

基于摄像头域内域间合并的无监督行人重识别方法

陈利文 叶 锋 黄添强 黄丽清 翁 彬 徐 超 胡 杰

(福建师范大学计算机与网络安全学院 福州 350117)

(福建省公共服务大数据挖掘与应用工程技术研究中心(福建师范大学)福州 350117)

(数字福建大数据安全技术研究所(福建师范大学)福州 350117)

(yefeng@fjnu.edu.cn)

An Unsupervised Person Re-Identification Method Based on Intra-/Inter-Camera Merger

Chen Liwen, Ye Feng, Huang Tianqiang, Huang Liqing, Weng Bin, Xu Chao, and Hu Jie

(College of Computer and Cyber Security, Fujian Normal University, Fuzhou 350117)

(Fujian Provincial Engineering Research Center of Big Data Analysis and Application(Fujian Normal University), Fuzhou 350117)

(Digital Fujian Institute of Big Data Security Technology(Fujian Normal University), Fuzhou 350117)

Abstract In criminal investigation, intelligent monitoring, image retrieval and other fields, person re-identification has always been a hot and significant research topic. Since most of the existing methods rely on labeled datasets, the lack of labels makes unsupervised person re-identification technology more challenging. In order to overcome the problem of the lack of labels, a new framework to generate reliable pseudo labels is proposed as supervision information for existing supervised person re-identification models. Assuming that images taken by the same camera mainly vary in pedestrian's physical appearances rather than backgrounds, then images taken by different cameras vary in backgrounds. To eliminate effects brought by the differences in image background, images are divided into several domains by camera Serial numbers as the first step. Then we construct undirected graphs for each camera with Euclidean distance of image pairs, and there will be an edge only when two images are close enough. The images in one maximum analogous set are regarded as the same person. Then we merge maximum cliques from different cameras simply with their cosine distance and assign pseudo labels. The proposed framework runs in an unsupervised manner, and the proposed method can obtain higher accuracy than the other related methods on Market1501 and DukeMTMC-ReID datasets, which further shows the effectiveness of the proposed method.

Key words person re-identification; unsupervised learning; max clique algorithm; pseudo labels; clustering

摘要 在刑事侦查、智能监控、图像检索等领域，行人重识别一直是研究的热点。由于现有的大部分方法依赖有标注数据集，因此标签的缺乏使得无监督的行人重识别技术变得更具挑战性。为了克服这一问题，提出了一个用于生成可靠伪标签的框架，这些生成标签可以为现有监督行人重识别模型提供监督信号。假设数据集内的大部分图片都满足同一个摄像头拍摄的图片差异主要在于前景（行人）、同一个行人被不同摄像头拍摄到的图片差异主要在于背景。为了消除图片背景带来的差异，首先把数据集中的图片依据摄

收稿日期：2021-07-06；修回日期：2022-02-23

基金项目：国家自然科学基金项目(62072106, 61070062); 福建省自然科学基金项目(2020J01168); 福建省科技厅创新战略研究项目(2020R0178);

福建省教育厅项目(JT180078)

This work was supported by the National Natural Science Foundation of China (62072106, 61070062), the Natural Science Foundation in Fujian Province (2020J01168), the Innovation Strategy Research Project of Science and Technology Department of Fujian Province (2020R0178), and the Project of Fujian Provincial Education Department (JT180078).

像头编号分成若干个域,通过计算每个域内的图片间的欧式距离,建立图模型,执行最大团算法寻找最相似的若干个图片并认为它们属于同一个行人;紧接着计算不同摄像头域间的团的相似度,据此进行合并;最终给出全局伪标签.所提的框架无需人为标注数据,以一种无监督的方式运行,并在Market1501和DukeMTMC-ReID数据集上进行实验,实验发现所提方法比其他相关方法具有更高的精度,从而进一步证明了所提方法的有效性.

关键词 行人重识别;无监督学习;最大团算法;伪标签;聚类

中图法分类号 TP391

行人重识别研究的目标是如何从多个摄像机视角下匹配同一个行人,该技术的研究对智能视频监控和图像检索、刑事侦查等公共安全方面有着重大意义.但是光照条件、行人姿势、背景环境等因素都会导致同一个行人在不同视角下可能有较大的差异,同时还存在行人被遮挡或者图片模糊的情况,所以目前行人重识别仍然是一个具有挑战性且热门的研究课题.近几年有诸多研究者尝试使用深度学习的方法^[1-4]来解决此问题,并取得了极大的进展.但是这些方法的性能通常受限于2个方面:一是依赖大量的数据标注,比如研究者们常用的Market1501数据集^[5]包括了1501个行人的32 668张图片,对这些图片进行标注是费时费力的;二是泛化能力有限,当把训练好的模型应用到另一个数据集上进行测试时性能急剧下降,比如文献[6]提出的BoT(bag of tricks)模型在Market1501数据集上训练之后测试的Rank-1是94.5%,如果在DukeMTMC-ReID^[7]数据集上进行测试,Rank-1下降到41.4%.这些方法需要有标签数据集作为监督信息来优化模型,所以也称为有监督行人重识别方法.而无监督行人重识别无需数据集标注行人的编号信息,能以较低的成本对新场景进行数据采集和训练,能有效避免泛化能力有限的问题,因此研究基于无监督模式的重识别算法是有重大意义的.本文提出了一个新的框架为无标注的行人重识别数据集构造可靠的伪标签,再结合有监督方法进行训练,最终得到一个鲁棒的行人重识别模型,并在Market-1501和DukeMTMC-reID数据集上测试,同时与相关方法进行比较,实验结果验证了本文算法的有效性和优越性.

本文的主要贡献包括3个方面:

1)提出了一个鲁棒的伪标签生成框架,首先依据图片样本的摄像头编号把数据集分成若干个域,依次构造摄像头域内的局部伪标签和域间的全局伪标签.

2)使用最大团算法作为强约束对摄像头域内的

样本进行聚类,相比于常用的聚类方法不仅聚类时间更短,而且有更优的聚类结果.

3)将本文方法在2个大规模公开数据集Market-1501和DukeMTMC-ReID上进行了实验,同时与现有的无监督行人重识别方法进行比较,实验结果表明,本文提出的方法性能是最佳的.

1 相关工作

基于大量标注数据和深度学习的有监督行人重识别方法已经取得了极大的进展,比如MGN^[1],PCB^[2],AignedReID^[3]等方法,它们通过结合行人图片的全局特征和局部特征,或者使用特征对齐的方式重组特征描述向量,使用三元组损失和交叉熵损失优化网络参数,最终MGN在Market1501数据集上的Rank-1达到了95.7%.文献[6]研究了行人重识别模型构建和训练过程中的一些技巧,比如随机擦除、标签平滑、预热学习率和损失函数等,并提出了一个仅使用全局特征且网络结构简单的模型BoT,该模型在Market1501数据集上的Rank-1是94.5%,仅比结构复杂的MGN低1.2%.因此本文使用BoT作为基础模型,用于提取无标签数据集的特征和训练伪标签数据集.

传统的无监督行人重识别方法尝试设计手工特征^[8]、发掘显著性信息^[9]等方式利用无标签数据集.这些算法可能在特定的数据集上表现良好,但并不适合处理复杂的真实场景,不具备通用性.还有一些基于深度学习的方法,比如BUC^[10]设计了一个通过发掘行人之间相似性的聚类合并方法来产生伪标签,但是数据集中的样本数目较大,聚类过程十分耗时,而且难以确定聚类中心个数.文献[11]提出把样本之间的相似性作为软限制构造软伪标签,起始阶段把每个样本当成一个类别,最终训练软伪标签数据集.显而易见,构造的伪标签越接近真实标签,训练出来的模型效果也就越好.与BUC构造硬标签类似,本文通过使用更强的约束范式来生成更可靠的伪标签,

从而取得了更好的结果。

无监督域自适应算法是无监督算法的一种，和BUC等算法相比，无监督域自适应算法要求有一个已知标签的数据集作为辅助域，通过迁移学习或者缩小辅助域和目标域之间的差异，从而在无标签的目标域上获得较好的结果。文献[12]通过把辅助域和目标域的特征映射到同一个特征空间中来解决辅助域和目标域的特征差异；CORAL^[13]尝试对齐2个域的特征分布的均值和方差；SSG^[14]提出一个挖掘无标签样本潜在相似性的自相似分组方法，进而构造伪标签。在辅助域的帮助下，文献[12–14]方法的表现优于没有辅助域的无监督方法，但是其性能会受到辅助域和目标域的差异程度的影响。

摄像头内监督(intra-camera supervision, ICS)假设已知每个摄像头拍摄的行人图片的标签是已知的，但这些标签是独立的，即各个摄像头之间的图片标签没有关联，ICS场景下的行人重识别是半监督学习问题。这样的标签在真实场景中能以较低的代价获取，比如使用目标跟踪方法来获取一个摄像头拍摄到的行人图片，并认为相邻帧内的行人是同一个行人。现有的ICS模型通过摄像头内学习和摄像头间学习来解决这个问题，比如MATE^[15]使用多分支网络来分别学习对应各个摄像头域的模型参数。ACAN^[16]使用多摄像头对抗学习把不同摄像头拍摄到的图片映射到一个公共空间。文献[15–16]方法的效果优于无监督方法，但是和经典的有监督方法相比仍有较大的差距。本文提出的方法借鉴了ICS模型，首先通过摄像头编号把训练集分成各个摄像头域，分别在每个摄像头域内构造伪标签，因此把无监督行人重识别问题转化为ICS问题；其次考虑到ICS模型已知的摄像头域内标签是真实标签，而本文的伪标签和真值存在差异，若直接使用ICS模型，效果不如ICS方法，所以我们通过计算不同摄像头域之间样本的相似度来构造全局标签，然后使用BoT进行训练。

部分半监督行人重识别方法假设仅有少部分的样本是有标签的，比如文献[17]假设每个行人仅有一个样本是有标注的，通过预测其他样本的标签，然后用这些伪标签数据来更新模型，迭代地预测标签和更新模型，最终得到一个鲁棒的模型。文献[18]假设数据集中的一部分标签是已知的，比如Market-1501数据集中有1/5的样本是已经标注的，最终能取得75.2%的Rank-1和53.2%的mAP；如果已标注数据达到了1/3，那么Rank-1和mAP将分别提升到83.9%和65.6%。本文对此实验进行了简化和扩展：一是直接

使用基础模型对有标签的这部分数据集进行训练；二是测试了更多不同比例的有标签样本基础模型的表现。结果如图1所示，从图1可以看出，随着样本比例的增加，Rank-1性能和mAP性能逐渐增加，并在后期趋于平稳，尤其是在样本比例比较少的情况下，增幅更明显。受图1实验结果的启发，本文通过逐步生成更多可靠的伪标签数据来获得性能的稳步提升。

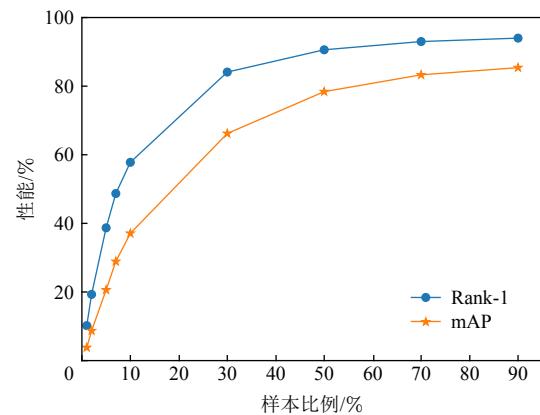


Fig. 1 Experimental results on Market1501 with labeled samples

图1 在Market1501数据集上的标签样本实验结果

2 基于摄像头域内域间融合的伪标签生成算法

本节主要介绍基于摄像头域内域间融合的伪标签生成算法，这是本文所提的无监督行人重识别算法框架的核心部分。

2.1 问题定义

首先给出本文要解决问题的形式化描述。本文的无监督行人重识别的设定是：已知数据集为 $X = \{X_1, X_2, \dots, X_{N_c}\}$ ，这里的 N_c 表示摄像头的个数， X_i 表示摄像头*i*捕捉到的行人图片。目标是重新建立 X 中的样本图片关系，即同一个行人的图片应该归为一类，用公式描述为 $X' = \{P_1, P_2, \dots, P_{N_n}\}$ ，这里的 N_n 表示整个数据集中的行人个数， P_i 表示行人*i*的所有图片。在实际情况中，因为行人个数是未知的，所以很难完成从 X 到 X' 的转换。我们的目标是使生成的伪标签尽可能接近真实标签，同时让基础模型在生成的伪标签数据集上能有较好的表现。

本文使用BoT作为基础模型，表示为 $H(\theta)$ ，用于提取图片特征 $F = H(X; \theta)$ ，算法的最终目标是学习到合适的参数 θ 使得基础模型在测试集上有较好的表现。

2.2 本文的算法框架

本文提出的算法框架如图2所示，这是一个迭代

的算法。首先无标签的数据集通过基础模型提取特征，然后分别经过摄像头域内合并、域间合并以及样本召回3个步骤得到伪标签，最后使用基础模型对伪标签数据集进行训练，这是一次迭代过程。下一次迭代过程使用上一次迭代时训练得到的模型参数提取特征，然后重复生成伪标签和训练的步骤。在每轮迭代

开始前，需要对数据集进行处理。为了避免摄像头间因背景不同、摄像头参数不同、拍摄角度不同等因素带来的影响，把数据集依据摄像头编号分成若干个域，对每个域中的样本图片 X_p 提取特征 $f_{p,i} \in \mathbb{R}^d$ ，其中 $f_{p,i}$ 表示摄像头 p 拍摄到的图片 i 经过基础模型的骨干网络提取得到的特征，并且经过了L2正则化。

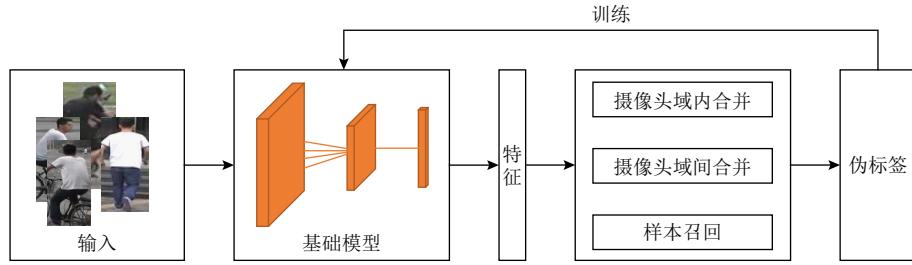


Fig. 2 The proposed algorithm framework

图2 本文算法框架

2.3 摄像头域内合并算法

摄像头域内合并算法分别对每个摄像头域进行操作，可以分为预合并、基于最大团的合并算法和样本清洗这3个步骤，最终得到各个摄像头域内伪标签。

1) 预合并。如图3所示，一些图片行人看起来很相似，它们很可能是同一个行人，比如摄像头拍摄到的视频中相邻两帧的行人图片。这种视觉上看起来很相似的图片，通过基础模型提取的特征也是相似的，

所以可以使用样本特征计算样本之间的相似度，进而预先合并视觉上相似的图片。具体做法是：使用特征 f_p 计算各样本之间的欧氏距离 d_1 ，设定阈值 t_1 ，若2个样本之间的距离小于 t_1 ，那就认为它们足够相似，可以预先合并。如图4(a)表示某个摄像头域内预合并前的样本，合并后的样本被当成整体组成一个新样本，如图4(b)所示，其特征由它包含的所有样本特征的平均值表示。为简单起见，摄像头域 p 下的图片特征仍记作 f_p 。



Fig. 3 Some samples captured by the same camera in Market1501

图3 Market1501 数据集中由同一个摄像头拍摄到的样本图片

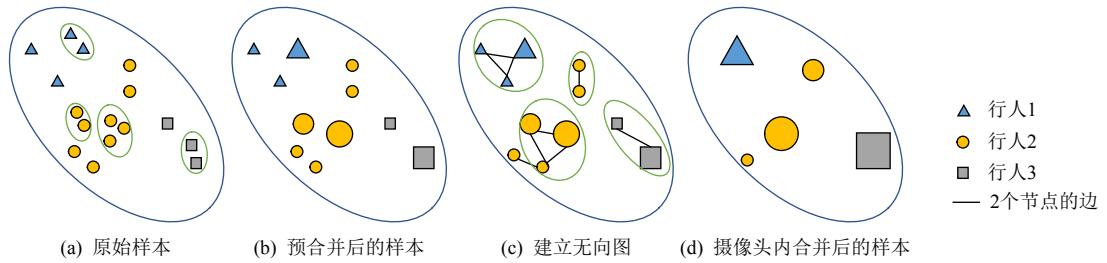


Fig. 4 Partial samples for intra-camera merger procedure

图4 某摄像头域内合并时的部分样本

2) 基于最大团的合并算法。预合并操作之后，建立图模型。把每个样本当成图中的一个结点，然后计

算各个样本之间的欧氏距离 d_2 ，并设定阈值 t_2 ，若2个结点之间的距离小于 t_2 ，则在它们之间连接一条无向

边。这样，摄像头域内的样本，可以看作是一个巨大的无向图，这个无向图由很多个小的无向图组成，每个小的无向图都是连通图，如图4(c)所示。对于每个小的无向图而言，它表示的物理意义是：若2个结点有边相连，则这2个结点对应的样本在一定程度上是相似的，但不能直接认定为属于同一个行人。考虑到最大团是一个图中的子图，并且这个子图中的所有结点两两相连，因此可以使用最大团作为强约束在每个小的无向图中找到可能表示同一个行人的样本。

一个需要注意的细节是，寻找最大团是一个NP难题，最大团算法耗时随着无向图中结点个数的增加而急剧增加，所以当小无向图中的结点个数过多时，最大团算法执行缓慢。鉴于实际数据集中，每个摄像头拍摄到的同一行人图片数量较少，因此可以限制小无向图的结点个数。如果某个小无向图的结点个数超出限制，那么我们使用更小的阈值来确定是否连接2个结点。如果结点个数仍然超出限制，那就继续减小阈值，直到满足要求。另一方面，如果小无向图中的结点个数太少，比如只有2或者3个结点，此时没必要使用最大团算法进行求解。我们使用一种简单可靠的方法：如果2个或者3个结点之间的平均距离小于更小的阈值，比如 $0.75t_2$ ，那就认为它们等价于1个最大团。最终的合并结果如图4(d)所示。

3) 样本清洗。经过前2步合并操作之后，各摄像头域内表示同一个行人的样本被合并在同一个集合内，但可能存在不完整的合并，如图4(d)所示，存在3个圆形结点，即同一个行人分别被分在了3个集合中，这会给后面的训练过程带来标签噪声，影响模型效果。因此，本文方法通过丢弃一部分集合来减少这种不完整的合并。具体做法是，使用式(1)计算每个集合的团内相似度：

$$sim_i = \frac{1}{N_{p,i}} \sum_{k \in S_i} f_{p,k} \cdot f_{p,k}^T, \quad (1)$$

式(1)中， $N_{p,i}$ 表示摄像头 p 内的第*i*个样本集合对应的样本个数， $f_{p,k}$ 表示该集合中的第*k*个样本特征， T 表示矩阵的转置。丢弃团内相似度最小的一部分集合，是因为若给这些集合里的样本分配伪标签，可能会有较多的错误，因此暂且丢弃这些样本，清洗后的结果如图5所示。

2.4 摄像头域间合并算法

摄像头域内合并之后，除了暂时无法分配标签的样本，其他样本都被赋予了摄像头内部的独立标签，这和ICS问题的设定一致。但我们并不直接使用ICS方法，因为ICS问题中各个摄像头域内的标签是

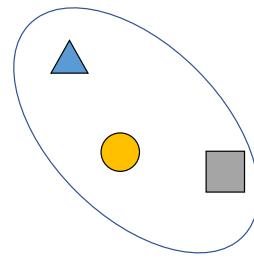


Fig. 5 The result after sample cleaning

图5 样本清洗后的结果

真实标签，若直接使用上述生成的伪标签，则性能会低于ICS方法。这里本文进一步构造全局伪标签，给不同摄像头域内表示同一个行人的样本图片分配相同的标签，然后结合现有的有监督模型进行训练。

如图6所示，Market1501数据集中的4个样本分别是摄像头2下的行人2(图6(a))和行人22(图6(b))以及摄像头5下的行人2(图6(c))和行人22(图6(d))。由神经网络提取得到的样本1和样本3的特征的相似程度高于样本1和样本4的，因为样本1和样本3的前景(行人)是相似的，而样本1和样本4的背景和前景都不同。基于这个前提假设，本算法合并来自不同摄像头的集合。首先通过式(2)计算来自摄像头域 p 和 q 的2个集合*i*和*j*的域间团的相似度：

$$sim_{i,j} = \frac{\exp(f_i \cdot f_j^T / \tau)}{\sum_{k=1}^{N_q} \exp(f_i \cdot f_k^T / \tau)}, \quad i \in [1, N_p], j \in [1, N_q], \tau = 0.05 \quad (2)$$

其中 N_p, N_q 分别表示摄像头 p 和 q 经过域内合并后的集合个数， τ 是一个超参数。如果 $sim_{i,j}$ 和 $sim_{j,i}$ 同时大于设定的阈值 η_1 ，那么就合并集合*i*和*j*，赋予它们相同的伪标签。

经过域间合并操作，无标签样本被分配了全局的伪标签，原数据集可表示为 $X^* = \{X_1, X_2, \dots, X_{N_n}\}$ 。这里的 N_n 表示合并后的集合个数，也就是行人个数。



(a) 样本1 (b) 样本2 (c) 样本3 (d) 样本4
(摄像头2, 行人2) (摄像头2, 行人22) (摄像头5, 行人2) (摄像头5, 行人22)

Fig. 6 Samples of person 2 and 22 under camera 2 and 5.

图6 行人2和22在摄像头2和5下的样本

2.5 样本召回

在摄像头域内合并操作中，有一部分样本暂时

无法分配合适的标签,所以这些样本被弃用.为了最大程度利用无标签样本,本文使用样本召回操作给之前弃用的样本重新分配标签.本文使用无参分类器^[19]计算样本*i*应该分配标签*k*的概率 $prob_{i,k}$.令 $k = \arg \max(prob_i)$,若 $prob_{i,k}$ 大于给定阈值 η_2 ,则给样本*i*分配标签*k*;否则样本*i*被弃用.

2.6 训练模型

本文采用BoT对生成的伪标签数据集进行训练,得到模型参数 θ .在下一轮迭代开始前,使用 $F = H(X; \theta)$ 提取特征,然后开始新一轮的迭代.更多的细节在3.2节详述.

3 实验和分析

3.1 数据集和评价指标

本文主要在Market1501和DukeMTMC-ReID这2个大规模行人重识别数据集上进行实验,鲁棒性实验和参数分析使用了MSMT17数据集^[20]. Market1501数据集由6个摄像头采集的图片组成,包括751个行人的12 936张图片作为训练集,750个行人的19 732张图片作为测试集; DukeMTMC-ReID数据集由8个摄像头进行采集,训练集包括1 404个行人的16 522张图片.各个摄像头捕捉到的行人个数和图片张数如表1所示.

Table 1 Person Numbers and Image Numbers Under Each Camera in Market1501 and DukeMTMC-ReID

表 1 Market1501 和 DukeMTMC-ReID 数据集中各个摄像头下的行人个数和图片张数

摄像头编号	Market1501		DukeMTMC-ReID	
	行人个数	图片张数	行人个数	图片张数
1	652	2 017	404	2 809
2	541	1 709	378	3 009
3	694	2 707	201	1 088
4	241	920	165	1 395
5	576	2 338	218	1 685
6	558	3 245	348	3 700
7			217	1 330
8			265	1 506

MSMT17数据集一共有15个摄像头采集到的126 441张图片,其中训练集有32 621张图片.本文使用此数据集分析部分参数对实验结果的影响.

本文采用累计匹配特征曲线(CMC)以及平均查准率(mAP)作为评价指标,累计匹配特征曲线衡量

的是查询图片出现在匹配结果列表中的概率,但是在实际中往往只考虑是否在前1个或前5个结果中匹配成功,本文关注前1个匹配成功的概率即Rank-1.

3.2 实现细节和相关的专家级方法

本文使用BoT这一有监督算法作为基础模型,除训练轮次,其他超参数设置以及训练方式和原算法保持一致.在第1次迭代开始前,使用在ImageNet^[21]预训练的参数 θ 提取无标签数据集的图片特征;之后的迭代使用基础模型在生成的伪标签数据集上训练得到的模型参数 θ 提取无标签数据集的图片特征.关于训练轮次的设置,在迭代初期,由于生成的伪标签样本数据量较小,所以训练轮次应该设置较小,以避免过拟合,本文设置第1次迭代时训练轮次为20;此后训练轮次固定为60.

本文使用的主要硬件资源为Intel® Core™ i7-7800X CPU和NVIDIA Titan Xp(12 GB显存);软件环境为Python3.7,PyTorch1.6,Scikit-Learn0.24.2.

目前基于深度学习的无监督行人重识别方法较少,无监督域自适应方法和半监督方法有时也被当作无监督方法,但它们的难度是低于无监督方法的.为了展示本文方法的优越性,本文比较了3个无监督方法:BOW^[5],BUC^[10],SSL^[11];4个无监督域自适应方法:MAR^[22],MMT^[23],ECN^[24],SSG^[14];2个ICS方法:MATE^[15],UGA^[25].

3.3 实验结果

如图7所示,Market1501和DukeMTMC-ReID随着迭代轮次逐渐增加,样本比例模型性能也随着迭代轮次逐步提升.在迭代初期,Rank-1和mAP的性能迅速增长,然后缓慢持续增长直至稳定.2个数据集都是经过15轮次迭代后性能达到最大稳定状态,在Market1501数据集的Rank-1和mAP分别是89%和74.9%,经过重排序^[26]操作,Rank-1和mAP分别达到90.5%和86.1%;在DukeMTMC-ReID数据集上Rank-1和mAP分别是76.9%和61.9%,重排序之后分别是79.1%和72.1%.

图7所示曲线中,随着迭代轮次的增加,赋予伪标签的样本占原总样本的比例逐渐增加,Rank-1和mAP也同步增加,这一结果与图1中小样本实验的结果相吻合,这表明本文所提框架生成的标签训练效果接近真实标签,也即所生成的伪标签是可靠的.为了直观地展示伪标签和真实标签的相似程度,在Market1501数据集中随机选择64个行人,使用第14轮次迭代得到的基础模型参数提取样本特征,然后用t-SNE对特征进行降维,可视化结果如图8所示,

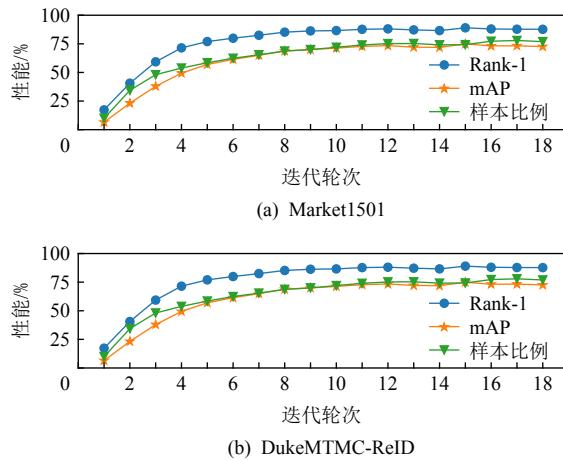


Fig. 7 The performance of our proposed method on Market1501 and DukeMTMC-ReID

图 7 本文方法在 Market1501 和 DukeMTMC-ReID 数据集上的性能

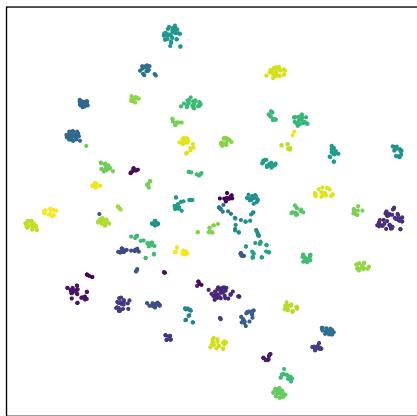


Fig. 8 Visualization of 64 randomly selected samples in Market1501.

图 8 Market1501 数据集中随机选择的 64 个样本的可视化图

不同的颜色表示不同的标签,可以看出绝大多数样本类内距离较小,类间界限明显,聚类结果较好。

表 2 和 **表 3** 是本文方法和 9 种当前方法比较的结果。BOW 在 Market1501 和 DukeMTMC-ReID 2 个数据集上的 Rank-1 和 mAP 值都比较低; BUC 是用自底向上逐步聚类的方法构造伪标签,和本文方法思路相近,但本文在 Market1501 和 DukeMTMC-ReID 2 个数据集上的 Rank-1 分别提升了 22.8% 和 29.5%; SSL 使用软伪标签作为监督信息训练模型,相比于 BUC 有了较大的提升,但依然没有超过本文方法。通过和当前无监督方法的比较,表明了本文构造的伪标签对于无监督行人重识别是可靠的。

无监督域自适应算法以一个有标签的数据集作为辅助域,希望在无标签的目标数据集上能有较好的表现。因为其难度低于无监督的方式,所以无监督

Table 2 Comparison of Our Proposed Method and the State-of-The-Art Methods on Market1501

表 2 本文方法和当前方法在 Market1501 数据集上的比较

方法	类别	Rank-1/%	mAP/%
BOW	无监督	35.8	14.8
BUC	无监督	66.2	38.3
SSL	无监督	71.7	37.8
MAR	域自适应	67.7	40.0
MMT	域自适应	87.7	71.2
ECN	域自适应	75.1	43.0
SSG	域自适应	86.2	68.7
MATE	ICS	85.3	65.2
UGA	ICS	63.7	41.2
本文方法	无监督	89.0	74.9

注: 黑体数字表示最优结果

Table 3 Comparison of Our Proposed Method and the State-of-The-Art Methods on DukeMTMC-ReID

表 3 本文方法和当前方法在 DukeMTMC-ReID 数据集上的比较

方法	类别	Rank-1/%	mAP/%
BOW	无监督	17.1	8.3
BUC	无监督	47.4	27.5
SSL	无监督	52.5	28.6
MAR	域自适应	67.1	48.0
MMT	域自适应	78.0	65.1
ECN	域自适应	63.3	40.4
SSG	域自适应	76.0	60.3
MATE	ICS	71.7	50.7
UGA	ICS	75.0	53.3
本文方法	无监督	76.9	61.9

注: 黑体数字表示最优结果

自适应算法效果普遍强于无监督方法,比如 MMT 在 DukeMTMC-ReID 数据集上表现最佳,本文方法结果位居第二,两者 Rank-1 相差 1.1%。

摄像头内监督行人重识别(ICS)是目前流行的半监督行人重识别问题,但是 MATE 和 UGA 的效果不如本文方法。这表明和半监督方法相比,本文所提出的方法仍有很强的竞争力。

3.4 参数分析

在摄像头域内域间合并以及样本召回的过程中需要设定一些阈值或者参数,比如每个小无向图的最大结点个数、预合并的距离阈值等。各个参数对伪标签质量影响的分析为:

1) 预合并时的距离阈值 t_1 。预合并通过提前把视觉上几乎没有差异的样本归为一类形成新样本来减少后续相似样本的个数,进而减少无向图中的结点

个数, 加速最大团算法的求解。然而为每个摄像头域确定一个合适的具体值作为阈值比较困难, 通过分析, 每个摄像头域内都存在这样一部分样本, 它们来自相邻帧的同一个行人, 它们之间的距离应该是小于和其他样本对的, 这表示在每个摄像头域内, 距离最小的那部分样本对应该被合并。因此首先把各个摄像头域内样本间的距离从小到大排序, 选择距离最小的前 r_1 百分比的样本对进行合并, 这样能针对不同的摄像头使用不同的距离阈值, 进而得到较好的合并结果。因此在实验中, 不直接设置 t_1 的值, 而是通过设置 r_1 计算得到。在实验中, 对于 2 个数据集 r_1 值均设置为 0.1。可以预见, 在大规模数据集中这类可以预先合并的样本更多, 因此应该设置更大的阈值。

2) 基于最大团的合并算法中建立图模型的距离阈值 t_2 , t_2 类似 t_1 , 但它用于衡量 2 个样本是否在一定程度上可以合并。为了把更多可能表示同一个行人的样本包括在同一个无向图中, t_2 值的设置应该稍大一些。如图 9 所示, 分别设置 r_2 为 0.05, 0.1, 0.15, 0.2, 0.3, 0.5, 1.0 等数值迭代 15 轮。当 r_2 值在 0.05~0.3 时, Rank-1 先增加后减小, 随着 r_2 值的继续增大, Rank-1 略微增加, 但可以视为实验误差而被认为基本不变, 因此在本实验中 r_2 值设置为 0.2, 和 t_1 一样, 使用 r_2 计算得到 t_2 。

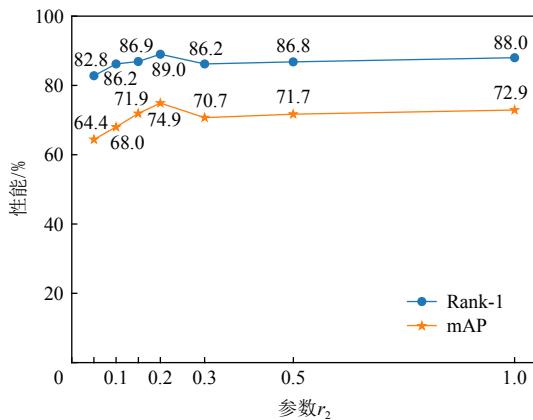


Fig. 9 The effects of r_2 on the performance of Market1501

图 9 r_2 对 Market1501 数据集性能的影响

3) 无向图的最大结点个数。同一个无向图中的结点表示对应的样本可能是同一个行人, 为了避免遗漏潜在样本, 本不应该限制结点个数, 但由于求解最大团是 NP 难题, 若结点过多, 则算法求解费时甚至无法求解。考虑到已经预先合并了相似样本, 以及最大团算法的运行效率, 本文设置结点个数最大为 50, 若数据集中同一个摄像头拍摄到的行人图片较多, 结点个数应和预合并阈值一样应设置更大一些。如

图 10 所示, 在 MSMT17 数据集上的实验结果表明了这样调整参数是有效的。

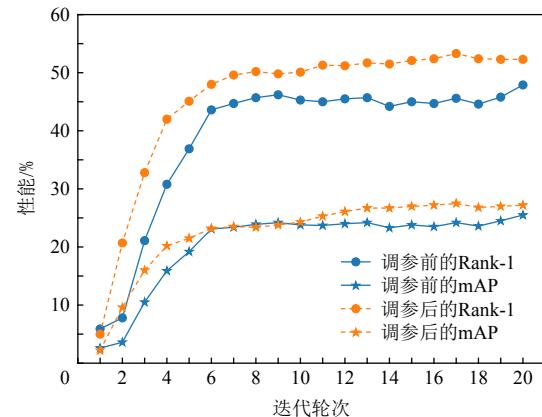


Fig. 10 The performance of our proposed method on MSMT17

图 10 本文方法在 MSMT17 数据集上的性能

4) 摄像头域内合并算法中清洗标签丢弃的集合个数。鉴于存在同一个行人在同一个摄像头下也可能差异较大的情况, 以及算法的误差带来的影响, 经过摄像头内合并后得到的团可能存在重复的情况, 我们丢弃掉一部分不可靠的团, 并在后续的算法中重新召回这些样本。本文尝试不同的清洗比例, 实验结果如图 11 所示, 随着清洗比例的增加, Rank-1 先增加后降低, 在不清洗时 Rank-1 仅有 82.9%; 清洗比例为 5%, 10%, 20% 时 Rank-1 都在 86% 以上; 而清洗比例太高时, 如 30% 则 Rank-1 降为 84.6%。由此结果, 本文设定清洗比例为 15%, 即只保留团内相似度最大的前 85% 的团。

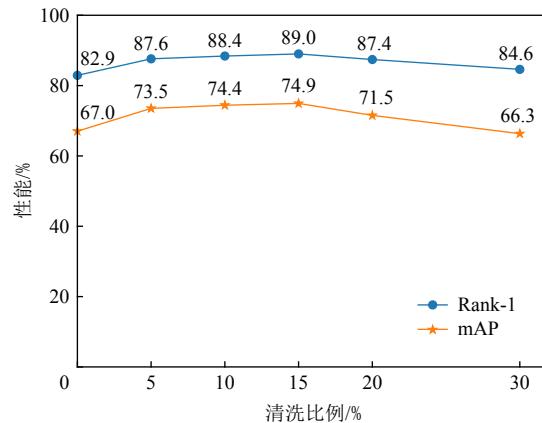


Fig. 11 The effects of clean ratios on the performance of Market1501

图 11 清洗比例对 Market1501 数据集性能的影响

5) 摄像头域间合并的阈值 η_1 。各个摄像头域内的样本通过计算域间团的相似度进行合并, 我们认为, 在迭代初期各个摄像头域间样本的特征差异较大,

此阈值应该设置较小；随着算法迭代，提取得到的特征逐渐可靠，使用更严格的约束限制合并，所以阈值应该设置较大。在实验中前 13 轮次的迭代 $\eta_1 = 0.7$ ，之后的迭代中 $\eta_1 = 0.85$ 。

6) 样本召回的阈值 η_2 。在摄像头域内合并算法中，有一部分样本暂时无法分配合适的伪标签被临时弃用，但在摄像头域间合并结束后可以通过无参分类器对这些样本进行分类，若属于某个类别的概率较大，那就分配对应的伪标签参与到最终的训练中。在本实验中 η_2 设置为 0.35。

3.5 聚类算法的消融实验

本文使用最大团算法作为强约束进行摄像头域内合并，为了验证是否优于传统的聚类算法，本文以 KMeans 聚类算法为例进行实验和分析。具体的实验设置为：当实验数据集是 Market1501 时，使用在 DukeMTMC-ReID 数据集上预训练的模型参数提取特征；然后分别使用摄像头域内合并算法和 KMeans 聚类算法生成摄像头域内的伪标签，在使用 KMeans 算法进行聚类时，聚类中心个数设置为各摄像头内真实的行人个数；紧接着使用摄像头域间合并算法和样本召回构造全局伪标签；最后使用 BoT 进行训练，结果如表 4 所示：

Table 4 Comparison of Our Proposed Method and the Clustering-Based Method

表 4 本文方法和基于聚类的方法的比较

数据集	方法	Rank-1 /%	mAP /%	运行时间 /s
Market1501	聚类方法	54	34.1	630
	本文方法	60.7	38.1	1.8
DukeMTMC-ReID	聚类方法	46.7	36.0	513
	本文方法	53.4	39.2	2.9

在 2 个数据集上，本文基于最大团的摄像头域内合并算法的表现均优于 KMeans 聚类算法。此外，基于最大团的域内合并算法的用时远低于聚类算法，两者相差上百倍，本文所提算法具有明显的优势。如表 1 所示，Market1501 数据集中有 6 个摄像头，但是每个摄像头内的行人个数都比较多，所以聚类算法在 Market1501 数据集上更费时，而本文提出的算法受此影响较小，可用于更大规模的数据集。

3.6 本文方法的鲁棒性实验

本文所提出的方法用于解决无监督行人重识别中的标签缺失问题，以插件的形式结合有监督方法起作用。2 个实验表明本文方法具备通用性：1) 在 MSMT17 数据集上的实验结果意味着本文方法可以

在更大规模的数据集上正常工作；2) 以 TransReID [27] 作为基准模型也能取得不错的结果，表示本文方法能适配基于 Transformer 模型的新型有监督行人重识别方法。2 个实验的详细描述为：

1) 在 MSMT17 数据集上的实验使用在 3.4 节中设置的参数迭代 20 次，本文方法在 MSMT17 数据集上的 Rank-1 是 47.9%，mAP 是 25.5%，每一次迭代的结果如图 11 中实线所示。当按照各参数对实验结果的影响分析调整参数后，其结果如图 11 中虚线所示，Rank-1 达到了 52.3%，而 mAP 达到了 27.2%。和相关当前方法的比较如表 5 所示，本文方法在更困难的 MSMT17 数据集上效果优于大部分域自适应方法或者其他半监督方法。

Table 5 Comparison of Our Proposed Methods and the State-of-The-Art Methods on MSMT17

表 5 本文方法和当前方法在 MSMT17 数据集上的比较

方法	类别	Rank-1 /%	mAP /%
MMT	域自适应	58.8	29.7
ECN	域自适应	30.2	10.2
SSG++	域自适应	41.6	19.3
UGA	ICS	49.5	21.7
本文方法 1	无监督	47.9	25.5
本文方法 2	无监督	52.3	27.2

注：本文方法 1 使用在 3.4 节设置的参数，本文方法 2 使用在 3.6 节设置的参数。

2) 本文所提出的伪标签生成框架可适用于基于 Transformer 模型的有监督行人重识别算法，本文以 TransReID 作为基础模型取代原本使用的 BoT 进行实验，其结果如图 12 所示，可以看出 Rank-1 和 mAP 在前 5 轮迭代中快速增加，在之后的轮次里缓慢增加直至 Rank-1 达到 87.4%，mAP 达到 72.9%，虽然

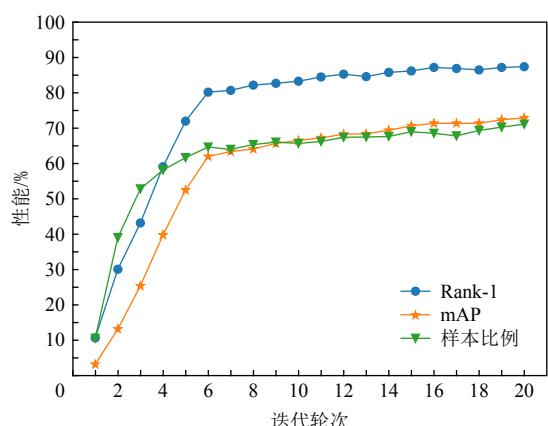


Fig. 12 The performance of TransReID on Market1501

图 12 TransReID 模型在 Market1501 数据集上的表现

TransReID 模型低于 BoT 模型的结果,但仍然超出大部分当前方法。

3.6 节 2 个实验结果表明,本文所提出的伪标签生成框架可以作为一个插件使现有的有监督模型可以处理大规模的无监督行人重识别数据集。

4 总 结

当前基于深度学习的行人重识别方法通常需要标注数据集,但对大量数据进行标注十分消耗人力物力。本文提出了一个框架用于生成伪标签数据集,然后使用有监督模型进行训练,实验结果证明了所提出方法的有效性。我们通过把无标签数据集依据摄像头编号分成若干个域,对每个域内的图片使用最大团算法作为强约束进行聚类,使得每个域内相似的图片被赋予同样的伪标签;然后通过计算域间团的相似度合并不同域的样本,进而构造全局伪标签。通过若干个实验,证明本文方法构造的伪标签优于基于聚类的方法,并且有更少的时间消耗,提升了无监督行人重识别的性能。

作者贡献声明:陈利文和叶锋提出研究思路和算法框架;黄添强和黄丽清负责设计实验以验证算法的有效性;陈利文、翁彬和胡杰共同编写代码进行实验测试;陈利文和徐超完成论文撰写工作。

参 考 文 献

- [1] Wang Guanshuo, Yuan Yufeng, Chen Xiong, et al. Learning discriminative features with multiple granularities for person re-identification[C] //Proc of the 26th ACM Int Conf on Multimedia. New York: ACM, 2018: 274–282
- [2] Sun Yifan, Zheng Liang, Yang Yi, et al. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)[C] //Proc of the 16th European Conference on Computer Vision. Berlin: Springer, 2018: 480–496
- [3] Zhang Xuan, Luo Hao, Fan Xing, et al. AlignedReID: Surpassing human-level performance in person re-identification[EB/OL]. 2017 [2021-10-26]. <https://arxiv.org/pdf/1711.08184.pdf>
- [4] Dai Chenchao, Wang Hongyuan, Ni Tongguang, et al. Person re-identification based on deep convolutional generative adversarial network and expanded neighbor reranking[J]. *Journal of Computer Research and Development*, 2019, 56(8): 1632–1641 (in Chinese) (戴臣超,王洪元,倪彤光,等. 基于深度卷积生成对抗网络和拓展近邻重排序的行人重识别[J]. *计算机研究与发展*, 2019, 56(8): 1632–1641)
- [5] Zheng Liang, Shen Liyue, Tian Lu, et al. Scalable person re-identification: A benchmark[C] //Proc of the 15th Int Conf on Computer Vision (ICCV). Piscataway, NJ: IEEE, 2015: 1116–1124
- [6] Wei Longhui, Zhang Shiliang, Gao Wen, et al. Person transfer GAN to bridge domain gap for person re-identification[C] //Proc of the 34th Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 79–88
- [7] Luo Hao, Gu Youzhi, Liao Xingyu, et al. Bag of tricks and a strong baseline for deep person re-identification[C] //Proc of the 35th Conf on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ: IEEE, 2019: 1487–1495
- [8] Zheng Zhenhong, Zheng Liang, Yang Yi. Unlabeled samples generated by gan improve the person re-identification baseline in Vitro[C] //Proc of the 16th Int Conf on Computer Vision (ICCV). Piscataway, NJ: IEEE, 2017: 3774–3782
- [9] Liao Shengcai, Hu Yang, Zhu Xiangyu, et al. Person re-identification by local maximal occurrence representation and metric learning [C] //Proc of the 31st Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2015: 2197–2206
- [10] Wang Hanxiao, Gong Shaogang, Xiang Tao. Unsupervised learning of generative topic saliency for person re-identification[EB/OL]. 2014 [2021-10-26]. <http://www.bmva.org/bmvc/2014/files/paper019.pdf>
- [11] Lin Yutian, Dong Xuanyi, Zheng Liang, et al. A bottom-up clustering approach to unsupervised person re-identification[C] //Proc of the 33rd Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2019: 8738–8745
- [12] Lin Yutian, Xie Lingxi, Wu Yu, et al. Unsupervised person re-identification via softened similarity learning[C] //Proc of the 36th Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2020: 3390–3399
- [13] Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation[C] //Proc of the 32nd Int Conf on Machine Learning. New York: ACM, 2015: 1180–1189
- [14] Sun Baochen, Saenko K. Deep coral: Correlation alignment for deep domain adaptation[C] //Proc of the 15th European Conf on Computer Vision. Berlin: Springer, 2016: 443–450
- [15] Fu Yang, Wei Yunchao, Wang Guanshuo, et al. Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification[C] //Proc of the 17th Int Conf on Computer Vision (ICCV). Piscataway, NJ: IEEE, 2019: 6112–6121
- [16] Zhu Xiangping, Zhu Xiatian, Li Minxian, et al. Intra-camera supervised person re-identification: A new benchmark[C] //Proc of the 17th Int Conf on Computer Vision Workshop (ICCVW). Piscataway, NJ: IEEE, 2019: 1079–1087
- [17] Qi Lei, Wang Lei, Huo Jing, et al. Adversarial camera alignment network for unsupervised cross-camera person re-identification[J/OL]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021 [2021-06-12]. <https://ieeexplore.ieee.org/document/9481292>
- [18] Wu Yu, Lin Yutian, Dong Xuanyi, et al. Progressive learning for person re-identification with one example[J]. *IEEE Transactions on Image Processing*, 2019, 28(6): 2872–2881

- [19] Chang Xinyuan, Ma Zhiheng, Wei Xing, et al. Transductive semi-supervised metric learning for person re-identification[J]. Pattern Recognition, 2020, 108: 107569
- [20] Wu Zhirong, Xiong Yuanjun, Yu S, et al. Unsupervised feature learning via non-parametric instance discrimination[C] //Proc of the 34th Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 3733–3742
- [21] Deng Jia, Dong Wei, Socher R, et al. ImageNet: A large-scale hierarchical image database[C] //Proc of the 25th Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2009: 248–255
- [22] Yu Hongxing, Zheng Weishi, Wu Ancong, et al. Unsupervised person re-identification by soft multilabel learning[C] //Proc of the 35th Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2019: 2148–2157
- [23] Ge Yixiao, Chen Dapeng, Li Hongsheng. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification[EB/OL]. 2020 [2021-10-26]. <https://arxiv.org/pdf/2001.01526.pdf>
- [24] Zhong Zhun, Zheng Liang, Luo Zhiming, et al. Invariance matters: Exemplar memory for domain adaptive person re-identification [C] //Proc of the 35th Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2019: 598–607
- [25] Wu Jinlin, Liu Hao, Yang Yang, et al. Unsupervised graph association for person re-identification[C] //Proc of the 17th Int Conf on Computer Vision (ICCV). Piscataway, NJ: IEEE, 2019: 8320–8329
- [26] Zhong Zhun, Zheng Liang, Cao Donglin, et al. Re-ranking person re-identification with k-reciprocal encoding[C] //Proc of the 33rd Conf on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2017: 3652–3661
- [27] He Shuting, Luo Hao, Wang Pichao, et al. TransReID: Transformer-based object re-identification[EB/OL]. 2021 [2021-10-26]. <https://arxiv.org/pdf/2102.04378.pdf>



Chen Liwen, born in 1996. Master candidate. His main research interest includes person re-identification.
陈利文, 1996 年生. 硕士研究生. 主要研究方向为行人重识别.



Ye Feng, born in 1978. PhD, associate professor. His main research interests include computer vision, video image processing.
叶 锋, 1978 年生. 博士, 副教授. 主要研究方向为计算机视觉、视频图像处理.



Huang Tianqiang, born in 1971. PhD, professor. His main research interests include multimedia information security, big data decision analysis.
黄添强, 1971 年生. 博士, 教授. 主要研究方向为多媒体信息安全、大数据决策分析.



Huang Liqing, born in 1991. PhD. Her main research interests include image and video super-resolution, restoration and detection.
黄丽清, 1991 年生. 博士. 主要研究方向为图像视频的超分辨率、重建和检测.



Weng Bin, born in 1981. PhD. His main research interests include machine learning and machine application.
翁 彬, 1981 年生. 博士. 主要研究方向为机器学习和机器应用.



Xu Chao, born in 1981. Master. His main research interests include multimedia security, machine learning.
徐 超, 1981 年生. 硕士. 主要研究方向为多媒体安全、机器学习.



Hu Jie, born in 1995. Master candidate. His main research interests include machine learning and machine application.
胡 杰, 1995 年生. 硕士研究生. 主要研究方向为机器学习和机器应用.