

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2023)05-1409-13

论文引用格式: Li K, Gong X and Fan J F. 2023. Spatiotemporal distance and multiple networks mutual learning-relevant pedestrian re-identification. Journal of Image and Graphics, 28(05): 1409-1421 (李宽, 龚勋, 樊剑锋. 2023. 结合时空距离的多网络互学习行人重识别. 中国图象图形学报, 28(05): 1409-1421) [DOI: 10. 11834/jig. 220668]

结合时空距离的多网络互学习行人重识别

李宽¹, 龚勋^{1,2,3,4*}, 樊剑锋¹

1. 西南交通大学唐山研究生院, 唐山 063000; 2. 西南交通大学计算机与人工智能学院, 成都 611756; 3. 可持续城市交通智能化教育部工程研究中心, 成都 611756; 4. 四川省制造业产业链协同与信息化支撑技术重点实验室, 成都 610031

摘要: 目的 在真实行人识别场景中, 获得准确的标注需要耗费大量人力, 因此无监督领域自适应成为行人重识别具有潜力的研究方向, 这类方法通常需要聚类生成伪标签, 往往会存在噪音。此外, 在行人搜索过程中, 好的排序算法也是取得更好识别性能的关键, 但寻常的 Re-Ranking 排序优化由于巨大的性能消耗, 限制了在真实场景下的应用。针对这两个问题, 本文提出了一个联合多网络、分摄像头训练的框架, 利用时空信息对排序进行优化。方法 对源域数据使用有监督进行预训练, 利用未标记的目标域样本进行多个网络模型的深度互学习无监督训练, 提高网络的泛化能力, 同时在训练过程中进行分摄像头处理, 减小跨摄像头的影响, 提升伪标签的质量。在排序匹配阶段利用时空信息对排序进行优化, 进一步提升匹配性能。结果 实验在 2 个跨域实验数据集上进行测试比较, 在源域为 DukeMTMC-ReID (Duke multi-tracking multi-camera re-identification) 数据集, 目标域为 Market-1501 数据集的实验中, 本文方法的平均精度均值 (mean average precision, mAP) 和 Rank1 分别为 82.5% 和 95.3%; 在源域为 Market-1501, 目标域为 DukeMTMC-ReID 数据集的实验中, mAP 和 Rank1 分别为 75.3% 和 90.2%。结论 提出的结合时空距离排序的分摄像头网络互学习模型, 提升了伪标签的精度, 并优化了匹配排序, 相比于其他优化算法大幅减少了计算量, 进一步提升了行人重识别性能。

关键词: 行人重识别; 互学习; 分摄像头; 跨域; 时空距离

Spatiotemporal distance and multiple networks mutual learning-relevant pedestrian re-identification

Li Kuan¹, Gong Xun^{1,2,3,4*}, Fan Jianfeng¹

1. Graduate School of Tangshan, Southwest Jiaotong University, Tangshan 063000, China; 2. School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China; 3. Engineering Research Center of Sustainable Urban Intelligent Transportation, Chengdu 611756, China; 4. Industry Chains Collaboration and Information Support Technology Key Laboratory of Sichuan Province, Chengdu 610031, China

Abstract: Objective Pedestrian re-identification can be focused on real-time target detection and matching. Due to labor-intensive to annotate accurate labels, unsupervised domain adaptation has become a potential solution. To generate pseudo labels, this method is required for clustering accompany with noise. Experimental analysis is demonstrated that camera-

收稿日期: 2022-07-07; 修回日期: 2022-11-25; 预印本日期: 2022-12-02

* 通信作者: 龚勋 xgong@swjtu.edu.cn

基金项目: 国家自然科学基金项目 (61876158); 四川省重点研发项目 (2023YFG0267); 中央高校基本科研业务费科技创新项目 (2682021ZTPY030, 2682022KJ045)

Supported by: National Natural Science Foundation of China (61876158); Science & Technology Program of Sichuan Province, China (2023YFG0267); Fundamental Research Funds for the Central Universities (2682021ZTPY030, 2682022KJ045)

cross is one of the key distorted factors for noise. Current eigenvector method is oriented to weaken the cross-domain and it is challenged for identifying camera ID-based information effectively. Hence, we design a camera module to resolve the problem of camera-cross. In addition, a single network is often used to extract features. Experimental analysis illustrate the single feature extraction ability of a single backbone network would also have more effective impact on the final performance. Therefore, learning-mutual is used to optimize the single network. For pedestrian searching, a good ranking algorithm is beneficial for a better recognition performance. We optimize traditional re-ranking algorithm using spatio-temporal information in the dataset because regular re-ranking optimization limits its application in real scenarios due to huge performance consumption. The time and space consumption close to the original ranking can reach the traditional re-ranking effect. To optimize the ranking, we develop a joint of spatio-temporal information-relevant multi-network and camera-splitting training framework. **Method** First, to improve the initial recognition performance, the network is pre-trained on the source domain dataset, and two of loss functions in relevance with label smoothing cross-entropy and triplet are used to pre-train the source domain. Second, due to the unique features extracted from a single backbone network, the single network model cannot be used to preserve good generalization ability in the ever-changing real scenarios. Therefore, we design a learning-mutual model to enhance its robustness. The pedestrian re-identification-oriented camera-split strategy is implemented to deal with recognition interference derived from cross-camera. For the pseudo-label generation, the dataset is split according to the camera ID, and the output vector is averaged after different networks-toward input. Additionally, we make full use of spatial information to optimize the pedestrian re-identification algorithm in another dimension because prior recognition analyses are originated from the distribution factors of pedestrians. For example, since the same camera is relatively close under the same timestamp, we use the timestamp information in the image. The one-hot-coded time stamps are spliced into feature vectors and it is then clustered to obtain pseudo labels. For training, to transfer knowledge from one network model to another, we use the class prediction of each network model as a soft label for training other related network models. For learning-mutual module, a time-averaging model is added, which can be updated iteratively during the training process. To suppress error-amplified, a large amount of prior information can be preserved. Furthermore, the learning-mutual correlation loss function is designed as well. Traditional classification loss and triplet loss are modified, and the loss function is designed on the basis of the integration of pseudo-labels and multiple backbone networks-related features. The network model training-based feature distribution can be constrained by multiple network models at the same time. For features-sorting, to optimize the traditional sort algorithm, pedestrian re-identification characteristics and spatiotemporal information of the dataset can be used according to the cameras of the same pseudo label number. The distribution of timestamp and statistics is used to generate the time distribution between different cameras, and a spatiotemporal score of camera is defined to fine-tune distance-between characteristics. This method is focused on a re-ranking spatially and the efficient and effective method can achieve similar spatio-temporal results close to the original ranking. **Result** The comparative analysis is carried out and popular 10 methods are compared on two cross-domain experimental datasets. For source domain-relevant Duke multi-tracking multi-camera reidentification (DukeMTMC-ReID) data set and target domain-related market-1501 dataset, the mean average precision (mAP) value can be reached to 82.5%, and the Rank1 is increased by 3.1% and reached to 95.3%. For the dataset in relevant to source domain market-1501 and target domain DukeMTMC-ReID, mAP and Rank1 can be reached to 75.3% and 90.2% of each. **Conclusion** To improve the accuracy of pseudo labels and optimize the matching ranking, the spatiotemporal distance ranking-coordinated learning-mutual model is developed in sub-camera network. Its computation is optimized more and pedestrian re-recognition performance is improved further.

Key words: pedstrian re-ID; mutual learning; multiple cameras; cross domain; time and space distance

0 引 言

行人重识别是指在不同监控摄像头下找到同一行人。随着人们对安全需求的增加,大量摄像头安

装在各种监控场景中,如学校、机场、商城和医院等。而人工搜寻效率低,结果不可靠,行人重识别能实时检测匹配行人,适应了大数据下对行人监测的需求,逐渐成为学术热点问题。

由于行人数据标记需要的巨大工作量,实际应

用存在较多困难。跨域重识别能利用有标注的源域数据信息辅助训练无标注的目标域数据,具有应用潜力。但由于跨域带来的背景、光照以及摄像头角度等差异,识别准确率往往也会急剧下降。

目前主流的无监督行人重识别主要由生成伪标签阶段和利用伪标签优化模型两个阶段组成,其中伪标签生成的可靠性是影响模型最终效果的主要因素。Zeng 等人(2020)利用 PK Sample (pharmacokinetic sample)的方法进行层次聚类,以提高聚类的精度。Zheng 等人(2021)提出了一种基于标签迁移的群感知特征学习策略,通过在线精炼伪标签,以改进多组伪标签优化。但这些方法没有考虑不同摄像头的影响可能导致不同摄像头下的类内距离大于相同摄像头下的类间距离,因此,即使是离聚类中心较近的数据,它的伪标签结果也并不可靠,这样便不可避免地使网络向着错误方向进行训练。因此,本文增加了一个分摄像头的维度,保证了生成标签的可靠性。此外,由于单骨干网络提取特征的单一性,在千变万化的实际场景中,以上方法并不能保证较好地泛化能力。Zhang 等人(2018)提出了多网络互学习,且用平均学生模型的方法防止多网络导致的误差放大。本文在该方法的基础上,提出了分摄像头多网络互学习行人重识别框架。

在重识别的匹配排序阶段,目前有很多排序优化的算法。Zhong 等人(2017)提出的互邻匹配 Re-Ranking 方法利用待搜寻行人图像和搜寻到的前几幅图像的邻域图像是否互相包含来判断结果的准确性。Sarfranz 等人(2018)将一幅图像与另一幅图像的相邻图像距离累加来当做新距离,它不需要对每个图像对都计算一次排名,减少了计算量。但将这些算法应用于真实场景仍对机器性能提出了巨大的挑战。上面的一些方法对排序的优化都基于已得到的图像间的距离,但数据中还包含了时间戳信息,这类信息没有得到有效利用。为此,本文利用数据集集中的时间信息,并增加了时空这两个维度,在时间消耗接近原始排序的前提下提升了行人搜索性能。

本文的贡献主要有以下方面:1)对数据集中每个单独的摄像头下图像进行聚类生成伪标签,以此作为优化标准之一,提升伪标签质量;2)利用教师学生网络互学习的方法学习更多特征,不仅接受来自自身标签的监督,还利用协同网络的学习经验来进一步提升泛化能力;3)利用数据的时间戳信息和摄

像头信息,在时空维度上优化行人搜索匹配,降低了时间消耗。同时,该方法在传统的有监督重识别场景中也能提高搜索性能。

1 相关方法

1.1 无监督跨域行人重识别

由于手动标注数据的高成本,目前无监督跨域行人重识别受到广泛关注。相关探索主要分为3个方面,即优化特征分布、生成对抗网络和自监督学习。

1)优化特征分布。该方法思想为减小源域和目标域特征分布的差距,从而减少跨域带来的模型性能损失。Mekhazni 等人(2020)提出一种基于最大均值差异(maximum mean discrepancy, MMD)的减小源域与目标域特征分布差距的方法, Jin 等人(2020)方法的思想是分离正样本对和负样本对距离的概率分布。但这类方法存在局限性,并不能保证源域、目标域的特征差距小于期望值,在跨域场景下,存在较大的性能下降。

2)生成对抗网络(generative adversarial network, GAN)。该方法使用 GAN 将源域图像转换到目标域。Deng 等人(2018)在 cycleGAN 的基础上进行改进,新增了两个约束条件,一个是生成的目标域风格图像应该与其相应的源域图像相似,这是为了能利用源域图像的 ID(identity document)信息;另一个是使生成的目标域风格图像与原有目标域中图像都不同,因为源域与目标域行人 ID 没有重叠。但基于 GAN 的方法对 GAN 生成图像的质量要求较高, GAN 训练慢,生成的目标域风格图像与真实数据集的相似度不可保证,实验结果与其他方法相比有一定差距。

3)自监督学习。自监督学习通常采用自监督和生成伪标签的方法,利用伪标签来不断微调模型,这是目前主流的研究方向。Fan 等人(2018)提出一种利用聚类来生成伪标签的重识别方法。对于生成的伪标签,它将距离聚类中心最近的数据伪标签看做可靠伪标签,并利用这些标签来进行训练。Ge 等人(2020)在判断可信伪标签的基础上做了改进,添加了一个记忆模块来存储源域和目标域的聚类中心,用以保留全局训练的特征,防止模型过度偏向错误方向。耿伟峰等人(2023)使用图神经网络融合了样

本特征与类别表示等多层次信息,增强了自监督信息的可靠性。

本文以自监督学习为基础,并针对以上方法没有考虑的在跨域问题中起主要影响的背景、视角变化等因素,添加了摄像头维度,约束了跨摄像头带来的伪标签预测错误影响。同时利用互学习来挖掘更多特征,使模型在各种场景都能保持良好性能。

1.2 重识别排序优化算法

在行人重识别场景中,最常见的排序优化算法为 Re-Ranking。该算法基于如下假设:对于某幅待搜索的行人图像,在搜索集中,假如某图像的邻近图像包含了待搜索图像,则该图像为正确匹配的可能性更大。

Jegou 等人(2010)引入了上下文相异度量,利用相邻向量的相似度更新图像间距离, Bai 等人(2009)利用流形距离在查询和被查询对象中找到最短路径,利用最短路径上的图像计算新的度量距离, Zhong 等人(2017)利用 k-reciprocal 编码对图像进行重排序。但以上研究都有一个共同的问题,即额外的计算量需要花费大量的内存空间和时间成本。为此, Zhang 等人(2020)利用 GPU (graphics processing

unit)并行计算的方法加快了重排序。但该方法额外的内存消耗仍然存在,而且没有从根本上降低算法复杂度。

与上述研究不同,本文从算法角度提出了一个适配于重识别的重排序算法,大幅减少了时间和空间代价。

2 算法描述

2.1 总体框架

如图 1 所示,本文算法框架主要由伪标签生成模块、多网络互学习模块和排序优化模块组成。

伪标签生成模块根据全局和摄像头分别生成伪标签 y_i 和 y'_i ,并在分摄像头阶段添加了时空信息;特征互学习模块利用教师学生网络的互学习训练出更具有泛化性的模型;时空距离排序模块利用时空距离对最终结果进行行人检索优化。

每次迭代开始之前,图像先通过伪标签生成模块进行全局伪标签 y_i 和摄像头伪标签 y'_i 的生成,然后在特征互学习模块中利用生成的伪标签 y_i 和 y'_i 进行特征互学习,最后利用时空距离排序模块进行行人检索。

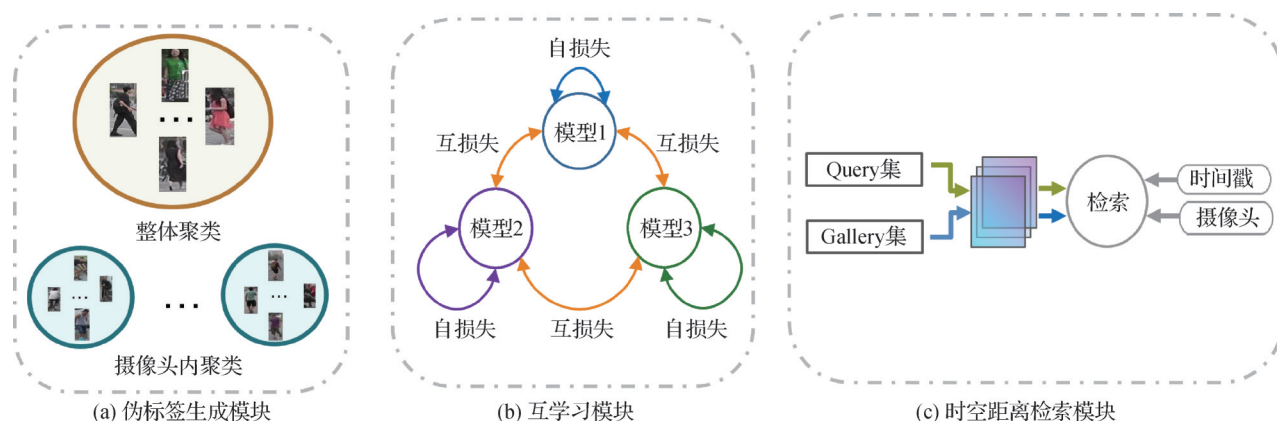


图1 本文算法框架模块图

Fig. 1 Framework of our method ((a) pseudo label generation module; (b) mutual learning module; (c) spatio-temporal distance retrieval module)

2.2 伪标签生成模块

在无监督领域自适应训练的每个迭代中,伪标签 y_i 和 y'_i 的生成如图 2 所示。

伪标签生成包括 3 个步骤,具体如下:

1) 对于目标域的样本图像,每个网络模型提取卷积特征,并将多个网络模型提取的特征进行平均以确定集成特征。具体为

$$f(\mathbf{x}_{i,i}) = \sum_{k=1}^K \frac{f(\mathbf{x}_{i,i} | \gamma^k)}{K} \quad (1)$$

式中, $\mathbf{x}_{i,i}$ 为输入图像, K 为模型数量, γ^k 为第 k 个模型的参数, $f(\mathbf{x}_{i,i})$ 为第 i 幅图像平均后向量。

2) 对 $f(\mathbf{x}_i)$ 进行 K-Means 聚类,将所有目标域样本划分为 U_i 个不同的类,生成作为训练样本 X_i 的伪标签 Y_i 。

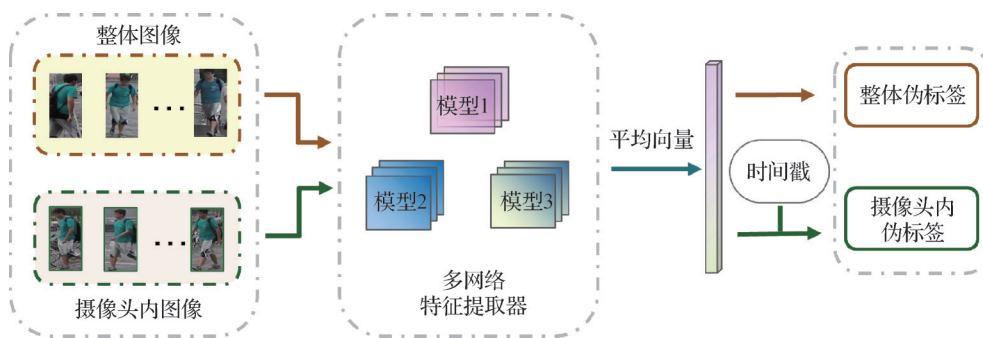


图2 伪标签生成模块

Fig. 2 Pseudo label generation module

3)对 X_l 中所有处于 U_c 个摄像头下的子集 X_l^i ,其中 $l \in [1, U_c]$,由于相同摄像头下相同行人的时间戳较为接近,因此将图像的特征 $f(X_l^i)$ 与时间戳信息 $Time_l$ 的拼接 $g(X_l^i)$ 分别进行K-Means聚类,将所有目标域样本在不同子集下划分为 U_l^i 个不同的类,上述 $Time_l$ 为时间戳的独热编码,分别生成所有子集训练样本 X_l^i 的伪标签 Y_l^i 。

2.3 多网络互学习模块

2.3.1 总体流程

本模块采用了两阶段的训练方案,包括在源域的有监督预训练学习和目标域的无监督自适应学习。在初始阶段,在源域数据集上以有监督的方式对具有不同网络架构的多个模型进行预训练。然后利用未标记的目标域样本进行多个网络模型的深度互学习无监督训练,使模型适应目标域。在目标域数据集中,利用伪标签生成模块的伪标签 y_l 和 y_l^i ,通过优化分类损失与三元组损失对各个网络模型进行微调。此外,将各个网络模型的时间平均模型的输出视为软伪标签作为优化基准,从而减少伪标签噪音的影响。本文采用3个骨干网络模型,两两之间互学习,以使网络模型优化。

2.3.2 源域有监督损失

对于每一个网络模型而言,首先以有监督的方式预训练一个深度神经网络模型 M^k ,该模型由预训练网络模型 γ^k 参数化,与行人重识别任务常用的损失函数一致,本文使用标签平滑的交叉熵损失函数与三元组损失函数进行源域预训练。

M^k 将每个样本图像 x_i 转换为特征, γ^k 表示网络模型的参数, $f(x_i|\gamma^k)$ 表示将样本图像 x_i 输入到网络模型参数设置为 γ^k 的网络模型中提取得到的特征,图像 x_i 属于标签 j 的预测概率,表示为 $p_j(x_i|\gamma^k)$ 。标

签平滑的交叉熵损失定义为

$$L_{id}^k = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^U q_j \times \log(p_j(x_i|\gamma^k)) \quad (2)$$

式中,如果 $j = y_i$,那么 $q_j = 1 - \varepsilon + \varepsilon/U$,否则 $q_j = \varepsilon/U$, ε 是一个小常数,在后续实验中设置为0.1, N 为图像个数, U 为聚类类别数。令

$$dis_n = \exp\left\|f(x_i|\gamma^k) - f(x_{i-}|\gamma^k)\right\|$$

$$dis_p = \exp\left\|f(x_i|\gamma^k) - f(x_{i+}|\gamma^k)\right\|$$

式中, x_{i+} 表示 x_i 的正样本, x_{i-} 表示负样本。 $\|\cdot\|$ 表示 L_2 范式。三元组损失定义为

$$L_{tri}^k = \frac{1}{N} \sum_{i=1}^N \log \frac{dis_n}{dis_n + dis_p} \quad (3)$$

最后的总损失计算为

$$L_s^k = L_{id}^k + L_{tri}^k \quad (4)$$

式中, k 代表第 k 个网络模型。

2.3.3 目标域互学习损失

互学习模块如图3所示。在每次迭代中,首先将目标域中的同一批图像输入到所有由 $\{\gamma^k\}$ 参数化的 $\{M^k\}$ 网络模型中,以预测分类置信度 $\{p_j(x_{i,i}|\gamma^k)\}$ 特征表示 $\{f(x_{i,i}|\gamma^k)\}$ 。为了将知识从一个网络模型转移到另一个网络模型,每个网络模型的类预测可以作为训练其他网络模型的软标签。为防止误差放大,本文添加了时间平均模型,网络模型 M^k 的时间平均模型在当前迭代 T 时的参数记为 Γ_T^k ,其更新式为

$$\Gamma_T^k = \alpha \cdot \Gamma_{T-1}^k + (1 - \alpha) \times \gamma^k \quad (5)$$

式中, $\alpha \in [0, 1]$ 为尺度因子,这里设置为0.999,初始时间平均参数为 $\Gamma_0^k = \gamma^k$ 。使用网络模型 M^k 的时间平均模型预测每个身份 j 的概率为 $p_j(x_{i,i}|\Gamma_T^k)$,计

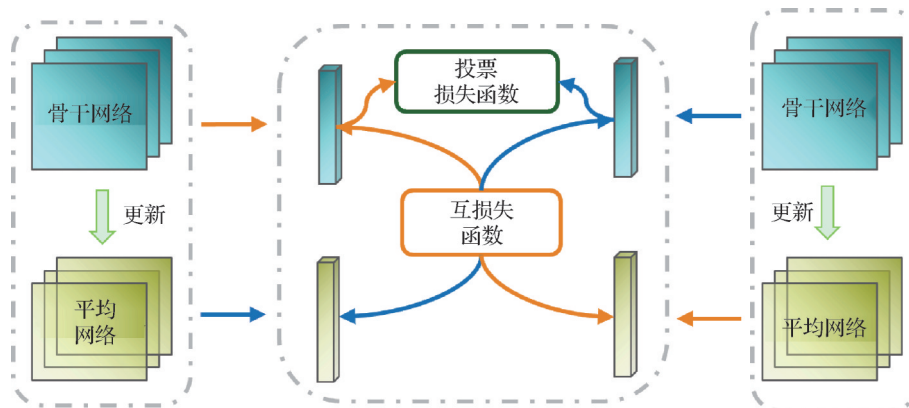


图3 互学习模块

Fig. 3 Ensemble learning module

算特征表示为 $f(\mathbf{x}_{t,i} | \mathbf{\Gamma}_T^k)$ 。

通过整合原始网络模型 $\mathbf{M}$ 与时间平均模型 $\mathbf{\Gamma}$ ，本文提出了互损失函数。将两个网络模型 $\mathbf{M}^k$ 和 $\mathbf{M}^m$ 的身份损失定义为 $\mathbf{M}^k$ 的类预测和 $\mathbf{M}^m$ 的时间平均模型类预测之间的交叉熵，以及各摄像头内标签预测与另一模型 $\mathbf{M}^m$ 的时间平均模型各摄像头内标签预测之间的交叉熵之和，前者称为互分类损失，后者称为摄像头内互分类损失。

互分类损失的函数计算式为

$$L_{\text{mid}}^{k \leftarrow m} = -\frac{1}{N_t} \sum_{i=1}^{N_t} \sum_{j=1}^{U_i} p_j(\mathbf{x}_{t,i} | \mathbf{\Gamma}_T^m) \times \log(p_j(\mathbf{x}_{t,i} | \mathbf{\Gamma}_T^k)) \quad (6)$$

式中, N_t 为图像总数, U_i 为聚类类别数。

摄像头内互分类损失的函数计算式为

$$L_{\text{minid}}^{k \leftarrow m} = \sum_{l=1}^L -\frac{1}{N_t^l} \sum_{i=1}^{N_t^l} \sum_{j=1}^{U_i^l} p_j(\mathbf{x}_{t,i}^l | \mathbf{\Gamma}_T^m) \times \log(p_j(\mathbf{x}_{t,i}^l | \mathbf{\Gamma}_T^k)) \quad (7)$$

式中, L 为目标域训练集摄像头数目, N_t^l 为某摄像头下图像总数, U_i^l 为该摄像头下聚类类别数。

设网络模型 $\mathbf{M}^k$ 的互分类损失为所有其他网络模型所学习的上述损失的平均值, 则

$$L_{\text{mid}}^k = \frac{1}{K-1} \sum_{m \neq k} (L_{\text{mid}}^{k \leftarrow m} + L_{\text{minid}}^{k \leftarrow m}) \quad (8)$$

与互分类损失函数类似, 对于每个网络模型 $\mathbf{M}^k$, 定义其他网络模型的时间平均模型 $\mathbf{M}^m$ 学习的互三元组损失为两个相似的二元交叉熵, 全局互三元组损失为

$$L_{\text{mtri}}^{k \leftarrow m} = -\frac{1}{N_t} \sum_{i=1}^{N_t} (\text{soft}(\mathbf{x}_{t,i} | \mathbf{\Gamma}_T^m) \times \log(\text{soft}(\mathbf{x}_{t,i} | \mathbf{\Gamma}_T^k)) + (1 - \text{soft}(\mathbf{x}_{t,i} | \mathbf{\Gamma}_T^m)) \times \log(1 - \text{soft}(\mathbf{x}_{t,i} | \mathbf{\Gamma}_T^k))) \quad (9)$$

式中, $\text{soft}(\cdot)$ 为样本对之间特征距离的 softmax 变换。

单个摄像头内互三元组损失为

$$L_{\text{mintri}_i}^{k \leftarrow m} = -\frac{1}{N_t^l} \sum_{i=1}^{N_t^l} (\text{soft}(\mathbf{x}_{t,i}^l | \mathbf{\Gamma}_T^m) \times \log(\text{soft}(\mathbf{x}_{t,i}^l | \mathbf{\Gamma}_T^k)) + (1 - \text{soft}(\mathbf{x}_{t,i}^l | \mathbf{\Gamma}_T^m)) \times \log(1 - \text{soft}(\mathbf{x}_{t,i}^l | \mathbf{\Gamma}_T^k))) \quad (10)$$

网络模型 $\mathbf{M}^k$ 的互三元组损失计算为所有其他网络模型的上述互三元组损失的平均值, 即

$$L_{\text{mtri}}^k = \frac{1}{K-1} \sum_{m \neq k} \left(L_{\text{mtri}}^{k \leftarrow m} + \sum_{l=1}^L L_{\text{mintri}_i}^{k \leftarrow m} \right) \quad (11)$$

互三元组损失函数所起到的作用与互分类损失函数类似, 都是为了使不同网络模型间进行相互学习, 互三元组损失函数减少了网络模型间三元组损失的差异, 从而使网络模型训练得到的特征分布能够同时被多个网络模型约束, 以增强模型的鲁棒性。

如图4所示, 互学习损失函数充分利用了网络模型的输出进行互相约束。

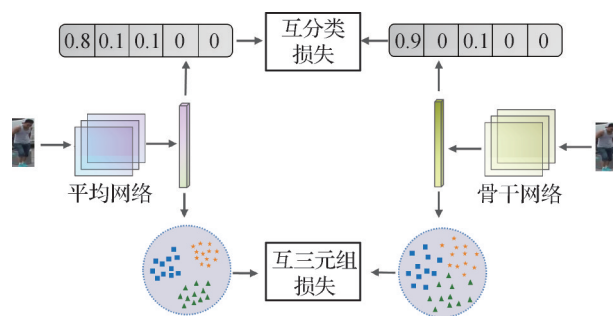


图4 互学习损失

Fig. 4 Mutual learning loss

2.3.4 目标域自身损失

为了从伪标签 y_t 与摄像头内标签 y_t' 中学习稳定和有所区别的知识, 本文引入了各个网络模型自身的投票损失, 包括投票分类损失和投票三元组损失。

在每个网络模型 $\mathbf{M}^k$ 上, 对于整体伪标签 y_t , 定

义投票分类损失为带有标签平滑的交叉熵, 具体为

$$L_{id}^k = \frac{1}{N} \sum_{t=1}^{N_t} \sum_{j=1}^{U_t} (q_j \log p_j(\mathbf{x}_{t,i} | \gamma^k)) \quad (12)$$

式中, 如果 $j = y_{t,i}$, 那么 $q_j = 1 - \varepsilon + \varepsilon/U_t$, 否则 $q_j = \varepsilon/U_t$, ε 是一个小常数, 设为 0.1, N_t 为图像个数, U_t 为聚类类别数。同样, 对于摄像头内伪标签, 定义单个摄像头下的投票分类损失为

$$L_{inid}^k = \frac{1}{N_t^l} \sum_{i=1}^{N_t^l} \sum_{j=1}^{U_t^l} (q_j \times \log(p_j(\mathbf{x}_{t,i}^l | \gamma^k))) \quad (13)$$

所有摄像头的投票分类损失 L_{inid}^k 为

$$L_{inid}^k = \sum_{l=1}^L L_{inid,l}^k \quad (14)$$

式中, 如果 $j = y_{t,i}^l$, 那么 $q_j = 1 - \varepsilon + \varepsilon/U_t^l$, 否则 $q_j = \varepsilon/U_t^l$, ε 是一个小常数, 设为 0.1。总体的投票分类损失 L_{allid}^k 为

$$L_{allid}^k = L_{id}^k + L_{inid}^k \quad (15)$$

为了更好地学习到稳定的知识, 令

$$dis_{t,n} = \exp \left\| f(\mathbf{x}_{t,i} | \gamma) - f(\mathbf{x}_{t,i}^- | \gamma) \right\|$$

$$dis_{t,p} = \exp \left\| f(\mathbf{x}_{t,i} | \gamma) - f(\mathbf{x}_{t,i}^+ | \gamma) \right\|$$

定义投票三元组损失为

$$L_{tri}^k = \frac{1}{N_t} \sum_{i=1}^{N_t} \log \frac{dis_{t,n}}{dis_{t,n} + dis_{t,p}} \quad (16)$$

同样, 为了模型能学习到单个摄像头下更为鲁棒的信息, 令

$$dis_{t,n}^l = \exp \left\| f(\mathbf{x}_{t,i}^l | \gamma) - f(\mathbf{x}_{t,i}^{l-} | \gamma) \right\|$$

$$dis_{t,p}^l = \exp \left\| f(\mathbf{x}_{t,i}^l | \gamma) - f(\mathbf{x}_{t,i}^{l+} | \gamma) \right\|$$

定义单个摄像头内部的投票三元组损失为

$$L_{intri,l}^k = \frac{1}{N_t^l} \sum_{i=1}^{N_t^l} \log \frac{dis_{t,n}^l}{dis_{t,n}^l + dis_{t,p}^l} \quad (17)$$

所有摄像头内部的投票三元组损失 L_{intri}^k 为

$$L_{intri}^k = \sum_{l=1}^L L_{intri,l}^k \quad (18)$$

总体的投票分类损失 L_{allid}^k 定义为

$$L_{allid}^k = L_{id}^k + \sum_{l=1}^L L_{inid,l}^k \quad (19)$$

投票损失定义为分类损失和三元组损失的总和, 具体为

$$L_{vot}^k = L_{allid}^k + L_{alltri}^k \quad (20)$$

上述投票损失使用聚类生成得到的伪标签优化交叉熵损失与三元组损失, 增强各个网络模型自身的特征表示能力, 从而在互学习过程中能够使用更

具备辨别能力的有效知识进行互相的学习。

2.3.5 总体损失

最终的总体损失定义为每个网络模型互相学习所造成的损失之和。即

$$L_{all} = \sum_{k=1}^K L_{vot}^k + L_{intri}^k + L_{mid}^k \quad (21)$$

式中, L_{all} 表示所有网络模型的损失函数之和, K 表示所训练骨干网络模型的个数。

2.4 时空距离检索模块

就行人重识别任务而言, 常见的公开数据集与真实场景下摄像头间的时空信息 (如拍摄图像时的时间戳与摄像头编号) 往往可以很容易获取。因此本文设计了一种通过时间戳与摄像头编号信息进行排序优化的方法, 以提升最终推理的性能。

本文算法对排序优化的流程如图5所示, 具体步骤如下:

1) 时间戳与摄像头信息获取。在经常用于行人重识别任务的 Market1501 与 DukeMTMC-ReID (Duke multi-tracking multi-camera re-identification) 数据集上, 可以通过文件名获取到当前图像的具体时间戳与摄像头编号信息。而在真实场景下, 时间戳与摄像头编号信息的获取也不需要大量的计算成本。

2) 目标域训练集特征聚类生成伪标签。对于本文的无监督领域自适应行人重识别而言, 关注的是目标域测试集的推理结果, 而源域的时间戳与摄像头信息并不适用于目标域, 所以排序优化所需要关注的仅是目标域的信息, 但是目标域训练集并没有标签, 因此需要对目标域训练集特征进行聚类生成伪标签, 本文采用 DBSCAN (density-based spatial clustering of applications with noise) 算法进行聚类。

3) 摄像头时间评分。根据相同伪标签的摄像头编号与时间戳的分布, 统计生成不同摄像头间的时间差分布 T_{ij} , 记摄像头 i 与摄像头 j 之间时间差的分布在第 k 次更新后为 T_{ij}^k , 其中 $k \in [0, N]$, N 为目标训练集图像个数, 它统计了 i 和 j 摄像头下在第 k 次更新后相同时间差的行人个数。对每个伪标签进行统计, 当其同时存在于摄像头 i 与摄像头 j 上时, 计算其平均帧的时间差 t , 并对 T_{ij} 进行更新, 具体为

$$T_{ij}^k(t) = T_{ij}^{k-1}(t) + 1 \quad (22)$$

式中, $T_{ij}(t)$ 初始为 0。

定义 $Num(\mathbf{x}_i, \mathbf{x}_j)$ 为同时位于图像 $\mathbf{x}_i$ 与图像 $\mathbf{x}_j$ 所

处的摄像头上,且处于同一时间差区间内的伪标签个数,定义 $Sum(cam_{x_i}, cam_{x_j})$ 为在图像 x_i 所处摄像头与图像 x_j 所处摄像头同时存在的伪标签的个数。摄像头时间评分为

$$D_{st}(x_i, x_j) = \frac{Num(x_i, x_j)}{Sum(cam_{x_i}, cam_{x_j})} \quad (23)$$

4)联合距离推理。将摄像头时间评分与余弦距离进行结合,虽然摄像头时间评分 $D_{st}(x_i, x_j)$ 在统计时是一个处于 $(0, 1)$ 的概率,与余弦距离所处区间 $(-1, 1)$ 的差异较小,但两者在推理过程中权重应该有所不同,令

$$sc_{cam} = 1 + \exp(-\omega D_{st}(x_i, x_j))$$

$$sc_{cos} = 1 + 2 \exp(-\omega \cos(x_i, x_j))$$

最终的联合距离为

$$J(x_i, x_j) = \frac{1}{sc_{cam} + sc_{cos}} \quad (24)$$

式中, $\cos(x_i, x_j)$ 表示图像 x_i 与图像 x_j 特征间的余弦距离, ω 为可进行人为调整的超参数。

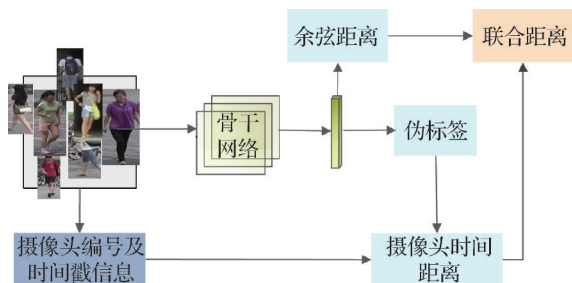


图5 时空距离检索流程图

Fig. 5 Spatio-temporal distance retrieval map

3 实验及分析

3.1 实验数据集与评估指标

由于现有的公开数据集中,对时间戳标注并没有给予足够的关注,仅有 Zheng 等人(2015)提出的 Market-1501 和 Ristani 等人(2016)提出的 DukeMTMC-ReID 数据集上有着明确的时间戳标注,因此为了验证根据时间戳与摄像头信息排序优化的有效性,本文在目标域设定为 Market-1501,源域设定为 DukeMTMC-ReID,以及目标域设定为 DukeMTMC-ReID,源域设定为 Market-1501 两种情

况下对所提出的方法进行评估。

Market-1501 数据集是于 2015 年夏季在清华大学校园拍摄采集的,包含 6 个不同摄像头的 32 668 幅图像,1 501 个行人身份。DukeMTMC-ReID 数据集是 DukeMTMC 的子集,是从美国杜克大学 2014 年校园监控录像中采集的,包含 8 个不同摄像头的 36 411 幅图像,1 812 个行人身份。

在评估中,采用累计匹配特征曲线(cumulative match characteristic, CMC)和平均精度均值(mean average precision, mAP)作为评估指标。

3.2 实验设置

本文中模型的训练分为两个阶段:源域的预训练和目标域的无监督自适应。

在源域的预训练阶段,首先在源数据集上使用 3 个网络模型进行有监督的预训练。本文采用 DenseNet-121 (densely connected convolutional network-121)、ResNet-50 (residual network-50) 和 inception-v3 (inception network-v3) 这 3 种架构作为骨干网络,并使用在 ImageNet 上预先训练的参数对其进行初始化。

批量大小设为 64,用 16 个随机选择的标签和 4 个随机抽样的图像对 64 幅图像进行采样,以计算三元组损失。本文使用 Adam 优化参数,并设置 weight decay 为 0.000 5。初始学习率设置为 0.000 35,在总共 80 个 epoch 中,在第 40 和第 70 个 epoch 时,初始学习率降低到之前值的 1/10。

3.3 实验参数分析

本文的超参数包括 K-Means 聚类簇个数、DBSCAN 的搜索半径,以及摄像头时间距离计算公式使用的 ω 。为了确定本文一系列超参数的选取对网络模型性能的影响,进行了详尽的参数分析实验。

在无监督领域自适应阶段进行聚类过程中,本文对聚类方法以及其中参数进行选取。本文分别使用了 K-Means 聚类方法与 DBSCAN 聚类方法,并对其中参数进行调整以进行实验。首先是 K-Means 聚类中类别数超参数的确定,基于 Market-1501 与 DukeMTMC-ReID 数据集其测试集原本类别量,设置 K-Means 聚类类别数分别为 500, 700, 900。其次是 DBSCAN 聚类方法中参数的确定,本文对目标域训练集输入源域预训练网络模型提取出的特征进行了 k 距离 k_i 的计算, k_i 计算结果如图 6 所示。可以看出,预训练网络模型所提取的目标域训练集特征其 k 距

离范围取值在 0.3~1.0 之间。因此将 DBSCAN 聚类的扫描半径 ϵ 设置在 0.4~1.0 间滑动选取, 间隔为 0.2。

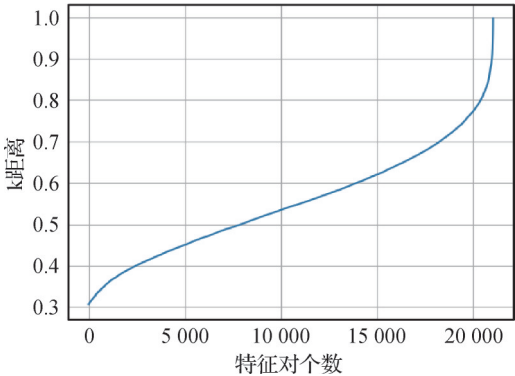


图6 无监督领域自适应聚类的k距离曲线
Fig. 6 Curve of k-distance of unsupervised neighborhood adaptive clustering

推理结果如表 1 所示。可以发现, 在使用 DBSCAN 聚类方法并将扫描半径设置为 0.8 时, 进行无监督领域自适应学习的聚类阶段产生的推理结果最好。但是其他聚类生成伪标签方法使用的是 K-Means 聚类方法, 且聚类数量设置为 500, 为了与其他方法进行公平对比, 因此后续实验使用 K-Means(500)聚类生成伪标签训练得到的模型作为对比实验使用的模型。

表1 无监督领域自适应学习聚类过程中不同聚类方法及参数下的实验结果

Table 1 Experimental results under different clustering methods and parameters in unsupervised domain adaptive learning clustering process				
方法	mAP	Rank1	Rank5	Rank10
K-Means(500)	75.3	90.2	93.5	94.8
K-Means(700)	75.8	90.6	93.8	95.1
K-Means(900)	75.2	90.1	93.3	95.0
DBSCAN(0.4)	73.6	89.3	92.0	93.4
DBSCAN(0.6)	74.8	89.7	93.4	94.6
DBSCAN(0.8)	76.3	91.4	94.2	95.2
DBSCAN(1.0)	75.3	90.4	93.6	94.7

注:加粗字体表示各列最优结果。

同时, 为了验证在摄像头时间距离构建中聚类方法对模型推理结果的影响, 同样对聚类方法的参数进行了分析。K-Means 聚类参数与前文一致, 设

置为 500, 700, 900。DBSCAN 聚类方法的扫描半径参数 ϵ 选取方式亦与前文一致, 本文对目标域训练集输入网络模型提取出的特征进行了 k_i 的计算, 结果如图 7 所示。

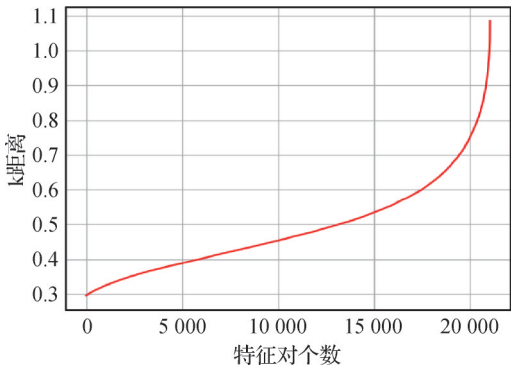


图7 摄像头时间距离构建聚类的k距离曲线
Fig. 7 Curve of k-distance of the camera time distance clustering

由图 7 可以得知, 扫描半径 ϵ 的合理取值范围大致在 0.3~1.1 之间。将扫描半径 ϵ 取值范围设置为 [0.3, 1.1], 间隔为 0.2, 整体实验结果如表 2 所示。不难发现, DBSCAN 聚类方法在摄像头时间距离构造过程中进行聚类效果更为优秀。

表2 摄像头时间距离构建过程中不同聚类方法及参数下的实验结果

Table 2 Experimental results under different clustering methods and parameters in camera time distance construction process				
方法	mAP	Rank1	Rank5	Rank10
KMeans(500)	72.1	88.4	91.3	93.6
KMeans(700)	73.3	89.2	91.8	93.9
KMeans(900)	72.6	88.9	91.4	93.8
DBSCAN(0.3)	72.4	88.8	91.7	93.8
DBSCAN(0.5)	73.4	89.3	92.0	93.8
DBSCAN(0.7)	73.9	89.7	93.2	94.5
DBSCAN(0.9)	74.8	90.4	93.7	94.5
DBSCAN(1.1)	76.3	91.4	94.2	95.2

注:加粗字体表示各列最优结果。

最后, 为了验证 ω 取值对网络模型的影响, 设置 ω 的取值范围为 [3, 7], 间隔为 1。结果如表 3 所示, 可以发现, 当 ω 设置为 5 时, 摄像头时间距离与余弦距离进行共同推理的效果最好。这个取值可以在特

征间余弦距离不失真的基础上有效使用摄像头时间距离。

表3 摄像头时间距离构建过程中不同聚类方法及参数下的实验结果

Table 3 Experimental results under different clustering methods and parameters in camera time distance construction process				
ω	mAP	Rank1	Rank5	Rank10
3	75.2	90.4	93.6	94.6
4	75.8	90.8	93.9	95.0
5	76.3	91.4	94.2	95.2
6	75.4	90.6	93.7	94.6
7	74.8	90.2	93.2	94.4

注:加粗字体表示各列最优结果。

为验证互学习中网络个数的影响,分别对单网络(即不采用互学习)、双网络、三网络进行对比实验,实验结果分别如表4和表5所示。由表4和表5可以看到,互学习能有效结合多网络的特征提取能力,提升模型性能,而本文采用的三网络性能也高于双网络。综合考虑计算资源耗费与模型性能提升,本文采用三网络模型。

表4 源域为 DukeMTMC-ReID 目标域为 Market-1501 设置下的不同网络个数实验结果表
Table 4 Experimental results of different number of networks with DukemtMC-ReID as the source domain and Market-1501 as the target domain

网络	mAP	Rank1	Rank5	Rank10
den	69.3	86.5	94.9	96.7
inc	63.0	76.4	87.6	90.9
res	65.8	77.5	88.6	91.3
den+inc	68.8	86.2	94.2	96.2
inc+res	66.4	86.0	93.9	95.9
res+den	72.2	86.4	94.7	96.4
den+inc+res	76.0	89.9	96.0	97.5

注:den表示DenseNet-121,inc表示inception-v3,res表示ResNet-50。

3.4 实验结果分析

为了评估所提方法的有效性,设计对比实验,将

表5 源域为 Market-1501 目标域为 DukeMTMC-ReID 设置下的不同网络个数实验结果

Table 5 Table of experimental results of different number of networks with Market-1501 as the source domain and DukemtMC-ReID as the target domain

网络	mAP	Rank1	Rank5	Rank10
den	60.4	75.3	86.0	89.5
inc	53.6	73.0	80.6	83.2
res	54.5	74.0	83.8	87.6
den+inc	60.4	75.7	85.6	88.1
inc+res	58.8	74.4	84.1	87.3
res+den	61.3	75.9	85.8	88.3
den+inc+res	66.1	79.6	88.3	92.2

注:den表示DenseNet-121,inc表示inception-v3,res表示ResNet-50。

本文方法与当前的一些具有代表性的方法进行比较,包括手工特征方法(unsupervised cross-dataset transfer learning, UMDL)(Peng等,2016)、基于特征对齐的方法(unsupervised camera-aware domain adaptation framework, UCDA-CCE)(Qi等,2019)、基于生成对抗网络的方法(similarity preserving generative adversarial network, SPGAN)(Deng等,2018)、Cam-Style(camera style generative adversarial network)(Zhong等,2018b)、HHL(hetero-homogeneous learning)(Zhong等,2018a)和ECN(exemplar camera neighborhood)(Zhong等,2019)、基于聚类进行伪标签预测的方法PTG(person transfer gan)(Wei等,2018)、UDAP(unsupervised domain adaptive re-identification)(Song等,2020)、MEB-Net(multiple expert brainstorming network)(Zhai等,2020)、GLT(group-aware label transfer)(Zheng等,2021a)、HCD(hierarchical cluster dynamics)(Zheng等,2021b)和SAT(self-adaptative techniques)(Bertocco等,2021),对比结果如表6和表7所示,其中,所有对比方法均采用公开发表的结果。

从表6可以看出,在源域为DukeMTMC-ReID、目标域为Market-1501数据集的实验中,本文方法的mAP和Rank1准确率分别为82.5%和95.3%,较对比方法中指标最高的方法HCD分别高出2.5%和3.4%。从表7可以看出,在源域为Market1501、目标

表 6 源域为 DukeMTMC-ReID 目标域为 Market-1501 设置下本文方法与其他方法的对比结果

Table 6 Comparison results between the proposed method and other methods under the setting of source domain DukemtMC-ReID and target domain Market-1501

方法	mAP	Rank1	Rank5	Rank10
UMDL	12.4	34.5	52.6	59.6
UCDA-CCE	30.9	60.4	—	—
SPGAN	26.7	57.7	75.8	82.4
CamStyle	27.4	58.8	78.2	84.3
HHL	31.4	62.2	78.8	84.0
ECN	43.0	75.1	87.6	91.6
PTG	20.5	45.5	60.7	66.7
UDAP	53.7	75.8	89.5	93.2
MEB-Net	76.0	89.9	96.0	97.5
SAT	78.4	92.9	96.9	97.8
GLT	79.5	92.2	96.5	97.8
HCD	80.0	91.9	—	—
本文	82.5	95.3	97.9	98.4

注:加粗字体表示各列最优结果,“-”表示无公开数据。

表 7 源域为 Market-1501 目标域为 DukeMTMC-ReID 设置下本文方法与其他方法的对比结果

Table 7 Comparison results between the proposed method and other methods under the setting of source domain Market-1501 and target domain DukemtMC-ReID

方法	mAP	Rank1	Rank5	Rank10
MDL	7.3	18.5	31.4	37.6
UCDA-CCE	31.0	47.7	—	—
SPGAN	26.2	46.4	62.3	68.0
CamStyle	25.1	48.4	62.5	68.9
HHL	27.2	46.9	61.0	66.7
ECN	40.4	63.3	75.8	80.4
PTG	16.4	30.0	43.4	48.5
UDAP	49.0	68.4	80.1	83.5
MEB-Net	66.1	79.6	88.3	92.2
GLT	69.2	82.0	90.2	92.8
HCD	70.1	82.0	—	—
SAT	72.6	85.0	92.1	93.9
本文	75.3	90.2	93.5	94.8

注:加粗字体表示各列最优结果,“-”表示无公开数据。

域为 DukeMTMC-ReID 数据集的实验中,本文方法的 mAP 和 Rank1 准确率分别为 75.3% 和 90.2%,较对比方法中指标最高的方法 SAT 分别高出 2.7% 和 4.8%。这些数据说明了本文使用的摄像头内伪标签构造与深度互学习框架能使网络从图像中提取出更多有用的信息。

同时,为了验证本文各模块的效果,分别在 4 种不同设置下进行消融实验,主要检测了摄像头内伪标签构造及基于时间戳和摄像头排序优化的作用,实验结果如图 8—图 11 所示。其中,-表示去除该模块,+表示添加该模块。可以看出,摄像头时空距离的引入不仅使无监督领域自适应设置下的行人重识别任务最终推理的性能大幅提升,并且适用于传统

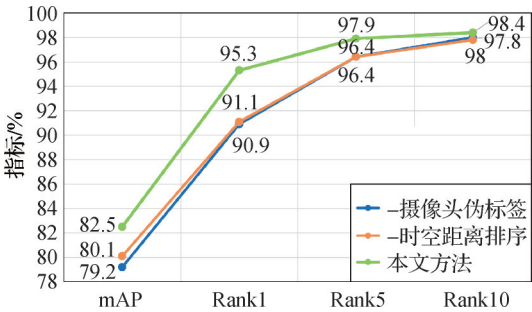


图 8 DukeMTMC-ReID 到 Market-1501 消融实验结果

Fig. 8 Ablation results of DukeMTMC-ReID to Market-1501

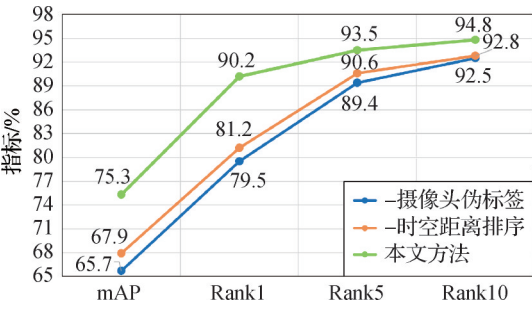


图 9 Market-1501 到 DukeMTMC-ReID 消融实验结果

Fig. 9 Ablation results of Market-1501 to DukeMTMC-ReID

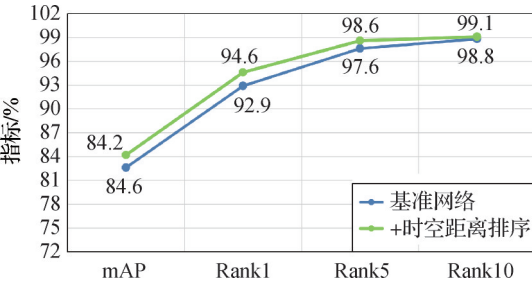


图 10 Market-1501 数据集消融实验结果

Fig. 10 Market-1501 ablation experiment results

的闭集设置下的行人重识别任务。而真实场景下摄像头标签与时间标签并不是难以获取的信息,因此本文方法可以广泛应用于大量场景。

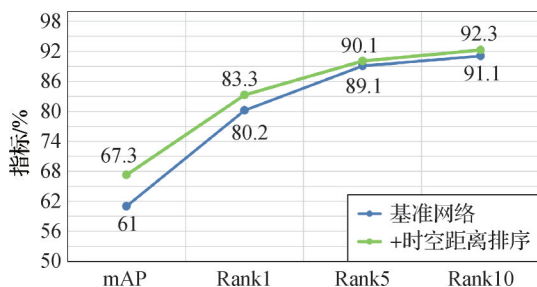


图 11 DukeMTMC-ReID 数据集消融实验结果

Fig. 11 DukeMTMC-ReID ablation experiment results

4 结 论

本文针对无监督领域自适应行人重识别任务,提出了一种结合时空距离的分摄像头互学习方法。其中,摄像头内伪标签的构造利用了数据的时间信息,这比直接使用特征聚类生成的数据集整体伪标签更为可靠,减少了模型对夹杂噪音的整体伪标签的依赖。同时,深度互学习的框架使模型的鲁棒性得到了提升。在此基础上,引入了数据集中摄像头信息及时间戳信息进行排序优化,相比于现有的排序优化算法大幅降低了内存和时间消耗,且进一步提高了行人重识别模型的准确度。

本文方法有效解决了无监督领域自适应问题中伪标签噪音的影响,拓展了数据集的使用方法,提出了一种新的排序搜索算法。但是,本文只用了互学习过程中的一个模型,没有充分利用多网络的丰富信息。未来的工作可从如何整合多网络提取一个聚合特征入手,提升最终特征抽取器的泛化性。

参考文献 (References)

Bai X, Yang X W, Longin Jan Latecki, Liu W and Tu Z W. 2009. Learning context-sensitive shape similarity by graph transduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5): 861-874 [DOI: 10.1109/TPAMI.2009.85]

Bertocco G C, Andaló F and Rocha A. 2021. Unsupervised and self-adaptive techniques for cross-domain person re-identification. *IEEE Transactions on Information Forensics and Security*, 16: 4419-4434 [DOI: 10.1109/TIFS.2021.3107157]

Deng W J, Zheng L, Ye Q X, Kang G L, Yang Y and Jiao J B. 2018. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 994-1003 [DOI: 10.1109/CVPR.2018.00110]

Fan H H, Zheng L, Yan C G and Yang Y. 2018. Unsupervised person re-identification: clustering and fine-tuning. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 14(4): #83 [DOI: 10.1145/3243316]

Ge Y X, Zhu F, Chen D, Zhao R and Li H S. 2020. Self-paced contrastive learning with hybrid memory for domain adaptive object Re-ID. *Advances in Neural Information Processing Systems*. 33: 11309-11321 [DOI: 10.48550/arXiv.2006.02713]

Geng W F, Wang X, Jing L P and Yu J. 2023. Consensus graph learning-based self-supervised ensemble clustering. *Journal of Image and Graphics*, 28(04): 1069-1078 (狄伟峰,王翔,景丽萍,于剑. 2023. 共识图学习驱动的自监督集成聚类. *中国图象图形学报*, 28(04): 1069-1078) [DOI: 10.11834/jig.210947]

Gong X, Zhang Z Y, Liu L, Ma B and Wu K L. 2022. A survey of human-object interaction detection. *Journal of Southwest Jiaotong University*, 57(4): 693-704 (龚勋, 张志莹, 刘璐, 马冰, 吴昆伦. 2022. 人物交互检测研究进展综述. *西南交通大学学报*, 57(4): 693-704) [DOI: 10.3969/j.issn.0258-2724.20210339]

Jegou H, Schmid C, Harzallah H and Verbeek J. 2010. Accurate image search using the contextual dissimilarity measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(1): 2-11 [DOI: 10.1109/TPAMI.2008.285]

Jin X, Lan C L, Zeng W J and Chen Z B. 2020. Global distance-distributions separation for unsupervised person re-identification//*Proceedings of the 16th European Conference on Computer Vision*. Glasgow, UK: Springer: 735-751 [DOI: 10.1007/978-3-030-58571-6_43]

Mekhazni D, Bhuiyan A, Ekladios G and Granger E. 2020. Unsupervised domain adaptation in the dissimilarity space for person re-identification//*Proceedings of the 16th European Conference on Computer Vision*. Glasgow, UK: Springer: 159-174 [DOI: 10.1007/978-3-030-58583-9_10]

Peng P X, Xiang T, Wang Y W, Pontil M, Gong S G, Huang T J and Tian Y H. 2016. Unsupervised cross-dataset transfer learning for person re-identification//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE: 1306-1315 [DOI: 10.1109/CVPR.2016.146]

Qi L, Wang L, Huo J, Zhou L P, Shi Y H and Gao Y. 2019. A novel unsupervised camera-aware domain adaptation framework for person re-identification//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE: 8079-8088 [DOI: 10.1109/ICCV.2019.00817]

Ristani E, Solera F, Zou R, Cucchiara R and Tomasi C. 2016. Perfor-

- mance measures and a data set for multi-target, multi-camera tracking//Computer Vision. Amsterdam, the Netherlands: Springer: 17-35 [DOI: 10.1007/978-3-319-48881-3_2]
- Sarfraz M S, Schumann A, Eberle A and Stiefelhofen R. 2018. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 420-429 [DOI: 10.1109/CVPR.2018.00051]
- Song L C, Wang C, Zhang L F, Du B, Zhang Q, Huang C and Wang X G. 2020. Unsupervised domain adaptive re-identification: theory and practice. Pattern Recognition, 102: #107173 [DOI: 10.1016/j.patcog.2019.107173]
- Wei L H, Zhang S L, Gao W and Tian Q. 2018. Person transfer GAN to bridge domain gap for person re-identification//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake, USA: IEEE: 79-88 [DOI: 10.1109/CVPR.2018.00016]
- Zeng K W, Ning M N, Wang Y H and Guo Y. 2020. Hierarchical clustering with hard-batch triplet loss for person re-identification//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 13654-13662 [DOI: 10.1109/CVPR42600.2020.01367]
- Zhai Y P, Ye Q X, Lu S J, Jia M X, Ji R R and Tian Y H. 2020. Multiple expert brainstorming for domain adaptive person re-identification//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 594-611 [DOI: 10.1007/978-3-030-58571-6_35]
- Zhang X M, Jiang M Y, Zheng Z D, Tan X, Ding E and Yang Y. 2020. Understanding image retrieval re-ranking: a graph neural network perspective. //Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: #7620 [DOI: 10.48550/arXiv.2012.07620]
- Zhang Y, Xiang T, Hospedales T M and Lu H C. 2018. Deep mutual learning//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 4320-4328 [DOI: 10.1109/CVPR.2018.00454]
- Zheng K C, Liu W, He L X, Mei T, Luo J B and Zha Z J. 2021a. Group-aware label transfer for domain adaptive person re-identification//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 5306-5315 [DOI: 10.1109/CVPR46437.2021.00527]
- Zheng L, Shen L Y, Tian L, Wang S J, Wang J D and Tian Q. 2015. Scalable person re-identification: a benchmark//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 1116-1124 [DOI: 10.1109/ICCV.2015.133]
- Zheng Y, Tang S X, Teng G L, Ge Y X, Liu K J, Qin J, Qi D L and Chen D P. 2021b. Online pseudo label generation by hierarchical cluster dynamics for adaptive person re-identification//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 8351-8361 [DOI: 10.1109/ICCV48922.2021.00826]
- Zhong Z, Zheng L, Cao D L and Li S Z. 2017. Re-ranking person re-identification with k-reciprocal encoding//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 3652-3661 [DOI: 10.1109/CVPR.2017.389]
- Zhong Z, Zheng L, Li S Z and Yang Y. 2018a. Generalizing a person retrieval model hetero- and homogeneously//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 176-192 [DOI: 10.1007/978-3-030-01261-8_11]
- Zhong Z, Zheng L, Luo Z, Li S and Yang Y. 2019. Invariance matters: exemplar memory for domain adaptive person re-identification//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 598-607 [DOI: 10.1109/CVPR.2019.00069]
- Zhong Z, Zheng L, Zheng Z D, Li S Z and Yang Y. 2018b. Camera style adaptation for person re-identification//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 5157-5166 [DOI: 10.1109/CVPR.2018.00541]

作者简介

李宽,男,硕士研究生,主要研究方向为行人重识别和深度学习。E-mail: 1046384422@qq.com

龚勋,通信作者,男,教授,主要研究方向为计算机视觉、人工智能、行人重识别及行为分析。E-mail: xgong@swjtu.edu.cn

樊剑锋,男,硕士研究生,主要研究方向为行人重识别和深度学习。E-mail: 350317121@qq.com