



研究与开发

双分支引导对比学习的无监督行人重识别

任航佳, 梁凤梅

(太原理工大学电子信息工程学院, 山西 晋中 030600)

摘 要: 现有无监督行人重识别算法使用残差网络, 仅能提取粗略的全局特征, 对细微的局部特征反映不足, 且聚类方法生成的伪标签会引入噪声, 影响特征判别。针对上述问题, 提出一种双分支引导对比学习的方法。首先, 引入一种有效的特征提取方式, 将提取的特征分为全局分支和局部分支, 提高对局部信息的利用; 其次, 通过全局特征和局部特征之间的一致性细化全局特征预测的伪标签, 充分利用局部特征和整体特征之间的互补关系, 有效降低伪标签聚类产生的噪声; 最后, 引入对比学习模块, 将细化的标签进行对比学习, 提高模型的鲁棒性。在 Market1501、DukeMTMC-ReID 以及 MSMT17 数据集上的实验结果验证了所提方法的有效性及高性能。

关键词: 无监督行人重识别; 全局特征; 局部特征; 标签细化; 对比学习

中图分类号: TP391

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2025125

Dual branch guided contrastive learning for unsupervised pedestrian re-identification

REN Hangjia, LIANG Fengmei

College of Electronic Information Engineering, Taiyuan University of Technology, Jinzhong 030600, China

Abstract: The current unsupervised pedestrian re-identification algorithms using residual networks can only extract rough global features, but it can't adequately reflect subtle local features. In addition, the pseudo labels generated by clustering methods introduce noise, which will affect the performance of feature discrimination. A dual branch guided contrastive learning method was proposed. Firstly, an effective feature extraction method was introduced, which divided the extracted features into global branches and local branches to improve the utilization of local information. Secondly, the consistency between global and local features was proposed to refine the pseudo labels for global feature prediction, utilizing the complementary relationship between local and global features, thereby effectively reducing the noise generated by pseudo label clustering. Finally, a contrastive learning module

收稿日期: 2024-12-14; 修回日期: 2025-04-28

通信作者: 任航佳, 270027679@qq.com

基金项目: 虚拟现实技术与系统全国重点实验室开放课题基金项目 (No.VRLAB2023A06); 山西省科技合作交流专项 (No.202104041101030)

Foundation Items: Open Project Fund of National Key Laboratory of Virtual Reality Technology and Systems (No.VRLAB2023A06), Shanxi Science and Technology Cooperation and Exchange Project (No.202104041101030)

was proposed to perform contrastive learning on refined labels and improve the robustness of the model. The experimental results on the Market1501, DukeMTMC-ReID, and MSMT17 datasets validate the effectiveness of the proposed method.

Key words: unsupervised pedestrian re-identification, global feature, local feature, label refinement, contrastive learning

0 引言

行人重识别^[1]是指在不同的视角、时间和环境条件下,对同一行人进行识别和匹配,该技术在视频监控、智能交通和个人行为识别等领域有着广泛的应用。随着深度学习的快速发展,相关领域已经取得了一定的进展,其中有监督的行人重识别方法^[2-3]已经实现了较满意的准确率。然而,有监督的方法需要大量的标签标注,这大大增加了人力、物力和财力的负担,在这种情况下,无监督的行人重识别方法越来越受到重视。目前的无监督方法主要有完全无监督方法^[4-6]和无监督域自适应方法^[7-9],其中纯粹的无监督行人重识别通常从未标注的数据中利用伪标签来训练,而无监督域自适应的行人重识别先在源标记数据集上预训练模型,然后在目标未标记数据集上对模型进行微调。由于引入了外部源域,域自适应的性能优于纯粹的无监督学习。然而,域自适应仍然受到复杂训练过程的困扰,并且其要求源域和目标域之间的差异不显著。因此,本文侧重没有任何数据标记情况下的行人重新识别,即纯无监督学习。

在无监督行人重识别网络训练过程中,基于伪标签的方法可以从未标记的数据中学习 to 拥有较好鲁棒性和判别性的表征特征,这些方法大多采用伪标签聚类的训练方案。具体来说,首先使用特征提取网络从数据集中提取人物图像特征,然后使用聚类算法,如基于密度的聚类(density-based spatial clustering of applications with noise, DBSCAN)算法^[10]、 k -means^[11]等,生成提供监

督信息的伪标签,最后通过一组损失函数来约束和优化训练过程。经过多次迭代,模型将提高特征质量,聚类方法将减少伪标签和真标签之间的差异。然而,在无监督任务中通常采用残差神经网络(residual neural network, ResNet)^[12]来提取全局特征,对细微的局部特征反映不足,并且在聚类训练过程中会不可避免地引入噪声,这将影响判别特征时的性能。

为了提高局部信息的利用率,文献[13]提出协同多特征聚类框架,利用全局特征和局部特征挖掘未标记数据集上潜在的身份相似性,该方法通过多特征提高了相似度学习的准确性,从而提升了网络性能,但其使用的局部特征是通过最终的全局特征图进行简单分块得到的,忽略了浅层局部特征中的细节信息,且全局特征和局部特征之间的关系并未得到充分利用。为了减少伪标签噪声对实验性能的影响,文献[14]通过相互学习并引入辅助网络的方式来细化伪标签,虽然该方法在一定程度上可以提高伪标签的质量,但在减少噪声影响的同时却忽略了局部的细微特征,导致在复杂环境下的特征鲁棒性不足。文献[4]通过利用聚类结果指导特征学习,并在不同的数据增强视图中分别针对实例级和聚类级对比学习来帮助孪生网络学习与聚类结果相关的判别特征,但其主要针对聚类指导的不对称进行光对比学习,具有一定的局限性。文献[5]通过动态混合对比学习方法结合身份和图像的对比学习,有效地探索了硬样本对之间的特征相似性,从而提高了无监督行人重识别的性能,但其在特征噪声和标签噪声的处理上略显不足。



为解决上述问题,本文的主要贡献如下。

(1) 为了增强对细节特征的提取,引入一种局部分支模块,提高对细节特征的挖掘程度。

(2) 为了降低伪标签聚类产生的噪声,提出一种新的伪标签细化模块,通过全局特征和局部特征之间的关联来进行优化,利用局部特征和整体特征之间的互补关系来降低噪声带来的影响。

(3) 为了提高识别的准确度,引入对比学习模块,通过全局特征和局部特征之间的互相对比,减少全局分支和局部分支之间的差异。

(4) 为了提高对比学习的效率,引入新的损失函数。

1 算法原理与网络结构

本文提出了一种双分支引导对比学习的无监督行人重识别方法,对 cluster-contrast 方法^[15]中提取特征的编码器进行了改进,网络结构整体框架如图1所示。为了捕获局部线索特征,本文使用双分支结构的特征提取器模型。首先,将主干网络之前的结构指定为两个分支的公共部分,在主干网络之后将网络划分为全局分支和局部分支,在全局分支中输入完整的图像,而在局部分支中将输入的图像分割成若干条,以获得更多的区域语义特征,在分割过程中确定不同区域的边

界,将图像或数据划分为多个较小的区域,使模型能够更加聚焦于局部的细节信息,为了避免分割过程不够准确而导致一些区域的语义信息丢失或错误,后续工作中采用优化算法来确保分割的准确性,并使用一致性约束函数优化模型;然后,通过 DBSCAN 算法对每个特征进行聚类,在此基础上,通过全局特征和局部特征之间的一致性来提升伪标签的质量,利用局部特征和整体特征之间的互补关系来降低噪声带来的影响;最后,设计一种一致性对比学习的方法,在对比学习中,利用聚类算法产生的聚类感知特征代替实例特征,并利用一致性损失函数减少全局分支与局部分支之间的差异,从而提高识别的准确度。

模型的整体框架主要由以下4个部分组成。

(1) 图像的特征提取。将基础网络结构划分为全局分支和局部分支,分别用来提取所需的全局特征和局部特征,并将其保存在对应的内存库中,所使用的网络结构只有部分基础结构共享。

(2) 伪标签的聚类。将提取到的全局特征和局部特征融合为新的融合特征,再通过聚类算法为融合特征其分配伪标签。

(3) 伪标签的细化。因为聚类算法在生成伪标签的同时会引入新的噪声,所以本文通过标签细化来减小噪声对实验带来的影响,利用局部特

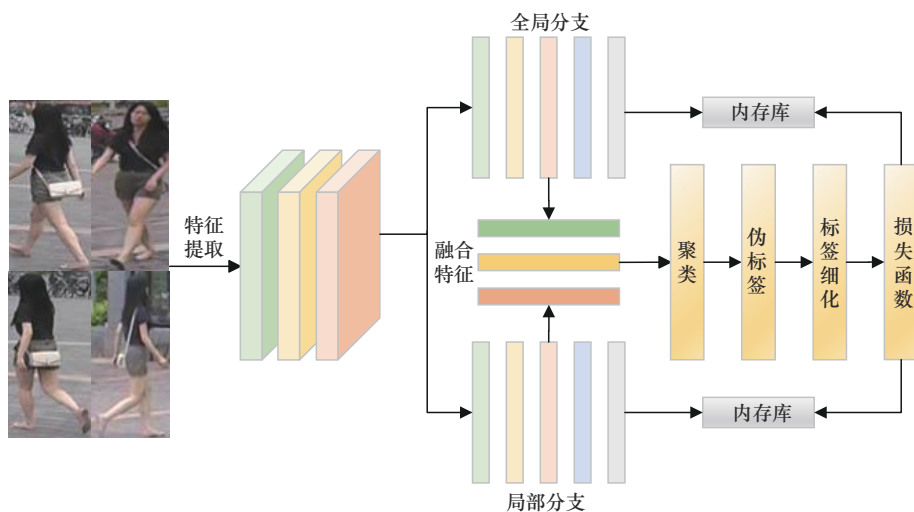


图1 网络结构整体框架

征和整体特征之间的互补关系来进行伪标签细化的一系列工作。

(4) 一致性对比学习。将细化后的融合特征通过一致性对比函数分别与两个内存库中的全局特征和局部特征进行对比, 并使用一致性约束函数来减少全局分支和局部分支之间的差异, 从而提高对比学习的效率。

1.1 特征提取器

本文使用的特征提取器基于 ResNet-50, 一个经典的特征提取网络结构。为了增强对局部信息的利用, 本文提出了双分支结构的特征提取器模型, 特征提取器结构如图 2 所示, 将 ResNet-50 的第四层卷积之后的机制划分为全局分支和局部分支, 将前面的卷积层指定为两个分支共享的基础结构。

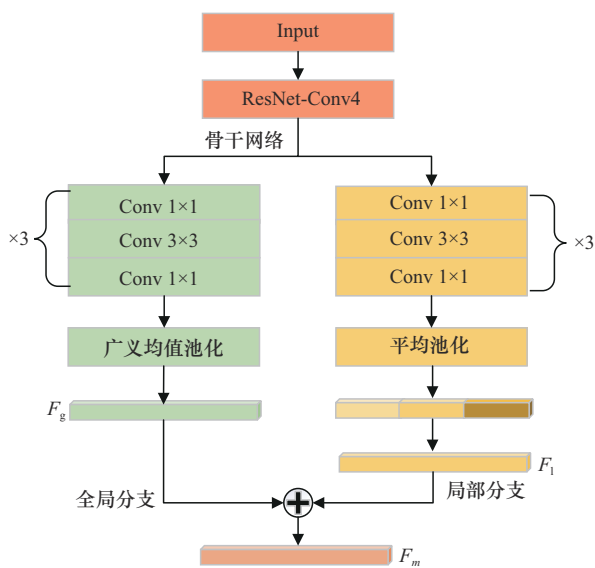


图2 特征提取器结构

(1) 全局分支

在全局分支中, 首先通过第四层卷积到第五层卷积来获得在 ImageNet 上预训练得到的视觉特征; 然后在第五层卷积后采用广义均值池化^[16]和批次归一化, 得到一个可训练的池化层, 以提高表示质量; 最后, 从批次规范层生成每个样本的 2 048 维嵌入, 表示为 F_g 。

(2) 局部分支

在局部分支中, 采用分区策略获取局部粒度特征, 在不引入额外辅助信息的情况下提高了局部信号响应; 在局部分支的初始阶段, 采用类似于全局分支的结构, 通过平均池化对后面的特征进行操作, 将其拆分为多个特征, 表示为 $F_{li}^3, i=1$, 不同的条纹特征将允许模型学习多个粒度的语义特征, 并提供互补信息, 增强特征的辨别能力。此外, 本文使用一组自适应权重来聚合局部分支的嵌入信息以将其作为局部特征的代表:

$$F_l = w_1 \times F_{l1} + w_2 \times F_{l2} + w_3 \times F_{l3} \quad (1)$$

其中, w_1 、 w_2 、 w_3 是可训练权重参数, 在每个反向传播期间被更新并且受到约束, 确保其总和等于 1。

在该模型中, 使用 ImageNet 预训练的参数重建来自 ResNet-50 的层间关系, 以确保稳定的初始参数集。输入图像大小调整为 128×384 , 由此得到全局特征 F_g 与局部特征 F_l 线性融合的最终特征 F_m , 其相互关系如下。

$$F_m = W_1 \times F_g + W_2 \times F_l \quad (2)$$

其中, W_1 、 W_2 为预训练权重。

1.2 伪标签聚类

对于融合特征 F_m , 本文计算了 Jaccard 距离并采用 DBSCAN^[10] 模型来生成伪标签, 然后将生成的伪标签送入标签细化模块进行处理。虽然伪标签在训练过程的初期是不准确的, 但随着学习阶段的提升聚类的精度会逐渐提高。

1.3 标签细化模块

文献[12]使用精细的伪标签来训练模型, 这些伪标签考虑了全局特征和局部特征之间的互补关系。然而, 由于全局特征和局部特征的相似度结构不同, 存在不可靠的互补信息。部分局部特征包含了与人无关的信息, 不适合对全局特征进行伪标签的提炼。此外, 全局特征只考虑全局上下文, 有时会忽略与部分特征相关的信息, 因此, 提高信息的互补性是增强性能的关键。



为了解决这个问题，本文引入交叉协议分数，它可以捕获全局特征和局部特征的 k 近邻相似度，将交叉一致分数定义为全局特征和局部特征的最近邻之间的Jaccard相似度。首先，对全局特征空间和每个局部特征空间执行 k 近邻搜索，生成每个图像上的排序列表，然后，计算图像的全局特征与第 n 部分特征空间 p_n 之间的交叉一致分数，如下。

$$C_i(g, p_n) = \frac{|R_i(g, k) \cap R_i(p_n, k)|}{|R_i(g, k) \cup R_i(p_n, k)|} \in [0, 1] \quad (3)$$

其中， $R_i(g, k)$ 、 $R_i(p_n, k)$ 分别为由 F_g 和 F_l 计算的排名表中的Rank- k 样本指标， $|\cdot|$ 为该集合的基数， g 为全局特征的标识， k 为近邻数。

本文在文献[12]中的平滑标签基础上，通过学习局部特征来提高对全局特征的预测，根据相应的交叉一致分数来细化每个身体部位的伪标签，反映了每个身体部位全局聚类结果的可靠性，伪标签细化结构如图3所示。给定局部特征 $p_n(n=1,2,3)$ 的伪标签为 y_t ，局部特征的标签平滑表示如下。

$$g_t = \gamma_t \times y_t + (1 - \gamma_t) \times p \quad (4)$$

其中， γ_t 为决定标签平滑强度的权值， y_t 为目标图像的伪标签， p 为一致向量。与传统的标签平滑对 γ_t 采用恒定权值不同，这里每个部分和 γ_t 的权值是根据交叉一致分数动态调整的。

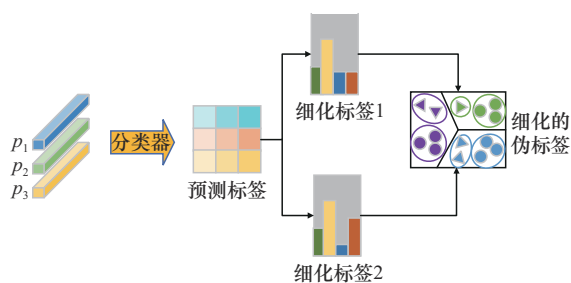


图3 伪标签细化结构

全局特征与局部特征伪标签相互精炼的核心是只保留与其样本伪标签一致的样本，在此基础上引入相互引导的伪标签细化策略，通过预测局

部特征的伪标签来细化全局特征的伪标签，同时对局部特征的伪标签利用全局特征的伪标签进行细化。考虑局部特征中差异较小的片段可能会产生误导性的影响，本文利用交叉一致分数对局部特征伪标签的预测进行汇总，以提高伪标签的准确性，精炼后的全局特征伪标签如下。

$$g_t = \left(y_t - \sum_{i=1}^N u_i \beta_i \right) + (1 - \alpha) u_i \beta_i \quad (5)$$

其中， u_t 为集合权重， $\alpha \in [0, 1]$ 为控制伪标签与集成预测比例的权值， N 为样本的数量， $\beta_t = h(f_t)$ 为第 n 个局部特征的预测向量，其中 $h(\cdot)$ 由一个全连接层和一个Softmax函数组成。若给定所获得的精细伪标签为 g_t ，则相互细化伪标签 L_m 表示如下。

$$L_m = - \sum_{i=1}^N [g_t \times \log(a_i)] \quad (6)$$

其中， a_t 是全局特征的预测向量，这种相互引导的标签细化可以帮助模型充分利用全局和局部特征的信息。

1.4 一致性对比学习模块

大多数对比学习的方法中，实例级对比学习机制在很大程度上忽略了一致性特征的交互。此外，这些方法还存在样本不平衡的问题，即样本较多的聚类比样本较少的聚类获得更多的更新机会，从而导致不同类别之间的更新率存在差异。

为了解决这个问题，可引入一致性对比学习来捕获全局和局部特征的簇感知一致性。具体而言，本文首先构建了两个存储库来保存聚类感知特征，并将 F_g 和 F_l 分别放入两个存储库中，为了更新内存库中的信息，本文引用ClusterNCE Loss^[15]将新生成的特征与两个内存库中的聚类特征进行比较，一致性对比结构如图4所示，计算如下。

$$L_q = -\log \frac{\exp(q \times \phi / \tau)}{\sum_{k=1}^K \exp(q \times \phi_k / \tau)} \quad (7)$$

其中, L_q 为一致性对比, q 表示从全局和局部分支执行的查询特性, 并将它们组合在一起, τ 用来控制分布集中度, ϕ 为 q 对应的表示向量, ϕ_k 是第 k 个簇的唯一表示向量, 它更新特征向量, 并计算两个聚类级别的损失。

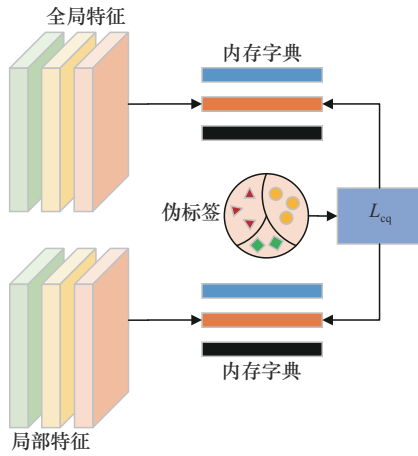


图4 一致性对比结构

为了最小化全局分支和局部分支之间的差异, 本文使用 JS 散度计算两个分支之间的一致性损失, 这种散度测量了两个概率分布之间的相似性, 并且是对称的, 可有效地解决单独使用 KL 散度时的不对称问题。

$$L_c = \text{JS}(h(F_g) \| h(F_l)) = \frac{1}{2} \left[\text{KL} \left(h(F_g) \left\| \frac{h(F_g) + h(F_l)}{2} \right\| \right) + \text{KL} \left(h(F_l) \left\| \frac{h(F_g) + h(F_l)}{2} \right\| \right) \right] \quad (8)$$

其中, L_c 为一致性损失, F_g 和 F_l 表示交叉一致分数的嵌入, 具有全局分支和局部分支信息, 函数 $h(\cdot)$ 由一个全连接层和一个 Softmax 函数组成, JS 和 KL 分别代表 JS 散度和 KL 散度。

模块通过引入由 ClusterNCE 损失和一致性约束损失 L_c 组成的一致性对比学习损失 L_{cq} 模型可以学习到更有效、鲁棒的表示:

$$L_{cq} = L_q + L_c \quad (9)$$

1.5 损失函数

本文采用标签平滑交叉熵损失^[17]和三元组损失^[18]对实验中的全局特征和局部特征进行优化, 给定改进后的伪标签的交叉熵损失如下。

$$L_s = \frac{1}{N} \sum_{t=1}^N (1 - \gamma_t) D_{\text{KL}}[(p \| \beta_t) + \gamma_t U(y_t, \beta_t)] \quad (10)$$

其中, L_s 为标签平滑交叉熵损失, p 为预测样本的概率分布, $U(\cdot)$ 和 $D_{\text{KL}}(\cdot)$ 分别为交叉熵和 KL 散度, 这两个指标用 γ_t 与交叉一致分数的值互相平衡。

改进后的平滑标签损失函数如下。

$$L_g = L_s + L_m \quad (11)$$

本文模型框架有 4 个阶段: 特征提取、伪标签聚类、伪标签细化和一致性对比。在训练过程开始时, 使用在 ImageNet 上预训练的参数初始化特征提取器, 提取训练数据集上的初始特征作为记忆特征, 由 DBSCAN 从 F_m 生成伪标签。在每个训练周期中, 用新的 F_m 作为训练数据集的标签来更新伪标签。在整个框架中, 将一致性对比损失 L_{cq} 和交叉熵平滑标签损失 L_g 结合起来, 形成整体损失如下。

$$L_{\text{all}} = \lambda \times L_{cq} + (1 - \lambda) \times L_g \quad (12)$$

2 实验结果与分析

2.1 实验设置

在训练中, 使用 ImageNet 上预先训练的参数初始化模型、Pytorch 框架和 1 个 RTX A4000 GPU 来进行训练和测试, 将批量大小设置为 128, 对于 Market1501 数据集和 DukeMTMC-ReID 数据集, 输入图像的大小设置为 128×384 , 对于 MSMT17 数据集输入图像的大小设置为 224×224 , 同时使用了几种数据增强策略, 如平移、水平翻转和随机擦除。模型优化采用自适应矩估计 (adaptive moment estimation, ADAM)^[19] 算法, 权重衰减设置为 5×10^{-4} 。聚类模型采用 DBSCAN 和具有 k 倒数近邻的 Jaccard 距离作为距离度量, 在



Market1501、DukeMTMC-ReID 和 MSMT17 数据集上, 最大样本距离参数 ϵ 分别设为 0.4、0.6 和 0.7, 每个人的实例数设为 16, 以确保批量样本的平衡, 每个实验的置信度设为 0.1, 损失函数权重 λ 设为 0.5, 周期设为 50。

2.2 数据集

为了验证方法的有效性, 本文在 Market1501^[20]、DukeMTMC-ReID^[21] 和 MSMT17^[22] 3 个大规模的行人重识别数据集上评估本文的方法。

Market1501 数据集包含 1 501 个身份, 32 668 张图像, 由 6 台不同的相机捕获。训练集包含 751 个身份, 12 936 张图像, 测试集包含 750 个身份, 19 732 张图像, 其中 query 集包含 3 368 张图像, gallery 集包含 16 364 张图像。

DukeMTMC-ReID 数据集包括 8 个 85 min 的 1 080p 视频, 每秒 60 帧, 由 8 台固定摄像机在杜克大学校园内放置, 在讲座间隙进行拍摄。其中, gallery 集包含 17 661 张图像, query 集包含 702 个身份, 2 228 张图像, 训练集包含 702 个身份, 16 522 张图像。

MSMT17 数据集由部署在校园内的 15 台摄像机构建而成, 它有 12 个室外摄像机和 3 个室内摄像机, 为了采集原始监控视频, 在一个月里选择了具有不同天气条件的 4 天, 每天采集 3 h 的视频, 涵盖了早上、中午、下午 3 个时间段。因此, 原始视频时长总共为 180 h, 训练集包含 1 041 个身份, 32 621 张图像, 测试集包括 3 060 个身份, 93 820 张图像, 对于测试集而言, 将 11 659 张图像随机选出来作为 query 集, 而其他的 82 161 张图像作为 gallery 集。MSMT17 数据集是一个更接近真实场景且更复杂的行人重识别数据集。

2.3 评估指标

为了与其他实验方法进行公平比较, 本文采用广泛使用的平均精度均值 (mean average precision, mAP) 和 k 次命中率 (Rank- k) 进行评估。mAP 是将所有类别的平均精度进行综合加权平均

得到的, 它反映了检索的人在数据集中所有正确的图片排在排序列表前面的程度。Rank- k 表示搜索结果中置信度最高的前 n 张图片, 通过计算 n 张图片中含有正确结果的概率, 反映检索的精度。

2.4 实验结果

为了检验所提方法的有效性, 本文在 Market1501、DukeMTMC-ReID 以及 MSMT17 3 个数据集上进行了对比实验, 将本文的方法与该领域近几年的主流方法进行了比较。本文主要研究的是纯无监督的行人重识别问题, 因此在进行实验对比时多数都是与纯无监督方法进行对比, 其中包括: 基于硬样本三元组损失的分层聚类 (hierarchical clustering with hard-batch triplet loss, HCT)^[23]、基于记忆的多标签分类损失 (memory-based multi-label classification loss, MMCL)^[24]、可学习重新识别描述的循环关联 (cycle association for learning re-identifiable descriptions, CycAs)^[25]、相机间的相似性 (intra-inter camera similarity, IICS)^[26]、自步对比学习 (self-paced contrastive learning, SPCL)^[27]、基于聚类一致性的伪标签优化 (refining pseudo labels with clustering consensus, RLCC)^[28]、聚类对比学习 (cluster contrast learning, CCL)^[15]、基于部件的伪标签优化 (part-based pseudo label refinement, PPLR)^[12]、混合对比模型 (hybrid contrastive model, HCM)^[29]、实例间对比编码 (inter-instance contrastive encoding, ICE)^[30]、聚类引导的非对称对比学习 (cluster-guided asymmetric contrastive learning, CACL)^[4]、动态混合对比学习 (dynamic hybrid contrastive learning, DHCL)^[5]、实时内存更新策略 (real-time memory updating strategy, RTMem)^[6]。

Market1501、DukeMTMC-ReID 以及 MSMT17 数据集上与先进方法的比较结果见表 1~表 3。本文方法在不同数据集上取得的结果始终优于其他方法的结果, 在 Market1501、DukeMTMC-ReID

和 MSMT17 数据集上的 mAP 分别为 84.8%、74.4% 和 35.1%，本文以文献[15]中的网络结构作为基线，相比于基线方法分别提高了 2.2、1.8 和 3.9 个百分点，Rank-1 的精度分别为 93.7%、86.2% 和 65.2%，相比于基线方法分别提高了 1.1、0.7 和 3.7 个百分点，该结果表明本文提出的双分支引导对比学习的方法可以有效降低伪标签噪声对实验结果的影响，同时可以利用局部的细微特征提高对比学习的效率和模型的鲁棒性，确保伪标签的准确性和可靠性，正确引导模型的训练过程，从而使得准确率可以有效提高。

表 1 Market1501 数据集上与先进方法的比较结果				
方法	mAP	Rank-1	Rank-5	Rank-10
HCT	56.4%	80.0%	91.6%	95.2%
MMCL	60.4%	84.4%	92.8%	95.0%
CycAs	64.8%	84.8%	—	—
IICS	72.9%	89.5%	95.2%	97.0%
SPCL	77.5%	89.7%	96.1%	97.6%
RLCC	77.7%	90.8%	96.3%	97.5%
HCM	79.0%	91.8%	96.7%	97.7%
ICE	79.5%	92.0%	97.0%	98.1%
CACL	80.9%	92.8%	97.4%	98.5%
PPLR	81.5%	92.8%	97.1%	98.1%
DHCL	81.7%	92.8%	97.2%	98.0%
CCL	82.6%	92.6%	97.1%	98.0%
RTMem	83.0%	92.8%	97.4%	98.3%
本文方法	84.8%	93.7%	97.6%	98.1%

表 2 DukeMTMC-ReID 数据集上与先进方法的比较结果				
方法	mAP	Rank-1	Rank-5	Rank-10
HCT	50.7%	69.6%	83.4%	87.4%
MMCL	51.4%	72.4%	82.9%	85.0%
CycAs	60.1%	77.9%	—	—
IICS	64.4%	80.0%	89.0%	91.6%
ICE	67.2%	81.3%	90.1%	93.0%
HCM	67.9%	82.3%	90.2%	92.8%
RLCC	69.2%	83.2%	91.6%	93.8%
CACL	69.6%	82.6%	91.2%	93.8%
DHCL	71.5%	84.4%	91.9%	94.1%
CCL	72.6%	85.5%	92.7%	94.3%
本文方法	74.4%	86.2%	92.7%	94.5%

表 3 MSMT17 数据集上与先进方法的比较结果				
方法	mAP	Rank-1	Rank-5	Rank-10
MMCL	16.2%	43.6%	54.3%	58.9%
CycAs	26.7%	50.1%	—	—
CACL	23.0%	48.9%	61.2%	66.4%
IICS	26.9%	56.4%	68.8%	73.4%
RLCC	27.9%	56.5%	68.4%	73.1%
HCM	26.9%	59.6%	70.1%	74.3%
ICE	29.8%	59.0%	71.7%	77.0%
PPLR	31.4%	61.1%	73.4%	77.8%
CCL	31.2%	61.5%	71.8%	76.7%
RTMem	32.8%	57.1%	70.0%	74.9%
本文方法	35.1%	65.2%	76.7%	80.2%

3 消融实验及分析

3.1 模块有效性分析

本文的方法总共包含 3 个消融因素，分别是双分支的特征提取模块、标签平滑模块和一致性对比模块，为了验证本文方法的有效性，本文设置了 4 种不同的模型，其中，模型 1 采用文献[15]中的网络结构作为基线，模型 2 在基线的基础上加入双分支的特征提取器，模型 3 在模型 2 的基础上加入标签平滑模块，模型 4 在模型 3 的基础上加入一致性对比模块，本文在 Market1501、DukeMTMC-ReID 和 MSMT17 数据集上的消融实验结果见表 4~表 6。

表 4 Market1501 数据集上的消融实验结果				
模型	mAP	Rank-1	Rank-5	Rank-10
1	82.6%	92.6%	97.1%	98.0%
2	83.1%	92.4%	97.3%	98.0%
3	84.2%	93.2%	97.4%	98.1%
4	84.8%	93.7%	97.6%	98.1%

表 5 DukeMTMC-ReID 数据集上的消融实验结果				
模型	mAP	Rank-1	Rank-5	Rank-10
1	72.8%	85.5%	92.7%	94.3%
2	73.2%	85.7%	92.4%	94.0%
3	74.0%	86.1%	92.7%	94.4%
4	74.4%	86.2%	92.7%	94.5%



表6 MSMT17数据集上的消融实验结果

模型	mAP	Rank-1	Rank-5	Rank-10
1	31.2%	61.5%	71.8%	76.7%
2	32.6%	60.8%	70.7%	75.9%
3	34.5%	64.2%	75.0%	78.8%
4	35.1%	65.2%	76.7%	80.2%

消融实验结果表明, 本文引入的3个模块对模型的性能均有不同程度的提高。其中, 双分支的特征提取器可以提高对信息的利用, 但性能提升不是很明显, 原因是在增加局部特征提取后不可避免地混入了新的噪声, 因此可加入平滑标签来降低伪标签噪声的影响; 从模型3中可以看出加入平滑标签对实验性能有较大提升; 加入一致性对比学习模块, 可以更进一步加强网络的识别能力。

在面对大规模数据集MSMT17时, 以往的无监督行人重识别算法往往在模型训练中出现欠拟合的问题, 因为该数据集包含大量行人、复杂背景和不同光线条件下的图像。然而, 本文引入的双分支对比学习方法有效增强了特征提取能力, 在MSMT17数据集上相比于基线方法性能有了显著提升。具体而言, 与基线方法相比, 本文方法在MSMT17数据集上的mAP提高了3.9个百分点, Rank-1提高了3.7个百分点, Rank-5提高了4.9个百分点, Rank-10提高了3.5个百分点。此外, 在Market1501数据集和DukeMTMC-ReID数据集上, 基于双分支对比学习的无监督行人重识别方法仍然取得了显著进步, 表明本文方法具有较强的通用性。

3.2 特征提取器方法比较

为了验证双分支的特征提取器对提高实验性能的有效性, 本文在Market1501数据集上进行了3组对比实验, 在残差网络ResNet-50的特征提取器上, 将网络分为仅有局部分支的特征提取器和仅有全局分支的特征提取器, 以及双分支的特征提取器, 选择mAP、Rank-1和Rank-5作为评价指标, 不同特征提取器对实验性能的影响见表7。

从表7中可以看出, 双分支的特征提取相比

于单分支其性能有了显著提升。其中, 双分支的mAP的精度比仅局部分支提高了5.6个百分点, 比仅全局分支提高了2.9个百分点, Rank-1均提升了2.3个百分点, 表明双分支的特征提取器可以有效提高信息的利用率, 从而提高实验的性能。同时也能看出, 只有局部分支的特征提取会使性能有较大幅度的下降, 说明仅提取局部特征会造成信息的不准确, 而仅有全局分支的特征提取具有一定的优势, 但加入局部分支后会使其有明显的性能提升, 表明了本文方法的有效性。

表7 不同特征提取器对实验性能的影响

方法	mAP	Rank-1	Rank-5
仅局部分支	79.2%	91.4%	96.4%
仅全局分支	81.9%	91.4%	96.6%
双分支	84.8%	93.7%	97.6%

3.3 一致性对比函数有效性分析

为了验证一致性对比函数 L_q 和 L_c 的有效性, 本文在Market-1501数据集上做了两组对比实验, 分别移除函数 L_q 和 L_c , 同样选择mAP、Rank-1和Rank-5作为评价指标, 不同损失函数实验结果对比见表8。同时, 为了验证标签平滑交叉熵损失函数 L_s 和相互细化标签函数 L_m 的有效性, 本文也在Market-1501数据集上分别移除了函数 L_s 和 L_m , 实验结果见表8。

表8 不同损失函数实验结果对比

方法	mAP	Rank-1	Rank-5
$L_c \& L_s \& L_m$	52.2%	76.5%	88.6%
$L_q \& L_s \& L_m$	84.5%	93.2%	97.1%
$L_q \& L_c \& L_m$	84.1%	92.6%	96.8%
$L_q \& L_c \& L_s$	84.3%	93.0%	96.7%
$L_q \& L_c \& L_s \& L_m$	84.8%	93.7%	97.6%

从表8可以看出, 移除函数 L_q 后, mAP从84.8%下降到52.2%, 性能损失严重, 说明在训练过程中ClusterNCE loss占主导地位。与移除一致性损失函数 L_c 相比, 在Market1501数据集上的

mAP从84.5%提升到84.8%，Rank-1从93.2%提升到93.7%，证明了在复杂环境下，保证全局和局部之间的一致性可以获得更鲁棒的特征。与移除标签平滑交叉熵损失函数 L_s 相比，mAP从84.1%提升到84.8%，提升了0.7个百分点，Rank-1从92.6%提升到93.7%，提升了1.1个百分点；与移除细化标签函数 L_m 相比，mAP从84.3%提升到84.8%，提升了0.5个百分点，Rank-1从93.0%提升到93.7%，提升了0.7个百分点，表明平滑标签和标签细化可以减小噪声的影响，从而提高特征嵌入的质量。

3.4 损失函数超参数设置

本文所提损失函数中有一个重要的超参数，即 L_{cq} 损失函数的权重因子 λ ，为了研究这个超参数对实验结果的影响，本文将权重因子 λ 分别设置为0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8、0.9进行大量实验，并以mAP和Rank-1作为评价指标。损失函数权重因子 λ 的参数分析如图5所示，当 $\lambda=0.5$ 时，实验效果最好，故本设计中损失函数的超参数选取该值。

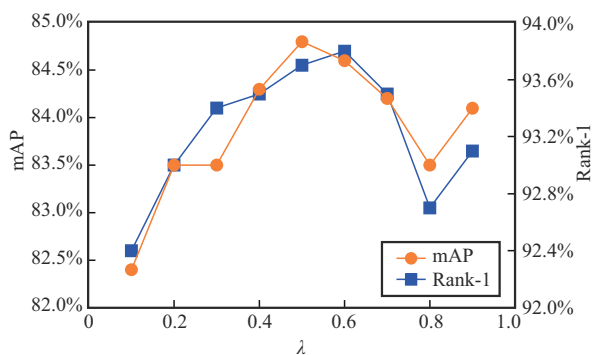


图5 损失函数权重因子 λ 的参数分析

4 结束语

本文针对无监督的行人重识别提出了一种双分支引导对比学习的新方法，通过引入局部分支来提高对细节特征的利用，同时使用标签细化模块来减小噪声对实验的影响，获得更平滑的语义

空间，最后通过一致性对比学习来指导聚类训练，从而更好地挖掘未标记数据中隐藏的信息。在Market1501、DukeMTMC-ReID和MSMT17数据集上的实验结果表明，本文方法具有较好的性能，可以有效控制伪标签方法产生的噪声。在未来的工作中，笔者将探索更多的噪声感知方法来增强模型的鲁棒性，并将这些方法应用于解决各种实际任务中的挑战。

参考文献：

- [1] WU D, ZHENG S J, ZHANG X P, et al. Deep learning-based methods for person re-identification: a comprehensive review[J]. *Neurocomputing*, 2019, 337: 354-371.
- [2] 许茹玉, 吴琳, 栗兴旺, 等. 多样细粒度特征与关系网络驱动的行人重识别[J]. *计算机工程与应用*, 2023, 59(19): 211-219.
XU R Y, WU L, SU X W, et al. Person re-identification driven by diverse fine-grained features and relation network[J]. *Computer Engineering and Applications*, 2023, 59(19): 211-219.
- [3] 王小檬, 梁凤梅. 融合有效掩膜和局部增强的遮挡行人重识别[J]. *计算机工程与应用*, 2024, 60(11): 156-164.
WANG X M, LIANG F M. Effective mask and local enhancement for occluded person re-identification[J]. *Computer Engineering and Applications*, 2024, 60(11): 156-164.
- [4] LI M K, LI C G, GUO J. Cluster-guided asymmetric contrastive learning for unsupervised person re-identification[J]. *IEEE Transactions on Image Processing*, 2022, 31: 3606-3617.
- [5] ZHAO Y, SHU Q Y, SHI X, et al. Unsupervised person re-identification by dynamic hybrid contrastive learning[J]. *Image and Vision Computing*, 2023, 137: 104786.
- [6] YIN J H, ZHANG X Y, MA Z Y, et al. A real-time memory updating strategy for unsupervised person re-identification[J]. *IEEE Transactions on Image Processing*, 2023, 32: 2309-2321.
- [7] 郑声晟, 殷海兵, 黄晓峰, 等. 基于GAN的无监督域自适应行人重识别[J]. *电信科学*, 2021, 37(2): 99-106.
ZHENG S S, YIN H B, HUANG X F, et al. GAN-based unsupervised domain adaptive person re-identification[J]. *Telecommunications Science*, 2021, 37(2): 99-106.
- [8] TIAN Q, DU X X. A plug-and-play noise-label correction framework for unsupervised domain adaptation person re-identification[J]. *The Visual Computer*, 2024, 40(6): 4493-4504.
- [9] ZHONG Z, ZHENG L, LUO Z M, et al. Invariance matters: exemplar memory for domain adaptive person re-identification[C]// *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE Press, 2019: 598-607.
- [10] 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 (KDD'96. AAAI Press, 1996: 226-231.
- [11] HOTELLING H. A generalized T test and measure of multivariate dispersion[M]//Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. Oakland: University of California Press, 1951: 23-42.
- [12] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 770-778.
- [13] TU Y P. Domain camera adaptation and collaborative multiple feature clustering for unsupervised person re-ID[C]//Proceedings of the 3rd International Workshop on Human-Centric Multimedia Analysis. New York: ACM Press, 2022: 51-59.
- [14] CHO Y, KIM W J, HONG S, et al. Part-based pseudo label refinement for unsupervised person re-identification[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 7298-7308.
- [15] DAI Z Z, WANG G Y, YUAN W H, et al. Cluster contrast for Unsupervised person re-identification[C]//Computer Vision-ACCV 2022. Cham: Springer Nature Switzerland, 2023: 319-337.
- [16] BERMAN M, JÉGOU H, VEDALDI A, et al. MultiGrain: a unified image embedding for classes and instances[J]. arXiv preprint, 2019, arXiv: 1902.05509.
- [17] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 2818-2826.
- [18] HERMANS A, BEYER L, LEIBE B. In defense of the triplet loss for person re-identification[J]. arXiv preprint, 2017, arXiv: 1703.07737.
- [19] KINGMA D P, BA J, HAMMAD M M. Adam: a method for stochastic optimization[J]. arXiv preprint, 2014, arXiv: 1412.6980.
- [20] 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 (ICCV). Piscataway: IEEE Press, 2015: 1116-1124.
- [21] RISTANI E, SOLERA F, ZOU R, et al. Performance measures and a data set for multi-target, multi-camera tracking[M]//Computer Vision-ECCV 2016 Workshops. Cham: Springer International Publishing, 2016: 17-35.
- [22] 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 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 79-88.
- [23] ZENG K W, NING M N, WANG Y H, et al. Hierarchical clustering with hard-batch triplet loss for person re-identification[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 13657-13665.
- [24] WANG D K, ZHANG S L. Unsupervised person re-identification via multi-label classification[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 10981-10990.
- [25] WANG Z D, ZHANG J W, ZHENG L, et al. CycAs: self-supervised cycle association for learning re-identifiable descriptions[M]//Computer Vision - ECCV 2020. Cham: Springer International Publishing, 2020: 72-88.
- [26] XUAN S Y, ZHANG S L. Intra-inter camera similarity for unsupervised person re-identification[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 11926-11935.
- [27] GE Y, CHEN D, ZHU F, et al. Self-paced contrastive learning with hybrid memory for domain adaptive object re-ID[J]. arXiv preprint, 2020, arXiv: 2006.02713.
- [28] 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 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 3436-3445.
- [29] SI T Z, HE F Z, ZHANG Z, et al. Hybrid contrastive learning for unsupervised person re-identification[J]. IEEE Transactions on Multimedia, 2022, 25: 4323-4334.
- [30] CHEN H, LAGADEC B, BREMOND F. ICE: inter-instance contrastive encoding for unsupervised person re-identification[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2021: 14940-14949.

[作者简介]



任航佳 (1998-), 男, 太原理工大学电子信息工程学院硕士生, 主要研究方向为深度学习、计算机视觉和行人重识别。



梁凤梅 (1969-), 女, 博士, 太原理工大学电子信息工程学院副教授、硕士生导师, 主要研究方向为图像处理与图像通信、计算机视觉和智能信息处理。