

一种基于注意力的无监督行人重识别方法^{*}

胡玉玲^{1,2} 王鑫依^{1,2} 张 一^{1,2} 邹伟光¹

(1. 北京建筑大学智能科学与技术学院 北京 102616; 2. 城市建筑超级智能技术北京市重点实验室 北京 102616)

摘 要: 行人重识别是跨摄像头非重叠域中对相同行人的检索和识别,针对跨摄像头不同域之间的特征差异以及聚类阶段生成的伪标签噪声问题,提出一种基于注意力的无监督行人重识别方法。在特征提取阶段,采用一种自适应图通道-空间注意力模块(AGCBAM),同时考虑通道和空间两个维度,通过自适应调整通道权重来适应跨域特征分布,同时关注到特定空间位置特征来实现细节信息的捕捉;在模型训练阶段,提出改进的类内邻近空间注意力(INSA)模块,将标签平滑和正实例之间的空间互补关系相结合,有效去除伪标签噪声,使模型更好地学习数据的真实分布。通过对 2 个主流数据集 Market-1501 和 MSMT17 进行实验,对比了现有的一些常用算法,模型在 mAP 和 Rank-1 精度上均有提升,验证了本文所提方法的有效性。

关键词: 行人重识别;无监督;注意力

中图分类号: TP391.41;TN919.8 **文献标识码:** A **国家标准学科分类代码:** 520.99

An attention-based unsupervised approach to pedestrian re-identification

Hu Yuling^{1,2} Wang Xinyi^{1,2} Zhang Yi^{1,2} Zou Weiguang¹

(1. College of Intelligent Science and Technology, Beijing University of Civil Engineering and Architecture, Beijing 102616, China;

2. Beijing Key Laboratory of Super Intelligent Technology for Urban Architecture, Beijing 102616, China)

Abstract: Pedestrian re-identification is used to retrieve and recognize the same pedestrian in the non-overlapping fields of cross-camera. Aiming at the feature differences between different fields of cross-camera and the pseudo-label noise generated in the clustering stage, this paper proposes an attention-based unsupervised pedestrian re-identification method. In the feature extraction stage, an adaptive graph channel-spatial attention module (AGCBAM) is proposed, which considers both channel and spatial dimensions, adapts to cross-camera feature distribution by adaptively adjusting channel weights, and pays attention to specific spatial location features to capture details. In training stage, an improved intra-class neighbor spatial attention module is proposed, which combines label smoothing and spatial-level connections of positive instances to better remove pseudo-label noise and enable the model to better learn the real distribution of data. Through experiments on two mainstream datasets, Market-1501 and MSMT17, some existing common algorithms are compared, and the accuracy of mAP and Rank-1 is improved, which verifies the effectiveness of the proposed method.

Keywords: person re-identification; unsupervised; attention

0 引 言

行人重识别 (pedestrian re-identification, pedestrian Re-ID) 的目的是在多个非重叠视域所拍摄的一组图像中检索到特定的行人,是图像检索中的一个重要问题。行人重识别可以和行人检测、行人跟踪等技术相结合,广泛应用于视频监控^[1]、智能交通^[2]等诸多领域。随着深度学习和深度卷积神经网络的蓬勃发展,行人重识别算法取得了

一系列研究成果。根据数据集标注方式和模型训练策略的不同,行人重识别算法可以分为有监督和无监督两大类^[3]。有监督行人重识别使用带有标注的数据集进行训练,取得了较为理想的性能,但是数量庞大的数据集图像标注工作非常消耗人力和时间,且针对特定的数据集训练,使得多种现实场景适应性不够理想。因此,避免高成本人工标注并应用于现实场景的无监督方法成为了研究热点。无监督行人重识别可分为无监督域自适应 (unsupervised domain

adaptation, UDA)和纯无监督学习(unsupervised learning, USL)两种。无监督域自适应,即从标记的源域数据集中学习丰富的信息,再将学习到的信息迁移到未标记的目标域数据集上,因此模型性能很大程度上依赖于源域数据集的规模和质量。与其相比,纯无监督学习不需要标签标注工作,更能灵活适用于多种现实场景。

目前,纯无监督行人重识别大多遵循两阶段交替训练的范式,即伪标签生成的聚类阶段和使用伪标签的训练阶段交替进行,其中,伪标签的质量直接关系到后续无监督重识别模型训练的好坏以及模型的性能。因此,生成更加准确的伪标签至关重要。在聚类阶段,即使用基于密度的聚类方法(density-based spatial clustering of applications with noise, DBSCAN)或 K 均值聚类方法(K-means clustering algorithm, K-means)对提取的实例特征进行聚类以生成每一实例的伪标签。但是,由于多个摄像头下所拍摄的同一实例其受光照、姿态等变化存在很大差异,仅仅使用单一实例特征生成伪标签,其同一类的样本点间的距离大,不同类的样本点间的距离小,会导致错误标记实例,从而导致标签噪声问题。为了解决这一问题,Liu 等^[4]提出多视图相似度聚合和多级间隙优化方法来提高聚类前后伪标签的准确性。Cho 等^[5]提出一致感知标签平滑,通过平滑标签分布来改进部分特征的伪标签。还有学者通过实例选择来改进聚类算法带来的噪声问题,如 Liu 等^[6]利用聚类边缘特征来更新存储库中属于同一伪标签的所有实例特征,使实例特征分布更加均衡。在训练阶段对带有伪标签的数据集进行内存特征存储和更新以获得更好的表示,不断优化模型参数以最小化实例特征与内存特征之间的对比损失。内存特征可以分为实例级内存^[7-8]和中心级内存^[9]。针对实例级内存,Yin 等^[10]提出一种实时内存更新策略,在无动量的小批次中使用随机采样的实例特征更新聚类质心。对于中心级内存,He 等^[11]提出类内邻近空间注意力模块,通过挖掘正实例的空间互补关系来解决伪标签存在噪声和中心特征在存储器中更新状态不一致等问题。然而,上述方法中均只考虑采用一种标签平滑操作来去除伪标签噪声,如只考虑了通过每一实例的空间互补关系对生成的中心级内存进行标签平滑,但忽略了聚类阶段生成的伪标签本身存在的噪声问题对中心级内存表示带来的影响,两种不同策略间存在着良好的互补关系。

此外,行人特征对于行人重识别同样至关重要,在聚类阶段和训练阶段均需要对图像进行特征提取。由于真实环境中摄像头角度等变化使得大多数行人图像中存在许多干扰因素,会导致特征提取不充分,进而带来行人的错误识别。为了解决上述问题,提高骨干网络的特征提取能力,马建红等^[12]通过引入双分支注意力特征融合结构,融合域不变特征和域特定特征,进而提升模型的泛化能力并减少聚类噪声。考虑到人体不同身体部位之间的局部特征联系对模型性能的影响,李萌等^[13]通过局部融合模块融合人体不

同部位间的关联特征,并结合最大池化操作和局部特征融合来提取更具辨别力的人体全局特征。陈元妹等^[14]则提出特征细化信息融合模块以准确定位存在行人的潜在区域。上述方法中重视局部和深层次的特征,可能会导致提取到的行人特征中包含较多的无关特征,难以充分表征行人,易导致聚类过程中引入标签噪声,存在对外观相似的行人难以区分的问题。此外,上述方法中的注意力机制大多采用固定的计算方式来确定关键区域,缺乏根据实时输入的数据进行动态调整权重以灵活应对复杂多变场景的能力。

因此,受到上述无监督行人重识别方法的启发,为了使得网络更好地表达跨域变化下图像的特征并更好地消弭伪标签噪声,本文在提出一种基于注意力的无监督行人重识别方法。

1 本文方法

1.1 整体框架

考虑到多摄像头的同一实例受光照、姿态等变化导致的巨大差异虽然会给聚类阶段生成的伪标签带来噪声,但也正是由于多摄像头拍摄角度的不同可以从多个角度为同一实例提供更好的互补信息,因此本文在纯无监督行人重识别方法^[7,9]基础上提出了先对聚类阶段生成的伪标签进行正则化以消除一定噪声,再利用空间互补信息进行伪标签的细化。此外,考虑到纯无监督行人重识别方法^[7,9]的特征提取阶段均采用 ResNet50 网络,而仅仅利用 ResNet50 网络处理这类复杂的视觉信息时存在局限性,如进行特征提取难以有效应对因摄像头角度变化而产生的实例外观、空间关系等变化的影响,对于因视角变化而引起的关键特征的的关注能力不足,导致特征提取的准确性和鲁棒性欠佳。为解决这一问题,本文将所提出的自适应通道-空间注意力模块引入 ResNet50 网络中。

本文所提出的一种基于注意力的无监督行人重识别方法的总体框架如图 1 所示,所提出的方法交替进行聚类阶段和训练阶段。在聚类阶段,使用 ResNet50 作为骨干网络,并利用所提出的自适应通道-空间注意力模块(adaptive graph convolutional block attention module, AGCBAM)得到具有较高表征能力和判别能力的特征,然后通过 DBSCAN 方法对获取到的特征进行聚类,为每个实例生成伪标签,如图 1(a)所示。在训练阶段,由主分支和改进的类内邻近空间注意力分支组成。在主分支中,计算包含聚类中心的特征进行内存记忆库的初始化;再将数据集的批次大小分为 $p \times k$,其中 p 是行人 ID 数, k 是批次中 ID 的正实例数;接着,使用 k 个正实例特征来更新每一批次的内存记忆库中对应的中心特征,即最优特征表示;通过对正实例特征和包含聚类中心的特征记忆进行乘法运算得到预测结果。在改进的类内邻近空间注意力分支中,通过计算正实例的邻近规则来获得精细标签,聚合预测以解决标签噪声问题,如图 1(b)所示。

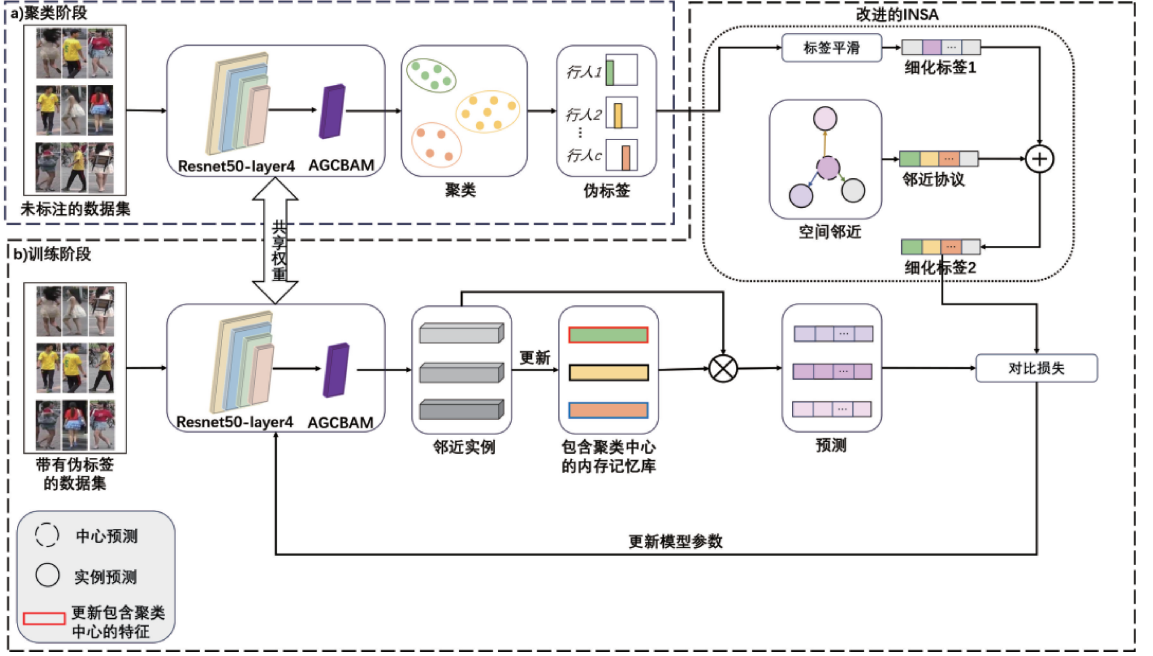


图 1 基于注意力的行人重识别研究框架

Fig. 1 Framework attention-based pedestrian re-identification

1.2 自适应通道-空间注意力模块

本文采用在 ImageNet^[15]上预训练的,性能和网络架构方面都表现出色的 ResNet50^[16]作为主干编码器提取特征,去掉 layer4 之后的所有层,在自适应通道-空间注意力模块(AGCBAM)后添加一个均值池化层、一个批归一化层^[17]和一个 l2 归一化层,以获得 2 048 维的特征。

为了使网络模型更好地应对摄像机角度变化、光照变化等带来的影响,在自适应图通道注意力模块(adaptive graph channel attention module, AGCAM)^[18]和卷积块注意力模块(convolutional block attention module, CBAM)^[19]的基础上,提出一种新的自适应通道-空间注意力模块 AGCBAM,对不同域特征分布自适应调整通道权重,同时关注特定空间位置的特征捕捉细节信息。其结构如图 2 所示,该模块分为通道和空间两部分,输入特征图 F_1 先经过 AGCAM,再经过空间注意力模块(spatial attention module, SAM)。



图 2 AGCBAM 结构

Fig. 2 Structure of AGCBAM

AGCBAM 由两部分组成,包括两个线性嵌入函数组成的特征映射层以及获取特征自适应图卷积权重的自适应图卷积模块(adaptive graph convolutional module, AGCM),其数学表达式如式(1)所示,结构如图 3 所示。

$$F_{AC} = F_1 \cdot \text{sigmoid}(F''(\text{ReLU}(\text{AGCM}(F'(\text{GAP}(x))), \mathbf{M}_i)))) \quad (1)$$

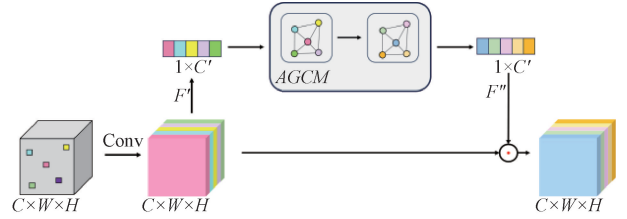


图 3 AGCAM 结构

Fig. 3 Structure of AGCAM

其中, $F'(\cdot)$ 和 $F''(\cdot)$ 为两个线性嵌入函数; \mathbf{M}_i 为包含特征之间关系的邻接矩阵; $X = F'(\text{GAP}(x))$ 是 AGCM 的输入。

AGCM 的结构如图 4 所示,输入特征映射 f_i 经过自适应平均池化和卷积层,生成对角矩阵 \mathbf{M}_d ; 结合固定的单位矩阵 \mathbf{M}_u 以及邻接矩阵 \mathbf{M}_a 计算最终的邻接矩阵 \mathbf{M}_i , 并将其与 f_i 相乘,通过卷积层和激活函数后输出特征映射 f_o , 这一过程的数学表达式如式(2)和(3)所示。

$$f_o = W_c(f_i(\mathbf{M}_u \times \mathbf{M}_d + \mathbf{M}_a)) \quad (2)$$

$$\mathbf{M}_d = T(\text{softmax}(W_c(f_i))) \quad (3)$$

其中, W_c 为一维卷积层的权重, T 为对角排列。 \mathbf{M}_a 为反向传播优化中生成的任意两个特征顶点之间的关系,其值无限制。

AGCM 输出的特征图 F_{AC} 作为空间注意力的输入,经 SAM 进行通道维度压缩,即 F_{AC} 在通道维度上分别进行平均池化和最大池化操作,将得到的两个 $H \times W \times 1$ 大

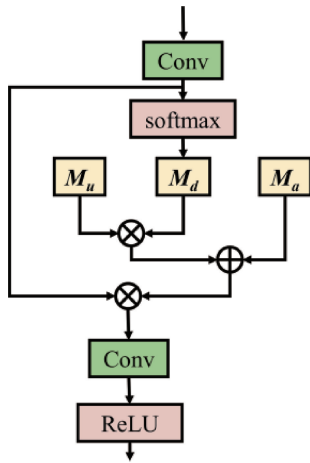


图 4 AGCM 结构

Fig. 4 Structure of AGCM

小的特征进行叠加,通过 Sigmoid 激活函数得到一个空间注意力权重 $Ms(F_{AC})$,并将 F_{AC} 与 $Ms(F_{AC})$ 相乘得到压缩后的新特征图 F_s ,这一过程的数学表达式如式(4)所示,结构如图 5 所示。

$$F_s = Ms(F_{AC}) \times F_{AC} = \sigma(f^{7 \times 7}([MaxPool(F_{AC}))]; AvgPool(F_{AC})) \times F_{AC} \quad (4)$$

其中, σ 为 Sigmoid 激活函数; $f^{7 \times 7}$ 为一个 7×7 卷积; $MaxPool$ 为最大池化; $AvgPool$ 为平均池化; F_{AC} 为经过 AGCAM 处理后的结果。

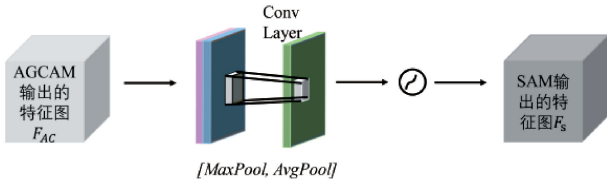


图 5 SAM 结构

Fig. 5 Structure of SAM

1.3 改进的类内邻近空间注意力

考虑到摄像头拍摄角度的不同,具有同一 ID 的不同实例提取到的语义特征会存在明显差异,因此,有相同伪标签的所有正实例中含有丰富的补充信息,可以有效地为每一实例提供更完备的特征信息,进而改进伪标签噪声带来的不准确聚类问题。此外,由于不同的伪标签噪声处理方法所考虑的因素不同,不同方法之间存在着互补作用。因此,受到 He 等^[11]在为单个实例提供补充信息方面工作的启发,在其所提出的类内邻近空间注意力(intra-class neighbor-based spatial attention, INSA)的基础上,本文引入标签平滑^[8]对聚类生成的伪标签进行特定噪声的去除,再把标签平滑后的伪标签作为后续操作的输入,挖掘正实例的空间连接来获得更加细化的标签,二者结合,通过双重标签平滑策略更好地解决了标签噪声问题。改进的类内邻近空间注意力算法为:

首先,对基于聚类方法生成的伪标签通过正则化技术,降低网络模型对一些特定标签的过度依赖,以更好的适应不同域的特征,减轻网络模型训练中数据的过拟合问题,其数学表达式如式(5)所示。

$$\hat{y}_a = y_{hot}(1 - \alpha) + \alpha / K \quad (5)$$

其中, y_{hot} 为 one-hot 编码形式的基于聚类方法生成的伪标签; $\alpha \in [0, 0.1]$ 为平滑因子, K 为总类别数。

接着,对查询实例的预测结果 p_a 进行计算,即查询实例特征 q_a 与内存中存储的所有中心特征 M 相乘,计算过程如式(6)所示。

$$p_a = q_a M \quad (6)$$

然后,计算查询实例 q_b 对应的中心预测 \bar{p}_b ,即查询实例特征对应的中心特征 c_b 与内存中存储的所有中心特征 M 相乘,如式(7)所示。

$$\bar{p}_b = c_b M \quad (7)$$

再计算每个查询实例的互补信息 S_a ,将正实例的互补关系进行连接,如式(8)所示; (\hat{N}_b 为第 b 类的一组正实例预测)。

$$S_a = \frac{p_a \bar{p}_b}{\sum_{p_a \in \hat{N}_b} p_a \bar{p}_b} \quad (8)$$

最后,基于上述互补信息 S_a ,细化经过预处理之后的伪标签 \hat{y}_a ,如式(9)所示。

$$\hat{y}_a = \beta \hat{y}_a + (1 - \beta) \sum_{p_a \in \hat{N}_b} S_a p_a \quad (9)$$

其中, $\beta \in [0, 1]$ 用来控制正则化处理后的伪标签与预测集合之间比例的权重参数。

通过上述正则化手段和正实例之间的空间连接进行信息互补,有效消除伪标签噪声,使得网络模型更好地学习到不同视域下真实的数据分布以及特征表示。

2 方法验证

2.1 数据集及评价指标

实验基于两个广泛使用的、大规模的数据集开展,分别是 Market-1501^[20]、MSMT17^[21],有关数据集的相关说明如表 1 所示。

表 1 Market-1501 和 MSMT17
Table 1 Market-150 and MSMT17

数据集	摄像头 个数	图片/ 张	ID	训练集	测试集
Market-1501	6	32 688	1 501	12 936	19 732
MSMT17	15	126 441	4 101	32 621	93 820

采用平均精度(mAP)和累积匹配特征(CMC)^[22] Rank-1(R1)、Rank-5(R5)、Rank-10(R10)来评价本文所提出的无监督重识别网络的性能。其中,mAP 的计算公式如式(10)所示;rank-n 为所有查询图像 q 的集合 Q 中,前 n

个结果中出现匹配正确的概率,其数学表达式如式(11)所示。

$$mAP = \frac{1}{Q} \sum_{q \in Q} \frac{1}{N(q)} \sum_{k=1}^{N(q)} \frac{k}{A_k(q)} \quad (10)$$

其中, $N(q)$ 为搜索到的与查询图像 q 相关的图片总数, $A_k(q)$ 为搜索到的与查询图像 q 的第 k 张相关图片的搜索精度。

$$Rank-i = \frac{\sum x_i}{N} \quad (11)$$

其中, N 为测试集的总数, x_i 为在测试集 N 中的第 i 张图片,若搜索结果中置信度从高到低排列的前 i 个图片匹配正确则为 1,反之为 0。

2.2 实验设置

实验基于 PyTorch 实现,训练使用 3 块 NVIDIA A800 GPU,测试使用 1 块 GPU。在聚类阶段,使用 DBSCAN 执行聚类以生成伪标签,将核心点的最小邻域点数设置为 4。根据 STDA^[11] 将 Market-1501 样本之间的距离阈值设置为 0.6, MSMT17 设置为 0.7。调整输入图像的大小, Market-1501 为 256×128 , MSMT17 为 384×128 。在训练阶段,使用随机水平翻转、填充像素、随机裁剪和随机擦除。batchsize 为 256,由 16 个伪 ID 和每个 ID 的 16 个图像组成。使用 Adam 优化器,训练 50 个 epoch。

2.3 方法对比

为了验证本文模型的泛化性,在 Market-1501 和 MSMT17 数据集上进行测试,将所提方法的实验结果与近几年常用的无监督行人重识别方法进行对比,其对比结果如表 2 所示,最优结果加粗表示,“—”表示原论文中没有该项结果。本文方法在 Market-1501 数据集上 Rank-1 达到了 93.5%, Rank-5 达到了 97.6%, Rank-10 达到了 98.5%, mAP 达到了 84.5%;在 MSMT17 数据集上 Rank-1 达到了 64.5%, Rank-5 达到了 75.6%, Rank-10 达到了

79.8%, mAP 达到了 36.9%。与近五年来的一些主流方法相比,本文方法在 2 个主流数据集上的识别效果都有较为明显的提高,说明了本文所提方法能够较为充分的表达行人特征且更加有效的消除伪标签噪声。其中,与在聚类阶段通过自底向上的分层聚类并在训练中使用硬批次三元组损失的 HCT^[23] 相比,本文在特征特取中引入 AGCBAM 并在训练阶段引入改进的 INSA 进行伪标签平滑,使得在 Market-1501 数据集上的 mAP 提高了 27.6%, Rank-1 提高了 13.5%;与改进训练阶段内存特征更新策略相比,即与利用时间动态变化改进伪标签的 RLCC^[24]、用同一速度更新内存特征的 ClusterContrast^[9] 和实时内存更新并改变对比损失的 RETM^[10] 相比较,本文所提方法在 Market-1501 数据集上的 mAP 分别提高了 6.8%、1.5% 和 1.5%, Rank-1 分别提高了 2.7%、0.6% 和 0.7%;在 MSMT17 数据集上的 mAP 分别提高了 9%、3.9% 和 4.1%, Rank-1 分别提高了 8%、2.5% 和 7.4%;与改进训练阶段内存特征的方法相比,即与混合存储特征的对比学习 SpCL^[7]、利用局部特征和全局特征来进行伪标签精细化的 PPLR^[5] 和全局关系感知对比学习方法 GRACL^[25] 相比较,本文所提方法在 Market-1501 数据集上的 mAP 分别提高了 11.4%、3% 和 0.8%, Rank-1 分别提高了 5.4%、0.7% 和 0.3%;在 MSMT17 数据集上的 mAP 分别提高了 17.8%、5.5% 和 2.3%, Rank-1 分别提高了 22.2%、3.4% 和 0.5%;与在训练阶段细化伪标签的方法相比,即与引入类内邻近空间注意(INSA)模块以及类间序列的时态注意(ISTA)模块的 STDA^[11] 和具有自适应选择策略的训练网络 FUREID^[26] 相比较,本文所提方法在 Market-1501 数据集上的 mAP 分别提高了 1.8% 和 2.8%, Rank-1 分别提高了 0.4% 和 0.8%;在 MSMT17 数据集上的 mAP 分别提高了 5.1% 和 10.7%, Rank-1 分别提高了 1.9% 和 11.9%。

表 2 与常用的无监督行人重识别方法对比

Table 2 Comparative experimental results with unsupervised pedestrian re-identification approaches

方法		Market-1501				MSMT17			
		mAP	R1	R5	R10	mAP	R1	R5	R10
HCT ^[23]	CVPR'20	56.9	80.0	91.6	95.2	—	—	—	—
SpCL ^[7]	NeurIPS'20	73.1	88.1	95.1	97.0	19.1	42.3	55.6	61.2
RLCC ^[24]	CVPR'21	77.7	90.8	96.3	97.5	27.9	56.5	68.4	73.1
PPLR ^[5]	CVPR'22	81.5	92.8	97.1	98.1	31.4	61.1	73.4	77.8
GRACL ^[25]	TCSVT'22	83.7	93.2	97.6	98.6	34.6	64.0	75.2	79.3
ClusterContrast ^[9]	ACCV'22	83.0	92.9	97.2	98.0	33.0	62.0	71.8	76.7
RETM ^[10]	TIP'23	83.0	92.8	97.4	98.3	32.8	57.1	70.0	74.9
STDA ^[11]	TITS'23	82.7	93.1	97.3	98.4	31.8	62.6	73.4	77.5
FUREID ^[26]	Pattern Recognite'24	81.7	92.7	97.2	98.9	26.2	52.6	64.1	68.6
本文		84.5	93.5	97.6	98.5	36.9	64.5	75.6	79.8

2.4 消融实验

为了验证本文所提出的 AGCBAM 以及改进的 INSA 模块的有效性,在 Market-1501 数据集上实施了消融实验。借鉴将对比学习应用到训练中的较为常用的 USL 行人重识别方法^[7,9],采用具有对比损失的中心级内存纯无监督行人重识别方法框架作为基线。实验结果如表 3 所示。

表 3 消融实验
Table 3 Results of ablation experiment

方法	Market-1501			
	mAP	R1	R5	R10
Baseline	80.0	91.0	96.2	97.3
Baseline+AGCBAM	84.4	93.4	97.6	98.5
Baseline+改进的 INSA	84.0	93.1	97.1	98.1
本文	84.5	93.5	97.6	98.5

1)改进的 INSA 的有效性:通过对表 3 中的 Baseline+改进的 INSA 与 Baseline 对比,改进的 ISNA 在 Market-1501 上 mAP 提高了 4.0%,Rank-1 提高了 2.1%,表明所提出的改进的 INSA 可以有效去除噪声,并为细化标签提供更为丰富的正实例间互补信息。通过对表 4 中 Baseline+改进的 INSA 与 Baseline+ INSA 的对比,改进的 ISNA 在 Market-1501 上 mAP 提高了 0.9%,Rank-1 提高了 0.8%,表明改进的 INSA 两种不同策略相互补充,可以有效提升去除标签噪声的能力。

表 4 INSA 与改进的 INSA 的对比
Table 4 Comparative experimental results with INSA and improve INSA

方法	Market-1501			
	mAP	R1	R5	R10
Baseline+ INSA	83.1	92.3	96.9	97.8
Baseline+改进的 INSA	84.0	93.1	97.1	98.1

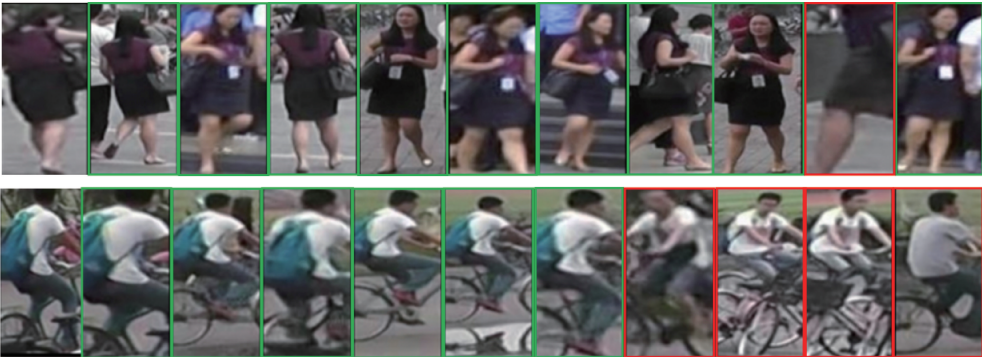
2)AGCBAM 的有效性:通过对表 3 中的 Baseline+AGCBAM 与 Baseline 对比,ResNet50 引入 AGCBAM 后,在 Market-1501 上 mAP 提高了 4.4%,Rank-1 提高了 2.4%,表明所提出的 AGCBAM 可以自适应地调整通道权重来适应不同视域的特征分布,同时还能聚焦特定空间位置的特征,更好地捕捉图片中的细节。此外,在 Baseline+改进的 INSA 的基础上,分别在 ResNet50 中引入 AGCBAM 和 CBAM。表 5 结果表明,所提出的 AGCBAM 在 mAP 和 Rank-1 上分别提高了 0.8% 和 0.5%。这表明,结合自适应卷积注意力和空间注意力,使得模型能更好的适应不同域的特征分布,更好的关注重点区域。

表 5 AGCBAM 与 CBAM 的对比
Table 5 Comparative experimental results with AGCBAM and CBAM

方法	Market-1501			
	mAP	R1	R5	R10
Baseline+改进的 INSA+CBAM	83.7	93.0	97.3	98.3
Baseline+改进的 INSA+AGCBAM	84.5	93.5	97.6	98.5

2.5 结果可视化

为了更加直观地分析所提出的重识别方法的效果,在 Market-1501 数据集上随机选取 2 位行人进行可视化实验,实验结果如图 6 所示。第 1 列为 query 图像,代表随机选取的待查询行人图像,右侧的重识别检索结果分别表示为 Rank-1~Rank-10。图像红色框代表图库实例与查询实例具有不同的标识,绿色框代表图库实例与查询实例具有相同的标识。背向行人特征难度较大,且存在背包、自行车等物体的干扰,在 Baseline 方法中出现 5 个错误识别结果,如图 6(a)所示;而本文方法中出现了 3 个错误识别结果,如图 6(b)所示。同时,通过四幅图的对比,所提方法可以更好地克服摄像头角度不同所带来的行人姿态变化和物体遮挡等情况,聚焦关键信息进而提取更具判别力的行人特征的同时,能够有效抑制噪声干扰。



(a) Baseline



图6 在Market-1501数据集上的可视化对比实验结果

Fig. 6 Visual comparison of experimental results on Market-1501 datasets

3 结 论

本文提出了一种基于注意力的无监督行人重识别方法,主要工作包括:在特征提取阶段引入自适应通道-空间注意力模块(AGCBAM),以级联的方式融合自适应图卷积注意力模块(AGCAM)和空间注意力模块(SAM),对不同视域的特征进行有选择性的增强,根据输入的数据特征动态调整通道权重使模型更好地应对不同域的特征分布,同时关注到特定空间位置的特征,捕捉细节信息。在训练阶段,引入标签平滑和类内邻近空间注意力(INSa),对聚类阶段生成的伪标签进行正则化处理以消除特定的标签噪声,再通过挖掘正实例间的互补信息进行空间连接,进一步细化伪标签,二者协同效应使得模型更好地学习到真实数据分布。通过在主流数据集上的对比实验,相较于现有的常用算法,在mAP和Rank-1精度上均有提升。该研究对提高行人重识别具有现实应用价值。

参考文献

- [1] JIN X, LAN C L, ZENG W J, et al. Global distance-distributions separation for unsupervised person re-identification [C]. Computer Vision-ECCV 2020, 2020: 735-751.
- [2] LU Z F, LIN R H, LOU X L, et al. Identity-unrelated information decoupling model for vehicle re-identification [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(10): 19001-19015.
- [3] 王素玉,肖塞.行人重识别研究综述[J].北京工业大学学报,2022,32(10):1100-1112.
WANG S Y, XIAO S. A review of research on pedestrian re-identification [J]. Journal of Beijing University of Technology, 2022, 32(10): 1100-1112.
- [4] LIU T, CHENG SH L, DU A Y. Multi-view similarity aggregation and multi-level gap optimization

for unsupervised person re-identification[J]. Expert Systems with Applications, 2024, 256(1): 124924.

- [5] CHO Y, KIM W J, HONG S, et al. Part-based pseudo label refinement for unsupervised person re-identification [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022: 7298-7308.
- [6] LIU Y X, GE H W, SUN L, et al. Camera-aware progressive learning for unsupervised person re-identification[J]. Neural Computing and Applications, 2023, 35: 11359-11371.
- [7] GE Y X, ZHU F, CHEN D P, et al. Self-paced contrastive learning with hybrid memory for domain adaptive object Re-ID [C]. The 34th International Conference on Neural Information Processing Systems (NIPS '20), 2020: 11309-11321.
- [8] WANG D K, ZHANG SH L. Unsupervised person re-identification via multi-label classification [J]. International Journal of Computer Vision, 2022, 130: 2924-2939.
- [9] DAI Z ZH, WANG G Y, YUAN W H, et al. Cluster contrast for unsupervised person re-identification[C]. Asian Conference on Computer Vision, 2022: 319-337.
- [10] YIN J H, ZHANG X Y, MA ZH Y, et al. A real-time memory updating strategy for unsupervised person re-identification [J]. IEEE Transactions on Image Processing, 2023, 32: 2309-2321.
- [11] HE Q L, WANG Z H, ZHENG ZH J, et al. Spatial and temporal dual-attention for unsupervised person re-identification[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(2): 1953-1965.
- [12] 马建红,靳岩,王亚辉,等.基于双分支注意力特征融合的跨域行人重识别[J].郑州大学学报(理学版),2025:

- 1-7, DOI: 10.13705/j.issn.1671-6841.2024031.
MA J H, JIN Y, WANG Y H, et al. Cross-domain pedestrian re-identification based on dual-branch attention feature fusion[J]. Journal of Zhengzhou University(Natural Science Edition), 2025:1-7, DOI: 10.13705/j.issn.1671-6841.2024031.
- [13] 李萌,徐麟,宋伟宁.视听资料中的目标人物重识别方法[J].电子测量技术,2022,45(19):19-24.
LI M, XU L, SONG W N. Target person re-identification method in audio-visual data[J]. Electronic Measurement Technology, 2022, 45(19): 19-24.
- [14] 陈元妹,王凤随,王路遥.细化特征引导对抗性解纠缠学习的无监督行人重识别[J].电子测量与仪器学报, 2024,38(5):130-138.
CHEN Y M, WANG F S, WANG L Y. Unsupervised pedestrian re-identification via refined feature-guided adversarial disentanglement learning[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(5): 130-138.
- [15] DENG J, DONG W, SOCHER R, et al. ImageNet: A large-scale hierarchical image database[C]. IEEE Conference on Computer Vision and Pattern Recognition,2009:248-255.
- [16] HE K M, ZHANG X Y, REN SH Q, et al. Deep residual learning for image recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR),2016:770-778.
- [17] LOFFE S, CHRISTIAN S. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]. International Conference on Machine Learning, 2015: 448-456.
- [18] XIN X, WANG Z, ZHANG J, et al. AGCA: An adaptive graph channel attention module for steel surface defect detection[J]. IEEE Transactions on Instrumentation and Measurement, 2023,72: 1-12.
- [19] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]. Computer Vision-ECCV 2018, 2018: 3-19.
- [20] ZHENG L, SHEN L Y, TIAN L, et al. Scalable person re-identification: A benchmark [C]. International Conference on Computer Vision(ICCV), 2015: 1116-1124.
- [21] WEI L H, ZHANG SH L, GAO W, et al. Person transfer GAN to bridge domain gap for person re-identification [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 79-88.
- [22] ZHENG ZH D, ZHENG L, YANG Y. A discriminatively learned CNN embedding for person reidentification[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2018, 14(1): 1-20.
- [23] ZENG K W, NING M N, WANG Y H, et al. Hierarchical clustering with hard-batch triplet loss for person re-identification[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 13654-13662.
- [24] ZHANG X, GE Y X, QIAO Y, et al. Refining pseudo labels with clustering consensus over generations for unsupervised object reidentification [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 3435-3444.
- [25] ZHANG H W, ZHANG G Q, CHEN Y H, et al. Global relation-aware contrast learning for unsupervised person re-identification [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(12): 8599-8610.
- [26] PENG J J, YU J Z, WANG CH J, et al. Adapt only once: Fast unsupervised person re-identification via relevance-aware guidance[J]. Pattern Recognition, 2024, 150: 110360.

作者简介

胡玉玲(通信作者),博士,教授,主要研究方向为应急疏散、目标检测、行人重识别、风险评估等。

E-mail:huyuling@bucea.edu.cn

王鑫依,硕士研究生,主要研究方向为目标检测、行人重识别、多摄像头行人跟踪等。

E-mail:2108550022037@stu.bucea.edu.cn

张一,硕士研究生,主要研究方向为目标检测、行人重识别等。

E-mail:1823373827@qq.com

邹伟光,硕士研究生,主要研究方向为目标检测、行人车辆识别、风险预测等。

E-mail:648999586@qq.com