



# 交叉注意互学习的跨领域重识别实验设计

云 霄, 褚 菲, 张晓光, 汪群群

(中国矿业大学 信息与控制工程学院, 徐州 221116)

**摘要:** 针对视频智能监控中跨领域行人重识别的研究热点, 将交叉注意互学习的无监督跨领域行人重识别设计为研究性实验现场硬件教学项目, 观察重要区域特征关注度和网络伪标签互监督两方面对模型跨领域能力的影响。设计多尺度交叉组合注意力机制, 分析其在通道和空间上对重要区域特征关注度的增强, 以及对区分度不高特征关注度的减少。在此基础上设计多尺度交叉组合注意力互学习实验网络, 采用硬标签互学习的方式更新训练过程, 消除错误伪标签对模型性能的影响。教学实践表明, 该实验项目有助于提升学生动手能力和独立思考问题的能力, 为科研成果转化为教学实验提供了借鉴。

**关键词:** 行人重识别; 无监督学习; 跨领域自适应; 注意力机制; 互学习

中图分类号: TP957.52

文献标志码: A

DOI: [10.12179/1672-4550.20220649](https://doi.org/10.12179/1672-4550.20220649)

## Design on Experiment of Cross Attention Mutual Learning for Unsupervised Cross-domain Re-identification

YUN Xiao, CHU Fei, ZHANG Xiaoguang, WANG Qunqun

(School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China)

**Abstract:** Aiming at the research hotspot of cross-domain person re-identification (Person Re-ID) in video intelligent surveillance, the cross attention mutual learning for unsupervised cross-domain person re-identification is designed as a research experimental teaching project, and the influence of both the attention of important regional features and the mutual supervision of network pseudo tags on the cross-domain ability of the model is observed. A multi-scale cross combination attention mechanism is designed, and its enhancement of attention to important regional features in channel and space, as well as the reduction of attention to features with low discrimination is analyzed. On this basis, a multi-scale cross combination attention mutual learning experimental network is designed, and the training process is updated by hard label mutual learning to eliminate the impact of false labels on the performance of the model. Teaching practice shows that the experimental project helps to improve students' practical ability and independent thinking ability, and provides a reference for the transformation of scientific research achievements into teaching experiments.

**Key words:** person re-identification; unsupervised learning; cross-domain adaption; attention mechanism; mutual learning

行人重识别 (person re-identification, Person Re-ID) 是监控摄像机在不同空间或时间所捕获的视图上检索同一身份行人的过程。对无法获取脸部图像的行人视图, 行人重识别技术可以通过对姿态、穿着等有鉴别力的行人身份信息进行提取和表达, 完成不同视图中同一行人的匹配任务, 实现对面脸识别技术的有力补充, 为公共安全保障、城市安防建设提供重要的技术基础。但是,

现阶段行人重识别技术大多基于有标注数据集进行预训练。而在实际应用中, 在预训练数据集下训练得到的行人重识别模型只能对具有相同风格的行人图像实现较好的识别效果, 对其他风格图像的识别精度往往差强人意。而由于实际应用场景的不确定性, 为每个应用场景专门预先训练有标注数据模型, 从时间和精力上都消耗巨大成本, 这不仅是目前视频监控领域面临的普遍棘手

收稿日期: 2022-11-17; 修回日期: 2023-02-12

基金项目: 中国矿业大学教改项目(2021YB20); 江苏省高教学会评估委员会课题(2021-C158); 国家自然科学基金(61902404); 教育部产学研合作协同育人项目(202002109018, 202002177005, 202002017012); 教育部高等学校自动化类专业教学指导委教改项目(202123); 中国矿业大学“动力中国·课程思政”示范项目(2021KCSZ31Y)。

作者简介: 云霄(1986-), 女, 博士, 副教授, 主要从事视频图像分析方面的研究。E-mail: [xyun@cumt.edu.cn](mailto:xyun@cumt.edu.cn)

的问题,也逐渐成为电子信息类研究性实验教学所亟待解决的问题。

无监督跨领域行人重识别,又称无监督领域自适应行人重识别(unsupervised domain adaptation person re-identification, UDA Re-ID),是指将在已标注数据集(称为源域)上训练得到的模型直接迁移到另一个未经过标注训练的数据集(称为目标域),实现行人重识别任务。在实际应用中,源域与目标域的行人身份、风格、光线、视角、成像机理等方面往往存在较大差异性,使得跨域模型迁移任务极具挑战。基于生成对抗网络(generative adversarial network, GAN)的 UDA Re-ID 方法能生成与源域数据集风格相同的目标域数据集,提高跨域模型的领域自适应能力。但是这种图像级样式转化生成的目标域样本与源域样本比较仍有差距,因此只能作为一种辅助学习方式,与其他 UDA Re-ID 方法结合使用。在此基础上,基于特征级分布对齐的 UDA Re-ID 方法提取不同领域风格图像中行人体型、穿着、装饰品、颜色等领域不变特征共享信息提高模型的识别准确率。文献[1]通过使用领域对抗神经网络寻找网络中的域不变特征实现模型迁移。文献[2]通过设计损失函数实现对不同摄像头之间的不变约束,从域内的角度增强了模型的鲁棒性。文献[3]通过使用跨数据集中级特征对齐正则化项进行联合优化的方式,实现较好的迁移效果。上述基于图像级样式转化和特征级分布对齐的 UDA Re-ID 方法大多通过将源域和目标域间数据的全局分布进行对齐,使源域模型能够在特征不可知的目标域使用,但可能会导致负方向转移。

为解决负转移的问题,基于伪标签预测的 UDA Re-ID 方法在源域和目标域的适应过程中考虑特定类的特定属性,让网络自适应地产生标签辅助网络训练过程。如文献[4]提出一种动态选择微调样本的方法,将聚类产生的伪标签样本有选择性地送入模型进行微调训练。文献[5]提出一种自相似分组方法改善伪标签质量,通过将行人特征进行分割,并对全局和局部特征分别聚类分配伪标签,改善模型迁移效果。文献[6]在聚类前通过交替训练图像生成器和特征编码器来增加样本空间的多样性以及减小新特征空间的距离,以达到增强目标域内行人聚类的目的。文献[7]提出非对称的协作网络,将聚类后置信度较低的样本直接丢弃,以降低用于再训练的样本多样性,增强样本的利用率并提升模型整体的迁移效果。文献[8]

提出一种标签互学习方法,使用两个网络让彼此的伪标签作为额外监督,减轻网络对噪声的拟合。在此基础上,文献[9]用教师学生双网络的竞争机制为网络选择更可信的目标伪标签微调模型。但是,由于聚类算法本身不完善,伪标签预测不能够保证在聚类过程中产生完全正确的伪标签,因此可能会出现利用错误伪标签指导网络更新学习的情况,严重影响模型对目标域样本的特征表达,降低模型性能。

为了将前瞻性科研成果融入实验教学中,本文结合电子信息专业数字信号处理、数字图像处理等课程中涉及的视觉行人重识别实验教学知识点,从提高模型对于重要区域特征的关注以及加强网络伪标签间的互相监督两个方面入手,为了降低噪声伪标签对 UDA Re-ID 方法的影响,设计交叉注意互学习的跨领域行人重识别实验,提升模型的领域自适应能力。首先,在实验初期样本初始化阶段采用自相似保留生成对抗网络(similarity preserving GAN, SPGAN)<sup>[10]</sup>扩充数据集。然后,设计多尺度交叉组合注意力机制,分析其在通道和空间上对重要区域特征关注度的增强,以及对区分度不高特征关注度的减少。在此基础上设计多尺度交叉组合注意力互学习实验网络,采用硬标签互学习的方式更新训练过程,消除错误伪标签对模型性能的影响。教学实验结果表明,该实验方案的设计在提升跨领域行人重识别性能方面优于其他同类经典方法。

## 1 实验设计网络

本实验设计的整体网络结构有源域有监督预训练和目标域无监督微调训练两个部分,如图1所示。

本实验设计将源域图像输入相似性保留生成对抗网络,获取具有目标域样本风格的生成图像,保留源域图像的原始身份(ID)信息不变,初始化迁移模型,增强模型训练的风格初始化认知。初始化之后的生成图像和原始图像一起输入本实验所设计的 MCCAM-Net 进行特征提取,所提取的特征经过 4096 维池化层、全链接层以及与源域数据集 ID 数量匹配的 ID 全链接层,实现 ID 维的尺度变换。最后, ID 维的特征通过 softmax 函数激活,输出样本属于每个类别的概率值,以样本真实标签作为参考计算源域损失  $L_s$ ,并反馈 MCCAM-Net 以更新网络参数,直到网络收敛或者达到最大迭代次数结束训练,实现源域模型迁移。

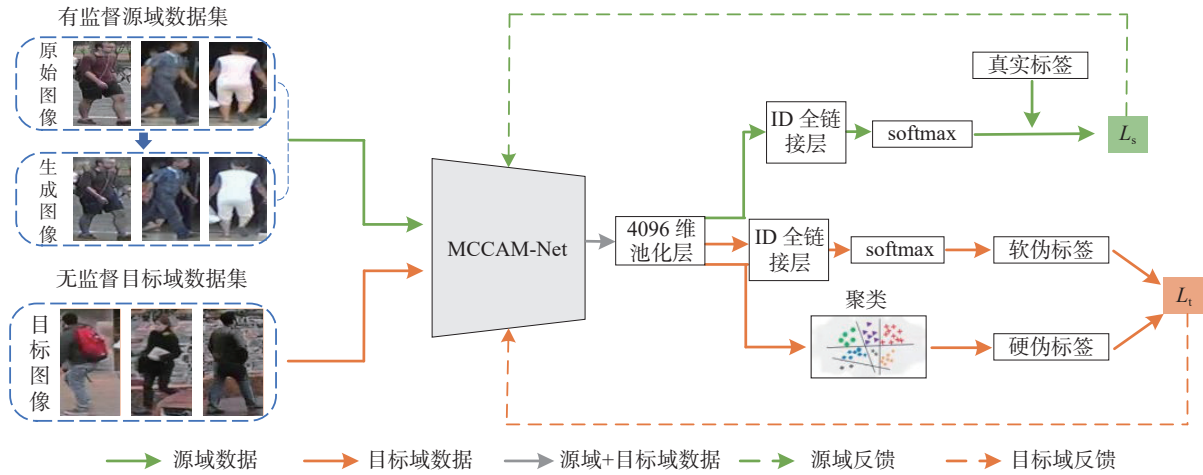


图1 本实验设计整体网络结构

本实验设计将无标注的目标域图像数据输入训练好的 MCCAM-Net 源域模型进行特征提取, 经过全链接层的 4096 维尺度变换输出特征, 同时送入聚类算法和 ID 全链接层、softmax 激活函数, 分别获取模型再训练的硬伪标签以及软伪标签。最后, 在硬软标签的联合监督下通过目标域损失  $L_t$  进一步微调网络, 依次迭代直到达到训练周期结束, 进一步增强目标域样本的识别能力, 提升行人重识别模型的跨领域迁移性能。

### 1.1 多尺度交叉组合注意力机制(MCCAM)

为了获取更多关注区域的通道和空间两个维度上的细节信息, 本实验设计多尺度交叉组合注意力机制 (multi-size cross combination attention, MCCAM), 包括多尺度交叉注意力机制和空间注意力机制两部分。其中空间注意力机制指导网络关注图像各区域的特征, 多尺度交叉注意力机制在空间注意力机制基础上使网络聚焦于有更具表达力的特征部分。

#### 1.1.1 多尺度交叉注意力机制(MCAM)

传统的通道注意力机制 (channel attention mechanism, CAM) 采用全局平均池化和全局最大池化对输入特征图的空间维度进行挤压并获得聚合后特征图的空间信息, 将生成注意力图与原始特征值做卷积, 实现对原特征图权重信息的调整。CAM 为了获取空间信息暴力的挤压特征图的尺寸, 更多地关注网络任务区域, 导致背景信息丢失; 同时, 由于 CAM 对不同池化空间信息的关注度不同, 注意力特征图与原始特征图的卷积可能会加大对区分度不高区域信息的关注度, 出现信息关注度偏移现象。

为解决以上问题, 本实验从多尺寸空间信息获取和交叉注意两方面出发, 首先设计多尺度交叉注意力机制 (multi-size cross attention mechanism, MCAM), 结构如图 2 所示。为了更大程度地保留原始信息, 保持特征图尺寸不变, 减小信息丢失, 首先采用自适应最大池化 (adaptive max pooling, AMP) 层和自适应平均池化 (adaptive average pooling, AAP) 层对尺寸为  $R^C \times H \times W$  的特征图  $X$  空间信息进行聚合, 生成了两个不同的空间背景描述特征图  $f_1$  和  $f_2$ 。 $f_1$  和  $f_2$  分别输入由不同的卷积层和激活层组成的交叉模块 (cross module, CM) 共享网络, 以加强对重要区域特征的关注度。为了减小参数开销, 交叉模块中的卷积层尺寸设定为  $R^r \times 1 \times 1$ , 其中  $r$  为通道缩减率。为了减小网络对关注度不高区域信息的关注, 本文将  $f_1$  与  $f_2$  经过交叉模块中的 sigmoid 激活函数生成的关注区域映射  $mask1$  和  $mask2$  做卷积运算, 分别生成两个特征图  $f_3$  与  $f_4$ , 使用连接操作合并两个输出向量得到尺寸为  $R^{2C} \times H \times W$  的特征图  $f_a$ , 并采用  $1 \times 1$  卷积  $Conv_{1 \times 1}$  进行降维, 最后通过 sigmoid 激活函数输出交叉关注后的尺寸为  $R^C \times H \times W$  的特征映射图  $X_{output}$ 。

交叉模块中 sigmoid 激活函数生成的关注区域映射图  $mask1$  和  $mask2$  可分别表示为:

$$\begin{cases} mask1 = \text{sigmoid}(CM(AMP(X))) \\ mask2 = \text{sigmoid}(CM(AAP(X))) \end{cases} \quad (1)$$

则输出特征图  $f_3$  和  $f_4$  分别为:

$$\begin{cases} f_3 = f_1 \otimes mask2 \\ f_4 = f_2 \otimes mask1 \end{cases} \quad (2)$$

式中,  $\otimes$  表示特征图矩阵元素相乘操作。MCAM 输出的特征图  $X_{output}$  定义为:

$$X_{output} = \text{sigmoid}(\text{Conv}_{1 \times 1}(f_3 + f_4)) \quad (3)$$

相比传统的通道注意力机制，本实验所设计的多尺度交叉注意力机制在不改变特征空间尺寸的情况下增强特征重要区域的关注度，避免压缩过程特征丢失；同时，由于两种不同池化空间信息的关注度存在差异，使用交叉注意的方式可以在卷积过程中过滤掉自身关注中区分度不高的干扰信息，进一步加强重要信息的网络关注度。

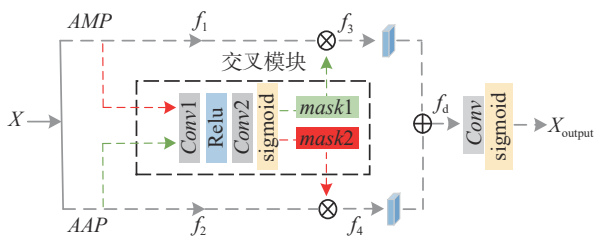


图 2 多尺度交叉注意力机制实现原理

### 1.1.2 多尺度交叉组合注意力机制(MCCAM)

本实验将前述所设计的多尺度交叉注意力机制 MCAM 与空间注意力机制 (spatial attention mechanism, SAM) 相结合，构建多尺度交叉组合注意力机制 MCCAM，利用特征间的关系从空间上

加强网络对关键信息的表达，利用空间转换模块保留所关注的空间信息，并将保留的感兴趣区域以权重掩膜(mask)的形式加权输出，根据掩膜完成对原始特征权重信息的调整，降低网络对不相关区域的关注度。

根据多尺度交叉注意力机制 MCAM 与空间注意力机制 SAM 先后顺序的不同，本实验设计两种多尺度交叉组合注意力机制 MCCAM 方式进行测试，即前交叉注意力机制与后交叉注意力机制，如图 3 所示。首先给定一个输入图像，经过原网络结构内部一系列的卷积池化层 Stage N 输出中间层特征  $f$ 。前后交叉注意力机制将中间层特征  $f$  经过 MCAM 和 SAM 获取特征映射，前交叉注意力机制将对应位置的输出特征图先经过 MCAM 计算互补注意力对输入特征图进行权重调整，后交叉注意力机制反之，获取的输出特征图与输入特征图进行卷积操作。然后，调整后特征与原始输入特征  $f$  加权得到注意力组合模块 MCCAM 的输出  $f^*$ ，以保证在关注重点区域特征的基础上结合原始特征防止丢失细节信息。最后，输出特征经过原有的网络层 Stage (N+1) 后进行最终的分类与识别任务。

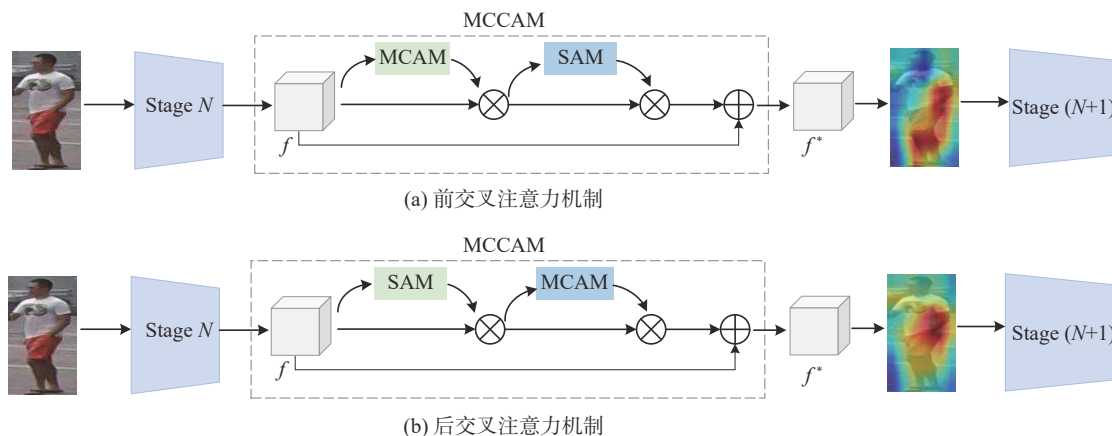


图 3 MCCAM 的两种组合实验测试方式

本实验所设计的多尺度交叉组合注意力机制 MCCAM 从通道和空间上引导网络对区域的关注度，对关键信息的把握更加具体准确。学生可以将 MCCAM 作为一个即插即用的模块集成到现有的网络框架中，观察 MCCAM 对现有网络框架的影响，选择不同的结合方法可以提升学生的动手能力，以及根据实验结果独立分析问题的能力。实验结果表明相比于后交叉注意力机制，前交叉注意力对全局重点信息的把握更全面，也更集

中，因此在后续实验中，多尺度交叉组合注意力机制 MCCAM 都默认为前交叉注意力机制。

### 1.2 MCCAM-Net

基于伪标签预测的 UDA Re-ID 方法一般通过对图像特征进行相似性聚类生成伪标签，作为监督信号监督网络更新，但有时会产生错误的伪标签，使模型朝反梯度方向更新。为解决错误伪标签的问题，本实验通过设计多分支结构获取多样化特征，实现生成伪标签的互相监督，在前述

多尺度交叉组合注意力机制 MCCAM 的基础上,设计多尺度交叉组合注意力互学习网络(multi-scale cross combination attention mutual learning network, MCCAM-Net),网络结构如图 4 所示。将有标注的源域样本输入保留残差网络 ResNet50 的 Layer1 至 Layer4 层,并在 Layer2 和 Layer3 层后分别插入本实验所设计的多尺度交叉组合注意力机制 MCCAM,以加强网络对重要区域特征的关注度。此外,为了缩小特征图参数矩阵的尺寸,减少参数计算量,在 Layer4 层后加入不同的池化操作,并将网络分为上下两个分支,由于不同的池化方式所产生的关注范围不同,上分支采用与骨干网相同的全局平均池化(global average

pooling, GAP)层获取池化特征;而相比于 GAP,全局最大池化(global max pooling, GMP)层能够引导网络在最具描述性的部分识别出相对较弱的显著特征,因此本实验设计在下分支采用 GMP 代替来 GAP,最终得到两种不同的池化后特征输出  $f_a$  与  $f_m$ 。最后,双分支网络的全局特征输出  $f_a$  与  $f_m$  分别经过全连接(fully connected, FC)层和 softmax 函数输出预测的类别概率值  $\{p_1, p_2, p_3, \dots, p_i, \dots, p_n\}$  与  $\{p_1, p_2, p_3, \dots, p_j, \dots, p_n\}$ , 其中  $n$  表示类别总数,概率最大值  $p_i$  与  $p_j$  对应的类别即为各种预测的类别。在源域损失与目标域损失的监督下完成 MCCAM-Net 模型自适应更新训练过程。

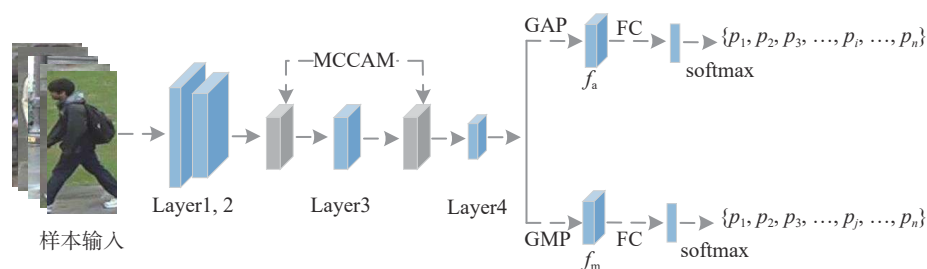


图 4 MCCAM-Net 结构图

### 1.3 源域有监督学习

本实验首先通过所设计的 MCCAM-Net 计算出概率最大值  $p_i$  与  $p_j$  对应的预测类别,然后以行人的真实身份标签  $q_i$  作为监督信号通过身份分类损失  $L_{id}$  对行人进行精确分类,同时将采用距离度量损失<sup>[11]</sup>  $L_{dm}$  度量特征之间的相似度,拉近相同行人间的距离,推远不同行人间的距离。损失函数计算方式具体如下:

身份分类损失  $L_{id}$ : MCCAM-Net 网络经过 softmax 函数输出概率值,通过  $L_{id}$  评估预测值与真实标签的差距,身份分类损失定义为:

$$L_{id} = - \sum_{i=1}^n q_i \log(p_i) \quad (4)$$

式中,  $q_i$  表示样本真实标签,  $p_i$  表示样本被预测为  $q_i$  的概率值,  $n$  表示类别总数。

距离度量损失  $L_{dm}$ : 分类损失在分离不同身份的行人时,能够尽可能地增大不同类别样本间的距离,但是无法拉近同类别样本距离,因此本实验在分类损失的基础上设计距离度量损失  $L_{dm}$ , 指导网络从类内和类间两个角度对距离阈值内的所有正负样本进行采样,从而使模型在小批量学习

中获取更具鉴别力的特征。距离度量损失  $L_{dm}$  定义为:

$$L_{dm} = \sum_i^N [\|f(x_{O}^i) - f(x_P^i)\|_2 - \|f(x_{O}^i) - f(x_N^i)\|_2 + \alpha]_+ \quad (5)$$

式中,  $\|\cdot\|_2$  为欧氏距离,  $f(x_{O}^i)$  表示原样本特征。  $x_P^i$  和  $x_N^i$  分别表示  $x_{O}^i$  对应的正负样本对, 阈值  $\alpha$  用来控制正负样本之间的距离。  $+$  表示  $[\ ]$  内的值大于零时, 取该值为损失, 小于零时损失为零。

源域联合损失  $L_s$ : 本实验设计将  $L_{id}$  与  $L_{dm}$  进行融合, 获取源域联合损失, 在联合监督下更新训练。源域联合损失函数定义为:

$$L_s = L_{id} + L_{dm} \quad (6)$$

学生通过独立实验并观察实验结果, 能够证明出源域联合损失有助于改善模型迁移前期噪声伪标签的影响, 获取具有一定识别及迁移能力的行人重识别模型。

### 1.4 目标域无监督自适应学习

基于伪标签预测的 UDA Re-ID 方法在预测过程时易产生错误的伪标签, 利用错误的伪标签指导网络更新学习将会造成模型的反梯度更

新。为解决该问题，本实验设计采用软硬标签互学习的方法更新目标域微调训练，减少噪声

伪标签的影响，目标域无监督自适应学习过程如图 5 所示。

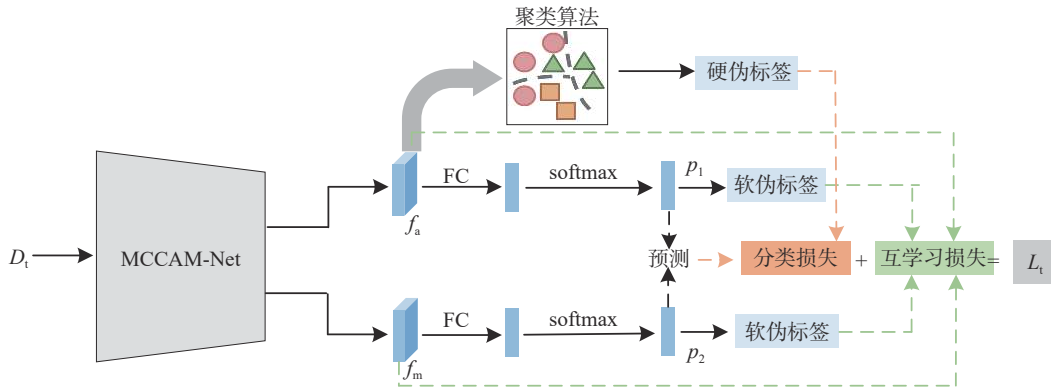


图 5 目标样本领域自适应学习过程

首先，无标注的目标域样本  $D_i$  输入本实验所设计的 MCCAM-Net，获取具有鉴别力的行人特征  $f_a$  与  $f_m$ ，并送入双分支网络的 FC 层与 softmax 函数，预测第  $i$  个训练样本  $x_i$  的类别概率值  $p_1(x_i)$  与  $p_2(x_i)$ ，定义为软伪标签。双分支网络的上分支将输出特征  $f_a$  送入基于密度的 DBSCAN 聚类算法，获取模型再训练硬伪标签。整个训练过程采用软硬标签互学习的方式，在分类损失  $L_c$  及互学习损失  $L_{ml}$  的共同监督下完成模型的微调训练过程。这种伪标签互相监督的更新方式有利于提高伪标签的质量，减小噪声伪标签反梯度更新网络的可能性。损失函数的具体计算方式如下：

分类损失  $L_c$ ：将聚类产生的硬伪标签  $y_i$  作为监督信号，用 MCCAM-Net 双分支的分类结果  $p_1(x_i)$  与  $p_2(x_i)$  计算分类损失为：

$$L_c = - \sum_i (y_i \log(p_1(x_i))) - \sum_i (y_i \log(p_2(x_i))) \quad (7)$$

互学习损失  $L_{ml}$ ：在互学习过程中，仅仅使用软分类损失作为样本预测依据，会使网络只关注双分支都正确预测的样本，任何一方错误预测损失都将不再被考虑，这种过分关注正确标签样本的微调方式可能会导致模型泛化力差，甚至出现过拟合现象。因此，为了充分利用所有样本信息，本实验设计在分类前采用双分支特征相互学习的方式拉近同一样本间的距离。因此，本实验设计的互学习损失将双分支网络软伪标签分类损失  $L_{mc}$  与 MCCAM-Net 输出特征距离度量损失  $L_{md}$  相结合，互学习损失函数定义为：

$$L_{ml} = L_{mc} + L_{md} \quad (8)$$

式中， $L_{mc}$  为软伪标签分类损失<sup>[12]</sup>， $L_{md}$  为 MCCAM-

Net 在 FC 层之前的输出特征  $f_{ar}$  与  $f_{ms}$  间的欧式距离度量。

目标域联合损失  $L_t$ ：分类损失  $L_c$  指导网络进行正确分类，互学习损失  $L_{ml}$  拉近同一样本距离。本实验设计在分类损失和互学习损失联合监督下实现模型的微调再训练，目标域联合损失定义为：

$$L_t = \lambda_1(L_c + L_{mc}) + \lambda_2 L_{md} \quad (9)$$

式中， $\lambda_1$  与  $\lambda_2$  分别为软硬分类损失与互学习中的度量损失在整个目标域损失中占据的比值。

## 2 实验结果与分析

为了验证本实验设计网络结构的有效性，在 3 个公共数据集 DukeMTMC、Market-1501 和 MSMT17 上进行了大量实验，并采用行人重识别任务中常用的平均精度均值 (mAP) 和前 1、2、5 位命中率 Rank1、Rank2、Rank5 评估算法性能。实验中将所设计的实验方法与其他同领域经典方法做了对比，并对多尺度交叉组合注意力机制 MCCAM 和损失函数等模块做了消融实验，以验证各模块的有效性。

### 2.1 实验环境与参数配置

本实验设计图像预处理阶段采用了 SPGAN 模型对源域样本的训练集图片进行风格转换，将原图像和转换图像一起输入 MCCAM-Net 网络进行有监督学习，输入图像统一设置为  $256 \times 128$  像素的固定尺寸。参数优化方式采用 Adam 优化器，权重衰减率为 0.0005，batchsize 设置为 64，学习率初始值为 0.00035，并随迭代次数不断更新。源域有监督训练共进行 80 轮，无监督微调训练环节共迭代训练 40 轮，其中微调环节使用的 DBSCAN

算法中最小聚类样本设置为 4, 密度半径  $r=0.002$ , 式(9)中的目标域联合损失函数权重  $\lambda_1$  与  $\lambda_2$  分别设置为 1 和 0.5。所有实验结果都是在 Pytorch 框架下的 4 个 Nvidia 1080Ti GPU 环境下得到的。

### 2.2 参数实验分析

参数实验对所设计的多尺度交叉组合注意力机制 MCCAM 的设置方式、放置位置以及目标域联合损失函数权重  $\lambda_1$ 、 $\lambda_2$  取值进行了实验结果分析, 测试在不同的模块参数配比下 mAP 和 Rank1 的变化, 从而选择最佳损失函数配置及比重。

#### 2.2.1 MCCAM 放置位置实验分析

本实验让学生将所设计的 MCCAM 放置于基准网络中的不同位置, 观察放置位置对网络重识别性能 mAP 和 Rank1 的影响。基准网络(baseline)是指图 4 中在 ImageNet 数据库上预训练的 Res-

Net50 网络(包括 Layer1 至 Layer4 层)与双分支结构。实验结果如表 1 所示, 其中 Duke-to-Market (即 D-M)表示在 Duke 数据集上进行有监督训练, 在 Market 数据集上进行无监督微调训练并测试, Market-to-Duke(即 M-D)表示在 Market 数据集上进行有监督训练, 在 Duke 数据集上进行无监督微调训练并测试, 加粗字体表示最高的识别性能。表 1 中 None 表示不放入 MCCAM 模块, Stage1~Stage3 分别表示将 MCCAM 模块放置于基准网络的 Stage1~Stage3 层之后, Stage1,2 表示在 Stage1 和 Stage2 层后都放置了 MCCAM 模块, Stage2,3 和 Stage1,2,3 含义类似。由实验结果可知, 将 MCCAM 放置在基准网络的 Layer2 与 Layer3 层之后(即 Stage2,3)能够获得最佳重识别性能, 因此后续实验设置遵循本节结论。

表 1 MCCAM 模块放置位置对网络性能影响

放置位置		None	Stage1	Stage2	Stage3	Stage1,2	Stage2,3	Stage1,2,3
D-M	mAP	70.8	72.1	72.5	72.3	72.6	<b>72.9</b>	72.7
	Rank1	88.4	90.3	90.8	90.2	90.4	<b>91.0</b>	90.8
M-D	mAP	59.7	61.8	62.3	61.9	62.4	<b>62.6</b>	62.2
	Rank1	75.6	77.1	77.6	77.4	77.6	<b>78.1</b>	77.6

#### 2.2.2 MCCAM 模块设置实验分析

为了验证所设计的 MCCAM 中组合注意力机制设置方式的有效性, 在本实验中分别将不同的注意力机制插入基准网络的相同位置, 如表 2 所示。表中基准网络+SAM+MCAM 是指在相同位置按照先后顺序依次添加 SAM 与 MCAM 模块, 即后交叉注意机制; 基准网络+MCAM+SAM 是指按照先后顺序依次插入 MCAM 与 SAM, 即前交叉注意机制。由实验结果可知, 由于组合注意能够指导网络把握多层次特征, 使网络对目标相关区域的关注更广, 采用组合注意力机制(即基准网络+SAM+MCAM 和 Baseline +MCAM+SAM)比使用单一注意力机制(即基准网络+SAM 和基准网络+MCAM)在网络性能提升方面更有优势。同时, 由于通道注意力从多个维度上进行, 而空间注意力仅从二维层面进行, 通道优先的学习方式有利于网络对全局主体特征的把握, 因此前交叉注意机制比后交叉注意机制的组合方式性能更好, 能够生成更精细的注意力图。

#### 2.2.3 损失函数权重取值实验分析

为了让学生理解式(9)中  $\lambda_1$ 、 $\lambda_2$  取值对网络重

识别性能的影响, 在 DukeMTMC 和 Market-1501 数据集上进行了损失函数权重取值实验。由图 6 可知, 当  $\lambda_1 < 0.4$  时性能下降明显; 当  $\lambda_1$  在 1 附近时, 它对重识别性能的影响基本保持稳定, 原因是分类损失能够发现样本的真实表现概率, 在进行分类任务的同时可以引导网络充分考虑每个样本的所有身份, 当给予分类损失较小的权重时可能会造成过拟合现象影响网络性能。 $\lambda_2$  控制互学习中的距离度量损失, 它的目的是拉近相同样本之间的距离, 引导网络在小批次内学习到更多的鉴别特征, 由实验结果可看出当  $\lambda_2 = 0.5$  时能够获得最优性能。

表 2 MCCAM 模块设置方式对网络性能影响

设置方式	D-M		M-D	
	mAP	Rank1	mAP	Rank1
基准网络	70.8	88.4	59.7	75.6
基准网络+SAM	71.2	89.1	60.3	76.4
基准网络+MCAM	71.6	89.8	61.2	76.9
基准网络+SAM+MCAM	72.1	90.3	61.9	77.5
基准网络+MCAM+SAM	<b>72.9</b>	<b>91.0</b>	<b>62.6</b>	<b>78.1</b>

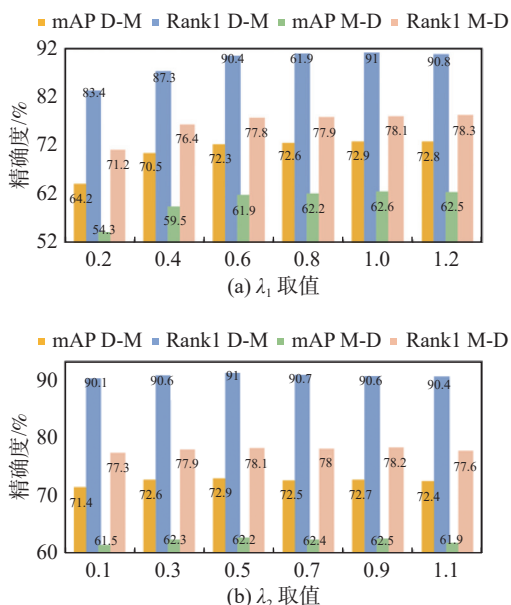


图 6 目标域联合损失函数权重 $\lambda_1$ 、 $\lambda_2$ 取值对网络性能影响

### 2.3 MCCAM-Net 有效性实验分析

本实验对所设计的 MCCAM-Net 的组成模块进行了消融分析，以验证 SPGAN 数据扩充方式和多尺度交叉组合注意力机制 MCCAM 模块的有效性，实验结果如表 3 所示。其中，本实验在相同的双分支网络中分别加入 MCCAM 模块与卷积注意力模块 CBAM<sup>[13]</sup>，以测试所设计 MCCAM 模块的性能提升。由实验结果可以看出，MCCAM 模块能描述更完善、更有鉴别力的网络特征，双分支差异网络促进伪标签之间相互学习，减小噪声伪标签的影响。同时，本实验所添加的 SPGAN 模

块不仅扩充了数据集，改善了模型迁移初始阶段对目标域样本的特征表达力，还在一定程度上提高了模型的泛化能力。

表 3 MCCAM-Net 有效性分析

方法	D-M		M-D	
	mAP	Rank1	mAP	Rank1
ResNet50	31.8	61.9	29.6	46.0
基准网络	70.8	88.4	59.7	75.6
基准网络+CBAM	72.2	90.5	61.5	77.3
基准网络+MCCAM	72.9	91.0	62.6	78.1
基准网络+MCCAM+SPGAN (本设计)	<b>73.6</b>	<b>91.6</b>	<b>63.3</b>	<b>78.5</b>

为了证明本实验所设计的组合注意力模块 MCCAM 相比于其他注意力机制的优越性，在识别网络相同位置分别加入 3 种注意力机制：只关注通道注意的压缩激励模块 (squeeze and excitation, SE)<sup>[14]</sup>、同时关注空间和通道的卷积注意力模块 CBAM<sup>[13]</sup>、本实验设计的交叉关注通道和空间的多尺度交叉组合注意力机制 MCCAM，并在图 7 中通过类激活映射图 Grad-CAM 可视化 layer4 后的输出特征，观察网络对目标区域预测的重要性(颜色饱和度越高，说明预测是目标的可能性越大)。从实验结果可以看出，相比于其他两种注意力模块，MCCAM 注意力机制特征图覆盖到了待识别人物的更多部位，对于行人主体特征的把握也更加全面准确，相比于其他注意力机制的确能够指导网络更准确地关注行人的重要区域信息。



图 7 不同注意力机制输出特征可视化结果

### 2.4 经典算法对比实验

本实验分别在 DukeMTMC、Market-1501、MSMT17 共 3 个数据集上将所设计的整体网络与其他同领域经典算法进行了对比实验，实验过程未使用任何后处理技术(如重新排序等)。对比经典算法包括基于生成对抗网络的 SPGAN<sup>[10]</sup>、CAC-CSP<sup>[15]</sup> 算法，基于单分支网络聚类的 CDS<sup>[4]</sup>、ACT<sup>[7]</sup> 以及 SILC<sup>[16]</sup> 算法，基于多分支网络聚类的 DAAM<sup>[17]</sup>、NRMT<sup>[8]</sup> 和 MMT<sup>[18]</sup> 算法。实验结果如表 4 所示，其中 Duke-to-MSMT(D-Ms) 表示在 Duke 上训练，在 MSMT 数据集上测试。Market-

to-MSMT(M-Ms) 表示在 Market 上训练，在 MSMT 数据集上测试。由实验结果可看出，本设计在 D-M、M-D、D-Ms、M-Ms 领域自适应任务中的识别精度分别达到 73.6%，63.3%，25.1%，23.0%，首次命中率也明显高于其他经典的领域自适应重识别算法，充分说明了本设计网络结构相比于其他领域自适应重识别算法的适用性及优势。

### 2.5 现场硬件实验结果

本实验所设计网络在英伟达人工智能边缘计算开发板 Jetson TX2(如图 8(a)所示)上进行了现场实验，在 Market 数据集上进行了有监督训练，

在新建数据集上进行了无监督微调训练并测试。新建数据集由实验楼走廊楼梯口真实监控摄像头拍摄, 数据集示例如图 8(b), 其中 15XX 为行人身份编号, 每个身份设置二十张训练图像。现场

实验结果如图 9 所示, 行人身份在领域自适应任务中的识别精度高达 88%。本现场实验的目的是让学生利用所学知识, 将实验室训练模型迁移至现场数据中进行验证, 增强学生的学习积极性和自主性。

表 4 MCCAM-Net 与其他经典算法对比

方法	D-M/%			M-D/%		
	mAP	Rank1	Rank5	mAP	Rank1	Rank5
SPGAN(2018)	22.8	51.5	70.1	22.3	41.1	56.6
CDS(2019)	39.9	71.6	81.2	42.7	67.2	75.9
ACT(2020)	60.6	80.5	—	54.5	72.4	—
SILC(2021)	61.8	80.7	90.1	50.3	68.5	80.2
DAAM(2020)	67.8	86.4	—	63.9	77.6	—
NRMT(2020)	71.7	87.8	94.6	62.2	77.8	86.9
MCCAM-Net	72.9	91.0	96.1	62.6	78.1	87.2
<b>MCCAM-Net+SPGAN(本设计)</b>	<b>73.6</b>	<b>91.6</b>	<b>96.5</b>	<b>63.3</b>	<b>78.5</b>	<b>87.5</b>

方法	D-Ms/%			M-Ms/%		
	mAP	Rank1	Rank5	mAP	Rank1	Rank5
CAC-CSP(2021)	13.3	37.0	49.9	10.5	29.3	40.2
NRMT(2020)	20.6	45.2	57.8	19.8	43.7	56.5
MMT(2020)	23.3	50.1	63.9	22.9	49.2	63.1
MCCAM-Net	24.2	51.0	64.5	21.4	47.3	61.9
<b>MCCAM-Net+SPGAN(本设计)</b>	<b>25.1</b>	<b>51.7</b>	<b>64.8</b>	<b>23.0</b>	<b>48.5</b>	<b>62.2</b>



(a) Jetson TX2 开发板

query: 1526



(b) 新建数据集示例

图 8 现场硬件实验设置



图 9 现场硬件实验结果示例

### 3 结束语

本实验设计将教师科研成果转化为学生综合研究型现场硬件实验, 通过设计基于多尺度交叉注意力互学习的无监督领域自适应行人重识别, 消除跨领域行人重识别中错误伪标签对模型性能的影响。首先, 构建差异双分支网络, 在通道注意力的基础上设计交叉注意机制, 并结合空间注意力设计多尺度交叉组合注意力机制, 从多个层面加强了网络对全局主体特征的把握; 然后, 为了减小模型迁移对图像风格的影响, 采用数据扩充策略增强模型对目标域样本风格的预先认知; 最后, 采用软硬标签互学习的方式联合分类损失和互学习损失监督网络的更新过程。

教学实践表明, 所设计的网络结构能够大大提高模型的跨领域行人重识别性能, 将本研究引入现场硬件教学实验, 将促进研究型实验教学与理论教学的结合, 在加强学生理解理论知识的同时, 提高其独立思考能力和科研创新意识。

### 参考文献

- [1] AJAKAN H, GERMAIN P, LAROCHELLE H, et al. Domain-adversarial neural networks[C]//Proceedings of the Second Workshop on Transfer and Multi-Task Learning: Theory meets Practice (NIPS). Montreal, Canada: MIT Press, 2014: 1–7.
- [2] ZHONG Z, ZHENG L, LUO Z, et al. Invariance matters: Exemplar memory for domain adaptive person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Angeles, USA: IEEE, 2019: 598–607.
- [3] SHAN L, LI H, LI C, et al. Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification[C]//Proceedings of the British Machine Vision Conference. Newcastle, UK: British Machine Vision Association, 2018: 1–13.
- [4] WU J, LIAO S, WANG X, et al. Clustering and dynamic sampling based unsupervised domain adaptation for person re-identification[C]//IEEE International Conference on Multimedia and Expo (ICME). Shanghai, China: IEEE, 2019: 886–891.
- [5] FU Y, WEI Y, WANG G, et al. Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, South Korea: IEEE, 2019: 6112–6121.
- [6] ZHAI Y, LU S, YE Q, et al. Ad-cluster: Augmented discriminative clustering for domain adaptive person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual: IEEE, 2020: 9021–9030.
- [7] YANG F, LI K, ZHONG Z, et al. Asymmetric co-teaching for unsupervised cross-domain person re-identification[C]//Proceedings of the AAAI Conference on Artificial Intelligence, New York, USA: AAAI, 2020, 34(7): 12597–12604.
- [8] ZHAO F, LIAO S, XIE G, et al. Unsupervised domain adaptation with noise resistible mutual-training for person re-identification[C]//European Conference on Computer Vision. Springer: Virtual, 2020: 526–544.
- [9] XIAO R, LIU Z, WU B. Teacher-Student Competition for Unsupervised Domain Adaptation[C]//International Conference on Pattern Recognition (ICPR). Milan, Italy: IEEE, 2021: 8291–8298.
- [10] DENG W, ZHENG L, YE Q, et al. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 994–1003.
- [11] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: A unified embedding for face recognition and clustering[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015: 815–823.
- [12] HINTON G, VINYALS O, DEAN J. Distilling the Knowledge in a Neural Network[J]. *Computer Science*, 2015, 14(7): 38–39.
- [13] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database[C]//IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009: 248–255.
- [14] WOO S, PARK J, LEE J, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: IEEE, 2018: 3–19.
- [15] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 7132–7141.
- [16] LI H, PANG J, TAO D, et al. Cross adversarial consistency self-prediction learning for unsupervised domain adaptation person re-identification[J]. *Information Sciences*, 2021, 559: 46–60.
- [17] AINAM J, QIN K, OWUSU J, et al. Unsupervised domain adaptation for person re-identification with iterative soft clustering[J]. *Knowledge-based Systems*, 2021, 212: 106644.
- [18] HUANG Y, PENG P, JIN Y, et al. Domain adaptive attention learning for unsupervised person re-identification[C]//proceedings of the AAAI Conference on Artificial Intelligence. New York, USA: AAAI, 2020, 34(7): 11069–11076.