软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software, [doi: 10.13328/j.cnki.jos.006108]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

基于近邻中心迭代策略的单标注样本视频行人重识别*

张云鹏[1], 王洪元[1], 张继[1], 陈莉[1], 吴琳钰[1], 顾嘉晖[1], 陈强[2]

[1](常州大学 信息科学与工程学院, 江苏 常州 213000)

[2](社会安全信息感知与系统工业和信息化部重点实验室(南京理工大学),江苏 南京 210094)

通讯作者: 王洪元, E-mail: hywang@cczu.edu.cn

摘 要: 为解决视频行人重识别数据集标注困难的问题,本文提出了基于单标注样本视频行人重识别的近邻中心迭代策略,该策略逐步利用伪标签视频片段迭代更新网络结构,以获得最佳的模型.针对预测无标签视频片段的伪标签准确率低的问题,提出了一个新的标签评估方法:每次训练后,将所选取的伪标签视频片段和有标签视频片段特征中每个类的中心点作为下一次训练中预测伪标签的度量中心点;同时提出了一个基于交叉熵损失和在线实例匹配损失的损失控制策略,使得训练过程更加稳定,无标签数据的伪标签预测准确率更高.在 MARS, DukeMTMC-VideoReID 这两个大型数据集上的实验验证了本文方法相比于最新的先进方法在性能上得到一个非常好的提升.

关键词: 视频行人重识别;近邻中心迭代策略;标签评估方法;单标注;损失控制策略

One-Shot Video-Based Person Re-Identification Based on Neighborhood Center Iteration Strategy

ZHANG Yun-Peng[1], WANG Hong-Yuan[1], ZHANG Ji[1], CHEN Li[1], WU Lin-Yu[1], GU Jia-Hui[1], CHEN Qiang[2]

[1] (School of Information Science and Engineering, Changzhou University, Changzhou 213000, China)

[2] (Key Laboratory of Information Perception and Systems for Public Security of MIIT (Nanjing University of Science and Technology), Nanjing 210094, China)

Abstract: In order to solve the problem of labeling difficulty in video-based person re-identification dataset, a neighborhood center iteration strategy based on one-shot video-based person re-identification is proposed in this paper, which gradually optimizes the network by using pseudo-labeled tracklets to obtain the best model. Aiming at the problem that the accuracy of predicting pseudo labels of unlabeled tracklets is low, a novel label evaluation method is proposed. After each training, the center points of each class in the features of the selected pseudo-labeled tracklets and labeled tracklets are used as the measurement center points for predicting the pseudo labels in the next training. At the same time, a loss control strategy based on cross entropy loss and online instance matching loss is proposed in this paper, which makes the training process more stable and the accuracy of the pseudo labels higher. Experiments are implemented on two large datasets: MARS and DukeMTMC-VideoReID, which demonstrate that our methods outperform the state-of-the-art methods.

Key words: Video-based person re-identification; Neighborhood center iteration strategy; Label evaluation method; One-shot; Loss control strategy

行人重识别（person re-identification）旨在解决跨摄像机检索匹配行人图像或视频的问题,主要有两种方法:基于图像的行人重识别和基于视频的行人重识别.前者利用行人图像匹配同一行人在不同摄像机视图下的行人图像[1-5],后者直接利用信息更加丰富的行人视频片段匹配同一行人在不同的摄像机视图下的行人视频

* 基金项目: 国家自然科学基金项目(61976028, 61572085, 61806026, 61502058);江苏省自然科学基金项目(BK20180956); 社会安全信息感知与系统工业和信息化部重点实验室(南京理工大学)创新基金(202004).

Fund items: the National Natural Science Foundation of China (61976028, 61572085, 61806026, 61502058); the Natural Science Foundation of Jiangsu Province (BK20180956); Key Laboratory Foundation of Information Perception and Systems for Public Security of MIIT (Nanjing University of Science and Technology) (202004)

收稿时间: 2020-01-15; 修改时间: 2020-04-19; 采用时间: 2020-05-21; jos 在线出版时间: 2020-10-12

[6-8].而基于视频的行人重识别与现实世界的应用更为贴切,从而在近期引起了极大的关注.现有的基于视频的行人重识别的方法主要依赖于完全标注的视频片段.由于标注数据的成本过于巨大,因此研究依赖少量标注的半监督视频行人重识别具有极大的应用价值.

单标注样本学习是半监督学习的一种.单标注样本视频行人重识别的关键在于如何准确的对大量无标签视频片段进行标签估计[9-11].其常见的方法是在迭代过程中先将数据嵌入特征空间,以每个行人唯一的有标签视频片段特征作为固定度量中心,无标签视频片段根据与固定度量中心的距离为其分配伪标签.初始有标签数据和每次选定的伪标签数据合并作为新的数据集,进行下一次训练.如图 1 所示（图中共有 3 类数据,实心圆表示无标签数据,颜色表示各自真实的分类,空心圆表示该类的初始有标签数据特征,虚线圆内与空心圆颜色不同的点则表示伪标签标注错误的数据,以空心圆为中心选取一定比例的伪标签数据用于下一次训练）,随着选取用作下次训练伪标签数据的增加,标注错误的伪标签的数量也极大地增加,因此以上这种固定度量中心的方法是有缺陷的.在这种情况下,当有标签数据在特征空间中处于类的边缘或者远离类的中心,随着选取伪标签数据的增加,将会得到大量的不准确的伪标签数据,而过多的不可靠的伪标签数据在迭代过程中将会严重影响模型的性能.

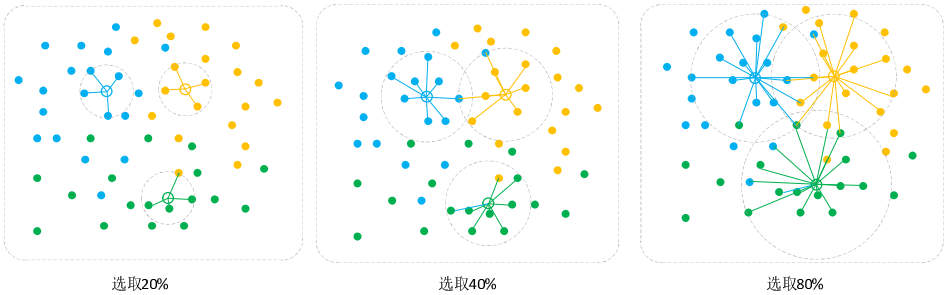

Fig.1 Common label evaluation methods

图 1 常见标签评估方式

为了在每轮训练过程中得到更多的正确伪标签视频片段用于下一次训练,本文提出了一个新的策略:近邻中心迭代策略(Neighborhood Center Iteration, NCI).每一次迭代训练后,在特征空间中找出所选取的伪标签视频片段和有标签视频片段特征每一类的中心点,作为其下一轮预测无标签视频片段的伪标签的度量中心点.随着选取伪标签视频片段的数量逐步增加,本文的策略能更加准确地加入复杂的无标签视频片段用于下一次训练.此外,传统的行人重识别特征学习主要依赖于三重损失[12]等函数,其计算量大,因此本文提出一个损失控制策略,联合训练交叉熵损失（CrossEntropy Loss）和在线实例匹配损失[13]（Online Instance Matching Loss,OIM Loss）,既能有效地缩小类内距离又能使得训练过程更加的稳定高效.

本文的主要工作如下:（1）提出一个新的训练策略 NCI,该策略中提出的新标签评估准则能有效地提升无标签视频片段的伪标签预测准确率和最终算法的精度.（2）提出一个新的损失控制策略,联合训练 CrossEntropy Loss 和 OIM Loss,使得训练过程更加的稳定.相对于最新的半监督和单标注学习方法,本文的方法在 MARS 和 DukeMTMC-VideoReID 两个大型数据集上都有很好的性能提升.

1 相关的研究工作

对于监督视频行人重识别,新出现了许多基于深度学习的方法[14-18].如文献[14]将细化循环单元模块和时空线索聚合模块用于恢复缺失帧和利用上下文信息,从而获得行人视频片段的特征表示;文献[17]提出时空注意力感知学习方法,旨在视频序列的时空上关注视频中行人的重要部分,以解决行人图像质量因不同的时间空间区域变化而变化的问题;文献[18] 提出了判别聚合网络方法,直接聚合原始视频帧,且结合度量学习和对抗学习的思想生成更多的判别图像,减少每个视频处理的图像帧数,误导性信息的低质量帧也可以得到很好的过滤和去噪.对于无监督的视频行人重识别,文献[13]中提出了半监督行人检测的 OIM Loss,它也可用于无监督的视

频行人重识别;文献[19]提出了一种自底向上聚类方法(Bottom-Up Clustering, BUC) 来联合优化 CNN 和无标签样本间的关系,并且在聚类过程中利用了一个多样性正则项来平和每个聚类的数据量.

以往的半监督行人重识别方法大多数是基于图像[20-23]行人重识别.近期出现了不少半监督视频行人重识别方法,如 Zhu 等[24]提出了一种基于半监督交叉视图投影的字典学习方法;也出现了一些单标注视频行人重识别任务的方法,如 Liu 等[10]用有标签的样本初始化模型,计算出与查询集样本最接近的 k 个样本并且删除其中的可疑样本,再将其余样本添加到训练集中,重复该过程直到算法收敛为止;Ye 等[11]提出了一种动态图匹配(Dynamic Graph Matching,DGM)方法,该方法迭代更新图和标签估计,以学习更好的特征空间;Wu 等[9]使用一个逐步利用无标签视频片段的策略(Exploit the Unknown Gradually,EUG),先用有标签视频片段初始化网络模型,再根据与有标签数据的距离将伪标签数据线性合并到训练集中进行后续的训练;文献[25]用了一个单标注样本渐进学习的方式(Progressive Learning,PL),将标签数据、伪标签数据和索引标签数据三个部分在迭代过程中联合训练模型.但是文献[10,11]中采用静态策略来确定每次训练所选择的伪标签数据的数量的方法是不合理的,因为初始模型可能不健壮,只有少数伪标签预测在初始阶段是可靠和准确的,如果选择与后期训练相同数量的数据,则不可避免地会出现更多错误的伪标签数据.而文献[9,25]中将有标签视频片段特征作为固定度量中心也会得到大量的不准确的伪标签数据.因此本文提出了近邻中心迭代策略,从一定程度上解决伪标签错误率低的问题.

2 近邻中心迭代策略

2.1 基本框架

本文将每个行人唯一有标签视频片段集合表示为 $\mathcal{L}=\{(x_1,y_1),...,(x_{n_l},y_{n_l})\}$,无标签的视频片段集合表示为 $\mathcal{U}=\{(x_{n_l+1}),...,(x_{n_l+n_u})\}$,其中 x_i, y_i 分别表示第 i 个视频片段和行人标签.因此有 $|\mathcal{L}|=n_l$ 和 $|\mathcal{U}|=n_u$,其中 $|\cdot|$ 表示集合内元素的个数. $s_i\in\{0,1\}$ 作为伪标签样本 x_i 选作下一次训练的选择指示器.

在迭代训练过程中,采用的是一种常见的渐进学习方式[9],每次训练选取一定比例可靠的伪标签视频片段用于下一次训练. $\mathcal{S}$ 表示选取下一次训练的伪标签数据的候选集

$$\mathcal{S}=\{(x_i,\hat{y}_i)|s_i=1,n_l+1\le i\le n_l+n_u\} \tag{1}$$

其中 $\hat{y}_i$ 表示第 i 个无标签视频片段的伪标签.

本文方法的具体框架如图 2 所示,采用 ResNet-50 结构的端到端模型作为特征提取网络,且在分类层前面加上了一个全连接层和一个时间平均池化层.对于每一个视频片段,当所有图片被提取为帧级特征后,时间平均池化层将所有的帧级特征合并作为视频片段的特征表示.

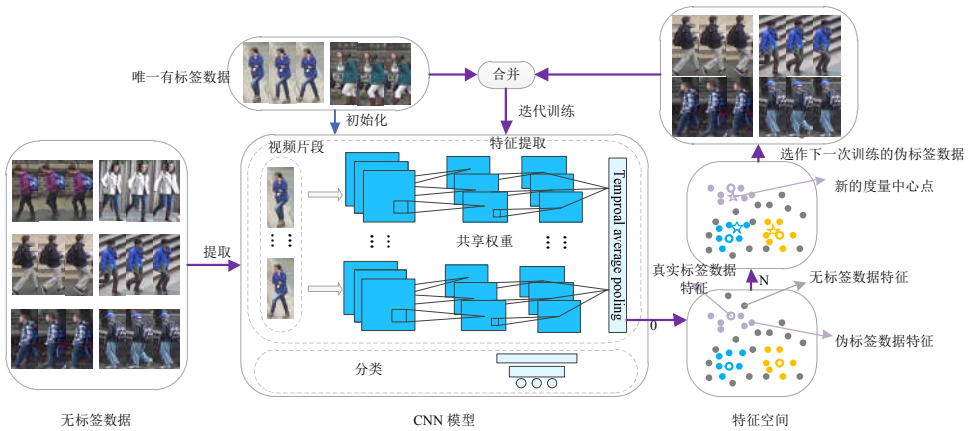

Fig.2 Overall framework of NCI strategy

图 2 NCI 策略整体框架

初始训练时使用唯一有标签视频片段集合 $\mathcal{L}$ 来初始化模型,再用训练好的模型提取 $\mathcal{U}$ 中无标签视频片段特征,每个无标签视频片段的伪标签由特征空间中距离最近的度量中心点的标签进行分配,然后产生每个无标签视频片段的选择指示器 s_i ,并根据公式(1)来得到候选集 $\mathcal{S}$.在之后的迭代中,每次候选集 $\mathcal{S}$ 和初始的标签数据 $\mathcal{L}$ 合并为新的数据集 $\mathcal{D}$, $\mathcal{D}=\mathcal{S}\cup\mathcal{L}$. $\mathcal{D}$ 则作为下一次训练用的训练集.且在训练过程中 $\mathcal{S}$ 随着训练次数的增加而不断地扩大.

2.2 标签评估标准

以往的标签评估方法[9,25]中,有标签数据作为固定度量中心在每轮训练中为最近的无标签数据进行伪标签分配.如图 1 所示,这一方法是有很大弊端的,原始有标签视频片段在特征空间内同类中的相对位置是固定的,且当原始有标签视频片段在特征空间中处于同类的边缘或者远离类中心的点时,每次训练会预测出更多错误的伪标签,随着选取伪标签数据 $\mathcal{S}$ 的增大（例如图中选取 80%）,选取到不可靠数据的概率变得更大.

针对这种情况,提出了一种新的标签评估标准.在迭代过程中,利用得到的可靠集合 $\mathcal{D}$ 中每个类的中心,作为下一次训练预测伪标签的度量中心点.具体来说,每次训练结束,训练完的模型提取无标签视频片段的特征并嵌入特征空间,此时无标签数据特征与上一次训练所得的集合 $\mathcal{D}$ 中每个类的中心（初次训练 $\mathcal{D}$ 中每个行人只有一个初始数据,则以此为类中心）依次计算距离,距离最近的类的标签则为该无标签视频片段的伪标签.然后无标签视频片段与为其分配伪标签的度量中心的距离排序,按比例选取距离较小并带有伪标签的无标签视频片段作为可靠伪标签数据候选集 $\mathcal{S}$,并与 $\mathcal{L}$ 合并为 $\mathcal{D}$ 作为下一次训练的数据集,依次迭代直至用完所有无标签视频片段.这样能够使得每次选取的度量中心更准确地反映出特征空间内每个类中的特征的集中趋势,能够更加接近类的真实中心,使得每次预测的伪标签更加准确.

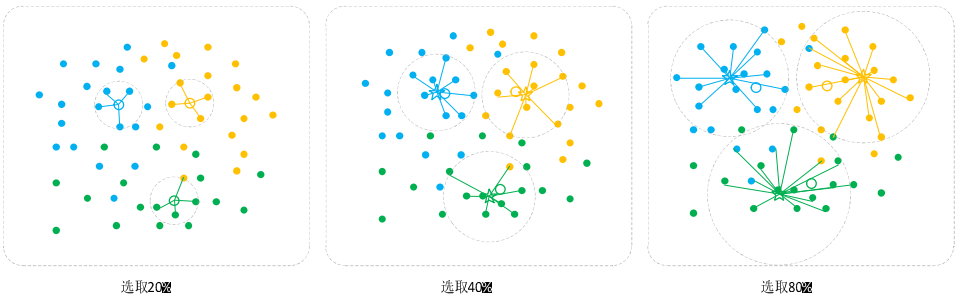

Fig.3 NCI label evaluation method

图 3 NCI 标签评估方式

如图 3 所示（图 3 共有 3 类,实心圆表示无标签数据特征,空心圆表示该类的初始有标签数据特征,五角星代表上一次训练所得集合 $\mathcal{D}$ 的类中心,虚线圆内与空心圆颜色不同的点则表示伪标签标注错误的数据,此时则以五角星为中心选取一定比例的伪标签数据用于下一次训练）,当初始训练后以唯一有标签样本为中心点选取 20%的数据,在之后训练中依次以新的中心（五角星）为度量中心点选取 40%、80%的数据,可以明显的看到,前一次迭代选取的伪标签数据与初始有标签数据合并之后产生的新的度量中心点,更加接近类的真实中心,而相比于图 1 预测出更多正确的伪标签.因此近邻中心迭代策略中的标签评估标准,能够极大地提高每次伪标签预测的准确率,进而提高最终结果.

数据样本的集中趋势描述有平均数、中位数等,本文分别用平均中心和中位数中心计算特征空间的样本中心.由于 MARS 数据集采样的摄像头较多且场景较为复杂,可能在特征空间中离群点较多,因此使用中位数中心更为合适.DukeMTMC-VideoReID 数据集场景相对简单,则使用平均中心更合适.用 R 表示 $\mathcal{D}$ 中所有类的中心的集合,其中平均中心公式可表示为

$$R_k = \frac{1}{N}\sum_{d_n \in \mathcal{D}_k} d_n \tag{2}$$

其中 R_k 表示第 k 类样本新的度量中心点, $\mathcal{D}_k$ 表示 $\mathcal{D}$ 中第 k 类样本的集合, N 为 $\mathcal{D}_k$ 中元素的个数.

2.3 动态抽样策略

由于前几次用于训练的数据较少,模型的性能较差,预测的无标签视频片段的伪标签可靠的数量较少,因此若前几次训练每次选取过多的伪标签数据会极大地影响最终的模型性能.因此本文采用了渐进的动态抽样策略.其中每个无标签视频片段与所有度量中心的距离的最小值可表示为

$$d(x_i) = \min_{R_k \in R} \left\| \phi(x_i) - R_k \right\|_2 \tag{3}$$

$x_i \in \mathcal{U}$, $R_k \in R$ 表示为新的度量中心点, $\phi(\cdot)$ 表示该无标签视频片段在特征空间中的特征.对于伪标签数据的选择,通过选择指示器 s_t 来将一定比例较小的 $d(x_i)$ 对应的无标签视频片段 x_i 作为可靠的伪标签数据采样到训练中

$$s_t = \arg\min_{\|s\|_0 = m_t} \sum_{i=nl+1}^{nl+nu} s_i d(x_i) \tag{4}$$

其中 m_t 表示当前轮次选取伪标签数据的数量.随着迭代次数 t 的增加,选取可靠伪标签数据的数量会逐步增加: $m_t = m_{t-1} + p \cdot n_u, p \in (0,1)$ 其中 p 表示迭代过程中选取伪标签数据数量的增长率.比较好的选择是将 p 设置为一个很小的值,这意味着 m_t 逐步增大,并且每一步的变化很小.这种设置随着迭代过程逐步优化,模型性能会非常稳定的提高,并最终获得令人满意的性能.

3 损失函数训练策略

常用的 OIM Loss 利用来自有标签行人视频数据的特征形成查询表,与批次样本之间进行距离比较,另外那些无标签视频片段可以被视为负样本,将它们的特征存储在循环队列中并进行比较.不仅适用于单标注视频行人重识别训练场景,并且相比于其它损失函数收敛的更快更稳定.OIM Loss 可以表示为

$$\mathcal{C} = X^{OIM} \cdot V^T \tag{6}$$

$$loss_{OIM} = -\frac{1}{N}\sum_i \log \frac{\exp(\mathcal{C}_i)}{\sum_j \exp(\mathcal{C}_{ij})} \tag{7}$$

其中 X^{OIM} 表示视频片段的特征矩阵, V 表示每个类代表性的特征, $\mathcal{C}$ 表示提取的特征 X 与每个类的余弦距离.而 CrossEntropy Loss 也是常用的损失函数,在深度训练中有着比较稳定和准确的效果.用 X^{Ce} 表示最终视频片段的特征矩阵,则 CrossEntropy Loss 可表示为

$$loss_{\mathrm{Ce}} = -\frac{1}{N}\sum_i \log \frac{\exp(X_i{}^{Ce})}{\sum_j \exp(X_{ij}{}^{Ce})} \tag{8}$$

基于以上两个损失函数,为了单标注视频行人重识别的训练过程更加稳定,模型性能更佳,本文提出了一个有效的损失函数训练策略,联合训练 OIM Loss 和 CrossEntropy Loss 两个损失函数

$$loss = \beta \cdot loss_{OIM} + (1-\beta) loss_{\mathrm{Ce}} \tag{9}$$

$$\beta=\begin{cases}0.5-0.5\cdot(p_{Ce}-p_{OIM}), & p_{Ce}\geq p_{OIM}\\ 0.5+0.5\cdot(p_{OIM}-p_{Ce}), & p_{Ce}<p_{OIM}\end{cases} \tag{10}$$

其中 p_{Ce} 和 p_{OIM} 表示训练过程中两个损失评估的精度,β 是一个可变参数,用于动态分配权重.损失函数的评估精度高则分配大一点的权重,评估精度低则分配小一些的权重,通过动态的调整训练权重,使得在训练过程中模型能够更加稳定,表现的更加鲁棒,无标签数据的伪标签精度更高.通过两个大型数据集上的实验也验证了本文的损失控制策略的有效性.

4 实验与分析

4.1 数据集

MARS[7]数据集是视频行人重识别任务中最大的数据集.数据集包含 1261 个行人,共有 17503 个视频片段和 3248 个干扰视频片段.其中 625 个行人用于训练,636 个行人用于测试.训练集中每个行人平均有 13 个视频片段,每个视频片段平均有 816 帧.

DukeMTMC-VideoReID[26]数据集包含 1812 个行人,共有 4832 个视频片段.并将行人分别划分为 702、702 和 408 份,分别用于训练、测试和干扰.总共 2196 个视频片段用于训练,以及 2636 个视频片段用于测试和干扰.每个视频片段平均有 168 帧.

本文使用累积匹配特征（Cumulative Matching Characteristic,CMC）曲线和平均准确率（Mean Average Precision,mAP）来评估每次迭代模型的性能,并且使用符号 $\mathcal{M}$ 表示最终预测无标签视频片段伪标签准确率.

4.2 实验设置

在两个数据集中,为每个行人随机选择摄像机 1 中的一个视频片段作为初始化有标签数据集 $\mathcal{L}$.如果摄像机 1 没有该行人,将在下一台摄像机中随机选择一个视频片段,以确保每个行人都有一个用于初始化的视频片段.

在实验中本文使用 ImageNet[27]预训练去掉最后的分类层的 ResNet50 作为 NCI 的初始模型.采用动量为 0.5 且权重衰减为 0.0005 的随机梯度下降（SGD）优化方法.整体学习率初始化为 0.1,并在最后 15 个周期衰减为 0.01.在用损失函数控制策略训练的时候,由于初始数据过少,本文使用 CrossEntropy Loss 来进行前几次迭代的训练,以获得稳定的伪标签数据,之后使用本文提出的损失函数控制策略,从而使得实验过程更加稳定,效果更好.

4.3 实验对比

4.3.1 参数分析

当训练循环到第 t 步,本文会选择 $t*p$ 比例的带有伪标签的无标签视频片段用作下一次的模型训练.其中增长率 p 的影响如表 1、表 2 所示,p 取 0.05 到 0.3 时,p 值越小 rank-1、mAP 的精度越高.且当 p=0.05 时,rank-1、mAP 和伪标签的精度最高,模型性能最好.如图 4 所示,当 p 取 0.05、0.10 和 0.20 时,前面几次迭代 3 张图曲线间的间隙不大,然而后面曲线间的间隙则越来越大,并且 p 取 0.05 时的曲线明显高于 0.10 和 0.20.原因是错误标签评估在迭代过程中会不断累积,选取伪标签越多错误的累积影响越大.因此增长率 p 扩大的越缓慢,选取的正确伪标签越多,从而模型精度 rank-1、mAP 越高.综合分析,p 值取小一些效果会更好.本文以下阐述以 p=0.05 和 p=0.1 的结果进行比较.

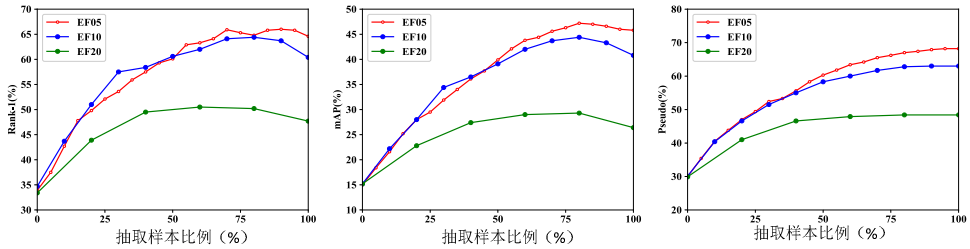

Fig.4 Results of different values of parameter p on the MARS dataset

图 4 参数 p 不同值在 MARS 数据集上的结果图

在选取特征空间的数据中心点时,本文使用了平均中心和中位数中心.结果如表 1 所示, p 取 0.05 到 0.3 时,在 MARS 数据集上中位数中心比平均中心伪标签精度明显更高.其中当 p=0.05 时,中位数中心比平均中心预测伪标签精度高 1.63%.当 $p=0.10$ 时,中位数中心比平均中心伪标签精度高 2.43%.而 p 取 0.05 到 0.3 时,在 DukeMTMC-VideoReID 数据集上平均中心比中位数中心伪标签精度明显更高.其中当 p=0.05 时,平均中心比中位数中心伪标签精度高 0.8%.当 $p=0.10$ 时,平均中心比中位数中心伪标签精度高 0.87%.因此本文实验选用中位数中心作为 MARS 数据集的标签评估方式,平均中心作为 DukeMTMC-VideoReID 数据集的标签评估方式.

Table 1 Comparison of center selection method correct rate (%)

表 1 中心选取方式正确率（%）的对比

	DukeMTMC-VideoReID		MARS	
Parameter	Mean	Median	Mean	Median
p=0.30	64.26	61.04	41.94	42.58
p=0.20	68.14	67.47	47.36	48.40
p=0.15	68.61	68.61	56.30	54.36
p=0.10	73.23	72.36	60.58	63.01
p=0.05	75.30	74.50	66.56	68.19

4.3.2 近邻中心迭代策略的有效性

如表 2、表 3 所示,表示 p 取 0.05 到 0.3 时,NCI 策略相比于 EUG 在 rank-1 accuracy(%)、mAP(%)、伪标签准确率 $\mathcal{M}$(%)有着全面性的提升.当两种方式均取 $p=0.10$ 时,在 DukeMTMC-VideoReID 数据集上,NCI 的 rank-1 精度提升 2.61%,mAP 精度提升 3.84%,伪标签的预测精度提升 1.61%.在 MARS 数据集上,NCI 的 rank-1 精度提升 2.78%,mAP 精度提升 6.12%,伪标签的预测精度提升 4.04%.均取 $p=0.05$ 时,在 DukeMTMC-VideoReID 数据集上,NCI 的 rank-1 精度提升 1.61%,mAP 精度提升 3.17%,伪标签的预测精度提升 1.13%.在 MARS 数据集上,NCI 的 rank-1 精度提升 1.93%,mAP 精度提升 3.35%,而伪标签的预测精度提升 1.97%.

综合以上分析能得出,增长率 p 取 0.05 到 0.3 时,无论是 rank-1、mAP 精度还是伪标签的准确率,均有了极大地提升.由此得出本文提出的 NCI 相比于最新的策略 EUG 有着全面的性能提升.

Table 2 Comparison of NCI and EUG results (%)

表 2 NCI 与 EUG 结果（%）对比

	DukeMTMC-VideoReID				MARS			
Methods	rank-1	rank-5	rank-20	mAP	rank-1	rank-5	rank-20	mAP
EUG[9](p=0.30)	63.82	78.64	87.04	54.57	42.77	56.51	67.17	21.12
EUG[9](p=0.20)	68.95	81.05	89.46	59.50	48.68	63.38	72.57	26.55

EUG[9](p=0.15)	69.08	81.19	88.88	59.21	52.32	64.29	73.08	29.56
EUG[9](p=0.10)	70.79	83.61	89.60	61.76	57.62	69.64	78.08	34.68
EUG[9](p=0.05)	**72.79**	**84.18**	**91.45**	**63.23**	**62.67**	**74.94**	**82.57**	**42.45**
NCI(p=0.30)	67.50	79.20	87.90	58.10	44.30	57.90	67.70	22.40
NCI(p=0.20)	71.20	83.80	91.00	62.80	47.70	60.80	70.00	26.40
NCI(p=0.15)	71.50	84.30	89.90	62.20	55.50	69.50	77.80	34.30
NCI(p=0.10)	73.40	86.80	93.20	65.60	60.40	76.00	84.30	40.80
NCI(p=0.05)	**74.40**	**88.50**	**93.40**	**66.40**	**64.60**	**78.10**	**84.40**	**45.80**

4.3.3 损失控制策略的有效性

表 3 是联合 NCI 和损失控制策略分别在 DukeMTMC-VideoReID 和 MARS 数据集上的实验结果,与 NCI 在 rank-1 accuracy(%)、mAP(%)、伪标签准确率 $\mathcal{M}$(%)的比较.以验证损失控制策略的有效性.如表 3 所示,NCI 和损失控制策略联合训练的结果与 NCI 进行比较可得,当均取 $p=0.10$ 时,DukeMTMC-VideoReID 数据集上在 rank-1 精度提升 6.1%,mAP 精度提升 7.5%,伪标签准确率提升 5.36%.在 MARS 数据集上在 rank-1 精度提升 0.7%,mAP 精度提升 0.6%,伪标签的准确率提升 0.51%.当 $p=0.05$ 时,DukeMTMC-VideoReID 数据集上 rank-1 精度提升 5.9%,mAP 精度提升 7.6%,伪标签的准确率提升 4.82%.在 MARS 数据集上 rank-1 精度提升 2%,mAP 精度提升 2.9%,伪标签的准确率提升 3.48%.

综合以上分析,本文提出的损失控制策略能有效地提升 NCI 的性能,最终提升模型的性能.同时表 3 在同等 p 值下的实验结果对比,能依次证明本文的 NCI 和损失控制策略提升效果明显.

Table 3 Comparison of Loss Control Strategy Results (%)

表 3 损失控制策略结果（%）的对比

	DukeMTMC-VideoReID					MARS				
Methods	rank-1	rank-5	rank-20	mAP	$\mathcal{M}$	rank-1	rank-5	rank-20	mAP	$\mathcal{M}$
EUG[9](p=0.10)	70.79	83.61	89.60	61.76	71.62	57.62	69.64	78.08	34.68	58.97
NCI(p=0.10)	73.40	86.80	93.20	65.60	73.23	60.40	76.00	84.30	40.80	63.01
NCI+Loss(p=0.10)	**79.50**	**90.20**	**95.20**	**73.10**	**79.59**	**61.10**	**76.80**	**83.40**	**41.40**	**63.52**
EUG[9](p=0.05)	72.79	84.18	91.45	63.23	74.17	62.67	74.94	82.57	42.45	66.22
NCI(p=0.05)	74.40	88.50	93.40	66.40	75.30	64.60	78.10	84.40	45.80	68.19
NCI+Loss(p=0.05)	**80.30**	**91.60**	**95.30**	**74.00**	**80.12**	**66.60**	**80.20**	**87.8**	**48.70**	**71.67**

4.3.4 与其他方法比较

表 4 是本文的方法 NCI 和损失控制策略分别在 DukeMTMC-VideoReID 和 MARS 数据集上,与其它方法在 rank-1 accuracy(%)和 mAP(%)的比较.表 4 中与本文的对比方法有 OIM、BUC、DGM、 Stepwise、EUG 和 PL 等方法.本文提出的方法相比其它方法对单标注视频行人重识别性能都有明显的提升.如表 4 中所示,本文提出的方法 NCI 在 DukeMTMC-VideoReID 数据集上,最高使 rank-1 达到 74.40%,mAP 达到 66.40%;在 MARS 数据集上,最高使 rank-1 达到 64.60%,mAP 达到 45.80%.而在 NCI 加上提出的损失控制策略之后,在 DukeMTMC-VideoReID 数据集上,最高使 rank-1 达到 80.30%,mAP 达到 74.00%;在 MARS 数据集上,最高使 rank-1 达到 66.60%,mAP 达到 48.70%.性能远超过 DGM、Stepwise、EUG 和 PL 等方法.

NCI 和损失控制策略联合训练的最终结果与无监督的方法 OIM 和 BUC 相比在 DukeMTMC-VideoReID 和 MARS 数据集上有着明显的优势.相比于单标注视频行人重识别最新的方法 EUG 和 PL 有很大提升.当 $p=0.05$ 时,在 DukeMTMC-VideoReID 数据集上,rank-1 分别提升了 7.51%、7.4%,mAP 上分别提升了 10.77%、10.7%.在 MARS 数据集上,rank-1 分别提升了 3.93%、3.8%,mAP 上分别提升了 6.25%、6.1%.而当 $p=0.10$ 时,

在 DukeMTMC-VideoReID 数据集上,rank-1 分别提升了 8.71%、8.5%,mAP 上分别提升了 11.34%、11.2%.在 MARS 数据集上分别提升了 3.48%、3.2%,mAP 上分别提升了 6.72%、6.5%.

综合以上分析,说明本文 NCI 和损失控制策略联合训练,相比于同类的方法有很大的提升,从而验证了本文提出的近邻中心迭代策略和损失控制策略的有效性和优越性.

Table 4 Comparison of accuracy (%) between NCI and other methods

表 4 NCI 与其它方法的结果（%）的对比

	DukeMTMC-VideoReID				MARS			
Methods	rank-1	rank-5	rank-20	mAP	rank-1	rank-5	rank-20	mAP
Baseline[9](one-shot)	39.60	56.84	66.95	33.27	36.16	50.20	61.86	15.45
OIM[13]	51.10	70.50	-	43.80	33.70	48.10	-	13.50
BUC[19]	74.80	86.80	-	66.70	55.10	68.30	-	29.40
DGM[11]	42.36	57.92	69.31	33.62	36.81	54.01	68.51	16.87
Stepwise[10]	56.26	76.37	79.20	46.76	41.21	55.55	66.76	19.65
EUG[9](p=0.10)	70.79	83.61	89.60	61.76	57.62	69.64	78.08	34.68
EUG[9](p=0.05)	72.79	84.18	91.45	63.23	62.67	74.94	82.57	42.45
PL[25](p=0.10)	71.00	83.80	90.30	61.90	57.90	70.30	79.30	34.90
PL[25](p=0.05)	72.90	84.30	91.40	63.30	62.80	75.20	83.80	42.60
NCI(p=0.10)	73.40	86.80	93.20	65.60	60.40	76.00	84.30	40.80
NCI(p=0.05)	**74.40**	**88.50**	**93.40**	**66.40**	**64.60**	**78.10**	**84.40**	**45.80**
NCI+Loss(p=0.10)	79.50	90.20	95.20	73.10	61.10	76.80	83.40	41.40
NCI+Loss(p=0.05)	**80.30**	**91.60**	**95.30**	**74.00**	**66.60**	**80.20**	**87.80**	**48.70**
Baseline[9](supervised)	83.62	94.59	97.58	78.34	80.75	92.07	96.11	67.39

5 结束语

单标注学习的错误标签估计会严重降低模型的鲁棒性,无标签视频片段的标签估计对于单标注视频行人重识别至关重要.针对这个问题,本文提出了一种近邻中心迭代策略,该策略从简单可靠的无标签视频片段样本开始,逐步更新用于预测伪标签的度量中心点,获取更加可靠的伪标签数据来更新模型.每次选取的可靠伪标签数据以较慢的速度增加.此外,本文提出了一种新的损失训练策略,能使得训练过程更加稳定又能缩小类内距离,从而获得可靠的伪标签数据和更鲁棒的模型.本文方法的有效性在 MARS 和 DukeMTMC-VideoReID 两个大规模数据集上得到了很好的验证.

References:

[1] Gray D, Tao H. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Proc. of the 10th European Conference on Computer Vision. 2008: 262-275. [doi: 10.1007/978-3-540-88682-2_21]

[2] 戴臣超,王洪元,倪彤光,陈首兵. 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. 计算机研究与发展, 2019, 56(8):1632-1641. [doi: 10.7544/issn1000-1239.2019.20190195]

[3] 丁宗元,王洪元,陈付华,倪彤光. 基于距离中心化与投影向量学习的行人重识别.计算机研究与发展, 2017,54(08): 1785-1794. [doi: 10.7544/issn1000-1239.2017.20170014]

[4] 叶钰,王正,梁超,韩镇,陈军,胡瑞敏.多源数据行人重识别研究综述.自动化学报,2020, 1: 1-16. [doi: 10.16383/j.aas.c190278]

[5] Fan H, Zheng L, Yan C, Yang Y. Unsupervised person reidentification: Clustering and fine-tuning. ACM Trans. on Multimedia Computing, Communications, and Applications. 2018, 14(4): 83. [doi: 10.1145/3243316]

[6] Hirzer M, Beleznai C, Roth PM, Bischof H. Person reidentification by descriptive and discriminative classification. In: Proc. of the 17th Scandinavian Conference on Image analysis. 2011: 91-102. [doi: 10.1007/978-3-642-21227-7_9]

[7] Zheng L, Bie Z, Sun Y, Wang J, Su C, Wang S. MARS: A Video Benchmark for Large-Scale Person Re-identification. In: Proc. of the 14th European Conference on Computer Vision. 2016: 868-884. [doi: 10.1007/978-3-319-46466-4_52]

[8] Wang T, Gong S, Zhu X, Wang S. Person reidentification by discriminative selection in video ranking. IEEE Trans. on Pattern Analysis and Machine Intelligence. 2016, 38(12): 2501-2514. [doi: 10.1109/TPAMI.2016.2522418]

[9] Wu Y, Lin Y, Dong X, Yan Y, Ouyang W, Yang Y. Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning. In: Proc. of the IEEE International Conference on Computer Vision and Pattern Recognition. 2018: 5177-5186. [doi: 10.1109/CVPR.2018.00543]

[10] Liu Z, Wang D, Lu H. Stepwise metric promotion for unsupervised video person re-identification. In: Proc. of the IEEE International Conference on Computer Vision. 2017: 2429-2438. [doi: 10.1109/ICCV.2017.266]

[11] Ye M, Ma AJ, Zheng L, Li J, Yuen PC. Dynamic label graph matching for unsupervised video re-identification. In: Proc. of the IEEE International Conference on Computer Vision. 2017: 5142-5150. [doi: 10.1109/ICCV.2017.550]

[12] Hermans A, Beyer L, Leibe B. In Defense of the Triplet Loss for Person Re-Identification. arXiv: 1703.07737.

[13] Xiao T, Li S, Wang B, Lin L, Wang X. Joint Detection and Identification Feature Learning for Person Search. In: Proc. of the IEEE International Conference on Computer Vision and Pattern Recognition. 2017: 3376-3385. [doi: 10.1109/CVPR.2017.360]

[14] Liu Y, Yuan Z, Zhou W, Li H. Spatial and temporal mutual promotion for video-based person re-identification. In: Proc. of the Thirty-Third AAAI Conference on Artificial Intelligence. 2019, 33: 8786-8793. [doi: 10.1609/aaai.v33i01.33018786]

[15] Li S, Bak S, Carr P, Wang, X. Diversity regularized spatio temporal attention for video-based person re-identification. In: Proc. of the IEEE International Conference on Computer Vision and Pattern Recognition. 2018: 369–378. [doi: 10.1109/CVPR.2018.00046]

[16] Liu CT, Wu CW, Wang YCF, Chien, SY. Spatially and Temporally Efficient Non-local Attention Network for Video-based Person Re-Identification. arXiv: 1908.01683.

[17] Chen GY, Lu JW, Yang M, Zhou J. Spatial-temporal attention-aware learning for video-based person re-identification. IEEE Trans. on Image Processing, 2019, 28(9): 4192-4205. [doi: 10.1109/TIP.2019.2908062]

[18] Rao YM, Lu JW, Zhou J. Learning discriminative aggregation network for video-based face recognition and person re-identification. International Journal of Computer Vision, 2019, 127(6-7): 701-718. [doi: 10.1007/s11263-018-1135-x]

[19] Lin Y, Dong X, Zheng L, Yan Y, Yi Y. A bottom-up clustering approach to unsupervised person re-identification. In: Proc. of the Thirty-Third AAAI Conference on Artificial Intelligence. 2019, 33: 8738-8745. [doi: 10.1609/aaai.v33i01.33018738]

[20] Bak S, Carr P. One-shot metric learning for person re-identification. In: Proc. of the IEEE International Conference on Computer Vision and Pattern Recognition. 2017: 2990-2999. [doi: 10.1109/CVPR.2017.171]

[21] Figueira D, Bazzani L, Minh HQ, Cristani M, Bernardino A. Semi-supervised multi-feature learning for person re-identification. In: Proc. of the 10th IEEE International Conference on Advanced Video and Signal Based Surveillance. 2013: 111-116. [doi: 10.1109/AVSS.2013.6636625]

[22] Liu X, Song M, Tao D, Zhou X, Chen C, Bu J. Semi-supervised coupled dictionary learning for person re-identification. In: Proc. of the IEEE International Conference on Computer Vision and Pattern Recognition. 2014: 3550-3557. [doi: 10.1109/CVPR.2014.454]

[23] Ma AJ, Li P. Semi-supervised ranking for re-identification with few labeled image pairs. In: Proc. of the 12th Asian Conference on Computer Vision. 2014: 598-613. [doi: 10.1007/978-3-319-16817-3_39]

[24] Zhu X, Jing XY, Yang L, You X., Chen D, Gao G, Wang Y. Semi-supervised cross-view projection-based dictionary learning for video-based person re-identification. IEEE Trans. on Circuits and Systems for Video Technology. 2017, 28(10): 2599-2611. [doi: 10.1109/TCSVT.2017.2718036]

[25] Wu Y, Lin Y, Dong X, Yan Y, Bian W, Yang Y. Progressive learning for person re-identification with one example. IEEE Trans. on Image Processing, 2019, 28(6): 2872-2881. [doi: 10.1109/TIP.2019.2891895]

[26] Ristani E, Solera F, Zou R, Cucchiara R, Tomasi C. Performance measures and a data set for multi-target, multicamera tracking. In: Proc. of the 14th European Conference on Computer Vision. 2016: 17-35. [doi: 10.1007/978-3-319-48881-3_2]

[27] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Proc. of the Neural Information Processing Systems 25. 2012: 1097–1105. [doi: 10.1145/3065386]