

DOI:10.16136/j.joel.2023.07.0422

联合特征细化和耐噪声对比学习的无监督行人重识别

钱亚萍^{1,2,3}, 王凤随^{1,2,3*}, 熊磊^{1,2,3}, 闫涛^{1,2,3}

(1. 安徽工程大学 电气工程学院, 安徽 芜湖 241000; 2. 检测技术与节能装置安徽省重点实验室, 安徽 芜湖 241000; 3. 高端装备先进感知与智能控制教育部重点实验室, 安徽 芜湖 241000)

摘要:针对无监督行人重识别(person re-identification, ReID)中行人特征表达不充分以及聚类过程产生噪声标签的问题,提出一种联合特征细化和耐噪声对比学习的无监督 ReID 方法。首先,为丰富无标记的行人表征,设计了非局部通道细化模块(non-local channel refinement module, NCRM)对关键特征信息进行加权强化,其融合了非局部通道的重要特征来捕获无标记数据的类间区别表征,形成更具有鉴别度的特征描述符。其次,考虑到特征的充分表达,采用广义均值(generalized mean, GEM)池化自适应调整参数来增强不同细粒度区域信息的提取能力。再次,为了减轻噪声标签对网络的负面影响,设计了耐噪声的动态对比均衡(dynamic contrastive equilibrium, DCE)损失函数进行无监督联合学习。最终,在两个公共数据集上的实验结果验证了所提方法的有效性和先进性, *mAP* 分别达到了 83.1% 和 71.9%, 优于其他先进方法。

关键词:行人重识别(ReID); 无监督; 深度特征网络; 非局部通道细化模块(NCRM); 耐噪声对比学习

中图分类号: TP391.4 文献标识码: A 文章编号: 1005-0086(2023)07-0762-09

Joint feature refinement and noise-tolerant comparative learning for unsupervised person re-identification

QIAN Yaping^{1,2,3}, WANG Fengsui^{1,2,3*}, XIONG Lei^{1,2,3}, YAN Tao^{1,2,3}

(1. School of Electrical Engineering, Anhui Polytechnic University, Wuhu, Anhui 241000, China; 2. Anhui Key Laboratory of Detection Technology and Energy Saving Devices, Wuhu, Anhui 241000, China; 3. Key Laboratory of Advanced Perception and Intelligent Control of High-end Equipment, Ministry of Education, Wuhu, Anhui 241000, China)

Abstract: An unsupervised person re-identification (ReID) method was proposed to solve the insufficient representation of person features and the noisy labels generated by the clustering process in the process of unsupervised ReID, which jointed feature refinement and noise-tolerant comparative learning. Firstly, a non-local channel refinement module (NCRM) was designed to enrich the unlabeled person representation by weighted reinforcement of key feature information, which fused the important features of non-local channel to capture the discriminative representation between classes of unlabeled data to form more discriminative feature descriptors. Secondly, generalized mean (GEM) pooling adaptive adjustment parameters were used to enhance the ability of extract information from different fine-grained regions to accomplish full expression of characteristics. Then, a noise-tolerant dynamic contrastive equalization (DCE) loss was designed for unsupervised associated learning to mitigate the negative impact of noisy

* E-mail: fswang@ahpu.edu.cn

收稿日期: 2022-06-06 修订日期: 2022-09-17

基金项目: 安徽省自然科学基金(2108085MF197, 1708085MF154)、安徽高校省级自然科学研究重点项目(KJ2019A0162)、检测技术与节能装置安徽省重点实验室开放基金(DTESD2020B02)、安徽工程大学国家自然科学基金预研项目(Xjky2022040)和安徽高校研究生科学研究项目(YJS20210448, YJS20210449)资助项目

label on the network. Finally, the experimental results on two public datasets verify the effectiveness and advancement of the proposed method. The *mAP* reaches 83.1 % and 71.9 % respectively, which is superior to other advanced methods.

Key words: person re-identification; unsupervised; deep feature network; non-local channel refinement module (NCRM); noise-tolerant comparative learning

0 引言

近年来,深度学习的发展使得有监督的行人重识别(person re-identification, ReID)方法获得了较高的识别精度^[1,2]。然而在现实场景中,有监督的 ReID 方法利用人工标注相机视图的成本较大,并且网络性能高度依赖于数据集中的标注样本,可扩展性不高。当现实场景中的数据量充足且庞大时,利用手动标注获取数据集的方式显得十分繁琐,而无监督的 ReID 方法不需要人为标注的身份信息,能够从未标记的数据集中学习行人身份类间区别表征和类内相似性,减少了人工对数据注释的繁重工作,在实际应用中更具备发展的潜力。

无监督 ReID 方法可以分为无监督领域自适应(unsupervised domain adaptive, UDA)方法^[3]和完全无监督(fully-unsupervised, FU)方法^[4]。其中 UDA 方法需要在标记的源域数据集上训练出行 ReID 模型,使其能够在未标记的目标域上获取优良的性能。与 UDA 方法相比, FU 方法不需要任何标记的数据集信息,因此更具挑战性。为了充分获取有用的特征信息, GE 等^[5]采用一种自主对比学习框架(self-paced contrastive learning, SPCL)来混合内存动态生成源域类级、目标域集群级和非集群实例级监督信号,用于学习特征表示。与传统的对比学习策略不同,该框架逐步创建更可靠的集群来细化混合记忆和学习目标,获得了良好的性能。但是,该框架过度依赖在源域数据集中的训练结果,限制了在实际应用中的发展。ZHANG 等^[6]提出一种基于软多标签和复合注意力模型的无监督 ReID 方法来逐步探索潜在的判别信息。在学习参考代理标签的基础上,通过构建目标和参考数据集的映射模型来构建软多标签。随后,将软多标签引入原始行人样本中,借助残差网络的训练,实现了多摄像机场景下目标的精确标记和特征的细粒度分类。与上述利用注意力的残差网络不同,张宝华等^[7]利用软多标签并结合深度特征融合网络来构建 ReID 模型,通过结合压缩激励网络和多层级深度特征融合改进残差网络的结构,提高训练速度和精度。与结合输入集的所有特征的融合方法不同, VALEM 等^[8]

采用对输入集的所有特征进行显式选择和随后融合特征的方法,通过从不同的评分器中编码多个特征的超图结构来学习表示,取得了显著的性能提升。由于该框架结合了无监督秩表排序的特征选择和鲁棒的超图模型,引入了较多的参数约束项,使得模型在复杂的现实环境中缺乏一定的竞争性。为了提升复杂监控环境中模型的表现能力, SRIDHAR 等^[9]提出一种基于时空关联规则的无深度注释聚类(spatio-temporal association rule based deep annotation-free clustering, STAR-DAC)框架,利用视觉特征重新识别未标记人图像,并进行增量聚类,通过挖掘的时空关联规则对网络进行聚类微调,为传统的无监督聚类方法注入了新思想。然而,大多数方法在聚类时将每个聚类视为一个伪身份类,忽略了由相机视图变化引起的类内方差问题。2021年, WANG 等^[10]提出了一种相机感知代理方法(camera-aware proxies, CAP),该方法基于相机感知代理设计了相机内和相机间的对比学习组件,通过利用相机感知代理处理较大的类内方差生成了可靠的伪标签进行无监督学习,有效降低了类内方差给模型带来的负面影响。

随着无监督 ReID 技术的发展,传统的聚类方法以及随机抽样策略无法充分引导行人有效信息的聚集。不同于传统的无监督聚类策略,张智等^[11]提出了一种基于风格转换的无监督聚类 ReID 方法,通过将生成对抗网络的风格转换方法引入聚类过程中,以此学习到更为丰富的行人表征。另外,针对无监督模型中随机抽样策略对模型产生的负面影响, HAN 等^[12]提出将群体抽样策略引入无监督 ReID 中,它将属于同一类的样本分组进行训练,有助于维护每个类内的相似性结构并减少单个样本的影响,获得了优越的性能。但是,在使用伪标签代替真实值作为监督信息时,由于无监督模型的特征表达能力不足和从样本中提取的特征的语义信息不足,容易引发噪声伪标签。并且,随着网络产生的错误监督信息(具有不同真实值,但拥有相同伪标签的行人样本)被持续地积累,最终模型的性能也随之受到影响。

针对上述问题,本文受 HAN 等的启发,提出了一种联合特征细化和耐噪声对比学习的完全无

监督 ReID 方法,旨在提升模型的特征提取能力和减轻聚类过程中噪声伪标签的负面影响。所提方法主要包含针对无监督模型特征表达的细化学习策略和网络损失计算阶段的耐噪声对比学习。通过在主流行人数据集中进行了一系列的实验,在完全无标注信息的情况下,实验结果表明所提方法总体性能优于对比方法。本文主要贡献包括如下:

1) 提出一种提升无监督模型特征表达的细化学习策略,该策略包括在骨干网络中设计了一种改进的非局部通道细化模块(non-local channel refinement module, NCRM),并替换原有网络中的池化方式来增强网络对非局部区域特征的细化学习,同时提升不同细粒度区域特征的信息提取能力,最终形成更为丰富的行人表征。

2) 为了减轻无监督聚类过程中噪声伪标签

的负面影响,设计了耐噪声的动态对比均衡(dynamic contrastive equilibrium, DCE) 损失函数进行无监督联合学习,帮助网络在适应每个聚类过程后行人类别变化的同时更好地监督约束网络,最终提升模型的整体性能。

3) 在两个公共的数据集上进行实验表明,本文方法具有良好的性能和可靠性。

1 特征细化及耐噪声对比学习

1.1 整体网络

本文借助分组采样策略,基于改进的 ResNet50 设计出适用于无监督 ReID 的网络结构,旨在提升伪标签的可信度,整体网络结构如图 1 所示。通过网络的逐步迭代和特征的不断更新来进行对比学习和聚类,生成伪标签。主要包含细化特征的提取、分组抽样和耐噪声对比学习过程,当训练周期达到预设

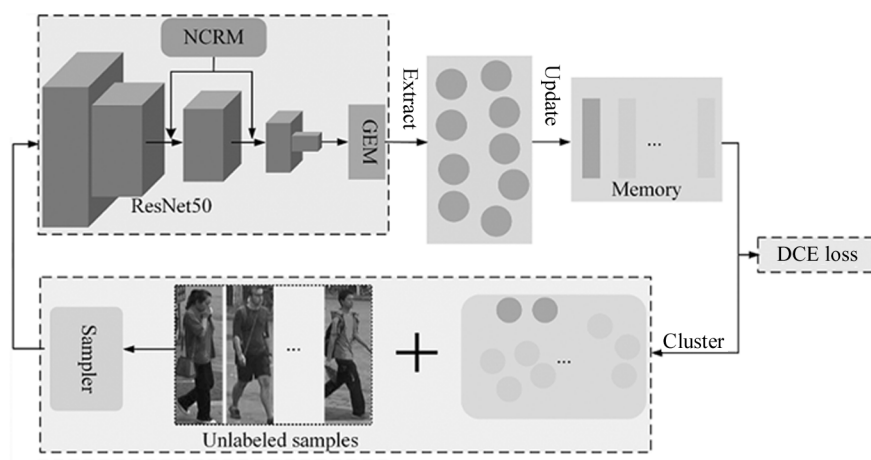


图 1 整体网络结构图

Fig. 1 Overall network structure

数值时,网络迭代训练结束。

1.2 特征细化

无监督 ReID 任务在于学习对不同人身份的分类能力。因此,充分学习行人特征之间的异同对提升模型性能十分关键。为了确保高质量的伪标签和高效的骨干优化,从增强原始 ResNet50 网络的信息提取能力出发,通过设计改进的 NCRM 并嵌入 ResNet50 网络的不同层来构建本文的深度特征细化网络,如图 2 所示。

细化学习策略具体如下:首先,输入行人特征设置为 $3 \times 256 \times 128$ 大小,将 ResNet50 原始网络保留至第二层(Layer2)处,经过 Layer2 后输出维度为 $512 \times 32 \times 16$ 的特征。其次,为了丰富行人特征的表达,将设计的 NCRM 嵌入 ResNet50 的 Layer2—

Layer3 和 Layer3—Layer4 之间关注潜在特征并加权,完成信息细化并融合基础全局特征,帮助网络自适应地关注到行人特征的细腻区域。通过此操作网络将挖掘原有结构中更细微的差异,从一般相似性中提取出包含每个行人身份且具有鉴别性的特征 f_2 。再次,考虑到原有网络中的全局平均池化(global average pooling, GAP)操作无法捕获特定区域的鉴别性行人特征,本文替换了原始网络中的平均池化层并采用了广义均值(generalized mean, GEM)池化层,其通过调整参数来增强不同细粒度区域的特征信息提取,将经 Layer4 网络输出的 $2048 \times 16 \times 8$ 大小的三维行人特征图转变为 $2048 \times 1 \times 1$ 大小,最终经过 BN 层产生了 2048 维的细化特征。

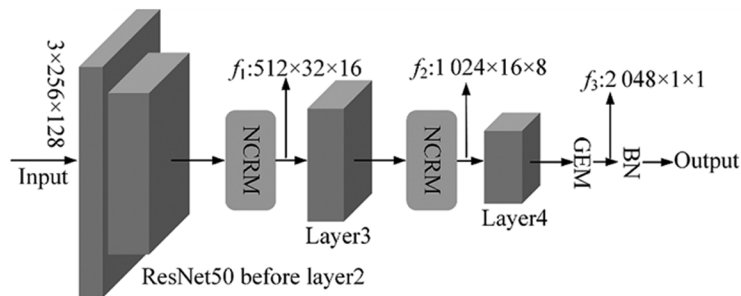


图2 深度特征细化网络结构图

Fig. 2 Structure diagram of deep feature refinement network

1.3 NCRM

非局部网络不仅能够捕获长距离依赖,并且可以作为卷积神经网络(convolutional neural networks, CNN)的插件嵌入到网络模型中,该种策略对多种图像分类任务的性能提升都具有一定的作用。非局部操作过程如下:

$$y_i = \frac{1}{C(x)} \sum_j f(x_i, x_j) g(x_j), \quad (1)$$

$$f(x_i, x_j) = \theta(x_i)^T \varphi(x_j), \quad (2)$$

$$z_i = w_z y_i + x_i, \quad (3)$$

式中, x 表示输入特征, y 表示输出信号, i 表示输出特征的位置(时间点、空间点、时空点), j 表示其余位置的特征点; $C(x)$ 表示归一化因子; $f(x_i, x_j)$ 表示计算 i 和所有可能的相关位置 j 之间的成对相似性关系; $g(x_j)$ 是一元输入缩放函数,用于获取在 j 位置的特征表示。其中成对相似性关系 $f(x_i, x_j)$ 计算过程如式(2)所示,其中 $\theta(x_j) = w_\theta x_j$, $\varphi(x_j) = w_\varphi x_j$, w_θ 、 w_φ 和 w_z 均为权重矩阵。

非局部操作能够深挖不同特征图中的不同像素点间的位置信息,以此获取特征在空间域上的全局信息,但在不添加任何约束时会造成特征冗余,并且未考虑到全局信息与特征通道之间的关系,不利于有用特征的聚集。

在无监督 ReID 中,希望模型学习的特征视角能够不受监督信息的指导,即模型在训练过程中需要持有良好的语义信息表达能力,才能捕获特征中足够的类间区别表示和类内相似性。针对原始骨干网络中缺乏特定模块来充分获得行人有用信息的表示,本文的 NCRM 从端到端的特征学习角度,将输入特征在特征通道层面和获得区域间的全局信息进行重新加权,强化特征之间的有用信息,抑制无关的冗余信息。用以此结构辅助构建本文的特征细化网络,可以增强整体模型迭代更新过程中的特征表达

能力,其结构如图3所示。具体工作如下:

首先,将 NCRM 的输入特征记为 $F \in R^{H \times W \times C}$, 输出特征记为 $F' \in R^{H \times W \times C}$, 利用 f 表示模块中像素与像素之间的关系,其定义如式(4)所示:

$$f(F_{c,i}, F_{c,j}) = \theta(F_{c,i})^T \varphi(F_{c,j}), \quad (4)$$

式中, $\theta(F_{c,i})$ 和 $\varphi(F_{c,j})$ 分别表示像素点 i 和 j 的激活值。

其次,本文在 NCRM 内部执行非局部操作,具体过程定义为式(5):

$$y_{c,i} = \frac{1}{C(F)} \sum_j f(F_{c,i}, F_{c,j}) g(F_{c,j}), \quad (5)$$

式中, $C(F)$ 表示归一化因子, $g(F_{c,i})$ 是一元输入缩放函数,用于获取在 j 位置的输入特征表示。

再次,考虑对经过降维操作的 F_1 特征图实施通道加权,使网络能够自适应地关注到图像中的行人特征重要区域,并抑制无关通道。具体来说,对 F_1 执行 GAP 操作,然后在 GAP 之后的特征上引入权重共享的一维卷积(卷积核大小 $K=5$)进行学习,并采取 Sigmoid 激活函数得到各个通道的权重 w_c 。然后,将权重与输入特征图 F_1 的对应元素相乘,得到 $g(F_{c,i})$ 。具体的实现过程式(6)和式(7)所示:

$$g(F_{c,j}) = w_c x_i, \quad (6)$$

式中, $g(F_{c,j})$ 用来计算位置 j 的输入信号表示, w_c 表示通过实施通道加权操作学习到的权重矩阵, x_i 表示原始输入 F_1 特征图的对应元素。

$$w_c = \sigma[C1D_K(F_{1,GAP})], \quad (7)$$

式中, $C1D_K$ 表示通过卷积核大小 $K=5$ 的一维卷积过程, $F_{1,GAP}$ 表示输入特征图 F_1 经 GAP 之后的特征表示, σ 表示非线性激活函数 Sigmoid。

最终,结合残差结构的优点, NCRM 将原始特征信息 F 与通道全局信息 f_c 进行叠加,以获得最终输出的增强特征 F' , 通过式(8)表示:

$$F' = w_{Fy_{c,i}} + F, \quad (8)$$

式中, w_F 表示权重矩阵。

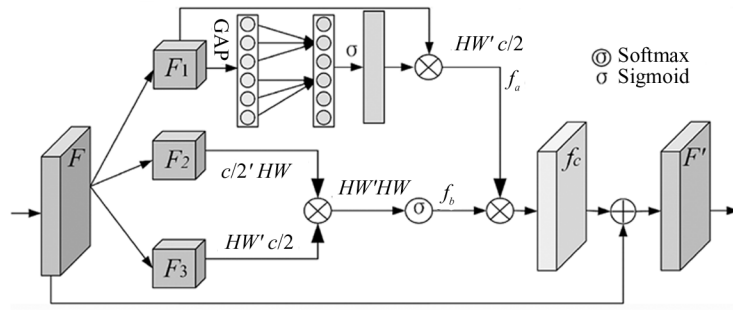


图3 非局部通道细化模块 NCRM 结构图

Fig. 3 NCRM structure diagram of nonlocal channel refinement module

1.4 分组抽样

给定一个未标记的数据集 $X = \{x_i\}_{i=1}^N$, 通过使用特征嵌入函数获得特征集 $V = \{v_0, v_1, \dots, v_{n-1}\}$, 如式(9)所示:

$$v = f_\theta(x), \quad (9)$$

式中, f_θ 是一个基于 CNN 的编码器, 参数为 θ , 用来将图像 x 映射到特征 v 。

为了获取伪标签, 利用伪标签生成器 G, 将特征集分为离群值集和聚类集 C。其中 C 中包含了 S 个聚类的行人样本, 聚类集 C 通过式(10)表示。式中每个集群至少包含两个行人样本。

$$C = \{c_0, c_1, \dots, c_{S-1}\}. \quad (10)$$

将经特征细化网络提取出的行人特征放入一个随机更新的记忆库 (Memory, M) 中进行储存, 在每个 epoch 的开始阶段, 将 M 中的特征作为对比损失的分器, 利用 M 中的特征来计算距离 (余弦相似度) 矩阵进行聚类并分配伪标签。在采样过程中, 采取组抽样策略实施分组抽样, 通过将样本划分为正确的类别来帮助网络生成可靠的伪标签。通过上述过程, 生成了带有伪标签的未标记数据集 $\{x_i, y_i\}_{i=1}^n$, 其中 $y_i \in \{1, 2, \dots, Y_i\}$, 表示聚类中簇的数量。同时, 通过对编码器进行优化, 并使用编码特征动态更新记忆库, 优化公式定义如式(11)所示:

$$M[x] = \alpha \cdot M[x] + (1 - \alpha) \cdot v, \quad (11)$$

式中, $M \in R^{N \times d}$, d 是特征维数, $\alpha \in [0, 1]$ 为记忆库中更新样本特征的动量系数, v 表示映射到特征空间的实例特征。

1.5 耐噪声对比学习

基于聚类的无监督 ReID 方法依赖于聚类算法和模型优化过程之间的迭代训练。而聚类过程产生的噪声标签会导致网络产生错误的优化过程, 影响无监督模型的精准度。近年来关于有噪声标签学习

的研究主要可分为样本重加权方法、标签校正方法和鲁棒损失设计方法, 无论是样本重加权还是标签校正方法, 都需要引入额外干净的数据来帮助网络纠正含有噪声的伪标签。为此, WANG 等^[13] 提出了一个可靠的对称交叉熵 (symmetric cross-entropy, SCE) 损失用于鲁棒学习。在本文中, 受 SCE 的启发, 在对比损失函数基础上设计了 DCE 损失函数用于耐噪声对比学习, 减轻噪声样本的负面影响的同时适应每个聚类过程后聚类质心的变化。具体工作如下:

1) 遵循基于伪标签的无监督 ReID 方法, 内存库中的特征利用伪标签生成器 G 聚类, 每个实例被分配一个伪标签来构造对比损失, 对比损失使一个实例接近其簇的质心, 并被推离其他的簇。将对比损失 L_V 定义为如式(12)所示:

$$L_V = -\log \frac{\exp(\langle v, c^+ \rangle / \tau_1)}{\sum_{k=0}^{K-1} \exp(\langle v, c_k \rangle / \tau_1)}, \quad (12)$$

式中, v 表示 f 映射到特征空间的实例编码特征, c^+ 表示 v 所属的簇的质心 (每个伪身份的特征中心); \langle, \rangle 表示计算两个特征向量之间的内积; c_k 表示 C_k 所属的簇的质心, k 是聚类中簇的数量, τ_1 为超参数。

2) 考虑到聚类算法中存在的噪声样本, 影响了模型的优化, 本文的目标是得到一个鲁棒的损失函数来抵抗噪声标签。首先通过直接平均 M 中具有相同伪标签的特征 v_i 来构建变化的质心, 为此引入了动态交叉熵损失 L_D 的计算, 如式(13)所示。其次, 在 L_D 的基础上, 受 SCE 损失的启发, 本文设计了式(14)中的 L_C 作为改进后的 SCE 损失。

$$L_D = -y_i^T \log [\text{softmax}(c_k^T v_i / \tau_2)], \quad (13)$$

$$L_C = L_D - [\text{softmax}(c_k^T v_i / \tau_2)]^T \log [\text{softmax}(y_i)], \quad (14)$$

式中, \hat{y}_i 表示赋予样本特征 v_i 的伪标签值, $c_k \in R^{k \times d}$

表示 v_i 所属簇的质心, v_i 表示当前模型提取的第 i 个训练样本的特征, $\text{softmax}(\cdot)$ 表示元素级的 softmax 函数计算; τ_2 为超参数。

3) 为了优化网络的训练, 本文将对损失 L_V 和改进的 SCE 损失 L_C 通过式(15)对网络进行联合监督学习, 定义为最终的 DCE 损失。其中, λ 是平衡两损失项之间的平衡因子。

$$L_{DCE} = \lambda L_V + L_C. \quad (15)$$

2 实验结果与分析

2.1 实验设置

本文实验环境配置为显卡 GeForce RTX 2080Ti, CPU 为 4 核 Intel(R) Xeon(R) Silver 4110 CPU @ 2.10 GHz, 操作系统为 Ubuntu 18.04, CUDA10.0 和深度学习框架 Pytorch1.1.0。使用 DBSCAN 作为伪标签生成器 G, 在每个 epoch 开始之前使用 Jaccard 距离进行聚类。行人图像预处理成 256×128 大小; 模型训练过程中, 同时引入了随机翻转、裁剪和擦除等策略实现数据增强; batchsize 大小设置为 64; epoch 大小设置为 50; 初始学习率设置为 0.00035, 每 20 个 epoch 后除以 10; k 设置为 30; 超参数 τ_1 和 τ_2 均设置为 0.05。在 Market-1501 和 DukeMTMC-reID 数据集上使用上述实验设置进行训练大约需要 3.5 h。

2.2 数据集描述及评价标准

Market-1501 数据集是从清华大学校内的 6 个不重叠的摄像机视图中收集的, 包含 1501 个行人, 训练集行人数量为 751, 共有 12936 张图像; 测试集

行人数量为 750, 共有 19732 张图像。DukeMTMC-reID 是从杜克大学校内的 8 个不重叠的摄像机视图中收集的, 包含 1404 个行人, 训练集行人数量为 702, 共有 16522 张图像; 测试集行人数量为 702, 共有 17661 张图像。

实验采用平均精度均值 (mAP) 和 CMC 曲线作为评估标准。 mAP 反应检索的行人在数据库所有正确的图片中处在顺序列表前面的程度, 计算过程如式(16)所示:

$$mAP = \frac{\sum_{i=1}^n AP_i}{C}, \quad (16)$$

式中, AP_i 表示每个类别的平均精度, C 表示类别数。

CMC 曲线可以通过 Rank- n 准确率的形式给出, 表示待查询图像与候选图像按照相似度从大到小排序的前 n 张候选图片中存在与查询图像属于同一身份的准确率。计算如式(17)所示:

$$\text{Rank-}n = \frac{1}{N} \sum_{i=1}^N \begin{cases} 1, r_i \leq n \\ 0, r_i > n \end{cases}, \quad (17)$$

式中, N 表示查询数据集中行人的数目, 总共 N 次查询和排序, r_i 表示候选图片中的首张图片与查询图片属于同一身份的行人图像的位置。

2.3 实验过程

2.3.1 对比精准度实验

将本文算法与基线算法在训练阶段保持相同网络结构的前提下, 在 Market-1501 和 DukeMTMC-reID 数据集中比较改进前后的 mAP 的变化关系, 实验结果如图 4 中的(a)和(b)所示。从图 4 可以看出, 在两大数据集的初始训练阶段, 本文所提网络的

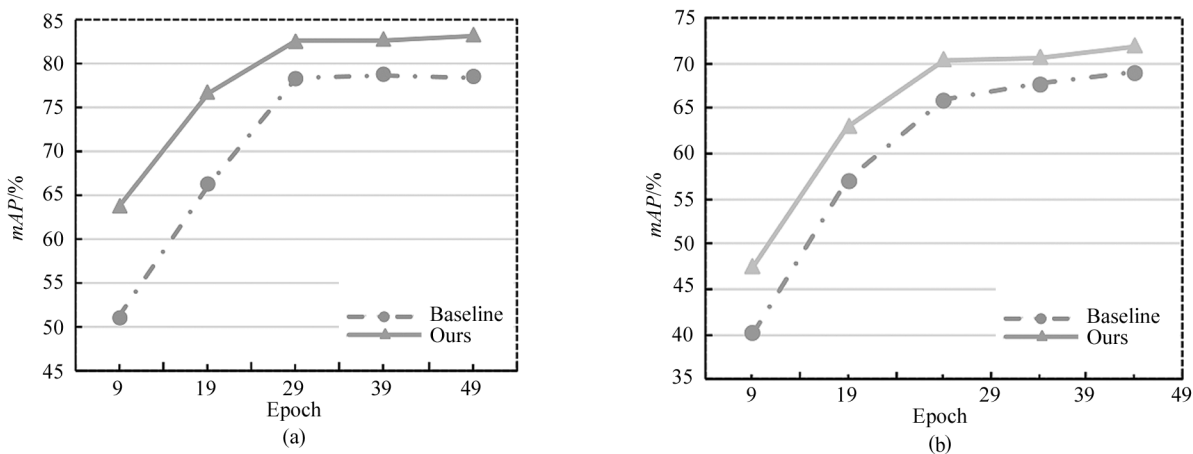


图 4 不同数据集训练阶段 epoch- mAP 变化图: (a) Market-1501 的 epoch- mAP 变化过程;

(b) DukeMTMC-reID 的 epoch- mAP 变化过程

Fig. 4 The epoch- mAP change map of different data sets in training stage: (a) The epoch- mAP change process of Market-1501;

(b) The epoch- mAP change process of DukeMTMC-reID

mAP 值大幅高于基线网络,并且在整个训练过程中,ours- mAP 曲线始终位于 baseline- mAP 上方,在训练的最终阶段,模型的精准度相比于基线网络均获得了有效的提升。

2.3.2 整体消融实验

为了验证特征细化和耐噪声学习过程对算法性能的影响,针对不同方法组合的情况,在 Market-1501 数据集下开展了消融实验,实验结果如表 1 所示。其中,Ours1 表示将原始网络中的平均池化层替换为 GEM 池化层;Ours2 表示在原基础网络 Res-Net50 中嵌入所设计的 NCRM;Ours3 表示耐噪声对比学习过程;Ours4 表示整个特征细化学习过程;Ours5 为本文的最终算法模型。

表 1 在 Market-1501 数据集下的消融实验

Tab.1 Ablation experiments under Market1501 dataset

Method	NCRM	GEM	DCE	Market-1501			
				mAP	Rank-1	Rank-5	Rank-10
Baseline				79.2	92.3	96.6	97.8
Ours1		✓		79.8	92.0	96.6	97.6
Ours2	✓			80.0	91.7	96.4	97.6
Ours3			✓	80.2	92.1	96.8	97.6
Ours4	✓	✓		81.0	92.5	97.0	98.0
Ours5	✓	✓	✓	83.1	93.5	97.3	98.5

首先,对基线网络进行测试,其 mAP 达到 79.2%,Rank-1 为 92.3%。其次,在基础网络 Res-Net50 中嵌入设计的 NCRM 和替换原有池化结构后,基线算法的精度均获得了不同程度的提升,充分验证了所提改进之处的有效性。接着,在改进网络

中同时引入二者, mAP 提升了 1.8%,Rank-1 提升了 0.9%,结果表明:在原始特征提取网络中联合 NCRM 和 GEM 池化方法后,增强了模型对融合多语义信息的捕获能力,丰富了特征描述符。在改进网络中单独引入 DCE loss,算法相比于基线 mAP 提升了 1.4%,Rank-1 提升了 0.7%,说明 DCE loss 能够减轻噪声样本产生的负面影响,并适应聚类后行人类别的变化。最终,在基线网络中同时引入三者,改进算法最终 mAP 达到了 83.1%,Rank-1 达到 93.5%,相比于基线网络, mAP 提升了 3.9%,Rank-1 提升了 1.2%,算法有效性得到了验证。

2.3.3 参数调优分析

为了探究 DCE 损失中 λ 取值对模型性能的影响,本文选取了 $\lambda=1.0,0.9,0.8,0.7,0.6$ 五个不同的数值,针对 λ 对 mAP 和 Rank-1 的影响分别在 Market-1501 和 DukeMTMC-reID 数据集上进行实验,实验结果如表 2 和图 5 所示。表 2 为不同参数值

表 2 参数 λ 在不同数据集上的实验结果

Tab.2 Experimental results of parameter λ in different datasets

λ	Market-1501		DukeMTMC-reID	
	mAP	Rank-1	mAP	Rank-1
1.0	82.0	92.5	71.1	84.2
0.9	82.4	93.3	71.6	85.3
0.8	83.1	93.5	71.9	85.6
0.7	82.6	92.3	71.2	84.5
0.6	81.2	92.0	70.9	84.1

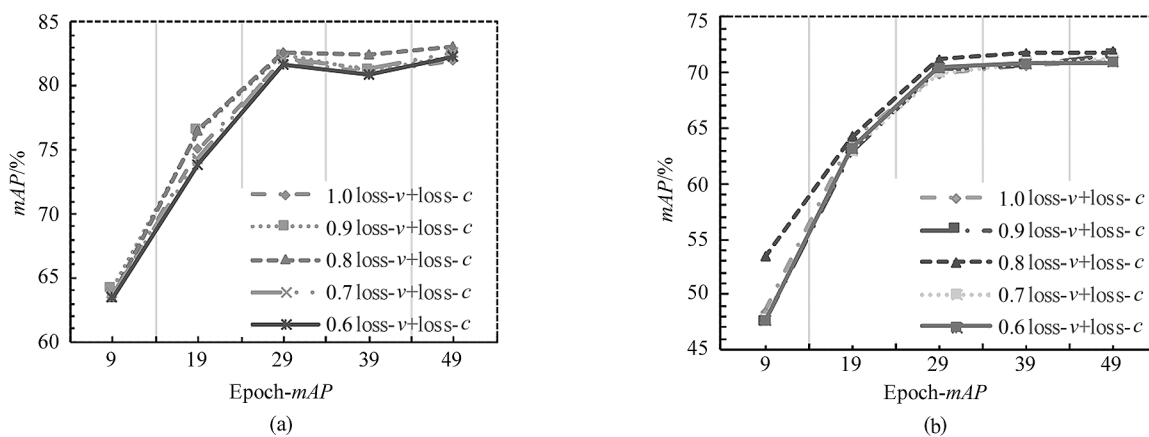


图 5 不同 λ 取值时 mAP 随 epoch 变化曲线图:(a) Market-1501 数据集上 mAP 变化过程;
(b) DukeMTMC-reID 数据集上 mAP 变化过程

Fig.5 Curves of mAP changing with epoch when λ values are different:(a) mAP change process on Market-1501 dataset;(b) mAP change process on DukeMTMC-reID dataset

在两大数据集中所取得的 mAP 和 Rank-1 值,图 5 为不同参数值在两大数据集中所取得 mAP 值的训练过程。表 2 的实验结果显示,在两个数据集中, λ 逐渐减小时, mAP 和 Rank-1 均呈先上升后下降趋势;当 λ 取 0.8 时,在两个数据集上均呈现较好的实验结果,表明参数 λ 对模型的影响在两个数据集上获得了一致性。观察图 5,在网络逐渐拟合的过程中,5 条曲线以不同的速度上升,在 epoch 到达 49 时, λ 的取值对 mAP 的影响相对较小,因为模型在优化到一定程度时,最容易匹配的首位行人已被全部找到。

因此, λ 取 0.8 时,训练网络效果达到最佳。

2.4 与其他无监督 ReID 方法对比

选取近 3 年其他先进的 FU 方法和 UDA 方法在两大数据集中与本文方法进行对比,并选取同一指标 mAP 和 Rank- n 进行评估,表 3 表示在 Market-1501 和 DukeMTMC-reID 数据集上与其他先进方法的比较结果。其中“—”表示原论文中没有报告的数据。

从表 3 中可以看出,本文算法的首位击中率和平均精度均优于其他经典的无监督 ReID 方法。本

表 3 在 Market-1501 和 DukeMTMC-reID 数据集上与其他先进的无监督行人重识别方法的对比

Tab. 3 Unsupervised ReID performance comparison with state-of-the-art methods on Market-1501 and DukeMTMC-reID datasets

Method	Source	Market-1501				Source	DukeMTMC-reID			
		mAP	Rank-1	Rank-5	Rank-10		mAP	Rank-1	Rank-5	Rank-10
Ref. [6]	None	46.7	71.6	86.1	—	None	52.3	69.4	81.5	—
ACAN-OCE ^[14]	None	47.7	72.2	86.3	90.4	None	45.1	67.6	81.2	85.2
ACAN-GRL ^[14]	None	50.6	73.3	87.6	91.8	None	46.6	65.1	80.6	85.1
HCT ^[15]	None	56.4	80.0	91.6	95.2	None	50.7	69.6	83.4	87.4
CVSE ^[16]	Duke	63.2	84.1	92.8	95.0	Market	56.1	75.3	82.9	85.4
MLC ^[17]	None	66.2	86.7	93.5	95.6	None	52.3	73.6	82.3	85.5
IICS ^[18]	None	72.9	89.5	95.2	97.0	None	64.4	80.0	89.0	91.6
SpCL ^[5]	None	73.1	88.1	95.1	97.0	None	65.3	81.2	90.3	92.2
MSC-GDC ^[19]	Duke	76.1	90.1	95.7	97.5	Market	66.4	80.1	89.9	92.3
CAP ^[10]	None	79.2	91.4	96.3	—	None	67.3	81.1	89.3	—
Baseline ^[12]	None	79.2	92.3	96.6	97.8	None	69.1	82.7	91.1	93.5
Ref. [20]	Duke	80.2	91.3	96.5	97.9	Market	70.4	82.4	90.6	92.5
Ours	None	83.1	93.5	97.3	98.5	None	71.9	85.6	91.8	94.3

文算法与 SPCL^[5] 算法相比,在 Market-1501 和 DukeMTMC-reID 数据集上的无域迁移学习的设置下, mAP 分别提升了 10.0% 和 6.6%, Rank-1 分别增长了 5.4% 和 4.4%。SPCL 算法与本文均不需要人为标记的数据和域迁移学习来提升模型的识别精度,且都采用特征聚类和伪标签优化来进行对比学习。不同的是,通过对比实验结果表明本文算法对深层细化特征的捕获能力更强,验证了本文算法的有效性。本文算法与 MLC^[17] 算法相比,在 Market-1501 和 DukeMTMC-reID 数据集上的无域迁移学习的设置下, mAP 分别提升了 16.9% 和 19.6%, Rank-1 分别增长了 6.8% 和 12.0%。与 IICS^[18] 算法相比,在 Market-1501 和 DukeMTMC-reID 数据集上,本文算法的 mAP 值分别高出 IICS^[18] 算法 10.2% 和 7.5%, Rank-1 分别高出其 4.0% 和 5.6%, 与其利用基础 ResNet50 骨干网络提取特征相比,本文算法的特征细化网络更能帮助无监督网络捕获

充分的行人表征,具有良好的识别能力。

3 结 论

本文提出了一种联合特征细化和耐噪声对比学习的无监督 ReID 方法。通过分析现有骨干网络提取行人表征的不足和聚类过程中噪声标签带来的负面影响,分别提出了改进的特征细化网络和耐噪声的对比损失函数。前者倚靠 NCRM 网络的特征强化和细粒度区域的激活响应来获取充分的特征表达,后者借助鲁棒损失的设计来减轻噪声样本对模型的干扰,最终联合两者来提升无监督重识别精度。在两个公共数据集中,围绕网络精度提升、消融、参数调优以及与其他先进方法对比等方面开展了实验,充分验证了所提方法的有效性和良好的识别性能。在下一步的工作中,将继续改进模型,考虑学习无标记数据特征之间的分布差异,思考如何构造约束条件以实现行人特征的分布对齐,获得更高的识

别精度。

参考文献:

- [1] CUI P, MA C. Person reidentification based on second-order mixed attention[J]. *Journal of Optoelectronics · Laser*, 2021, 32(6):645-652.
崔鹏, 马超. 基于二阶混合注意力的行人重识别[J]. *光电子·激光*, 2021, 32(6):645-652.
- [2] ZHANG X H. Person reidentification based on attention mechanism and local association feature[J]. *Journal of Optoelectronics · Laser*, 2022, 33(9):984-991.
张晓涵. 基于注意力机制与局部关联特征的行人重识别[J]. *光电子·激光*, 2022, 33(9):984-991.
- [3] CUI P, FAN Z X. Person re-identification based on domain discriminative network and domain adaptation[J]. *Journal of Optoelectronics · Laser*, 2019, 30(6):632-639.
崔鹏, 范志旭. 基于域鉴别网络和域自适应的行人重识别[J]. *光电子·激光*, 2019, 30(6):632-639.
- [4] AINAM J P, QIN K, OWUSU J W, et al. Unsupervised domain adaptation for person reidentification with iterative soft clustering[J]. *Knowledge-Based Systems*, 2021, 212:106644.
- [5] GE Y X, ZHU F, CHEN D P, et al. Self-paced contrastive learning with hybrid memory for domain adaptive object reid[C]//34th Conference on Neural Information Processing Systems (NIPS), December 6-12, 2020, Online, Cambridge: MIT Press, 2020:11309-11321.
- [6] ZHANG B H, ZHU S Y, ZHOU Y F, et al. A novel unsupervised person re-identification algorithm based on soft multi-label and compound attention model[J]. *Multimedia Tools and Applications*, 2022, 81(17):24081-24098.
- [7] ZHANG B H, ZHU S Y, LV X Q, et al. Soft multilabel learning and deep feature fusion for unsupervised person re-identification[J]. *Opto-Electronic Engineering*, 2020, 47(12):15-24.
张宝华, 朱思雨, 吕晓琪, 等. 软多标签和深度特征融合的无监督行人重识别[J]. *光电工程*, 2020, 47(12):15-24.
- [8] VALEM L P, PEDRONETTE D C G. Person Re-ID through unsupervised hypergraph rank selection and fusion[J]. *Image and Vision Computing*, 2022, 123:104473.
- [9] SRIDHAR R S, MUNAGA V. N. K P, RAMADOSS B. Spatio-temporal association rule based deep annotation-free clustering (STAR-DAC) for unsupervised person re-identification[J]. *Pattern Recognition*, 2022, 122:108287.
- [10] WANG M L, LAI B S, HUANG J Q, et al. Camera-aware proxies for unsupervised person re-identification[C]//35th AAAI Conference on Artificial Intelligence (AAAI), February 2-9, 2021, Online. New York: AAAI Press, 2021, 4A:2764-2772.
- [11] ZHANG Z, BI X J. Pedestrian re-recognition by unsupervised clustering based on style transformation[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(1):48-56.
张智, 毕晓君. 基于风格转换的无监督聚类行人重识别[J]. *智能系统学报*, 2021, 16(1):48-56.
- [12] HAN X M, YU X H, NAN J, et al. Rethinking sampling strategies for unsupervised person re-identification [EB/OL]. (2021-7-7) [2022-9-17]. <https://doi.org/10.48550/arXiv.2107.03024>.
- [13] WANG Y S, MA X J, CHEN Z Y, et al. Symmetric cross entropy for robust learning with noisy labels[C]//IEEE International Conference on Computer Vision (ICCV), October 27-November 3, 2019, Seoul, South Korea. New York: IEEE, 2019:322-330.
- [14] QI L, WANG L, HUO J, et al. Adversarial camera alignment network for unsupervised cross-camera person re-identification[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(5):2921-2936.
- [15] ZENG K W, NING M N, WANG Y H, et al. Hierarchical clustering with hard-batch triplet loss for person reidentification[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 16-18, 2020, Seattle, WA, USA. New York: IEEE, 2020:13654-13662.
- [16] ZHOU S R, WANG Y, ZHANG F, et al. Cross-view similarity exploration for unsupervised cross-domain person re-identification[J]. *Neural Computing and Applications*, 2021, 33(9):4001-4011.
- [17] LI Q, PENG X J, QIAO Y, et al. Unsupervised person re-identification with multi-label learning guided self-paced clustering[J]. *Pattern Recognition*, 2022, 125:108526.
- [18] XUAN S Y, ZHANG S L. Intra-inter camera similarity for unsupervised person reidentification[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 21-24, 2021, Online. New York: IEEE, 2021:11921-11930.
- [19] PANG Z Q, GUO J F, MA Z Q, et al. Median stable clustering and global distance classification for cross-domain person re-identification[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(5):3164-3177.
- [20] LU Y C, DENG W H. Transferring discriminative knowledge via connective momentum clustering on person re-identification[J]. *Pattern Recognition*, 2022, 126:108569.

作者简介:

王凤随 (1981—), 男, 博士, 教授, 硕士生导师, 主要从事图像与视频信息处理和计算机视觉等方面的研究。