on Computer Vision. Munich, Germany: Springer, 2018: 172-188.
- [10] XUAN Shiyu, ZHANG Shiliang. Intra-inter camera similarity for unsupervised person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE Press, 2021: 11926-11935.
- [11] WANG Menglin, LAI Baisheng, HUANG Jianqiang, et al. Camera-aware proxies for unsupervised person re-identification [C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2021: 2764-2772.
- [12] SHU Jun, XIE Qi, YI Lixuan, et al. Meta-weight-net: Learning an explicit mapping for sample weighting[J]. Advances in Neural Information Processing Systems, 2019, 32: 1919-1930.
- [13] CHENG Hao, ZHU Zhaowei, LI Xingyu, et al. Learning with instance-dependent label noise: A sample sieve approach[EB/OL]. (2021-05-22)[2022-03-21]. <https://arxiv.org/abs/2010.02347v2>.
- [14] LEE K H, HE Xiaodong, ZHANG Lei, et al. Cleannet: Transfer learning for scalable image classifier training with label noise [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE Press, 2018: 5447-5456.

- [15] VAHDAT A. Toward robustness against label noise in training deep discriminative neural networks[J]. *Advances in Neural Information Processing Systems*, 2017, 30: 5601-5610.
- [16] GHOSH A, KUMAR H, SASTRY P S. Robust loss functions under label noise for deep neural networks[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. San Francisco, USA: AAAI Press, 2017.
- [17] WANG Yisen, MA Xingjun, CHEN Zaiyi, et al. Symmetric cross entropy for robust learning with noisy labels[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Seoul, Korean: IEEE Press, 2019: 322-330.
- [18] ZHANG Zhilu, SABUNCU M. Generalized cross entropy loss for training deep neural networks with noisy labels[J]. *Advances in Neural Information Processing Systems*, 2018, 31: 8792-8802.
- [19] 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*. Portland Oregon: AAAI Press, 1996: 226-231.
- [20] ZHONG Zhun, ZHENG Liang, CAO Donglin, et al. Re-ranking person re-identification with k-reciprocal encoding[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Hawaii, USA: IEEE Press, 2017: 1318-1327.
- [21] LIN Yutian, XIE Lingxi, WU Yu, et al. Unsupervised person re-identification via softened similarity learning[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE Press, 2020: 3390-3399.
- [22] LI Jianing, ZHANG Shiliang. Joint visual and temporal consistency for unsupervised domain adaptive person re-identification [C]//*Proceedings of European Conference on Computer Vision*. Glasgow, UK: Springer, 2020: 483-499.
- [23] ZENG Kaiwei, NING Munan, WANG Yaohua, 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*. Seattle, USA: IEEE Press, 2020: 13657-13665.
- [24] LI Mingkun, LI Chunguang, GUO Jun. Cluster-guided asymmetric contrastive learning for unsupervised person re-identification[J]. *IEEE Transactions on Image Processing*, 2022, 31: 3606-3617.
- [25] LEE C Y, GALLAGHER P W, TU Z. Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree[C]//*Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*. Cadiz, Spain: IEEE Press, 2016: 464-472.

作者简介:



白梦林(1998-),女,硕士研究生,研究方向:行人重识别、计算机视觉, E-mail: baimenglin2022@163.com。



周非(1977-),通信作者,男,教授,研究方向:图像处理、通信信号处理。



舒浩峰(1998-),男,硕士研究生,研究方向:行人重识别、计算机视觉。

(编辑:张黄群)