第52卷第5期
2020年10月

南京航空航天大学学报
Journal of Nanjing University of Aeronautics & Astronautics

Vol. 52 No. 5
Oct. 2020

DOI:10.16356/j.1005-2615.2020.05.011

基于多样性约束和离散度分层聚类的无监督视频行人重识别

曹　亮，王洪元，戴臣超，陈　莉，刘　乾

（常州大学阿里云大数据学院，常州，213164）

摘要：视频行人重识别是一项应用非常广的计算机视觉任务。目前的视频行人重识别方法通常是基于监督学习的，该方法需要手工标记大量的数据，代价非常高且并不适用于现实场景。本文提出了一种从底向上的基于多样性约束和离散度分层聚类的无监督视频行人重识别方法。该方法首先将每个样本当作是一个不同的类，然后结合类内间离散度进行从底向上的分层聚类，类间和类内离散度都小的类别将被优先合并，同时在聚类准则中加入一项多样性约束来平衡每类中的样本数量，最后，利用线性变化的特征存储器动态更新模型。在Mars和DukeMTMC-VideoReID两个大型视频数据集上的实验结果表明，相比于目前先进的无监督视频行人重识别方法，本文方法在性能上有一定的提升。

关键词：无监督视频行人重识别；离散度；聚类；特征存储器；多样性约束

中图分类号：TP391.41　　**文献标志码：**A　　**文章编号：**1005-2615(2020)05-0752-08

Unsupervised Video-Based Person Re-identification Based on Diversity Constraint and Dispersion Hierarchical Clustering

CAO Liang, *WANG Hongyuan*, *DAI Chenchao*, *CHEN Li*, *LIU Qian*

(Aliyun School of Big Data, Changzhou University, Changzhou, 213164, China)

Abstract: Video-based person re-identification (Re-ID) is a widely used task in computer vision. At present, most video-based Re-ID methods are based on supervised learning, which requires intensive manual annotation, and is very expensive and not suitable for real-life scenarios. In this work, an unsupervised video-based person Re-ID method based on diversity constraint and dispersion hierarchical clustering is proposed. First, each sample is regarded as a single cluster, and both the inter and the intra-cluster dispersions are combined to perform bottom-up hierarchical clustering. Second, clusters with small inter-cluster and intra-cluster dispersions can be prioritized for merging. At the same time, the diversity constraint is added to the clustering criterion to balance the number of samples in each cluster. Finally, the model is dynamically updated by using a linear feature memory. Experimental results on two public benchmark datasets, including Mars and DukeMTMC-VideoReID, show that compared with the state-of-the-art unsupervised video-based person Re-ID methods, the proposed method has some improvement in performance.

Key words: unsupervised video-based person re-identification; dispersion; clustering; feature memory; diversity constraint

基金项目：国家自然科学基金(61976028, 61572085, 61806026, 61502058)资助项目；江苏省自然科学基金(BK20180956)资助项目。

收稿日期：2020-06-05；**修订日期：**2020-07-10

通信作者：王洪元，男，教授，硕士生导师，E-mail: hywang@cczu.edu.cn。

引用格式：曹亮，王洪元，戴臣超，等. 基于多样性约束和离散度分层聚类的无监督视频行人重识别[J]. 南京航空航天大学学报，2020，52(5)：752-759. CAO Liang, WANG Hongyuan, DAI Chenchao, et al. Unsupervised video-based person re-identification based on diversity constraint and dispersion hierarchical clustering[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2020, 52(5): 752-759.

图片行人重识别是给定一张由一个相机拍摄的图像，在另一个不同的非重叠相机拍摄的一组图像中去匹配相同的行人[1-2]。而视频行人重识别的目标是从两个视频片段中，判断是否包含相同的行人。相比基于图片的行人重识别严重依赖于与衣服颜色相关的外观特征，视频行人重识别可以融合时间和空间两域的信息，从而可以提取更丰富的时空特征，提高拥有相似外观行人的识别精度。因此基于视频的行人重识别近些年引起了研究者的广泛关注[3-5]。在大规模的复杂场景下，手工标记数据会消耗大量的人力和财力，从而限制了监督学习在实际案例中的应用，而无监督学习并不需要身份标签，可以避免标记大量数据。

传统的无监督方法主要是基于手绘特征[6-8]、显著性分析[9-10]和字典学习[11-12]等等，这些方法相比于监督学习，精度存在一定的差距。随着深度学习的兴起，一些基于迁移学习[13]的域适应方法获得了不错的效果。这些方法主要是利用源域中有标记的数据，并从中学习判别性的特征表示，然后将预学习到的模型适应到无标记的目标域中。比如Fan等[14]提出在有标记的源数据集上预训练一个卷积神经网络（Convolutional neural network，CNN）模型，然后对目标域中提取的特征采用K-均值聚类逐步更新模型。Zhong等[15]在有标记的源数据集和无标记的目标数据集上同时微调CNN模型。这些方法的前提是源域和目标域有着相同的数据分布，然而这种假设在行人重识别任务中是不成立的。同时这些域适应的方法仍然是需要有标记数据的，并不完全是无监督学习。

本文提出一种不采用任何身份标签的完全无监督方法。通过一种从底向上的基于多样性约束和离散度的分层聚类方法，将相似行人按步骤合并为一个类别，并根据聚类结果设置类别标签。分层聚类的过程如图1所示。具体来说，在一开始网络训练时，将每一个样本看成一类，然后每次会按照聚类准则合并一定数量的样本，将距离相近的样本优先聚成一类，直至最后所有的相似样本都能被合并到一类中。本文的聚类准则是将类内间离散度和多样性约束相结合，可以尽可能地将相似样本聚在一类中，避免错误聚类，从而提升了聚类效果。本文方法在Mars和DukeMTMC-VideoReID两个大型视频数据集上，进行多次实验，都取得了很好的效果。

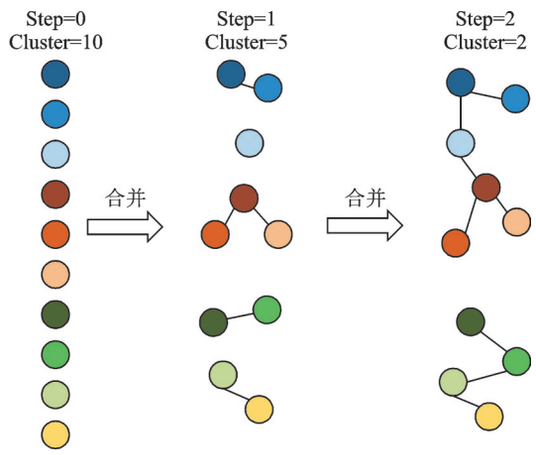

图1　从底向上的分层聚类

Fig.1　Bottom-up hierarchical clustering

1　相关工作

1.1　无监督行人重识别

为了避免对有标记数据的依赖，一些研究者提出了无监督跨数据集的行人重识别方法。现有的部分方法是迁移源域中的标签信息，或者是假设有很强的先验知识（比如假设目标域中的数据有特定的簇结构）。还有一些工作是减少标记，将目标域上的标记预算最小化，这种大部分是基于聚类的方法。例如，Yu等[16]提出了一种非对称度量聚类来发现未标记目标数据中潜在的标签。文献[17]中使用了聚类技术来推断关于目标身份的信息，并将估计的身份标签数据合并到训练过程中。在文献[18]中，为了学习数据集的共享表示，提出了基于字典学习的方法。另外还有一些方法是把生成对抗网络（Generative adversarial network，GAN）当做一种数据增强的技术来学习。在文献[19]中通过平滑相机风格的差异增强数据，扩充训练集的样本数量。戴臣超等[20]利用GAN扩充数据集并采用重排序的方法提高行人重识别的检索精度。

最近也有一些研究者专注于无监督视频行人重识别。比如Ma等[21]提出了一种基于自适应动态时间规整的无监督视频匹配方法，以选择更多的判别帧，提高序列测量的准确性。Ye等[22]提出了一种动态图匹配的方法来挖掘相机标签，以迭代的方式学习一种判别性的距离度量模型。Liu等[23]通过一种逐步度量学习方法来估计视频轨迹标签，但是它需要严格的视频过滤，以获得每个相机中每个身份的轨迹来初始化模型。上述这些方法均是需要对数据集做一些非常有用的注释，如文献[24]中所说，这些方法实际上是单样本方法。与这些方法不同，本文的工作是完全无监督的设置，没有使用任何身份标签。

1.2 聚类分析

聚类是机器学习中一个传统的无监督学习方法，该方法是比较不同特征之间的距离，然后将相似的特征看成同一身份。随着深度学习的不断发展，一些研究者[25]开始将聚类分析和深度学习相结合，来弥补传统方法的不足。Fan等[14]提出了一种结合K-均值聚类的渐进无监督学习方法。首先用外部有标记的数据初始化模型，之后，根据其可信度（定义为与聚类中心的距离）逐步选择未标记的数据进行训练。不过这种方法的前提是要知道身份类别的具体数量，在实际场景中没有很好的扩展性。对于完全无监督的行人重识别，Lin等[17]提出了一个从底向上的聚类框架，该框架根据预先定义的最小距离准则分层组合聚类，然后通过聚类结果微调模型，继续合并直至模型收敛，但是这种方法由于早期一些错误聚类导致后期识别率会急剧下降。在这基础之上，Ding等[26]提出了基于离散度的距离准则进行分层聚类。本文方法结合了上面两种方法的优势，同时改进了特征存储器的更新策略，进一步提高了无监督视频行人重识别的性能。

2 基于多样性约束和离散度的分层聚类

本文方法的整体框架如图2所示，可以分成3步：首先将无标签的样本送入网络训练提取视频特征，其中动态分类器会存储每类的中心特征，每次训练时会线性动态更新；每次训练完后会将提取的特征通过聚类准则进行聚类，从而产生新的类别ID；最后将更新后的数据重新训练。通过每次在迭代中微调模型和更新聚类信息，能逐渐提高类别标签的质量和模型的性能。

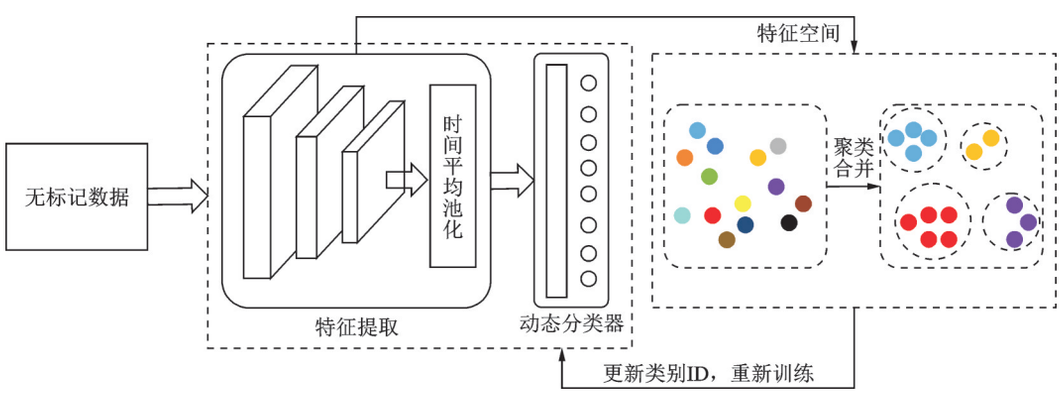

图2 整体框架图

Fig.2 Overall framework

2.1 特征存储的动态分类器

本文方法的目标是尽量将相似的样本聚成一类。通常可以使用对比损失或者三元组损失来优化距离，但是当数据集很大时，这些损失的效果并不是很好，因此提出在一个分类的框架下优化距离，开始时将每个样本看成一个单独的类进行初始化。然而对于这种类别数较多且每次变化的情况，一般的分类器很难收敛。为了缓解这一问题，本文利用了一个动态分类器进行分类，其作用相当于一个特征存储器。

假设共有N个无标记的样本$\{x_1,x_2,\cdots,x_N\}$，$\hat{y}_i$表示其类别索引$\{\hat{y}_i=i|1\leqslant i\leqslant N\}$，特征提取后，每个样本会变成一个$D$维的向量。特征存储器$\boldsymbol{M}\in\boldsymbol{R}^{N\times D}$存储每一类的中心特征，然后每次训练迭代后会自动更新。定义样本x属于第i类的概率为

$$p(i|x)=\frac{\exp(\boldsymbol{M}^{\mathrm{T}}[i]v/\tau)}{\sum_{j=1}^{C}\exp(\boldsymbol{M}^{\mathrm{T}}[j]v/\tau)} \tag{1}$$

式中：$v=\dfrac{\phi(\theta;x)}{\|\phi(\theta;x)\|}$，为提取的视频序列特征；$\boldsymbol{M}[j]$表示特征存储器$\boldsymbol{M}$的第$j$列；$C$是当前阶段类别的数量，在第一次训练时$C=N$，在后面的阶段相似的样本会逐步合并为一类，然后$C$会逐渐减少；$\tau$为平衡分布的温度参数，控制每个类上的概率分布。在前向传播时，通过$\boldsymbol{M}^{\mathrm{T}}\cdot v_i$计算样本$x_i$和其他样本之间的相似性，在反向传播时，更新$\boldsymbol{M}$为$\boldsymbol{M}[i]\leftarrow\alpha\boldsymbol{M}[i]+(1-\alpha)v_i$，超参数$\alpha\in[0,1]$表示特征存储器$\boldsymbol{M}$的更新率。

本文没有固定α的值，而是随着合并步骤线性增长。由于$\boldsymbol{M}$在训练开始时不够可靠，需要一个较小的α来更新$\boldsymbol{M}$。到了合并后期阶段，$\boldsymbol{M}$逐渐变得有判别性，并且$\boldsymbol{M}$也更加稳定。因此，此时使用较大的α来减慢更新速度。最终的损失定义为

$$L=-\log(p(\hat{y}_i|x_i)) \tag{2}$$

2.2 聚类准则

在聚类过程中，如何度量特征空间中类别之间

的特征距离决定着聚类效果的好坏。本文采用一种基于离散度[26]的度量方法，当度量类内关系时，离散度反映的是类内样本的紧凑程度，而当度量类间关系时，离散度反映的是两个类别之间的分离程度，同时基于类内样本数量的考虑，加入了多样性约束以提高聚类的准确性。

给定一个类 C，定义类内离散度为平均样本对距离

$$d(C)=\frac{1}{n}\sum_{i,j\in C}\mathrm{dist}(C_i,C_j) \tag{3}$$

式中：n 表示类别 C 中的样本个数；dist(•)为欧式距离。类内离散度低的可以优先考虑合并，这样能防止合并之后的类内离散度过高，提高聚类的有效性。

考虑类与类之间的差异性，定义类间离散度为

$$d(C_a,C_b)=\frac{1}{n_a n_b}\sum_{i\in C_a j\in C_b}\mathrm{dist}(C_{a_i},C_{b_j}) \tag{4}$$

类间离散度越低，表示两类越相似，越容易被合并成一类。

基于类内间离散度的聚类准则可以定义为

$$D=d_{ab}+\lambda(d_a+d_b) \tag{5}$$

式中：λ 为平衡参数；d_{ab} 为类 C_a 和类 C_b 之间的离散度，用来衡量两类之间的差异。因为同一身份的特征在特征空间中会比较靠近，所以离散度低的可以被认为是同一身份从而进行合并。d_a 和 d_b 为类 C_a 和类 C_b 的类内离散度，可以防止类内离散度高的类被合并，从而保持每类的平衡，提高聚类效果。

随着聚类的进行，类别数越来越少，类中的样本数越来越多。尽管不知道每一类中具体的样本数量，但可以假设样本是均匀分布在各个类别中的，这就说明一个类相比于其他类不会包含过多的样本，这种特性称之为多样性[17]。为了避免一些外观相似但身份不同的样本聚在一起，在聚类准则中加入了一项多样性约束，即有

$$d_{\mathrm{diversity}}=|a|+|b| \tag{6}$$

式中 $|a|$ 和 $|b|$ 表示类 C_a 和类 C_b 中的样本数量。在类内间离散度差不多的情况下，该约束能优先合并包含少量样本的类，所以最终聚类准则可以表示为

$$D=d_{ab}+\lambda_1(d_a+d_b)+\lambda_2 d_{\mathrm{diversity}} \tag{7}$$

式中 λ_1 和 λ_2 分别表示平衡类内间离散度和多样性约束的参数。在聚类过程中，该准则能够防止类内离散度高的或者样本数量多的类别被合并，减少错误聚类。

2.3　网络动态更新

本文方法会迭代训练网络并进行聚类，整个更新过程的算法流程如算法 1 所示，刚开始类别数目初始化为训练样本的个数，然后每次迭代会根据式(7)的聚类准则合并 m 个类别，$m=N\times p_m$，其中 $p_m\in(0,1)$ 用来控制聚类合并的速度，在 t 次合并后，类别数会减少为 $C=N-t\times m$。每次聚类合并后，更新样本所属的类别标签，然后重新训练，当类别数量 C 小于要合并的数量 m 时，停止迭代。

算法 1

输入：无标记数据 $X=\{x_i\}_{i=1}^{N}$，合并速率 p_m，平衡超参数 λ_1 和 λ_2，原始 CNN 模型 $\phi(\cdot;\theta_0)$，温度参数 τ，特征存储器更新率 α，合并次数 step=1

输出：优化的模型 $\phi(\cdot;\theta)$

(1) 初始化：类别标签 $Y=\{y_i=i\}_{i=1}^{N}$，类别数量 $C=N$，合并的类别数量 $m=C\times p_m$
(2) while $C>m$ do
(3)　用 X 和 Y 训练模型
(4)　通过式(7)计算类内间离散度和多样性约束进行聚类，根据计算结果选择 m 类进行合并，合并后类别数 $C\leftarrow C-m$
(5)　根据聚类结果更新类别标签 Y，$Y=\{y_i=j,\forall x_i\in C_j\}_{i=1}^{N}$
(6)　更新式(1)中的特征存储器 $\boldsymbol{M}$ 及其更新率 α，$\alpha=\mathrm{step}\times(\alpha\div(1/p_m))$，$\boldsymbol{M}[i]\leftarrow\alpha\boldsymbol{M}[i]+(1-\alpha)v_i$
(7)　进行下一次训练，step $\leftarrow$ step + 1
(8)　在验证集上评估性能
(9)　if $\mathrm{mAP}_i>\mathrm{mAP}_{\mathrm{best}}$ then
(10)　　$\mathrm{mAP}_{\mathrm{best}}=\mathrm{mAP}_i$
(11)　　优化的模型为 $\phi(\cdot;\theta)$
(12)　end if
(13) end while

3　实验分析

3.1　数据集和评价指标

MARS 数据集[3]是目前用于行人重识别任务的最大视频数据集，在清华校园中拍摄。它一共包含 1 261 个行人的 17 503 个视频序列和 3 248 个干扰视频序列，其中 625 个身份用于训练，636 个身份用于测试，训练集中的每个身份平均有 13 个视频序列。

DukeMTMC-VideoReID[24]是 DukeMTMC[27]的一个子集，是一个较新的基于视频的行人重识别数据集。它由 702 个训练身份、702 个测试身份和 408 个干扰身份组成。训练视频序列有 2 196 个，测试视频序列有 2 636 个。

本文使用 rank-k 和 mAP 评估方法的性能，rank-k 表示在排名前 k 个列表中正确匹配的概率，反映的是检索精度，平均精度均值 mAP(mean val-

ue of average precision，mAP）反映的是召回率。

3.2 实验细节

在本文实验中，采用在ImageNet上预先训练好的Resnet-50模型作为骨干网络提取特征。最后一层分类层随着式(1)中类别数目的改变而改变，将一个视频序列全部帧的平均特征作为视频序列特征用于聚类和最后的评估。对于模型训练，在第一阶段设置训练批次为30，后面阶段设置为3进行微调，采用随机梯度下降优化模型，此外设置初始学习率为0.1，15个迭代轮次（epoch）后衰减为0.01，批次数设为64，视频序列包含的图片帧数设为8。温度参数τ设置为0.1，在聚类阶段合并参数p_m设置为0.05，式(7)中λ_1和λ_2分别设置为0.06和0.003。

3.3 聚类准则的有效性分析

从表1可以看出类内间离散度和多样性约束的有效性，精度得到了明显提升，其中λ_1表示加入类内离散度，λ_2表示加入多样性约束。当仅考虑类间离散度时（baseline），Mars和Duke上的rank-1分别是60.8%和70.8%，mAP分别是39.2%和63.6%；加上类内离散度后，rank-1分别提升了0.9%和0.6%，因为它能够防止高离散的类被合并，从而保持类别平衡。加上多样性约束后，性能得到了进一步的提升，因为它能够优先合并数量较小的类别，最终rank-1达到了63.2%和75.4%，mAP达到了41.3%和67.4%，这表明类内间离散度和多样性约束能够提高聚类的准确性，提升视频行人重识别的匹配精度。

表1 在Mars和DukeMTMC-VideoReID上不同设置的性能比较

Table 1 Performance comparison of different settings on Mars and DukeMTMC-VideoReID

方法	Mars		DukeMTMC-VideoReID	
	rank-1	mAP	rank-1	mAP
baseline	60.8	39.2	70.8	63.6
baseline+λ_1	61.9	40.1	71.4	64
baseline+λ_1+λ_2	63.2	41.3	75.4	67.4

本文还对整个学习过程中聚类迭代的性能变化进行了分析，以研究其鲁棒性。图3(a)，(b)显示了在Mars和DukMTMC-VideoReID数据集上迭代次数对rank-1和mAP性能的影响，从图中可以看出两个数据集的rank-1精度和mAP都是随着聚类合并次数的增加而增加，最后随着类别数越来越少，增长幅度会变缓，逐渐趋于平稳。这表明了本文提出的聚类准则的有效性。

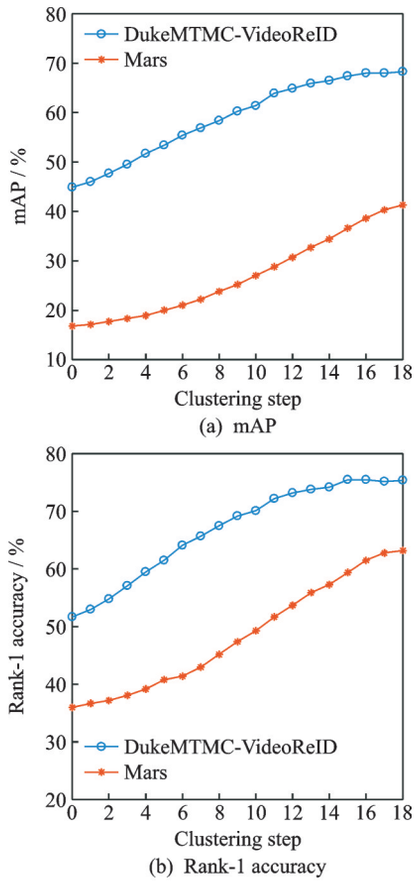

图3 在Mars和DukeMTMC-VideoReID数据集上迭代次数对rank-1和mAP性能的影响

Fig.3 Rank-1 and mAP performance with respect to clustering steps on Mars and DukeMTMC-VideoReID datasets

3.4 参数分析

首先分析本文方法中4个重要的超参数，平衡类内离散度的参数λ_1、平衡多样性约束的参数λ_2、特征存储器的更新率α和温度参数τ。改变一个参数的值，并保持其他参数不变进行实验分析。

在图4(a)，(b)中，保持λ_2不变，改变式(7)中λ_1的不同值，以研究其对rank-1和mAP精度的影响。实验发现当$\lambda_1=0.06$时，在两个数据集上的效果最好，过大或过小的值都会导致精度下降。

在图5(a)，(b)中，比较不同λ_2的值（保持$\lambda_1=0.06$）对结果的影响。当λ_2为0.003时，效果最好：Mars上的rank-1精度提升到了63.2%，mAP提升到了41.3%，Duke上的rank-1精度提升到了75.4%，mAP提升到了67.4%。如果λ_2的权重继续变大，rank-1和mAP的精度就会降低，表示太大的多样性约束可能会产生消极影响。

为了验证特征存储器中更新速率α的影响，还对比了线性增长的α和固定值α下的精度。从表2的结果可发现，线性变化α的值比固定α时精度更高，因为刚开始特征存储器$\boldsymbol{M}$是不可靠的，需要较

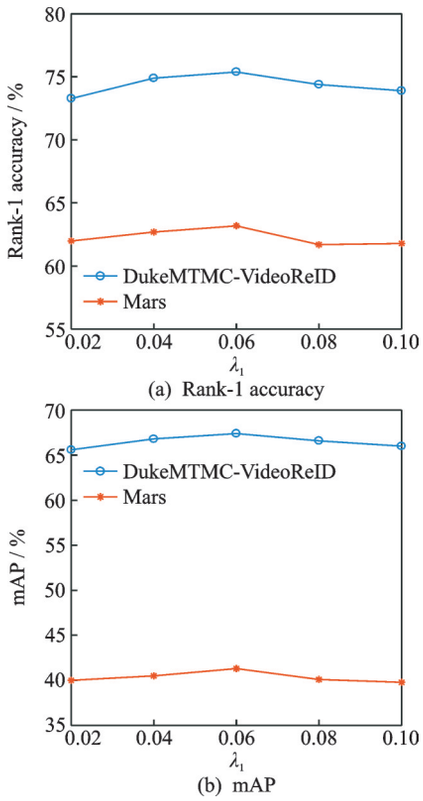

图 4　在 Mars 和 DukeMTMC-VideoReID 数据集上 λ_1 对 rank-1 和 mAP 性能的影响

Fig.4　Rank-1 and mAP performance with respect to λ_1 on Mars and DukeMTMC-VideoReID datasets

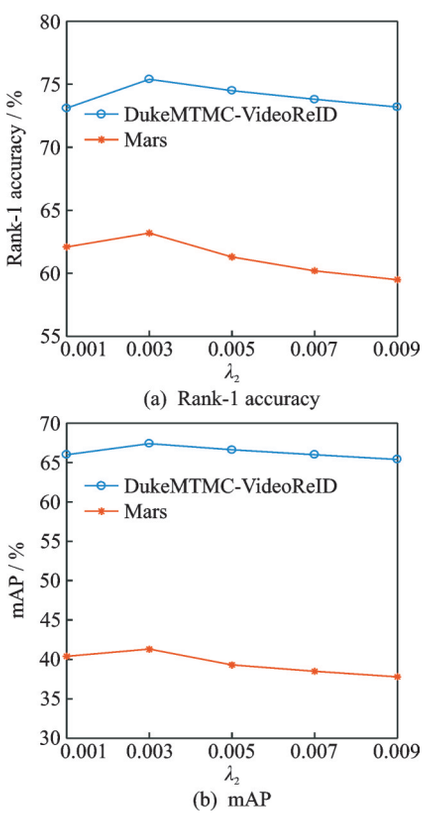

图 5　在 Mars 和 DukeMTMC-VideoReID 数据集上 λ_2 对 rank-1 和 mAP 性能的影响

Fig.5　Rank-1 and mAP performance with respect to λ_2 on Mars and DukeMTMC-VideoReID datasets

表 2　固定 α 和线性变化 α 性能的比较

Table 2　Performance comparison of a fixed α and a linear α　%

α	Mars		DukeMTMC-VideoReID	
	rank-1	mAP	rank-1	mAP
固定的 α	61.3	39.5	74.4	67.2
变化的 α	63.2	41.3	75.4	67.4

小的值进行更新，到后面阶段逐渐趋于稳定时就需要一个稍微大的值。对于 Mars 数据集，设置 α 线性增长到为 0.5 时效果最好，对于 DukeMTMC-VideoReID 数据集，设置 α 线性增长到为 0.4。

表 3 分析了式(1)中温度参数 τ 对精度的影响。不合适的值会导致网络不收敛，例如 τ=0.3。当 τ=0.1 时，可以达到最好的结果。

表 3　式(1)中不同 τ 值的评估

Table 3　Evaluation of different τ values in Eq. (1)　%

τ	DukeMTMC-VideoReID	
	rank-1	mAP
0.05	72.6	63.6
0.08	74.4	66.8
0.1	75.4	67.4
0.3	33.9	25.5

3.5　与现有方法对比

将本文方法与现有的一些无监督方法在 Mars 和 DukeMTMC-VideoReID 两个大型视频数据集上进行了比较，如表 4、5 所示。结果表明，本文的

表 4　在 Mars 数据集上的性能比较

Table 4　Performance comparison on Mars dataset　%

方法	Label	rank-1	rank-5	rank-10	mAP
OIM[28]	无	33.7	48.1	54.8	13.5
DGM+IDE[29]	单样本	36.8	54	—	16.8
Stepwise[23]	单样本	41.2	55.5	—	19.6
RACE[30]	单样本	43.2	57.1	62.1	24.5
DAL[31]	单样本	49.3	65.9	72.2	23.0
EUG[32]	单样本	62.6	74.9	—	42.4
BUC[17]	无	55.1	68.3	72.8	29.4
本文方法	无	63.2	76	79.9	41.3

表 5　在 DukeMTMC-VideoReID 数据集上的性能比较

Table 5　Performance comparison on DukeMTMC-VideoReID dataset　%

方法	Label	rank-1	rank-5	rank-10	mAP
OIM[28]	无	51.1	70.5	76.2	43.8
DGM+IDE[29]	单样本	42.3	57.9	69.3	33.6
Stepwise[23]	单样本	56.2	70.3	79.2	46.7
EUG[32]	单样本	72.7	84.1	—	63.2
BUC[17]	无	74.8	86.8	89.7	66.7
本文方法	无	75.4	87.7	90.3	67.4

方法优于目前一些无监督视频行人重识别方法，其中label列中“单样本[22]”是指使用了一些身份标签初始化模型，并不是完全无监督的，而label列中“无”指的是不使用身份标签，是完全无监督的。在Mars数据集上，比传统方法DGM+IDE[27]和Stepwise[21]，本文方法在rank-1上高了约20%，相比完全无监督的深度方法BUC[15]，本文方法在rank-1精度和mAP上分别高了8.1%和11.9%，同时相比单样本的深度方法EUG[30]，本文方法也有竞争力。在DukeMTMC-VideoReID数据集上，比BUC[15]在rank-1精度和mAP上分别高了0.6%和0.7%，同时还要优于单样本的EUG[30]方法，在rank-1精度和mAP上分别高了2.7%和4.2%。

4 结 论

本文提出一种基于多样性约束和离散度的分层聚类方法来解决深度无监督的视频行人重识别问题。首先使用卷积神经网络提取视频序列特征，然后使用由类内间离散度和多样性约束构成的聚类准则逐步对类别进行合并，最后在训练过程中利用一个具有特征存储功能的分类器优化模型。在两个大型视频数据集Mars和DukeMTMC-VideoReID上的实验表明，本文方法和目前先进的算法相比，在提升无监督视频行人重识别的性能上有一定的优越性。但本文提出的方法只是采用平均池化方法提取了视频序列的平均特征，并不具有判别性。因此，如何学习更有效的特征来提高无监督视频行人重识别的精度是进一步研究的目标。

参考文献：

[1] ZHENG Liang, YANG Yi, HAUPTMANN A G. Person re-identification: Past, present and future[EB/OL]. (2016-10-10)[2019-12-31]. https://arxiv.org/abs/1610.02984.pdf.

[2] 丁宗元，王洪元，陈付华，等.基于距离中心化与投影向量学习的行人重识别[J].计算机研究与发展，2017，54(8)：1785-1794.
DING Zongyuan, WANG Hongyuan, CHEN Fuhua, et al. Pedestrian weight recognition based on distance centralization and projection vector learning[J]. Computer Research and Development, 2017, 54(8): 1785-1794.

[3] ZHENG Liang, BIE Zhi, SUN Yifan, et al. Mars: A video benchmark for large-scale person re-identification[C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 868-884.

[4] GU Xinqian, MA Bingpeng, CHANG Hong, et al. Temporal knowledge propagation for image-to-video person re-identification[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2019: 9647-9656.

[5] YANG Jinrui, ZHENG Weishi, YANG Qize, et al. Spatial-temporal graph convolutional network for video-based person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 3289-3299.

[6] FARENZENA M, BAZZANI L, PERINA A, et al. Person re-identification by symmetry-driven accumulation of local features[C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos, USA: IEEE Computer Society, 2010: 2360-2367.

[7] LIAO Shengcai, HU Yang, ZHU Xiangyu, et al. Person re-identification by local maximal occurrence representation and metric learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2015: 2197-2206.

[8] LISANTI G, MASI I, BAGDANOV A D, et al. Person re-identification by iterative re-weighted sparse ranking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(8): 1629-1642.

[9] ZHAO Rui, OUYANG Wanli, WANG Xiaogang. Unsupervised salience learning for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2013: 3586-3593.

[10] WANG Hanxiao, GONG Shaogang, XIANG Tao. Unsupervised learning of generative topic saliency for person re-identification[C]//Proceedings of the British Machine Vision Conference. Cambridge, UK: BMVA Press, 2014.

[11] WANG Hongyuan, DING Zongyuan, ZHANG Ji, et al. Person reidentification by semisupervised dictionary rectification learning with retraining module[J]. Journal of Electronic Imaging, 2018, 27(4): 043043-1-043043-9.

[12] YAN Caixia, LUO Minnan, LIU Wenhe, et al. Robust dictionary learning with graph regularization for unsupervised person re-identification[J]. Multimedia Tools and Applications, 2018, 77(3): 3553-3577.

[13] NI Tongguang, GU Xiaoqing, WANG Hongyuan, et al. Discirminative deep trasfer metric learning for cross-scenario person re-identification[J]. Journal of Electronic Imaging, 2018, 27(4): 1-10.

[14] FAN Hehe, ZHENG Liang, YAN Chenggang, et al.

Unsupervised person re-identification: Clustering and fine-tuning[J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2018, 14(4): 1-18.

[15] ZHONG Zhun, ZHENG Liang, LUO Zhiming, et al. Invariance matters: Exemplar memory for domain adaptive person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2019: 598-607.

[16] YU Hongxing, WU Ancong, ZHENG Weishi. Cross-view asymmetric metric learning for unsupervised person re-identification[C]//Proceedings of the IEEE International Conference on Computer. Piscataway, USA: IEEE, 2017: 994-1002.

[17] LIN Yutian, DONG Xuanyi, ZHENG Liang, et al. A bottom-up clustering approach to unsupervised person re-identification[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Menlo Park, USA: AAAI, 2019: 8738-8745.

[18] KODIROV E, XIANG Tao, GONG Shaogang. Dictionary learning with iterative laplacian regularisation for unsupervised person re-identification[C]//Proceedings of the British Machine Vision Conference. Cambridge, UK: BMVA Press, 2015.

[19] ZHONG Zhun, ZHENG Liang, ZHENG Zhedong, et al. Camera style adaptation for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2018: 5157-5166.

[20] 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.
DAI Chenchao, WANG Hongyuan, NI Tongguang, et al. Person re-identification based on deep convolutional generative adversarial network and expanded neighbor reranking[J]. Computer Research and Development, 2019, 56(8): 1632-1641.

[21] MA Xiaolong, ZHU Xiatian, GONG Shaogang, et al. Person re-identification by unsupervised video matching [J]. Pattern Recognition, 2017, 65: 197-210.

[22] YE Mang, MA A J, ZHENG Liang, et al. Dynamic label graph matching for unsupervised video re-identification[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2017: 5142-5150.

[23] LIU Zimo, WANG Dong, LU Huchuan. Stepwise metric promotion for unsupervised video person re-identification[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2017: 2429-2438.

[24] 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.

[25] CARON M, BOJANOWSKI P, JOULIN A, et al. Deep clustering for unsupervised learning of visual features[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 132-149.

[26] DING Guodong, KHAN S, TANG Zhenmin, et al. Towards better VALIDITY: Dispersion based clustering for unsupervised person re-identification[EB/OL]. (2019-06-14)[2019-12-31]. https://arxiv. org/pdf/1906.01308.pdf.

[27] RISTANI E, SOLERA F, ZOU R, et al. Performance measures and a dataset for multi-target, multi-camera tracking[C]//Proceedings of the European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 17-35.

[28] XIAO Tong, LI Shuang, WANG Bochao, et al. Joint detection and identification feature learning for person search[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2017: 3415-3424.

[29] YE Mang, LI Jiawei, MA A J, et al. Dynamic graph co-matching for unsupervised video-based person re-identification[J]. IEEE Transactions on Image Processing, 2019, 28(6): 2976-2990.

[30] YE Mang, LAN Xiangyuan, YUEN P C. Robust anchor embedding for unsupervised video person re-identification in the wild[C]//Proceedings of the European Conference on Computer Vision. Berlin, Germany: Springer, 2018: 170-186.

[31] CHEN Yanbei, ZHU Xiatian, GONG Shaogang. Deep association learning for unsupervised video person re-identification [EB/OL]. (2018-08-22) [2019-12-31]. https://arxiv.org/pdf/1808.07301.pdf.

[32] WU Yu, LIN Yutian, DONG Xuanyi, et al. Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2018: 5177-5186.

（编辑：张蓓）