DOI: 10.13382/j.jemi.B2407706

基于深度聚类学习的无监督行人重识别*

邓子文 段 勇
(沈阳工业大学信息科学与工程学院 沈阳 110870)

摘 要:无监督行人重识别是一种在没有任何标签的情况下,通过特征提取和聚类算法对行人进行识别和匹配的计算机视觉方法。针对当前无监督行人重识别方法普遍存在的特征提取不足、聚类不准确、计算复杂度高以及模型缺乏鲁棒性等问题,提出了一种基于深度聚类学习的无监督行人重识别方法。首先,研究了结合 GeM 池化方法的实例批量归一化网络(IBN-Net)作为特征提取网络,使得提取出的行人特征更具判别性;其次,针对聚类算法对于超参数较为敏感的问题,提出通过有序点识别聚类结构(OPTICS)辅助基于密度的聚类算法(DBSCAN)选取超参数,进一步降低了 DBSCAN 对超参数的敏感度;此外,为了更加充分利用训练集的所有数据,将离群值也视为单独的聚类参与到记忆字典的初始化与更新过程中;最后,针对记忆字典更新过程中各个聚类更新速率不一致的问题,提出了聚类级别的记忆字典,消除了聚类更新速率不一致的问题。实验结果验证了研究工作的有效性,提出的方法在无监督行人重识别任务中的精度与准确度均有明显的提升。

关键词:行人重识别;无监督学习;对比学习;深度学习;聚类算法
中图分类号:TP391;TN911.7 **文献标识码**:A **国家标准学科分类代码**:510.40

Unsupervised person re-identification based on deep clustering learning

Deng Ziwen Duan Yong
(School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, China)

Abstract: Unsupervised person re-identification is a computer vision method that identifies and matches pedestrians without any labeled data, utilizing feature extraction and clustering algorithms. To address common issues in current unsupervised person re-identification methods, such as insufficient feature extraction, inaccurate clustering, high computational complexity and lack of model robustness, this paper proposes a deep clustering learning-based approach for unsupervised person re-identification. First, we investigate the use of IBN-Net combined with generalized mean pooling as the feature extraction network, which enhances the discriminative power of the extracted features. Second, to mitigate the sensitivity of clustering algorithms to hyperparameters, we introduce the OPTICS algorithm to assist DBSCAN in selecting hyperparameters, thus reducing DBSCAN's dependency on those hyperparameters. Additionally, to fully utilize all the data in the training set, outliers are treated as separate clusters and included in the initialization and updating process of the memory dictionary. Finally, to address the inconsistency in update rates among clusters in the memory dictionary, we propose a cluster-level memory dictionary that eliminates this issue. Experimental results validate the effectiveness of our approach, demonstrating significant improvements in both precision and accuracy in unsupervised person re-identification tasks.

Keywords: person re-identification; unsupervised learning; contrastive learning; deep learning; clustering algorithm

0 引 言

行人重识别又称行人再识别,其主要目标是确定一个特定的人是否出现在由不同摄像机所拍摄到的图像中[1]。行人重识别旨在弥补固定摄像头的视觉局限,并且能够在行人图像清晰度不足,无法进行面部识别的情况下进行行人身份的检测。目前,行人重识别方法已经

收稿日期:2024-07-22 Received Date: 2024-07-22
*基金项目:辽宁省高校优秀科技人才支持计划(LR15045)、辽宁省教育厅科学研究经费面上项目(LJKZ0139)资助

广泛应用于智能化监控系统、智慧城市建设以及智能安保等诸多领域。

传统的有监督行人重识别方法通常以深度学习为主要框架,包含特征提取、特征度量和排序优化 3 个部分。这类方法依靠已经被标签标注了的训练集数据来对模型进行训练,但由于对数据进行标注会耗费大量的人力与财力,导致有监督行人重识别算法在实际场景中的应用与部署存在一定的局限性。

无监督行人重识别方法旨在训练一个在没有任何标签标注的数据集中也能够跨摄像头检索出行人身份的模型。由于对实际场景中视频监控的需求在与日俱增,加之对数据标注的成本较为昂贵,使得无监督行人重识别方法在近年来受到了越来越多的关注[2]。

对于无监督的行人重识别问题,目前主要有两种主流的解决方法。一种是完全无监督的行人重识别方法,这种方法通常利用聚类算法为无标签的数据集生成伪标签[3]。另一种是领域自适应的行人重识别方法,该方法首先在有标记的源域数据集上对模型进行预训练,之后在无标记的目标域数据集上对模型进行微调[4]。由于引入了有标签的源域数据集,导致领域自适应的行人重识别算法的性能普遍优于完全无监督的方法。但领域自适应的行人重识别方法需要繁琐的训练机制,同时还需要源域与目标域数据之间的差异不能过大,使得无需借助任何标注信息而仅需一个预训练好的特征提取模型的完全无监督行人重识别方法在实际场景中的应用更具普适性。因此,本文也将主要关注完全无监督的行人重识别方法。

对于大多数完全无监督的重识别方法,其算法流程可以概括为伪标签的生成、记忆字典的初始化以及深度神经网络的训练 3 个部分。Ge 等[5]所提出的自步对比学习机制通过不断生成更加可靠的聚类来细化记忆字典中所存储的实例特征。Wang 等[6]提出的 MMCL 方法采用了相似性计算与循环一致性原理来预测高质量的伪标签。程思雨等[7]提出了基于 ViT 的细粒度特征增强无监督行人重识别方法,利用视觉-语言模型生成图像中人体局部区域的掩码,并采用 Transformer 网络将图像编码成块标记序列,并利用掩码引导计算全局与局部自注意力。贺晓东等[8]提出了一种基于局部特征与视点感知的车辆重识别算法,利用语义分割算法将车辆结构为 4 个部分,通过视点感知网络输出视点的概率信息,据此概率信息动态平滑地呈现车辆视点感知效果。

近年来,针对基于聚类的无监督行人重识别方法,一些工作致力于优化聚类算法以提升整体的聚类性能,从而进一步提高重识别的准确度。Li 等[9]提出了一种聚类引导的孪生对比学习方式来进行无监督行人重识别。该文章提出了一种聚类细化方法,按照一定比例去除掉较大规模的簇中的噪声样本,从而帮助网络更好地在聚类级别上学习特征信息。熊福明等[10]提出了一种基于二次聚类的无监督行人重识别方法。该方法主要包括全局二次聚类的无监督学习模块与基于聚类结果的有监督学习模块。通过这两个模块之间彼此的协同训练来共同抑制跨摄像头采集的图像所产生错误伪标签的问题,进一步的细化聚类结果。

现阶段的无监督行人重识别方法依旧存在一些不完善之处。首先,受训练阶段批次大小的限制,导致在实例特征级别存储的记忆字典中的特征在更新速率上难以与更新之后的特征网络所重新提取出的查询实例特征相匹配。其次,聚类所产生的离群值通常被视为噪点而被直接丢弃,导致训练集数据无法被充分利用。最后,大多数聚类算法对超参数较为敏感,不合适的超参数将会严重影响聚类算法的性能。针对上述问题,本文提出了一种基于深度聚类学习的无监督行人重识别算法。首先,利用实例批量归一化网络(instance-batch normalization network,IBN-Net)[11]提取训练集所有行人图像的特征。其次,使用有序点识别聚类结构算法(ordering points to identify the clustering structure,OPTICS)结合基于密度的空间聚类算法(density-based spatial clustering of applications with noise,DBSCAN)对这些行人特征进行聚类生成伪标签,从而减弱聚类算法对超参数的敏感度。之后,利用聚类质心与离群值共同初始化记忆字典,构建聚类级别的记忆字典从而保证聚类更新速率一致,同时充分利用训练集的所有数据。最后,采用动量更新的方式更新记忆字典,并使用本文提出的扩展对比损失更新特征提取网络。

1　本文方法

为了解决之前的研究工作在聚类之后没有充分利用被判定为离群值训练集数据的问题,以及由于聚类结果中每个簇的规模大小不同而导致的特征提取网络更新速率不一致的问题和聚类算法对于超参数的选取较为敏感的问题。本文提出了一种基于深度聚类学习的无监督行人重识别算法,该算法框架的详细结构如图 1 所示。基于深度聚类学习的无监督行人重识别算法框架主要由特征提取模块、聚类模块以及记忆字典的存储与更新模块 3 个模块构成。特征提取模块主要利用 IBN-Net 对无标签的训练集行人图像进行特征提取。随后,在聚类模块中对这些行人图像特征进行聚类操作并分配伪标签,本研究使用 OPTICS 辅助 DBSCAN 为其自动选取合适的超参数,从而降低了 DBSCAN 对于超参数的敏感程度。之后,计算每个聚类的平均值作为聚类质心连同离群值特征存储在记忆字典中完成记忆字典的初始化操作。最

后,从训练集行人图片数据中随机选取若干张不同身份的行人图片经过特征提取网络得到查询特征向量,利用动量更新机制对记忆字典中存储的聚类质心及离群值进行更新,同时利用本文提出的扩展对比损失函数计算查询特征与记忆字典中存储的与其相对应的聚类质心或离群值特征之间的对比损失更新特征提取网络。

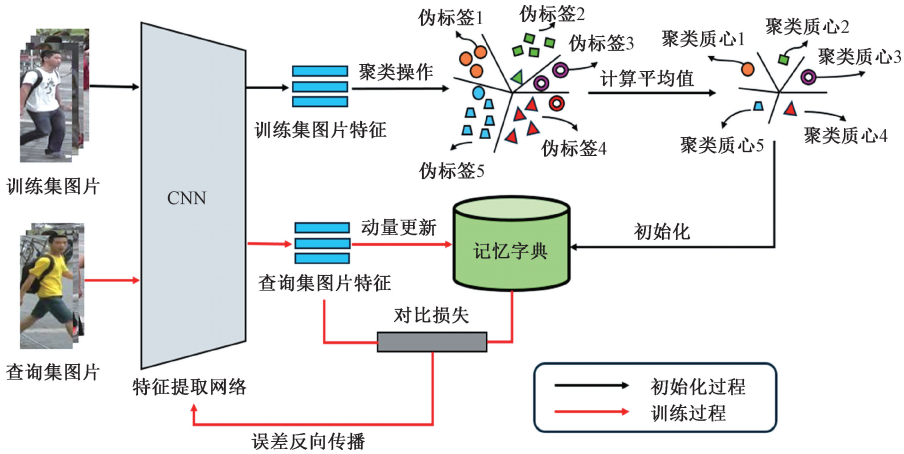


图1　基于深度聚类的无监督行人重识别方法整体结构

Fig. 1　Overall structure of deep clustering-based unsupervised person re-identification method

1.1　行人图像特征提取网络

对于无监督行人重识别任务,首先需要通过特征提取网络获取行人图像的特征。在以往的无监督行人重识别工作中,通常使用标准的50层残差网络(residual network 50-layer,ResNet50)[12]进行特征提取。但为了进一步提升模型性能,本文将ResNet50中所有批量归一化层(batch normalization,BN)全部替换为实例批量归一化层(instance-batch normalization,IBN),并结合广义均值池化方法(generalized-mean pooling,GeM)[13]作为特征提取网络。

1)IBN-Net

ResNet50主要由4部分残差块构成,IBN-Net将实例归一化模块(instance normalization, IN)整合到了ResNet50的前3个残差模块中而对最后1个残差模块不进行改动。针对1个具体的残差块,在第1个卷积层之后,一半的通道采用BN而另一半通道采用IN。

对于BN而言,在网络训练的过程中,通过减小内部协变量的偏移能够让模型适应更大的学习率并获得更快的收敛速度。而IN则直接针对单个样本图像进行标准化。BN在保留了单个样本之间具有判别性特征的同时使整个神经网络更容易受到样本图片中行人外观变化的影响,而IN消除了每个样本之间的差异性,同时也丢失了每个样本的判别性特征。IBN-Net将IN与BN结合在一起,充分发挥了二者的优势,同时尽可能减少了二者的劣势对网络造成的影响。

IBN-Net已经在领域自适应任务中取得了显著的成效。对于完全无监督的行人重识别任务,IBN-Net同样能够使模型整体性能得到提升,这是因为其具有的外观不变性。通过IN学习到的特征对于外观上的变化具有很好的稳定性,而BN则更善于保留特征中那些更具判别性的信息。因此,IBN-Net在设计中融合了IN和BN,使其在无监督任务中能够更好地保留内容信息的同时,实现了对行人外观变化的鲁棒性。这种内置的外观不变性使得IBN-Net即使在不使用源域数据集的情况下也能够取得显著提升。

2)广义平均池化方法

本文将IBN-Net网络结构中的平均池化层替换成了GeM池化层,并舍弃了其后所有的网络结构。

对于GeM池化方法,当参数$\boldsymbol{p}_k \to \infty$,GeM的计算公式便会成为最大池化的计算公式,如式(1)所示;当$\boldsymbol{p}_k = 1$,GeM的计算公式又会成为平均池化的计算公式,如式(2)所示;假定当前输入到GeM池化层中的特征为$\boldsymbol{\chi}$,该特征经过GeM池化操作之后输出一个向量$\boldsymbol{f}_k^{(g)}$,GeM具体的计算方法如式(3)所示。

$$\boldsymbol{f}^{(m)} = [\boldsymbol{f}_1^{(m)} \cdots \boldsymbol{f}_k^{(m)} \cdots \boldsymbol{f}_K^{(m)}]^{\mathrm{T}}, \boldsymbol{f}_k^{(m)} = \max_{x \in \boldsymbol{\chi}_k} \boldsymbol{x} \tag{1}$$

$$\boldsymbol{f}^{(a)} = [\boldsymbol{f}_1^{(a)} \cdots \boldsymbol{f}_k^{(a)} \cdots \boldsymbol{f}_K^{(a)}]^{\mathrm{T}}, \boldsymbol{f}_k^{(a)} = \frac{1}{|\boldsymbol{\chi}_k|} \sum_{x \in \boldsymbol{\chi}_k} \boldsymbol{x} \tag{2}$$

$$\boldsymbol{f}^{(g)} = [\boldsymbol{f}_1^{(g)} \cdots \boldsymbol{f}_k^{(g)} \cdots \boldsymbol{f}_K^{(g)}]^{\mathrm{T}}, \boldsymbol{f}_k^{(g)} = \left(\frac{1}{|\boldsymbol{\chi}_k|} \sum_{x \in \boldsymbol{\chi}_k} \boldsymbol{x}^{p_k} \right)^{\frac{1}{p_k}} \tag{3}$$

式中:$\boldsymbol{\chi}$代表输入到池化层中的图像特征;$\boldsymbol{\chi}_k$代表所有输入图像特征的集合;$\boldsymbol{f}_k^{(m)}$ $\boldsymbol{f}_k^{(a)}$和$\boldsymbol{f}_k^{(g)}$分别代表经过最大池化方法,平均池化方法与广义平均池化方法后所得

到的图像特征；$f^{(m)}$ $f^{(a)}$ 和 $f^{(g)}$ 代表由 3 种池化方法操作得到的图像特征集合。

由此可以得出一个结论，平均池化与最大池化实际上是 GeM 池化的两个特例。

GeM 能够通过调整 p_k 在最大池化与平均池化之间进行调整，使其兼具两者优点。在无监督行人重识别任务中，由不同摄像头从不同角度采集到的行人图像通常会变得多样且复杂。因此，单一的池化方法可能不足以应对所有情况。GeM 的灵活性使其能够在捕获局部细节和保持全局统一性之间找到平衡，从而适应不同风格的行人图像特征。

1.2　行人身份伪标签聚类模块

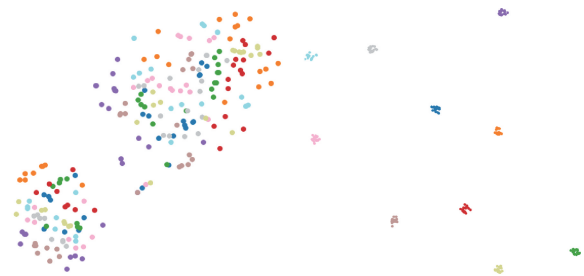
针对由 IBN-Net 提取出的行人特征，需要对其聚类以获得行人身份的伪标签并分配给无标签的训练集行人数据。

基于密度的 DBSCAN 聚类算法能够识别并处理聚类过程中产生的离群值。在行人重识别任务中，行人图像可能会受到遮挡、光照变化、环境因素等诸多因素的影响，从而导致产生许多噪声样本。DBSCAN 能够有效地识别这些噪声样本，从而避免它们对聚类结果产生负面影响。除此之外，由于 DNSCAN 无需事先指定具体聚类的数量，并且能够处理任意形状的簇，使得 DBSCAN 聚类算法被广泛的应用在诸多基于聚类的无监督行人重识别方法中。

但 DBSCAN 算法十分依赖通过人工手动选择的超参数。如果未能选择合适的超参数将会在很大程度上影响 DBSCAN 聚类算法最终的整体性能。具体来说，DBSCAN 聚类算法中有两个关键的超参数：邻域半径和最小样本点数。在 DBSCAN 聚类算法中规定：当在邻域半径所规划的范围内包含的样本点数量大于或等于最小样本点数时，该范围内的样本分布就是较为密集的，可以被划分为一个聚类。其中，邻域半径对于 DBSCAN 最终聚类准确性的影响尤为明显。因为，如果邻域半径的值设定过大，则几乎所有的样本点都会被包含在同一个聚类中，从而导致聚类的数量变得很少。而如果邻域半径设定的过小，则很多样本点都会被标记为噪声点，使得形成聚类的数量非常多且每个聚类包含的样本点十分稀少，最终导致聚类效果较差，无法有效地分辨真正的聚类。

为了降低 DBSCAN 聚类算法对于超参数的敏感度，本文提出了使用 OPTICS 聚类算法辅助 DBSCAN 选取超参数的方式。OPTICS 本质上也属于基于密度的聚类算法。但不同于 DBSCAN 算法直接生成聚类结果，OPTICS 能够通过分析样本点之间的可达性结构生成样本点的可达距离图从而帮助 DBSCAN 自动选取合适的邻域半径参数。

DBSCAN 对于输入参数较为敏感，即 DBSCAN 在不同超参数的定义下得到的聚类结果差异很大。当人工直接给定了邻域半径与最小样本点数时，实际上已经直接定义了生成聚类的最小密度，那么密度较小的聚类则会在实际聚类的过程中被忽略掉。而 OPTICS 针对 DBSCAN 的问题进行了改进，OPTICS 能够对任意密度的数据进行聚类，而非像 DBSCAN 基于全局密度参数进行聚类。随机从 PersonX 数据集中选取 10 个行人身份 ID，对这些身份 ID 所对应的特征向量使用 t-SNE 方法进行降维之后的可视化分布情况如图 2(a)所示。经过使用 OPTICS 结合 DBSCAN 聚类之后的行人特征分布经 t-SNE 降维之后的可视化效果如图 2(b)所示。从聚类效果图能够观察到，本文所提出的使用 OPTICS 结合 DBSCAN 的聚类方式能够有效的拉近具有相同聚类伪标签的行人特征之间的特征距离，同时拉远具有不同聚类伪标签的行人特征之间的特征距离，从而进一步提升整体的聚类效果。



(a) 原始数据特征分布　　(b) OPTICS+DBSCAN

(a) Original data feature distribution

图 2　OPTICS+DBSCAN 聚类效果

Fig. 2　OPTICS+DBSCAN clustering visualization

经由 IBN-Net 网络所提取出的特征相较于使用 ResNet50 网络提取的特征来说通常较为集中，特征向量之间的距离分布也更为均匀。由于特征分布较为集中，在本文所提出的方法中，经过 OPTICS 聚类算法得到各个特征样本点的可达距离集合之后，通过对其排序并选取集合中的最大值作为 DBSCAN 的邻域半径。通过这种方式为 DBSCAN 所选取的邻域半径可以确保所有的样本点都被包含在聚类范围内，有效避免了过度分割现象的产生。

1.3　记忆字典

针对经过聚类之后的行人特征向量，需要一个对其进行存储并在训练过程中动态更新的数据结构。本文提出了一种在聚类级别上进行存储与更新的记忆字典。记忆字典在无监督行人重识别任务中主要用于存储实例特征向量，以提高行人特征的一致性。

1) 记忆字典的初始化

对于不同的行人重识别数据集，其各自的数据分布特点与空间密度均有所差异。这种差异不可避免的会导

致聚类结果中每个簇的规模也会有所不同。对于在实例特征级别上进行存储与更新的方式来说,在每次训练的迭代过程中,受训练批量大小的限制,会使得规模较大的簇中只有小部分的实例特征能够被更新,而规模较小的簇中所有的实例特征则均能被更新。最终,导致大规模的簇与小规模的簇更新速率出现差异,致使模型在训练过程中更容易偏向更新速率较快的类别。

为了解决上述问题,本文提出了在聚类级别上进行存储与更新的记忆字典。在每一次训练迭代过程开始之前,首先利用 IBN-Net 对训练集中的所有数据进行特征提取,并对这些特征进行聚类操作。之后,需要从每一个聚类中挑选出一个具有代表性的实例特征来代表整个聚类存储在记忆字典中。为了挑选出这种具有代表性的实例特征,研究了一种简单但高效的方法。即对于每一个聚类中的所有实例特征,计算其平均值作为当前聚类的聚类质心代表整个聚类参与到记忆字典的初始化与更新过程。聚类质心的计算方法如式(4)所示。

$$\boldsymbol{c}_k = \frac{1}{|\boldsymbol{\chi}_k|}\sum_{\boldsymbol{u}_i \in \boldsymbol{\chi}_k} \boldsymbol{u}_i \tag{4}$$

式中:$\boldsymbol{\chi}_k$ 代表了第 k 个聚类;$|\boldsymbol{\chi}_k|$ 代表了第 k 个聚类中所包含实例特征的个数;$\boldsymbol{u}_i$ 代表第 k 个聚类中的第 i 个实例特征。

为了充分利用训练集所有的数据样本,同时进一步挖掘难样本数据中的关键信息,本文在初始化记忆字典的过程中,不仅使用了每个聚类的聚类质心,还将离群值视为单独的聚类参与到了记忆字典的初始化过程。

简单来说,在每次迭代训练开始之前,都要进行记忆字典的初始化操作。而记忆字典的初始化操作实际上就是将行人图像的特征聚类之后,将经过计算得到的聚类质心与离群值一并存储在记忆字典中,以便后续在记忆字典的更新操作中使用。

2)扩展对比损失函数

在通过聚类之后的行人特征向量完成了记忆字典初始化操作之后,完成了模型整体的初始化阶段,进入了模型的训练阶段。

在模型的训练过程中,本文从训练集中随机选取 P 个行人身份 ID,对于每个 ID 随机选取 K 张图像。将这 $P\times K$ 张图像作为每次训练过程中的查询样本,之后经由 IBN-Net 进行特征提取之后参与到记忆字典的更新与损失值的计算过程中。为了方便后文叙述,将这 $P\times K$ 张查询样本图片的特征简称为 q。

针对上文所述的记忆字典结构,本文设计了一种扩展对比损失函数(extended contrastive loss,ECL)。该损失函数是在传统对比损失函数的基础上,将离群点作为一种特殊类别纳入损失计算的框架中扩展而来的,故称其为扩展对比损失函数。给定一组查询特征 q,ECL 的计算方法如式(5)所示。

$$\boldsymbol{L}_q = -\log\frac{\exp(<\boldsymbol{q},\boldsymbol{z}^+>/\boldsymbol{\gamma})}{\sum_{k=1}^{n_c}\exp(<\boldsymbol{q},\boldsymbol{c}_k>/\boldsymbol{\gamma}) + \sum_{k=1}^{n_o}\exp(<\boldsymbol{q},\boldsymbol{\phi}_k>/\boldsymbol{\gamma})} \tag{5}$$

式中:$\boldsymbol{z}^+$ 为记忆字典中相对于输入样本特征 $\boldsymbol{q}$ 而言的正样本特征;$\boldsymbol{n}_c$ 代表聚类的总数;$\boldsymbol{n}_o$ 为离群值的总数;$\boldsymbol{c}_k$ 为第 k 个聚类的聚类质心;$\boldsymbol{\phi}_k$ 为第 k 个离群值;$<\cdot,\cdot>$ 是计算两个特征向量之间的相似度操作;$\boldsymbol{\gamma}$ 为温度系数。具体来说,如果 $\boldsymbol{q}$ 属于第 k 个聚类,则有 $\boldsymbol{z}^+=\boldsymbol{c}_k$,即 $\boldsymbol{z}^+$ 为第 k 个聚类的聚类质心;如果 $\boldsymbol{q}$ 是一个离群值,则有 $\boldsymbol{z}^+=\boldsymbol{\phi}_k$,即 $\boldsymbol{z}^+$ 为与 $\boldsymbol{q}$ 相关的第 k 个离群值。本文所提出的扩展对比损失函数会促使行人特征向量 $\boldsymbol{q}$ 靠近其所属的聚类或离群值,同时远离其他的负样本。

本文提出的扩展对比损失函数的计算过程中,使用的是聚类质心来代替整个聚类中所包含的样本特征来进行损失值的计算,而非像传统的对比损失函数直接在实例特征的级别上来逐一与查询特征进行对比计算对比损失值,这也进一步保证了各个聚类与离群值的更新速率能够尽可能的一致,从而有效避免了模型在训练的过程中偏向更新速率较快的聚类。

3)记忆字典的更新策略

在模型训练的过程中,除了利用本文提出的扩展对比损失函数更新特征提取网络,还需对记忆字典中存储的聚类质心及离群值进行动态更新。因为随着特征提取网络的不断更新,通过该网络提取出的行人特征相较于初始阶段可能会属于不同的聚类。

训练过程中,行人图像的查询特征将参与到记忆字典的更新操作中。本文设计的动量更新策略结合查询特征对记忆字典中存储的聚类质心与离群值特征向量进行动态更新。在迭代训练的过程中,聚类质心将由与其同属一个聚类的查询特征所更新,而离群值也将由与其相对应的查询特征更新。具体的更新计算方法如式(6)与(7)所示。

$$\boldsymbol{C}_k \leftarrow \boldsymbol{m}^c\boldsymbol{C}_k + (1-\boldsymbol{m}^c)\,\boldsymbol{q},\ \forall \boldsymbol{q} \in \boldsymbol{Q}_k \tag{6}$$

$$\boldsymbol{O}_k \leftarrow \boldsymbol{m}^o\boldsymbol{O}_k + (1-\boldsymbol{m}^o)\,\boldsymbol{q} \tag{7}$$

式中:$\boldsymbol{Q}_k$ 是属于第 k 个聚类的查询特征集合;$\boldsymbol{C}_k$ 是第 k 个聚类的聚类质心;$\boldsymbol{O}_k$ 是第 k 个离群值特征;$\boldsymbol{m}^c$ 与 $\boldsymbol{m}^o$ 均为动量系数且被设置为 0.2。动量系数能够保证查询特征 $\boldsymbol{q}$ 与被更新的聚类质心 $\boldsymbol{C}_k$ 或离群值 $\boldsymbol{O}_k$ 之间的一致性。具体来说,当动量系数趋近于 0 的时候,更新后的聚类质心或离群值将更加趋近于查询特征;而当动量系数趋近于 1 的时候,则更新之后的聚类质心或离群值将更加趋近于它们自身。

聚类级别的记忆字典主要用于存储与动态更新经过

聚类之后的行人特征向量。通过记忆字典的初始化、扩展对比损失函数和动量更新策略,提升了行人特征更新速率的一致性与重识别模型的准确性。

2　实验结果及分析

为了验证本文方法的有效性,分别在 3 个常见的大规模行人重识别数据集上进行了测试。同时,通过对比试验与其他无监督行人重识别算法进行比较并设计消融实验,验证了本文模型框架中各个模块的有效性。

2.1　数据集与评价标准

本文方法分别在 3 个大规模的行人重识别数据集上进行了实验测试,分别为 Market-1501[14]、MSMT17[15] 以及 PersonX[16]。其中,PersonX 数据集中的行人图片是基于 Unity 引擎合成的。其中包含了许多人工设置的干扰因素,如随机遮挡、分辨率差异以及光照因素差异等。

本文使用了平均精度均值(mean average precision, mAP)以及 rank1、rank5 和 rank10 精度作为最终的评价标准。

其中,mAP 是一种综合评价行人重识别检索系统性能的指标,常用于多标签任务。mAP 计算的是在行人重识别检索任务中,每个查询行人图片对应的所有候选图片中,正确匹配图片的平均准确率。反映了检索的行人在数据库中所有正确的图片排在排序列表前面的程度,能够更加全面的衡量行人重识别算法的性能。

对于 rank-k 准确率,指的是在检索结果中,正确匹配图片出现在前 k 个候选图片中的概率。例如 rank1 指的是正确匹配的图片出现在第 1 个候选图片中的概率。该指标能够反映行人重识别系统在前 k 个结果中找到正确匹配的能力。

2.2　实验细节

本文采用在 ImageNet 数据集上预训练好的 IBN-Net 残差神经网络作为特征提取网络并使用该网络提取出的训练集行人数据特征对记忆字典进行初始化操作。本文移除了 IBN-Net 网络第 4 个卷积模块之后的所有结构,并添加 GeM 池化层以及 1 个 BN 层和 L2 归一化层来进一步提升网络的特征提取性能。在模型的测试阶段,直接提取 GeM 输出的特征向量与查询特征计算余弦相似度。经过优化后的网络模型参数总大小约为 90 MB,相较于其他常规深度学习方法(如 ResNet-50 网络约为 97 MB)显著减小,有助于提高计算效率与内存占用性能。此外,模型参数量的缩减也侧面验证了本文所提出结构调整和优化策略的有效性,在保证高性能特征提取的同时,进一步降低了模型的冗余性与复杂度。

在每个迭代训练的初期,本实验提出使用 OPTICS 结合 DBSCAN 的聚类方法为无标记的数据集生成伪标签。对于 OPTICS 聚类算法函数,本实验将最小样本数(min-samples)设置为 32,该参数定义了形成一个核心点所需的最小样本数。在应用 OPTICS 算法后,实验通过计算每一个样本数据点的可达距离来估计数据的密度结构,并选择数据中最大的可达距离作为聚类的阈值,从而得到一个合适的邻域半径(epsilon, eps)用于后续 DBSCAN 聚类。之所以选择最大的可达距离作为 DBSCAN 聚类的邻域半径参数,是因为在 OPTICS 算法对于可达距离的排序中,最大可达距离通常位于密度变化较大的区域的末尾,代表了数据密度分布的分界线。较大的可达距离,通常对应了簇与簇之间的间隙,即密度变化明显的地方。因此,选择最大可达距离作为 DBSCAN 的邻域半径,能够帮助其定义出具有较强连通性与合理密度的簇。对于 DBSCAN 聚类函数,其 eps 参数使用了通过 OPTICS 聚类算法所得到的结果,min-samples 参数设置为 4,相似度矩阵参数(metric)本实验设置为“precomputed”,即使用了预先计算好的基于图像重排序的距离矩阵。

输入模型的图片大小规定为 256×128,对于训练集的行人图片采取了随机水平翻转、随机裁剪以及随机擦除等常见的数据增强方法。模型训练时,每次迭代过程中的查询图片总数为 256,其中包含了随机选取的 16 个行人身份 ID,对每个行人 ID 随机选取其中的 16 张行人图像。动量系数默认设置为 0.2,损失函数中的温度系数默认设置为 0.05。本文采用了权重衰减速率为 5×10^{-4} 的 AdamW 优化器来进行模型训练。模型的学习率初始设置为 0.000 35,并在 50 轮的迭代训练中每隔 20 轮衰减为自身的 0.1 倍。

2.3　对比实验

将本文提出的方法与近年来典型的无监督行人重识别方法以及传统的无监督行人重识别方法在 Market-1501 数据集上进行比较的结果如表 1 所示。

表 1　Market-1501 对比实验结果

Table 1　Comparative experimental results on Market-1501

(%)

模型	mAP	Rank-1	Rank-5	Rank-10
MMCL[6]	45.5	80.3	89.4	92.3
JVTC+[17]	47.5	79.5	89.2	91.9
JNL[2]	61.7	83.9	92.3	–
ACT[18]	66.7	83.9	91.4	93.4
RLCC[19]	77.7	90.8	96.3	97.5
ICE[20]	79.5	92.0	**97.0**	**98.1**
本文	**81.9**	**92.6**	96.9	98.0

从表 1 能够看出,在 Market1501 数据集上,将本文方

法与其他的无监督行人重识别算法在 mAP 以及 rank1、rank5 和 rank10 评价标准上进行了比较。本文方法取得了 *mAP* = 81.9% 以及 *rank*-1 = 92.6%、*rank*-5 = 96.9% 和 *rank*-10 = 98.0% 的准确率。相较于近年提出的较为先进的无监督行人重识别方法 ICE[20] 而言，本文的方法虽然在 rank-5 与 rank-10 上的评分均以 0.1% 的差距略低于 ICE 算法。但是本文方法在 mAP 评分与 rank-1、rank-5 以及 rank-10 准确率上相较于其他作为对比的无监督行人重识别方法而言均有着显著的提升，充分证明了本文方法的有效性。

为了进一步验证本文方法的有效性，分别在另外两个数据集上进行了对比试验。这两个数据集分别是数据规模更大且更加具有挑战性的 MSMT17 数据集以及 PersonX 数据集。实验结果如表 2 所示，可以明显看出，本文方法在 MSMT 数据集上取得了 *mAP* = 35.2%，*rank*-1 = 62.4% 的优秀性能。相较于目前较为先进的 ICE 方法而言分别在 mAP 与 rank-1 两个评价指标上提升了 5.4% 与 3.4%。

表 2　MSMT17 对比实验结果

Table 2　Comparative experimental results on MSMT17

(%)

模型	mAP	Rank-1	Rank-5	Rank-10
MMCL[6]	11.2	35.4	44.8	49.8
JVTC+[17]	17.3	43.1	53.8	59.4
SPCL[5]	19.1	42.3	55.6	61.2
RLCC[19]	27.9	56.5	68.4	73.1
ICE[20]	29.8	59.0	71.7	77.0
本文	**35.2**	**62.4**	**73.5**	**77.8**

而对于在 PersonX 数据集上进行的对比实验，本文除了利用完全无监督的行人重识别方法进行对照，还使用了若干领域自适应的行人重识别方法作为对照组。从表 3 中能够看到，源域(source)为 None 意味着其对应的算法为完全无监督的行人重识别算法，如果源域为 Market 则意味着其对应的方法是从 Market-1501 数据集向 PersonX 数据集进行迁移学习的领域自适应行人重识别算法。实验结果如表 3 所示，能够发现，本文方法相较于对照组中效果最好的使用了领域自适应方法的 SPCL 算法而言，在 mAP 与 rank-1 两个评价指标上分别提升了 8.6% 与 3.1%。

表 3　PersonX 对比实验结果

Table 3　Comparative experimental results on PersonX

(%)

模型	Source	mAP	Rank-1	Rank-5	Rank-10
SPCL[5]	None	72.3	88.1	96.6	98.3
SPCL[5]	Market	78.5	91.1	97.8	99.0
本文	**None**	**87.1**	**94.2**	**98.4**	**99.2**

通过上述在不同数据集上与不同方法的对比实验结果来看，本文方法在 3 个数据集上均拥有较为优秀的性能表现。

2.4　消融实验

本文在 Market-1501 数据集上进行消融实验来探讨各模块对本文方法的影响，从而进一步验证方法的合理性。

1)特征提取网络消融实验

本文使用 IBN-Net 结合 GeM 池化进行特征提取。

为验证这种结合方式的有效性，本文将使用 ResNet50 替换 IBN-Net 网络并使用 IBN-Net 结合平均池化作为对比实验参照组。如表 4 所示，相较于其他方法，本文方法不论是在 mAP 评价标准上还是在 rank-1、rank-5 以及 rank-10 精度上均取得最高的分数。这是因为 IBN-Net 网络能够同时在 IN 和 BN 之间取得平衡，而 GeM 池化方法能够有效结合最大池化与平均池化两者的优势，从而提升特征提取的质量。

表 4　特征提取网络消融实验

Table 4　Ablation study on feature extraction network

(%)

网络	mAP	Rank-1	Rank-5	Rank-10
ResNet50 + avg	76.0	88.6	94.7	96.0
ResNet50 + gem	79.8	91.1	95.9	97.4
IBN-Net + avg	78.6	90.8	96.1	97.2
IBN-Net + gem	**81.9**	**92.6**	**96.9**	**98.0**

为了更直观的体现本文方法的有效性，将不同对比方法在训练过程中产生的聚类个数进行了可视化如图 3 所示。图 3 中横坐标表示迭代训练的次数，纵坐标表示聚类产生簇的总数。Ground Truth 直线代表实际标签的个数，从图 3 能够看出本文所提出的方法最接近实际标签个数。从而验证了 IBN-Net 结合 GeM 提取出的特征质量更佳，更有利于提高后续聚类算法的准确度。

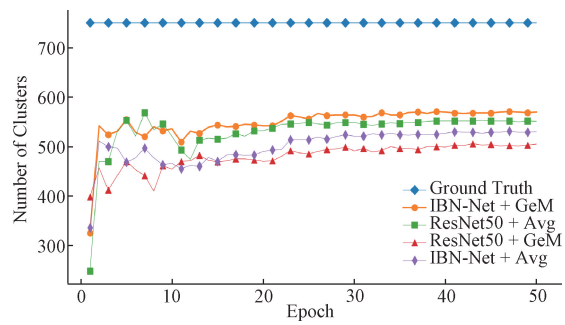


图 3　不同方法训练过程中聚类数量变化曲线

Fig. 3　Cluster number variation curves during training with different methods

2)聚类方法消融实验

为降低 DBSCAN 聚类算法对于人工选择的超参数的

敏感程度,本文使用 OPTICS 辅助 DBSCAN 自动选取超参数的方式。为了验证这种方法的合理性与有效性,本文将仅使用 DBSCAN 算法的方式作为对照组与使用了 OPTICS 结合 DBSCAN 的方式进行对比如表 5 所示。

由表 5 的实验结果可以看出,使用了 OPTICS 辅助 DBSCAN 聚类的方式能够有效提升无监督行人重识别网络整体的准确度。这是因为 DBSCAN 算法的性能依赖于人工设置的超参数,在不同的数据分布下容易出现不稳定的聚类结果。而 OPTICS 算法通过密度可达性进行排序,能够自适应不同密度的数据分布,从而帮助 DBSCAN 自动选择最优的超参数,减少了人工干预,有效提升了聚类性能。

表 5　聚类方法消融实验

Table 5　Ablation study on clustering methods　(%)

聚类方法	mAP	Rank-1	Rank-5	Rank-10
DBSCAN	80.4	91.9	96.8	97.9
DBSCAN+OPTICS	**81.9**	**92.6**	**96.9**	**98.0**

3)记忆字典更新方式消融实验

为了确保在聚类更新的过程中,对于规模不同的簇,其内部行人特征的更新速率能够保持一致。本文提出了一种在聚类级别上对记忆字典进行初始化与更新的方法。本文使用了在实例级别上对内存字典进行初始化与更新的方式作为实验对照组进行消融实验。实验结果如表 6 所示,本文的记忆字典的初始化与更新方式更有助于提升无监督行人重识别模型的整体性能。这是因为相较于实例级别的更新方式,在聚类级别上进行更新时,能够确保不同规模的簇内部行人特征的更新速率保持一致,减少了由簇的大小不同而导致规模较小的簇过度更新,规模较大的簇更新不足的问题。

表 6　记忆字典更新方式消融实验

Table 6　Ablation study on memory dictionary update methods　(%)

更新方式	mAP	Rank-1	Rank-5	Rank-10
Instance level	70.7	86.7	94.3	96.2
Cluster level	**81.9**	**92.6**	**96.9**	**98.0**

4)记忆字典初始化方式消融实验

本文在对记忆字典初始化的过程中除了考虑聚类质心,还将离群值视为单独的聚类加入到了记忆字典的初始化过程。这样做的目的是为了更加充分的利用训练集中的所有数据,同时充分发掘隐藏在这些离群值中的难样本数据信息,从而使得最终的无监督行人重识别模型更具鲁棒性。

将只利用聚类质心来进行记忆字典初始化的方法作为对照组与同时利用聚类质心与离群值进行记忆字典初始化的方法进行比较,结果如表 7 所示。可以看出在将离群值也同样视为单独的聚类连同聚类质心一并纳入到记忆字典初始化过程中的方法对于无监督行人重识别网络准确率的提升起到了一定的帮助。这是因为离群值往往代表复杂的难样本数据,这些样本通常在特征空间中分布较远,难以被简单的聚类方法捕获。将离群值视为单独的聚类,能够有助于重识别模型更好地学习这些难样本数据的特征。

表 7　记忆字典初始化方式消融实验

Table 7　Ablation study on memory dictionary initialization methods　(%)

初始化方式	mAP	Rank-1	Rank-5	Rank-10
Cluster centroids	80.8	91.3	95.3	96.7
Cluster centroids+Outliers	**81.9**	**92.6**	**96.9**	**98.0**

3　结　论

本文提出了一种基于深度聚类学习的无监督行人重识别方法。首先,本文提出了使用 IBN-Net 网络对行人图像进行特征提取以提升特征提取的质量。其次,利用在聚类级别上进行初始化与更新的记忆字典消除了聚类更新速率不一致的问题,并使聚类产生的离群值也参与到了记忆字典的初始化与更新操作中,使得训练集数据能够被充分利用。同时,利用 OPTICS 辅助 DBSCAN 进行聚类,使其不再过度依赖于人工选择超参数。本文方法在多个行人重识别数据集上均取得了良好的实验结果,证明了本文方法在无监督行人重识别任务中的有效性。在未来的工作中,将考虑使用一些辅助信息,如相机标签来细化聚类过程,进一步减弱同一标签的行人由不同摄像机拍摄所得到的图像可能具有较大风格差异的现象,从而获得更好的聚类效果。

参考文献

[1] YE M, LIANG CH, WANG ZH, et al. Specific person retrieval via incomplete text description[C]. Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, 2015: 547-550.

[2] YANG F X, ZHONG ZH, LUO ZH M, et al. Joint noise-tolerant learning and meta camera shift adaptation for unsupervised person re-identification[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 4855-4864.

[3] ZHAI Y, LU S, YE Q, SHAN X, et al. Ad-cluster: Augmented discriminative clustering for domain adaptive person re-identification[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020: 9021-9030.

[4] ZHONG CH Y, QI G Q, MAZUR N, et al. A domain

adaptive person re-identification based on dual attention mechanism and camstyle transfer[J]. Algorithms, 2021, 14(12): 361.

[5] GE Y, ZHU F, CHEN D, et al. Self-paced contrastive learning with hybrid memory for domain adaptive object re-id[C]. Advances in Neural Information Processing Systems, 2020: 11309-11321.

[6] WANG D K, ZHANG SH L. Unsupervised person re-identification via multi-label classification [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020: 10981-10990.

[7] 程思雨, 陈莹. 基于 ViT 的细粒度特征增强无监督行人重识别方法[J]. 电子测量与仪器学报, 2024, 38(9): 24-35.
CHENG S Y, CHEN Y. Fine-grained feature enhancement unsupervised person re-identification method based on ViT[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(9): 24-35.

[8] 贺晓东, 王春艳, 孙昊,等. 基于局部特征与视点感知的车辆重识别算法[J]. 仪器仪表学报, 2022, 43(10): 177-184.
HE X D, WANG CH Y, SUN H, et al. Local-features and viewpoint-aware for vehicle re-identification [J]. Chinese Journal of Scientific Instrument, 2022, 43(10): 177-184.

[9] LI M K, LI CH G, GUO J. Cluster-guided asymmetric contrastive learning for unsupervised person re-identification [J]. IEEE Transactions on Image Processing, 2022, 31(14): 3606-3617.

[10] 熊明福, 肖应雄, 陈佳, 等. 二次聚类的无监督行人重识别方法[J]. 计算机工程与应用, 2024, 60(1): 227-235.
XIONG F M, XIAO Y X, CHEN J, et al. An unsupervised pedestrian re-identification method based on secondary clustering [J]. Computer Engineering and Applications, 2024, 60(1): 227-235.

[11] PAN X, LUO P, SHI J, et al. Two at once: Enhancing learning and generalization capacities via ibn-Net[C]. Proceedings of the European Conference on Computer Vision, 2018: 464-479.

[12] HE K M, ZHANG X, REN SH Q, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[13] RADENOVIC F, TOLIAS G, CHUM O. Fine-tuning CNN image retrieval with no human annotation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(7): 1655-1668.

[14] ZHENG L, SHEN L Y, TIAN L, et al. Scalable person re-identification: A benchmark[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1116-1124.

[15] WEI L H, ZHANG SH L, GAO W, et al. Person transfer GAN to bridge domain gap for person re-identification[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 79-88.

[16] SUN X, ZHENG L. Dissecting person re-identification from the viewpoint of viewpoint[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 608-617.

[17] LI J N, ZHANG SH L. Joint visual and temporal consistency for unsupervised domain adaptive person re-identification[C]. Computer Vision – ECCV, 2020: 483-499.

[18] CHEN H, WANG Y H, LAGADEC B, et al. Joint generative and contrastive learning for unsupervised person re-identification[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2021: 2004-2013.

[19] ZHANG X, GE Y X, QIAO Y, et al. Refining pseudo labels with clustering consensus over generations for unsupervised object re-identification[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2021: 3436-3445.

[20] CHEN H, LAGADEC B, BREMOND F. ICE: Inter-instance contrastive encoding for unsupervised person re-identification[C]. Proceedings of the IEEE International Conference on Computer Vision, 2021: 14960-14969.

作者简介

邓子文,2022 年于沈阳工业大学获得学士学位,现为沈阳工业大学硕士研究生,主要研究方向为计算机视觉、深度学习。
E-mail: dzw@ smail. sut. edu. cn

Deng Ziwen received his B. Sc. degree from Shenyang University of Technology in 2022. Now he is a M. Sc. candidate at Shenyang University of Technology. His main research interests include computer vision and deep leaning.

段勇(通信作者),沈阳工业大学教授,博士生导师,主要研究方向为自主机器人、机器学习、计算机视觉。
E-mail: duanyong0607@ 126. com

Duan Yong (Corresponding author) is a professor and Ph. D. supervisor at Shenyang University of Technology. His main research interests include autonomous robot, machine learning and computer vision.