

# 基于注意力和标签自适应的跨域行人重识别<sup>①</sup>

陈思文<sup>②</sup> 吴怀宇<sup>③</sup> 陈 洋

(武汉科技大学机器人与智能系统研究院 武汉 430081)

**摘 要** 针对现有的大多数行人重识别算法都依赖于监督训练,而监督训练中人工标注的数据需要昂贵的资源开销从而限制了其在新场景中拓展应用的问题,提出了基于注意力和标签自适应的跨域行人重识别方法。该方法首先对深度卷积神经网络(DCNN)中不同深度的特征层嵌入注意力机制和 BNNeck 模块,增强模型在不同数据集下对行人的特征表示能力;其次针对没有任何标签的目标数据集,提出了无监督标签自适应方法,将标签信息逐渐扩展至目标数据集中;最后采用知识蒸馏(KD)的方法不断对模型进行微调,使模型逐渐适应新的场景。该方法在 Market-1501 数据集上的平均精度均值(mAP)为 33.1%,在 DukeMTMC-reID 数据集上的 mAP 为 36.1%,与 PTGAN、IPGAN 等跨域行人重识别算法相比性能有明显提升。

**关键词** 跨域行人重识别;注意力;标签自适应;知识蒸馏(KD);深度卷积神经网络(DCNN)

## 0 引 言

近年来,随着社会公共安全的需要,监控摄像头大量普及,随之而来的是庞大的视频监控数据,通过人工观察不同监控视频来查找指定行人需要耗费大量的人力物力,因此,基于跨摄像头的行人身份重识别成为视频分析工作中一个重要的研究课题。目前,大量的行人重识别方法<sup>[1-4]</sup>都是基于深度学习模型来实现的,旨在通过在目标场景下大量的带标签数据进行监督训练,获得在目标场景下行人外观特征的最佳视觉表示,从而实现对行人身份的有效识别。

然而,当模型在没有训练过的目标场景中应用时,其性能会出现大幅度下跌,这是因为不同场景下的拍摄视角、相机参数、光照环境以及图片质量等因素的不同,导致行人数据分布在域间存在巨大差异。例如,公开数据集 Market-1501<sup>[5]</sup>采集的行人图像分辨率较高、光照条件好、色彩鲜明,而 DukeMTMC-

reID<sup>[6]</sup>采集的行人图像分辨率较低、色彩相对较暗、背景复杂,这种显著的图像差异导致了同一模型在两个不同数据集上性能的巨大落差。

为增强行人重识别模型在跨域任务中的性能,文献[7]提出了源域相机到目标域相机的图像转换方法。该方法不直接在源域和目标域之间进行图像样式转换,而是将每一个相机看成是一个独立的子域,利用 StarGAN<sup>[8]</sup>的多域图像到图像的转换方法,生成带有源域身份关联的目标相机域样式的图像,通过在生成的带有身份关联的图像上学习目标域行人身份特征表示。但这种方法生成的图像并不能完全替代真实的图像,可能导致生成的图像细节不利于行人重识别任务。文献[9]通过 KMeans 对目标数据集进行样本聚类,然后选择离聚类中心较近的样本来训练模型,但是该方法需要目标数据集的类别数量。文献[10]使用了最大均值差异(maximum mean discrepancy, MMD)来寻找域不变特征空间表示,以此提高模型在目标域上的性能。文献[11]结

① 国家自然科学基金(62073250,62003249)和湖北省重点研发计划(2020BAB021)资助项目。

② 男,1996年生,硕士;研究方向:计算机视觉;E-mail: chsw1319268@163.com。

③ 通信作者,E-mail: wuhy@wust.edu.cn。

(收稿日期:2021-01-22)

合源域的有标签数据和目标域的无标签数据,提出了基于生成对抗网络架构的无监督域自适应方法,该方法通过无监督的方式学习从一个域到另一个域的映射关系。但是,这些方法仅考虑了源域和目标域之间的差异,而没有考虑源域与目标域之间的相关性。

为充分挖掘源域数据集与目标域数据集之间的相关性,进一步提高模型在跨域问题上的性能,本文提出了融合标签自适应和注意力机制的跨域行人重识别方法。通过引入注意力机制和 BNNeck 模块改进模型,以提取行人更深层次的域不变语义特征,从而提高模型在不同数据集上的相关性表达能力。同时提出了无监督标签自适应方法(unsupervised label adaptation, ULA),利用不同数据集之间的相关性,挖掘出目标域数据集上的可用数据对模型进行微调,从而进一步提高模型在跨域行人重识别任务上的性能。

## 1 方法设计

### 1.1 系统框架

提出的融合注意力机制和标签自适应的跨域行

人重识别系统架构如图 1 所示,主要分为 3 个部分,即基于注意力机制的行人重识别模型、无监督标签自适应以及基于知识蒸馏(knowledge distillation, KD)的参数更新。在训练期间,完全标注的源域数据集  $\{X_s, L(X_s)\}$  被用来预训练基于注意力机制的行人重识别模型,提出的无监督标签自适应方法在没有任何标注的目标域数据集  $X_t$  上通过选择可信度高的样本来微调中间模型,接着通过知识蒸馏的方法将中间模型参数加权并更新到原模型中,使原模型在获得新知识的同时,不至于出现对旧知识的灾难性遗忘。随着更多的样本不断加入到模型的训练中,模型在目标数据集中的性能将逐步得到提升。其中,  $X_s$  和  $X_t$  分别表示源域和目标域的图片,  $L(X_s)$  表示源域图像的身份,源域数据集中任意行人  $x_{s,i}$  与其身份标签  $l_{s,i}$  相对应。

具体来说,首先,在行人重识别模型的不同深度特征层中嵌入注意力机制和 BNNeck 模块,注意力机制可以提高模型在不同尺度特征图上对通道和空间的感知能力,从而学到更深层次的域不变语义特征,增强模型在不同数据集上的特征表示能力;而 BNNeck 模块通过在三元组损失和标签分类损失之间加入批归一化层,可以有效抑制标签分类损失对

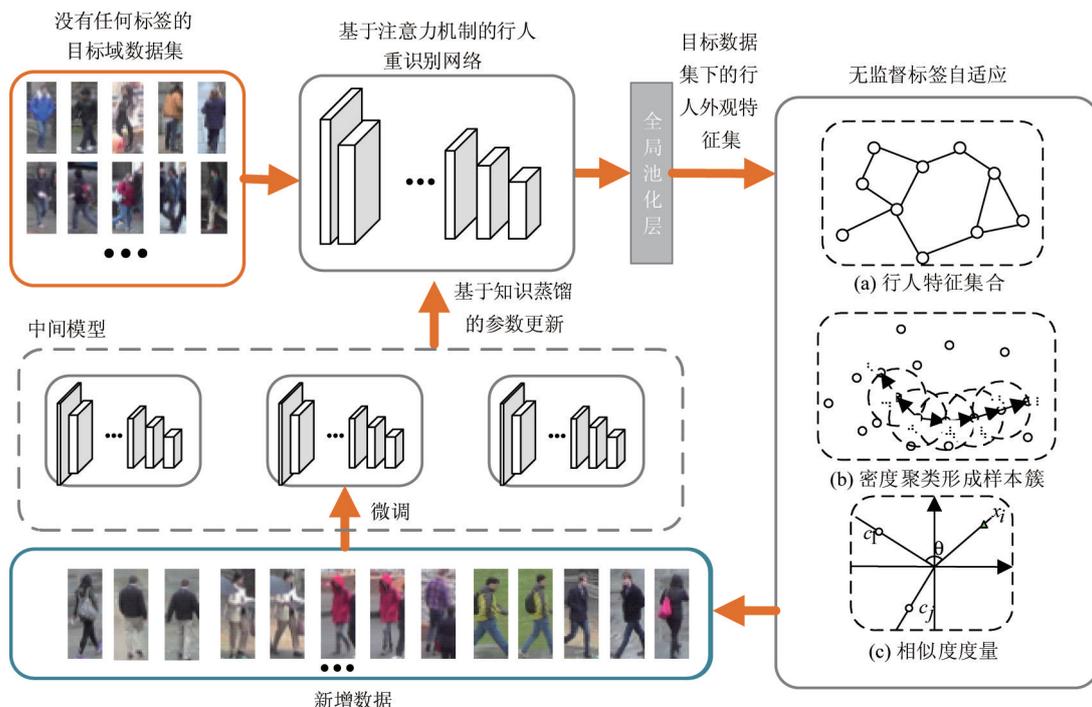


图 1 基于注意力和无监督标签自适应的跨域行人重识别系统框架

三元组损失的影响,使最终得到的特征分布更能满足相同身份相互靠近、不同身份互相远离的特性。其次,由于源域与目标域数据集之间的相关性,模型在目标域数据集上有一定的特征表示能力,因此提出标签自适应方法,通过在目标数据集上提取行人特征,构建行人特征库,再选择出行人特征库中分布稠密且形状任意的样本簇,并为其分配伪标签,从而可以获得目标数据集中一部分可信度高的样本数据。接着,通过中心损失,在选择的样本数据上学习相同行人在不同摄像头下的共性特征作为身份区分特征模板,对行人与身份区分特征模板的相似度比对,将标签信息自适应扩展至满足阈值的行人上,形成新的数据集。最后,通过知识蒸馏的方法,在保留旧知识不会出现灾难性遗忘的同时,学习新数据集上的知识,逐步改善模型在目标数据集上的性能。随着模型性能的改善,更多的样本数据将通过标签自适应方法添加到模型的训练中,从而进一步提高模型在目标数据集中的性能。

### 1.2 基于注意力机制的行人重识别网络

现有研究已经表明单域行人重识别模型在跨域问题上表现较差,受文献[4]启发,本文通过引入融合通道注意力和空间注意力机制的注意力模块和BNNeck<sup>[12]</sup>模块对常用的基线模型(Resnet50<sup>[13]</sup>)进行改进,以提高模型在跨数据集任务上的初始性能,从而进一步挖掘源域和目标域数据集之间的相关

性。如图2所示,常用的Resnet50网络主要由5个模块组成,分别是Conv1、Conv2\_x、Conv3\_x、Conv4\_x和Conv5\_x。其中,Conv1模块的卷积核大小为 $7 \times 7$ ,卷积步长为2,以使用较大的感受野对原图像进行下采样,得到长宽均为原始图像1/2的特征图,然后再经过最大池化层进一步对图像进行下采样处理;Conv2\_x、Conv3\_x、Conv4\_x、Conv5\_x模块则通过残差网络结构提取图像的深度特征,避免了随着网络层数的加深会出现梯度消失的问题。

为提高模型在不同数据集上的特征表示能力,分别在Conv3\_x、Conv4\_x和Conv5\_x的输出层嵌入通道注意力机制(channel attention module, CAM)和空间注意力机制(spatial attention module, SAM),如图2所示。输入特征经过CAM学习不同通道之间特征的相关性,经过SAM学习每个特征点在空间上的相关性。由于不同深度的特征层其感受野大小不同,浅层特征会保留更多的细节信息,深层特征会有更好的语义信息,对不同深度的特征层嵌入注意力机制,可以提高模型对不同深度特征层中通道和空间的感知能力,从而增强模型在不同数据集中的特征表示能力。此外,在特征提取网络的全局池化层与全连接层之间添加了BNNeck模块,用来抑制来自全连接层的平滑标签损失对三元组损失的影响,使不同身份行人在特征空间中的分布更具有区分性。

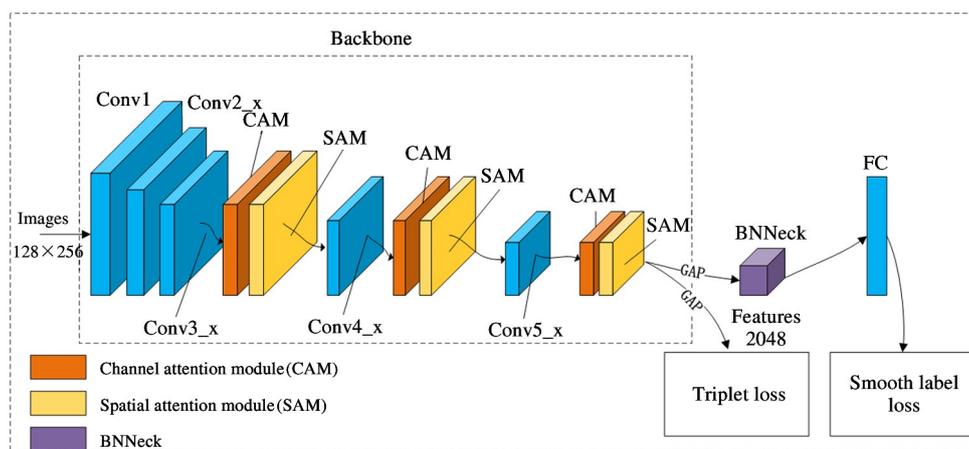


图2 基于注意力机制的行人重识别网络结构图

### 1.3 无监督标签自适应

针对目标域上没有任何标签的数据,受文献

[9,14]在无标签数据集中使用聚类算法选择样本微调模型成功案例的启发,提出无监督标签自适应

方法,充分利用源域数据集和目标域数据集之间的相关性,使用在源域数据集上预训练的模型提取目标域数据集的行人特征,将分布稠密的样本簇的共性特征作为身份区分模板,通过计算与共性特征的

相似度,将标签信息扩展至目标数据集中,从而选择出目标域数据集上最可信的样本来微调模型,以提高模型在目标域数据集上的性能,如图3所示。

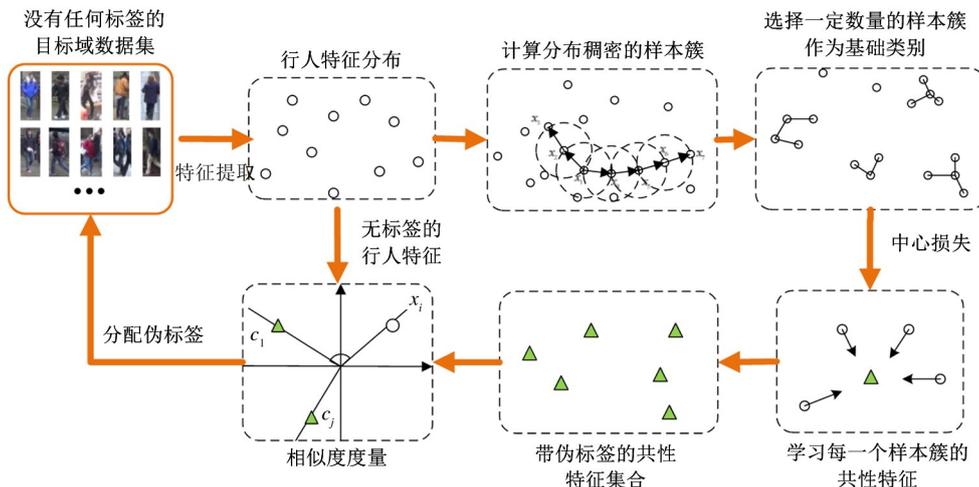


图3 无监督标签自适应方法示意图

具体来说,首先使用欧式距离来描述目标域数据集行人特征之间的分布关系为  $dist(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$ ,再通过基于密度的聚类方法,对于每一个目标域数据集上的样本  $\mathbf{x}_j \in \mathbf{X}_t$ ,计算其  $\varepsilon$ -邻域中包含的所有距离  $dist(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$  不大于  $\varepsilon$  的样本,如式(1)所示,从而选择出目标域数据集中分布稠密的样本形成样本簇。一般而言,分布稠密的样本属于同一身份的可能性较大。

$$N_\varepsilon(\mathbf{x}^{(j)}) = \{\mathbf{x}^{(i)} \in \mathbf{X}_t \mid dist(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) \leq \varepsilon\} \quad (1)$$

其中,  $N_\varepsilon(\mathbf{x}^{(j)})$  表示特征  $\mathbf{x}^{(j)}$  周围小于  $\varepsilon$  距离的所有样本,  $\mathbf{x}^{(i)}$  和  $\mathbf{x}^{(j)}$  分别表示目标数据集中第  $i$  个样本和第  $j$  个样本,  $dist(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$  表示目标数据集中第  $i$  个样本和第  $j$  个样本之间的欧氏距离,  $\varepsilon$  表示 2 个样本之间的密度阈值,这里  $\varepsilon$  取值 1.0。

其次将每一个样本簇视为独立类别的行人样本,选择一定数量的样本簇作为基础类别,并为其分配伪标签,样本簇中的每一个样本共享相同的伪标签,从而形成在目标域上带伪标签的数据集。为了能够在算法执行过程中不断将标签自适应扩展至新的数据上,通过中心损失<sup>[14]</sup>,如式(2)所示,最小化样本簇中行人到其共性特征的距离,以此学习目标

域数据集下每一类样本簇的共性特征  $\mathbf{C}_l$  来作为身份区分特征模板。

$$L_{\text{Common}} = \frac{1}{2} \sum_{j=1}^B \|\mathbf{f}_{j,t} - \mathbf{c}_{j,t}\|_2^2 \quad (2)$$

其中,  $L_{\text{Common}}$  表示每一类样本簇到其共性特征的距离,  $B$  表示样本簇和共性特征的数量,  $\mathbf{f}_{j,t}$  表示目标数据集样本簇上的第  $j$  类样本特征,  $\mathbf{c}_{j,t}$  表示目标数据集上第  $j$  类样本的共性特征。

最后,对于每一个新的来自目标域数据集的行人  $\mathbf{x}_{i,t} \in \mathbf{X}_t$ ,通过相似度距离度量,如式(3)所示,计算  $\mathbf{x}_{i,t}$  与每一个身份区分特征模板  $\mathbf{C}_l$  的相似度  $\theta(\mathbf{x}_{i,t}, \mathbf{c}_j)$ ,将相似度  $\theta(\mathbf{x}_{i,t}, \mathbf{c}_j)$  大于设定阈值的样本视为可信样本,为其分配与模板相同的伪标签并加入到模型的训练中。

$$\theta(\mathbf{x}_{i,t}, \mathbf{c}_j) = \frac{\sum_{k=1}^n \mathbf{x}_{ik,t} \mathbf{c}_{jk}}{\sqrt{\sum_{k=1}^n \mathbf{x}_{ik,t}^2} \sqrt{\sum_{k=1}^n \mathbf{c}_{jk}^2}} \quad (3)$$

其中,  $\theta(\mathbf{x}_{i,t}, \mathbf{c}_j)$  表示目标数据集上的行人  $\mathbf{x}_{i,t}$  与第  $j$  个身份区分特征模板  $\mathbf{c}_j$  的相似度。如果相似度  $\theta(\mathbf{x}_{i,t}, \mathbf{c}_j)$  大于 0.8,则判定该行人与模板身份一致,为该行人分配伪标签并添加至数据集中用来微调模型。  $\mathbf{x}_{ik,t}$  和  $\mathbf{c}_{jk}$  分别表示目标数据集上的行人  $\mathbf{x}_{i,t}$  与

身份区分特征模板  $c_j$  在  $k$ -维度上的特征值,  $n$  表示特征维度。

#### 1.4 基于知识蒸馏的参数更新

由于标签自适应方法依赖于模型在目标数据集下的初始性能,为了保证在目标数据集上样本选择的准确性和一致性,在适应到新场景的过程中不仅要学习新的知识,也要保留之前获得的知识。基于这样的考虑,借鉴一种新的知识蒸馏方法<sup>[15]</sup>,通过采用这种基于知识蒸馏的策略,使模型不仅能够适应到目标域数据集下,同时也能保留在源域数据集上的性能。

如图 1 所示,通过一个在带伪标签的目标数据集上微调的中间模型和在源域数据集中的预训练模型,将微调好的中间模型参数通过式(4)来逐步完成原模型权重参数的更新。

$$\mathbf{W}_{\text{target}} = (1 - \mu) \mathbf{W}_{\text{target}} + \mu \mathbf{W}_{\text{intermediate}} \quad (4)$$

其中,  $\mathbf{W}_{\text{target}}$  表示原模型的权重参数,  $\mathbf{W}_{\text{intermediate}}$  表示中间模型的权重参数,  $\mu$  表示蒸馏因子。通过改变  $\mu$  值来平衡旧知识的遗忘和新知识的学习速率。

#### 1.5 损失函数

为提高模型在目标数据集上的性能,通过联合平滑标签损失 (smooth label loss)<sup>[16]</sup> 和三元组损失 (triplet loss)<sup>[17]</sup> 来优化模型。对于平滑标签损失,根据模型的预测得分和图像标签,使用式(5)计算平滑标签损失。

$$L_D = \sum_{i=1}^N -q_i \log(p_i)$$

$$q_i = \begin{cases} 1 - \frac{N-1}{N} \varepsilon & y \neq i \\ \frac{\varepsilon}{N} & y = i \end{cases} \quad (5)$$

其中,  $y$  表示真实标签,  $i$  表示第  $i$  个类别,  $p_i$  表示标签预测得分,  $N$  表示训练集中的总类别数,  $\varepsilon$  是一个常数,用来鼓励模型不去相信标签是正确的,在这里  $\varepsilon$  值为 0.1。对于三元组损失,选择每个批次内相同身份最远的距离  $d_p$  和不同身份之间最近的距离  $d_n$  的样本形成三元组损失,并通过式(6)计算三元组损失:

$$L_{\text{Tri}} = [d_p - d_n + m]_+ \quad (6)$$

其中,  $m$  表示距离裕度,运算符  $[z]_+$  表示取  $z$  和 0 的

集合中的最大值,在这里  $m$  取值为 0.3。最后,总损失可以用式(7)表示。

$$L = L_D + L_{\text{Triplet}} + \beta L_{\text{Common}} \quad (7)$$

其中,  $\beta$  取值为 0.0035,表示共性特征在学习过程中所占的比重。

## 2 实验数据与结果分析

### 2.1 数据集介绍

为了验证方法的有效性,在 2 个大型的公开行人数据集 Market-1501 和 DukeMTMC-reID 上评估了本文提出的方法。

Market-1501 数据集包含来自 6 个摄像机的 1501 个身份的 32 668 张图像。其中,751 个身份的 12 936 幅图像用于训练,750 个身份的 19 732 幅图像用于测试。

DukeMTMC-reID 数据集是从 8 个摄像机收集的大规模 re-ID 数据集。其中,702 个身份的 16 522 幅图像用于训练,另外 702 个身份的图像用于测试。

### 2.2 性能评估指标

在行人重识别任务中,常用的性能评估指标有首位命中率 ( $Rank-1$ ) 和平均精度均值 ( $mAP$ )。其中,  $Rank-1$  为首次预测正确的概率,  $mAP$  的定义如式(8)所示。

$$mAP = \frac{AP}{Q} \quad (8)$$

其中,  $Q$  表示查询集的数量,  $AP$  由式(9)定义。

$$AP = \frac{\sum_{k=1}^n p(k) \cdot rel(k)}{R} \quad (9)$$

其中,  $k$  表示预测图像的排名,如果预测排名为  $k$  的图片与查询图片是同一 ID,则  $rel(k)$  为 1,否则  $rel(k)$  为 0,而  $p(k)$  被定义为

$$p(k) = \frac{\sum_{i=1}^k rel(i)}{k} \quad (10)$$

### 2.3 实施细节

实验环境如下: Windows 10 的操作系统,深度学习框架为 Pytorch1.6.0,编程语言版本为 Python。

对于行人重识别模型,本实验使用 Resnet 50 作为主干网络 (Baseline),在 Conv3\_x, Conv4\_x,

Conv5\_x输出上分别添加了通道注意力模块和空间注意力模块,并在最后的全局池化层和全连接层之间添加了 BNNeck 模块,输入图像的大小调整为 128 × 256 像素,输出维度为 2048 的特征向量。对于模型的预训练,本实验使用了随机擦除<sup>[18]</sup>、色彩抖动、随机裁剪以及用于训练的随机水平镜像等数据增强方法。采用 Adam 优化器,批处理大小设置为 64,学习率从  $3.5 \times 10^{-5}$  开始,每 50 个 epoch 衰减为原来的 0.1 倍,训练 120 个周期。针对模型的微调,设置密度聚类阈值  $\varepsilon$  为 1.0,相似度阈值为 0.8,每 5 个 epoch 使用知识蒸馏方法对原模型进行一次参数更新,并对无标签的目标数据集使用无监督标签自适应方法(ULA)以获得新的训练数据,微调学习率为  $1 \times 10^{-7}$ ,训练 80 个周期。

## 2.4 实验结果

### 2.4.1 无监督标签自适应性能分析

实验分别以 Market-1501 和 DukeMTMC-reID 作为无标注的测试数据集,验证无监督标签自适应(ULA)在不同设定类别数量下对模型在跨域任务上性能的影响,对比结果如表 1 所示。首先将 Resnet50 作为基础网络(Baseline),并在此基础上添加了注意力机制模块和 BNNeck 模块得到基于注意力机制的行人重识别模型(CSAM-ReID)。从表中可以看出,与基础网络(Baseline)相比,CSAM-ReID 在 Market-1501 数据集上 *mAP* 提升了 5.6%,*Rank-1* 提升了 4.6%;在 DukeMTMC-reID 数据集上 *mAP* 提升了 8.4%,*Rank-1* 提升了 11.1%。其次在 CSAM-ReID

表 1 不同预设类别数量的 ULA 方法对模型在目标数据集上性能的影响

方法	DukeMTMC-reID→ Market-1501		Market-1501→ DukeMTMC-reID	
	<i>Rank-1</i> /%	<i>mAP</i> /%	<i>Rank-1</i> /%	<i>mAP</i> /%
	Baseline	27.2	14.5	25.5
CSAM-ReID	31.8	20.1	36.6	22.3
+ ULA(700)	44.1	28.5	48.9	31.6
+ ULA(750)	48.2	32.0	54.7	35.1
+ ULA(800)	48.5	32.2	53.9	34.8
+ ULA(850)	45.4	30.7	52.9	34.6

注:“→”表示在左侧的数据集下训练,在右侧未训练过的数据集下测试。

基础上,使用 ULA 在无标签的目标数据集中选择高质量样本以微调模型,表 1 中展示了不同预设类别数量 ULA 方法的实验结果。当预设类别数量为 750 时,在 Market1501 数据集上 *mAP* 达到了 32.0%,*Rank-1* 达到了 48.2%;在 DukeMTMC-reID 数据集上 *mAP* 达到了 35.1%,*Rank-1* 达到了 54.7%。实验结果表明,改进的 CSAM-ReID 模型可以有效提升模型在目标域数据集上的初始性能,而标签自适应方法通过选择无标签的目标数据集上可信度高的样本微调模型,大大改善了行人重识别模型在目标数据集上的行人外观特征表示性能。

### 2.4.2 知识蒸馏性能分析

为验证基于知识蒸馏的参数更新方法在平衡旧知识的遗忘和新知识的学习任务中的有效性,在 CSAM-ReID 模型和 ULA 的基础上分别使用不同的知识蒸馏因子,实验结果如表 2 所示。可以看出,当知识蒸馏因子  $\mu$  为 0.5 时,相比于 CSAM-ReID + ULA(750),在 Market-1501 数据集上 *mAP* 提高 1.1%,*Rank-1* 提高了 0.4%;在 DukeMTMC-reID 上 *mAP* 提高了 1.0%,*Rank-1* 提高了 1.8%。实验结果表明,添加知识蒸馏的参数更新策略有助于模型对旧知识的保留,保证了在目标数据集上样本选择的准确性和一致性,从而提高了模型在目标数据集上的性能。

表 2 不同知识蒸馏因子对模型在目标数据集上性能的影响

方法	DukeMTMC-reID→ Market-1501		Market-1501→ DukeMTMC-reID	
	<i>Rank-1</i> /%	<i>mAP</i> /%	<i>Rank-1</i> /%	<i>mAP</i> /%
	CSAM-ReID + ULA(750)	48.2	32.0	54.7
+ KD(0.1)	40.3	28.2	44.6	30.6
+ KD(0.3)	44.3	31.2	54.6	34.6
+ KD(0.5)	48.6	33.1	56.5	36.1
+ KD(0.7)	48.4	32.5	55.9	35.8

注:“→”表示在左侧的数据集下训练,在右侧未训练过的数据集下测试。

### 2.4.3 所提方法与前沿算法的实验数据对比

如表 3 所示,所提算法与 Market-1501 和 Duke-

MTMC-reID 数据集上的 6 种最新的跨域行人重识别方法进行比较,分别是基于聚类的方法 PUL<sup>[9]</sup>,基于属性共同训练的方法 TJ-AIDL<sup>[19]</sup>、基于图像转换的方法 IPGAN<sup>[7]</sup>、PTGAN<sup>[20]</sup>、SPGAN<sup>[21]</sup>以及基于域鉴别网络和域自适应的 DDNDA<sup>[22]</sup>。从实验结果可

以看出,所提方法在 Market-1501 数据集上 Rank-1 为 48.6%,*mAP* 为 33.1%,在 DukeMTMC-reID 上的 Rank-1 为 56.5%,*mAP* 为 36.1%,相比于基于图像转换方法 IPGAN,在 Market-1501 和 DukeMTMC-reID 数据集上,其 *mAP* 分别提高了 6.1% 和 8.1%。

表 3 所提方法与其他先进算法的性能对比

方法	DukeMTMC-reID→Market-1501				Market-1501→DukeMTMC-reID			
	Rank-1/%	Rank-5/%	Rank-10/%	<i>mAP</i> /%	Rank-1/%	Rank-5/%	Rank-10/%	<i>mAP</i> /%
PUL <sup>[9]</sup>	30.4	44.5	50.7	16.4	44.7	59.1	65.6	20.1
TJ-AIDL <sup>[19]</sup>	44.3	59.6	65.0	23.0	<b>58.2</b>	74.8	81.1	26.5
IPGAN <sup>[7]</sup>	47.0	<b>63.0</b>	<b>68.1</b>	27.0	57.2	<b>76.0</b>	<b>82.7</b>	28.0
PTGAN <sup>[20]</sup>	27.4	-	50.7	-	38.6	-	66.1	-
SPGAN <sup>[21]</sup>	41.1	56.6	63.0	22.3	51.5	70.1	76.8	22.8
DDNDA <sup>[22]</sup>	38.1	55.0	62.2	17.1	28.4	41.5	49.9	17.6
本文方法	<b>48.6</b>	58.7	66.3	<b>33.1</b>	56.5	69.4	75.4	<b>36.1</b>

注:“→”表示在左侧的数据集下训练,在右侧未训练过的数据集下测试。

## 2.5 训练过程可视化

为了分析训练过程中标签自适应方法选择的样本和目标数据集上行人特征分布的变化,在训练期间随机选择 DukeMTMC-reID 数据集上 6 个带伪标签的行人和 8 位不同身份的行人分别对其样本选择过程和特征分布变化过程进行可视化,特征分布和样本选择结果分别如图 4 和图 5 所示。从图 4 中可以观察到,随着训练数的增加,在特征空间中相同身份的行人相互靠近,不同身份的行人彼此远离,从而提高模型对不同身份的行人辨别能力。同时,每一次对模型进行参数更新并使用标签自适应方法后都有大量的新样本加入到训练数据中。从图 5 中可以看出,由于直接将模型应用在目标域数据集上其性能会较差,因此刚开始选择的样本都是在同一摄像

头下姿态相似的行人,随着模型在目标域数据集下性能的改善,越来越多的样本通过标签自适应被关联了伪标签,加入到模型的进一步微调中。

图 6(a)和(b)分别展示了模型在 DukeMTMC-reID 数据集上不同阶段的损失函数和性能指标的曲线变化。从图 6(a)可以看到,每 5 个 epoch 损失函数都会陡然升高。这是因为通过标签自适应加入了新的样本,随着模型的不训练,损失函数逐渐接近最小值,达到平稳状态,此时模型的性能是最优的。从图 6(b)可以看到,模型在 epoch 为 160 时性能指标达到最高,随着模型的继续迭代,模型的性能开始显示下降的趋势。这是因为在连续迭代的过程中,先前选择的错误样本逐渐累积,模型接收了大量错误的样本训练而导致的。

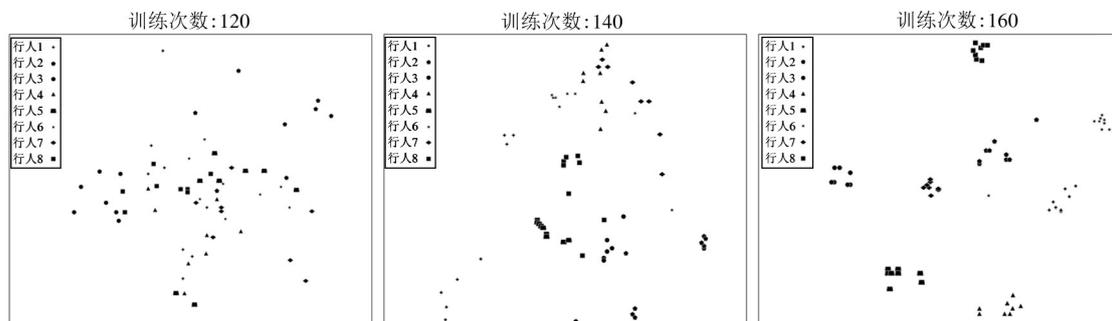


图 4 在训练过程中样本在特征空间中的分布变化图

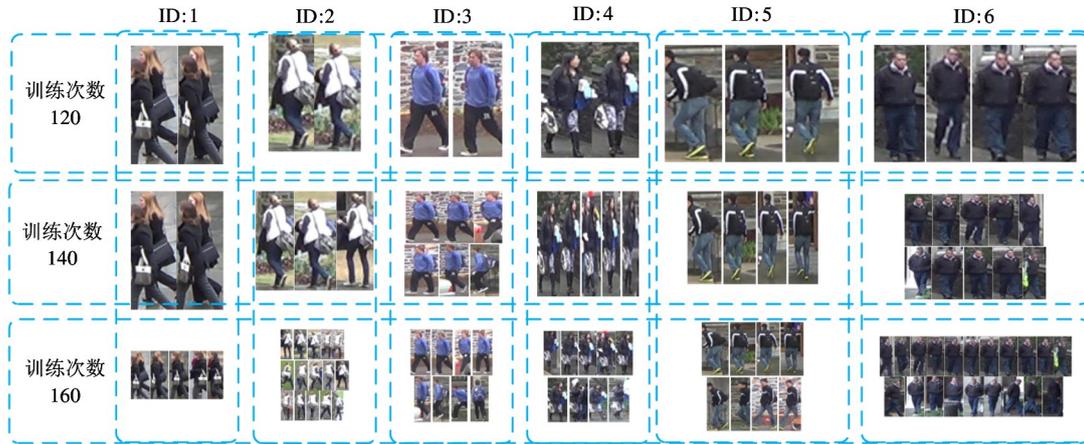


图 5 在训练过程中通过标签自适应选择样本的过程

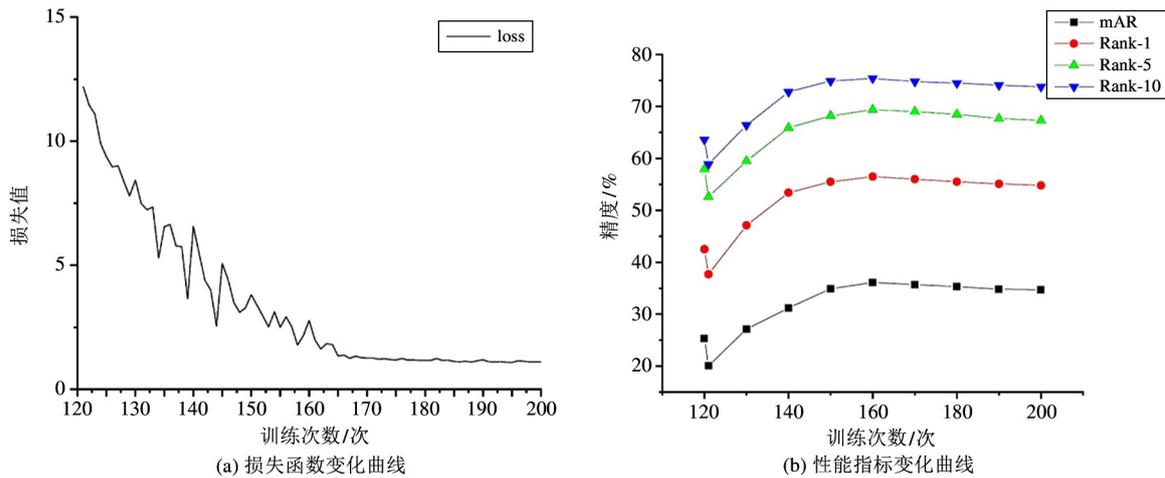


图 6 在 DukeMTMC-reID 数据集上训练的损失函数和性能指标变化曲线图

### 3 结论

本文提出了一种基于注意力机制和标签自适应的跨域行人重识别方法。通过在不同深度特征层嵌入注意力机制和 BNNeck 模块可有效提高行人重识别模型在跨域任务中的初始性能,进而提高行人重识别模型对源域数据集和没有任何标签的目标域数据集之间相关性的表达能力。无监督标签自适应方法可以充分挖掘没有任何标签的目标域数据集上的可用数据,并将标签信息扩展至新的数据上。基于知识蒸馏的方法可以使模型在学习新的知识的同时保留旧知识,保证样本选择的一致性和准确性,从而提高模型的稳定性。实验结果表明,所提方法能有效提高行人重识别模型在跨域任务中的性能。

### 参考文献

- [ 1 ] 罗浩,姜伟,范星,等. 基于深度学习的行人重识别研究进展[J]. 自动化学报, 2019,45(11):2032-2049
- [ 2 ] 徐龙壮,彭力. 基于多尺度卷积特征融合的行人重识别[J]. 激光与光电子学进展, 2019,56(14):221-227
- [ 3 ] 潘通,李文国. 基于特征拼接的行人重识别[J]. 激光与光电子学进展, 2019,56(16):193-199
- [ 4 ] CHEN T, DING S, XIE J, et al. ABD-Net: attentive but diverse person re-identification[C] // IEEE International Conference on Computer Vision, Seoul, Korea, 2019: 8351-8361
- [ 5 ] ZHENG L, SHEN L, TIAN L, et al. Scalable person re-identification: a benchmark [C] // Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 1116-1124
- [ 6 ] BEWLEY A, GE Z, OTT L, et al. Simple online and realtime tracking[C] // 2016 IEEE International Conference on Image Processing, Phoenix, USA, 2016: 3464-3468
- [ 7 ] LIU J, LI W, PEI H, et al. Identity preserving generative adversarial network for cross-domain person re-identification[J]. IEEE Access, 2019, 7: 114021-114032

- [ 8 ] CHOI Y, CHOI M, KIM M, et al. Stargan: unified generative adversarial networks for multi-domain image-to-image translation [ C ] // IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 8789-8797
- [ 9 ] FAN H, ZHENG L, YAN C, et al. Unsupervised person re-identification: clustering and fine-tuning [ J ]. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2018, 14(4): 1-18
- [ 10 ] LONG M, ZHU H, WANG J, et al. Deep transfer learning with joint adaptation networks [ C ] // International Conference on Machine Learning, Sydney, Australia, 2017: 2208-2217
- [ 11 ] 赵文仓,袁立镇,徐长凯. 基于鉴别模型和对抗损失的无监督域自适应方法[J]. 高技术通讯, 2020,30(7): 698-706
- [ 12 ] LUO H, GU Y, LIAO X, et al. Bag of tricks and a strong baseline for deep person re-identification [ C ] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, USA, 2019: 4321-4329
- [ 13 ] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [ C ] // Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, USA, 2016: 770-778
- [ 14 ] WANG P, LIN W H, CHAO K M, et al. A face-recognition approach using deep reinforcement learning approach for user authentication [ C ] // IEEE 14th International Conference on e-Business Engineering, Shanghai, China, 2017: 183-188
- [ 15 ] 廖星宇. 基于无监督增量学习的跨域行人重识别 [ D ]. 合肥:中国科学技术大学数学科学学院,2019
- [ 16 ] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [ C ] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 2818-2826
- [ 17 ] CHENG D, GONG Y, ZHOU S, et al. Person re-identification by multi-channel parts-based CNN with improved triplet loss function [ C ] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 1335-1344
- [ 18 ] ZHONG Z, ZHENG L, KANG G, et al. Random erasing data augmentation [ C ] // AAAI Conference on Artificial Intelligence, New York, USA, 2020: 13001-13008
- [ 19 ] VARIOR R R, HALOI M, WANG G. Gated siamese convolutional neural network architecture for human re-identification [ C ] // The 14th European Conference on Computer Vision, Amsterdam, Netherlands, 2016: 791-808
- [ 20 ] WEI L, ZHANG S, GAO W, et al. Person transfer GAN to bridge domain gap for person re-identification [ C ] // IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 79-88
- [ 21 ] DENG W, ZHENG L, YE Q, et al. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification [ C ] // IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 994-1003
- [ 22 ] 崔鹏,范志旭. 基于域鉴别网络和域自适应的行人重识别[J]. 光电子·激光,2019,30(6):632-639

## Cross-domain person re-identification based on attention and label adaptation

CHEN Siwen, WU Huaiyu, CHEN Yang

(Institute of Robotics and Intelligent Systems, Wuhan University of Science and Technology, Wuhan 438001)

### Abstract

In view of the fact that most of the existing person re-identification algorithms rely on supervised training, and the manually annotated data in supervised training require expensive man power and material resources, which limits expanding the application of person re-identification algorithms in new scenarios, a cross-domain person re-identification based on attention and label adaptation is proposed. Firstly, the attention mechanism and BNNeck module are embedded in the feature layers of deep convolution neural network (DCNN) to enhance the feature representation ability of the model to person under different datasets. Secondly, the label adaptation method is proposed to expand the label information to the target dataset. Finally, the method of knowledge distillation (KD) is used to continuously fine-tune the model, so that the model gradually adapts to the new scenarios. The proposed method achieves 33.1% mean average precision (mAP) accuracy on Market-1501 dataset and 36.1% mAP accuracy on DukeMTMC-reID dataset. Compared with the advanced PTGAN, IPGAN and other cross-domain person re-identification algorithms, the performance of the proposed method is obviously improved.

**Key words:** cross-domain person re-identification, attention, label adaptation, knowledge distillation (KD), deep convolutional neural network (DCNN)