DOI：10. 13471/j. cnki. acta. snus. ZR20250070

知识蒸馏与掩码重构的域泛化行人重识别*

郑昊天，胡海峰

中山大学电子与信息工程学院，广东 广州 510006

摘　要：域泛化行人重识别的挑战源于当前基准方法的2个固有局限性：1）数据集之间存在明显的域间隙，2）数据集域内多样性不足。现有一些多领域联合训练方法，往往无法充分学习跨域数据集间潜在的身份线索。为了克服上述局限，本文通过一种双分支策略来增强模型泛化性能。首先针对大规模预训练的扩展模型进行知识蒸馏，同时针对现有多域训练数据进行掩码图像特征挖掘。常用的域泛化行人重识别协议基准上的实验证明了本文方法的性能。在以Market-1501为目标域的留一法测试中，本文方法相对于基准方法提高了16.2%的Rank-1准确度，相对现存最佳方法则在Rank-1准确度上实现了3.6%的提升。

关键词：行人重识别；域泛化；知识蒸馏；掩码图像

中图分类号： TP391.41　　**文献标志码：** A　　**文章编号：** 2097－0137（2025）05－0043－07

A distillation and masked approach for domain generalizable person re-identification

ZHENG Haotian，*HU Haifeng*

School of Electronics and Information Technology，*Sun Yat-sen University*，*Guangzhou 510006*，*China*

Abstract：The challenge of domain generalization stems from two inherent limitations in current person re-identification benchmarks：1）significant inter-dataset domain gaps，and 2）insufficient intra-dataset diversity. While existing multi-domain joint training approaches attempt to address these issues，they often fail to fully exploit latent discriminative identity cues across datasets. To address the aforementioned limitations，our framework enhances network generalization capabilities through a dual-branch strategy：knowledge distillation employed from a large-scale pre-trained model along with mask image feature mining performed on existing multi-domain training data. Extensive experiments on popular domain generalization person ReID benchmarks demonstrate that our method can achieve superior performance. Notably，our approach achieves a 16.2% Rank-1 accuracy gain over the baseline and a 3.6% improvement over existing state-of-the-art methods under the leave-one-out protocol using Market-1501.

Key words：person re-identification；domain generalizable；knowledge distillation；masked image

行人重识别主要解决视野不重叠的多摄像头网络下的行人身份关联问题，通过提取具有判别性的视觉特征，实现在不同时空、拍摄设备、各种光照条件的差异化场景中对特定个体的准确匹配与追踪。当前，行人重识别方法主要基于深度学习框架，通过优化特征提取网络和嵌入空间映射，在主流基准数据集上取得了优异的性能表现（冯展祥等，2020）。然而，因其优异性能建立在目标域已具

* **收稿日期：**2025-04-12　　**录用日期：**2025-05-07　　**网络首发日期：**2025-07-04
基金项目：国家自然科学基金（62076262）
作者简介：郑昊天（1999年生），男；**研究方向：**行人重识别；E-mail：henght26@mail2.sysu.edu.cn
通信作者：胡海峰（1977年生），男；**研究方向：**深度学习；E-mail：huhaif@mail.sysu.edu.cn

备充足标注训练数据的强假设上，这些方法具有关键性局限。具体地说，现有方法受到2个根本假设的约束：数据可获得性约束和标注成本约束。前者要求目标场景已预先采集足够数量的监控图像数据；后者需要投入大量工作对目标域数据进行身份标注。这种依赖特定领域标注数据的工作特点，严重限制了ReID系统在真实场景中的部署适用性。因此，突破域依赖实现有效的域泛化行人重识别，是当前研究的重点突破方向。

提升泛化性能的研究主要分为2个方向，首先是无监督领域自适应（UDA），即在源域训练的基础上，利用大量目标域无标注数据进行模型微调。由于利用了目标域数据分布，UDA能获得更广泛的性能提升空间。但因仍需要采集一定规模的目标域数据，存在相应的数据获取成本（朱锦雷等，2023；郭迎春等，2022）。为突破这一局限性，研究者提出了域泛化（DG）行人重识别方法（Dai et al.，2021）。该方法旨在仅利用有限已知源域数据进行模型训练后，在完全零样本条件下直接应用于新领域，从而提升模型在未知且规模可能无限的目标域上的识别性能。后者与前者的区别在于，DG完全不需要接触目标域数据，更具实际应用价值。DG方法的核心在于如何从有限源域中提取具有跨域不变性的特征表示，这对算法设计提出了更高要求（叶钰等，2020）。

DG行人重识别研究主要分为3类：一是领域差异最小化方法，通过消除源域间差异来提升泛化能力。但其面临的根本挑战是目标域完全不可知，这一约束条件下难以准确建模和消除未知的域间差异，使得现有方法只能基于有限源域进行近似优化，导致泛化性能受限。其二是域不变特征解耦方法，采用特征解耦技术分离域不变与域特定特征，然而解耦过程严重依赖已知源域的分布特性，面对分布差异显著的未知域时，特征解耦的有效性会受到劣化。其三是多专家集成方法，其优势在于特征表示学习实现了多维度的决策空间拓展。但该类方法既需要精心设计子模型间的交互机制，对模型集成权重的优化又极大提升了复杂度。

本文提出了一种掩码重构和知识蒸馏解决方案。与传统的模型架构改进不同，本文引入了掩码图像建模作为辅助任务实现自监督学习，通过缺失部分推理增强模型泛化性能提高鲁棒性，令模型能从数据本身发现潜在模式；同时，在多域联合训练背景下模型能发现并强化那些在不同域中保持稳定的身份特征。

1 方 法

1.1 模型框架

提出的多源域学习框架如图1所示，采用双并行分支transformer架构设计。其中，学生与教师模型均为transformer神经网络，交叉注意力解码层和自注意力解码层为附加的一层transformer单元层。

在网络初始化阶段，学生分支加载通用图像分类的预训练权重作为初始化训练基础，教师分支则加载大规模预训练数据的模型参数作为数据源辅助。模型的优化目标融合了多目标的监督信息，包括身份识别任务的分类交叉熵与判别特征三元组损失、实现知识迁移的泛化蒸馏损失和掩码图像建模的重建损失。特别地，架构中引入了一对不对称的多头注意力机制层，按照以下训练方案将知识蒸

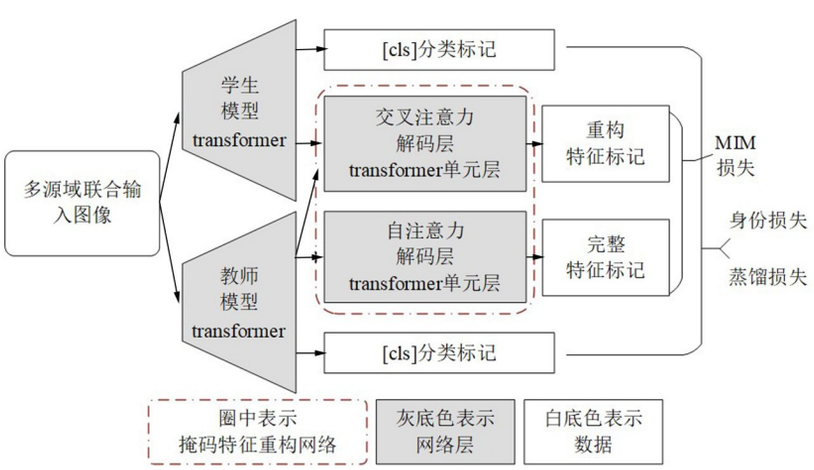

图1 模型总体结构图

Fig. 1 Overall architecture of model in this paper

馏的双分支模型和掩码图像的重构方法结合：

1）教师分支的自注意力层根据完整的末层特征输入生成特征标记，学生分支则处理经过随机掩码屏蔽的注意力输入，这种设计强制模型从大规模源域外数据支撑的特征分布中学习鲁棒的特征重建能力。

2）新增网络层通过掩码图像建模（MIM）损失进行端到端优化，通过该MIM损失函数量化评估学生网络在部分掩码条件下重建完整特征表示的能力，提升模型在域泛化场景中的重识别性能。

1.2　知识蒸馏方法

知识蒸馏是一种迁移学习技术，在本文中是通过知识迁移机制将复杂教师模型中蕴含的与行人身份有关的提取方式紧凑地转移至学生模型中。给定输入数据样本x，设教师模型为M_t，学生模型为M_s，知识蒸馏的标准范式为优化损失

$$\mathcal{L}_{KD} = D\left(M_t(x), M_s(x)\right),$$

其中$D(\cdot)$表示差异度量函数。

教师模型在产生最终分类预测输出的过程中，还输出了更重要的隐含知识表征和注意力分布，一次推理中细分的各阶段可蒸馏输出知识包括各层的注意力知识$\text{attention}\left(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}\right)$、中间层映射特征$\text{feat}_n$和输出层logits分类标签$y_{\text{logit}}$。其中

$$\text{attention}\left(\boldsymbol{Q},\ \boldsymbol{K},\ \boldsymbol{V}\right) = \text{softmax}\left(\frac{\boldsymbol{Q}\cdot\boldsymbol{K}}{\sqrt{d_k}}\right)\boldsymbol{V},$$

$$\text{feat}_n = \text{block}\left(\text{feat}_{n-1}\right),$$

$$y_{\text{logit}} = \text{classifier}\left(\text{feat}_n\right),$$

$\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}$分别表示transformer的多头注意力层中的查询矩阵、键矩阵、值矩阵，d_k是矩阵的维度，$\text{block}(\cdot)$表示transformer的一层编码块，$\text{classifier}(\cdot)$表示模型后端的分类器层。利用多层次知识迁移，在特征提取网络层中，学生网络能够继承教师模型通过大规模预训练获得的部分感知能力；在分类器网络层中，则更有效地整合学习跨域的高级语义表征。而且，采用灵活融合不同层次表征的策略，学生网络能适应特定域辨识需求。本文的知识蒸馏中，采用了教师网络最后层的映射特征feat_{n-1}和输出分类层最终标签分类y_{logit}。

在量化不同类型的知识差异时，使用的差异量化函数也不同。对于logit分类标签的蒸馏，通常使用交叉熵损失

$$\mathcal{L}_c = \sum\nolimits_i y_i \log\left(\frac{\mathrm{e}^{\hat{y}}}{\sum_j \mathrm{e}^{y_j}}\right),$$

其中i、j分别为某一行人分类和不属于该类的其他行人分类，y、$\hat{y}$分别表示真实分类标签和网络推理的分类标签。

对于映射特征的蒸馏，则使用三元组损失挖掘困难样本

$$\text{L}_{\text{tri}} = \text{ReLU}\left(\alpha + D\left(d_a,\ d_p\right) - D\left(d_a,\ d_n\right)\right),$$

d_p, d_n分别为锚样本d_a的困难挖掘正样本与负样本。蒸馏的总体损失函数是使用教师模型作为以上损失中的监督部分形成的复合损失

$$\mathcal{L}_{c(KD)} = \sum y_t \log\left(\frac{\mathrm{e}^{\hat{y}_s}}{\sum_j \mathrm{e}^{y_j}}\right), \tag{1}$$

$$\mathcal{L}_{\text{tri}(KD)} = \text{ReLU}\left(\alpha + D\left(d_a, d_{tp}\right) - D\left(d_a, d_{tn}\right)\right), \tag{2}$$

其中y_t, $\hat{y}_s$, d_{tp}, d_{tn}分别表示教师模型、学生模型输出的logits分类标签、教师模型映射特征挖掘的困难正样本和负样本。

$\mathcal{L}_{\text{tri}(KD)}$挖掘困难样本时，由教师模型基于大规模预训练获得跨域表征对齐能力，通过其嵌入空间选取具有域不变性的困难三元组。这种设计使得学生网络能根据教师网络在大规模预训练数据集中习得的鲁棒特征不变性，学习到跨域且对齐的身份线索提取策略。该策略不仅避免了直接使用多域数据训练导致的特征分布冲突问题，同时解决了多域训练受局限的泛化。

如图2所示，在传统训练中跨域困难样本与单一域困难样本对于网络而言不存在更显著的区分差异。图2中，后2个样本都属于不同标签的样本，但缺乏进一步的区分监督信息。传统训练中，投入额外域数据学习时，量化更细粒度的差异挑战性很大，对于存在跨域差异与其他差异的困难样本仅能进行定性区分和模糊判断。蒸馏训练中，教师网络会对各个样本输出不同的特征、标签分布，以及具备更多维度且含有更细致的语义的信息。如：对衣着颜色敏感的特征维度，能提供图2中后2个样本的区分性定量差异。因此，通过预训练获得的度量空间挖掘的困难样本能提供定量的差异评估，这是传统方法或直接使用大量域数据无法实现的优势。

直接对教师与学生网络的知识进行损失收敛学习容易造成学生网络对教师网络的过拟合。针对这一问题，Wang et al.（2021）表明引入温度系数

图2 传统训练与本文样本学习的差异对比

Fig. 2 Contrast between traditional training and the proposed sample learning

能调整损失函数的训练倾向；当温度系数较大时，模型会生成更加平滑的概率分布；而较小的温度系数促使模型表现出确定性的分类决策。对式(1)引入温度系数τ，得：

$$\mathcal{L}_{c(KD-\tau)}=\sum_i y_t \log\left(\frac{e^{\hat{y}_t/\tau}}{\sum_j e^{y_j/\tau}}\right). \tag{3}$$

通过调节τ控制知识蒸馏的软性容错。具体来说，在训练中采用了随训练逐渐缩小的τ，实现在网络训练初期提高稳定性、在训练后期提高准确率的效果。通过以上设计，本文完善了蒸馏师生网络的训练知识提炼。其中，教师模型通过其输出的软标签概率分布和困难样本筛选，为学生模型传递了大规模预训练数据含有的跨域泛化所需结构化身份辨识知识，使学生模型在较短训练时长下就能保持较好的泛化能力。知识蒸馏的总损失为

$$\mathcal{L}_{KD}=\mathcal{L}_{c(KD-\tau)}+\mathcal{L}_{\mathrm{tri}(KD)}.$$

1.3 掩码图像建模方法

掩码图像建模方法(Zhao et al., 2021)作为一种新兴的自监督学习范式，通常用于提升视觉特征的鲁棒性和泛化能力。该方法首先对输入图像实施随机分块或标记级别的掩码操作，迫使模型仅基于可见部分的上下文图像进行推理；其次，通过像素或特征级重建任务驱使网络理解图像各部分间的结构关联；最后，由关联预测和重构完整的视觉表征，显著增强模型对不完整输入的适应能力。受掩码图像建模方法的启发，本文transformer架构采用对输入图像的分块标记随机掩码的策略，它通过多层自注意力机制学习可见标记与掩码标记间的相互依赖关系。

1）输入图像在transformer输入处理下自然地完成规则网格分块；随后，为了更合理地构建掩码输入，类似于transformer中常用的位置嵌入标记，引入一个可学习的掩码嵌入标记x_{mask}。对学生网络提取的标记序列实施概率性x_{mask}替换，替换概率遵循二项分布。该过程可表述为

$$\mathrm{feat}_n=\left[x_{[\mathrm{cls}]},\ x_1,\ x_2,\ \cdots\right],$$

$$x_i=rx_i+(1-r)x_{\mathrm{mask}},\ r\in\{0,\ 1\},$$

其中$x_1, x_2, \cdots$是transformer中一层特征含有的多个标记。本节的掩码策略融合了自监督学习与知识蒸馏，并利用了上节中双分支蒸馏中的教师模型，即在学生网络负责重建被掩码标记的同时，经过大规模预训练的教师网络则提供完整标记的监督信号。相比传统MIM方法，本节方法利用教师网络的强表征能力引导掩码区域的特征学习，显著提升了掩码特征重建的语义一致性。

2）在教师和学生transformer架构末端均引入重构交叉注意力解码层，其结构与transformer中各单元层一致，通过异构双分支实现掩码特征显式重建。其中，教师分支对完整图像标记执行自注意力运算，生成完整的全局特征表示；学生分支则基于掩码后标记，以自身特征为查询序列和键序列、教师特征为值序列输入进行交叉注意力计算。注意力解码层使用的多头注意力架构为常用的768维向量共8组并行注意力头，包含归一化处理和残差旁路连接。这个设计有效利用了注意力层的交叉运算特性，结合了蒸馏教师网络丰富的语义先验知识，从而优化重建的训练效果。以上的双分支注意力重构表述为

$$\boldsymbol{y}_s=\mathrm{softmax}\left(\frac{\boldsymbol{f}_s\cdot\boldsymbol{f}_s}{\sqrt{d}}\right)\boldsymbol{f}_t,$$

$$\boldsymbol{y}_t=\mathrm{softmax}\left(\frac{\boldsymbol{f}_t\cdot\boldsymbol{f}_t}{\sqrt{d}}\right)\boldsymbol{f}_t,$$

其中$\boldsymbol{f}_s, \boldsymbol{f}_t$分别表示学生、教师网络的末层特征，$d$为特征序列长度。最后，由学生网络重构的特征标记与教师网络的完整标记进行softmax对齐，计算重建掩码图像建模损失

$$\mathcal{L}_{\mathrm{MIM}}=\sum^{i}\mathrm{softmax}\left(\boldsymbol{y}_t(i)\right)\log\left[\mathrm{softmax}\left(\boldsymbol{y}_s(i)\right)\right].$$

在MIM的相关研究中，对于网络结构、监督方法、掩码方法均存在不同的多样性设计。相比而言，本文方法与现有研究的差异主要体现在：1）监

督数据来源。现有方法多采用常规的自监督训练策略，这一策略在特定域的性能效果表现良好，但对于域泛化问题则暴露出泛化能力不足的缺陷；2）掩码信息混淆性。一些研究采用了直接将掩码标记清空的策略，而另一些研究则使用已知的规则将掩码填充到掩码标记中。如：Devlin et al.（2019）的策略中一定概率将掩码标记替换为任一有意义的标记，Ergasti et al.（2024）在一步掩码后额外重建每个预测的标记并二次预测其掩码与否。本文方法中针对复杂的域泛化图像标记，采用了较为简单的可学习掩码嵌入标记方式；3）额外附加的网络结构。一些研究（Ma et al.，2023；Yang et al.，2023）使用了较为复杂的解码器完成重构任务，本文只采用了两个额外的多头注意力层，在计算复杂度和模型参数效率之间实现了平衡。

本节提出的框架结合了师生网络与掩码图像建模。一方面，教师网络根据自身知识构建域泛化的特征空间映射，建立了对局部特征掩蔽不敏感的鲁棒性标记；另一方面，在掩码重构中，学生网络重构的目标设置为教师网络提供的具有广泛跨域判别力的完整身份表示。通过这一师生监督方式，学生网络无需耗费大量训练时间在大规模数据中进行微调训练，就能学习到更泛化的身份特征判别线索。同时，与蒸馏学习领域中特权蒸馏的概念类似，仅基于掩码信息的重构未掩码信息任务双重优化了学生网络的表征能力：增强对图像域特异性特征的理解以及强化对身份判别特征的有效捕捉。

2　实验结果与分析

2.1　数据集与基准协议

本文采用常用的 Market1501（Zheng et al.，2015）、DukeMTMC-reID（Zheng et al.，2017）、CUHK03（Li et al.，2014）和 MSMT17（Wei et al.，2018）行人重识别数据集进行实验。实验采用"留一法"协议：将其中 1 个数据集的测试集作为测试集，其余 3 个数据集的训练集的并集用于模型训练。需说明的是，由于 DukeMTMC-reID 数据集已停止公开，本文仅将其用于训练而不报告其测试结果。为保证与现有方法的公平对比，遵循标准图像预处理流程，包括归一化、图像尺寸随机裁剪、水平翻转等数据增强处理。在模型评估方面，采用 Rank-1 准确率和平均精度均值（mAP）等评估指标，以量化模型在不同数据集上的性能表现。

2.2　实验基准与设置

以 He et al.（2021b）为基础建立了 2 个独立的 transformer 网络分别用作学生和教师网络，并附加了一层 transformer 单元块作为交叉注意力解码层和自注意力解码层。对于学生网络，使用在 ImageNet（Russakovsky et al.，2015）上预训练的通用权重初始化；而对于教师网络，采用在 Luperson（Fu et al.，2021）上大规模预训练的权重初始化。所有的实验均在 2 个 GTX 1080 Ti GPU 上完成。输入图像均被调整为 256 × 128 大小，并参考 He et al.（2021b）应用了数据增强手段，包括正则化、随机裁剪、水平翻转等。训练批次大小设为 120，每个批次中包含来自 3 个源域数据集的各 40 张图像，每个数据集的图像由 10 个行人、每人 4 张图像组成。初始学习率设为 4×10^{-4}，在前 10 个训练周期中使用余弦预热法，并在随后的每 40 个周期训练后将学习率降为原本的 0.1，共训练 120 个周期。

2.3　性能对比

表 1 为本文方法与现有方法的性能对比。表格中的数据集名表示“留一法”协议中的目标域。本文方法在所有目标测试数据集上均稳定超越各类同领域方法。在 Rank-1 准确率和 mAP 2 个常用指标上取得显著提升。在最具挑战性的 MSMT17 数据集上，本文方法 Rank-1 准确率和 mAP 分别提升了 1.1%、1.0%，显示了其在复杂多样化场景下的强鲁棒性。此外，在 Market1501 数据集上，本方法的 Rank-1 准确率和 mAP 分别提升了 3.6%、2.2%。所提出的方法在多域数据集训练上的稳定优势，证明了其具有优异的鲁棒性和泛化能力。

2.4　消融性实验

通过去除模型各组成模块，评估了不同配置对域泛化行人重识别性能的影响。消融实验设计如下：通过控制大规模预训练知识蒸馏与掩码图像建模组件相应损失的启用状态，分析各模块的贡献度；其中，“学生自监督”表示在掩码图像建模中，被掩码输入和完整输入均由学生网络处理与重建；“学生挖掘”表示在知识蒸馏中仅使用教师网络蒸馏分类标签，而不包含挖掘样本与温度调节策略。实验数据见表 2，因 DukeMTMC-reID 数据集停止公开，未纳入结果统计。

结果表明，知识蒸馏方法和掩码图像建模方法引入能显著提升视觉 transformer（ViT）基准网络在行人重识别任务中的性能。其中，掩码图像建模模

表1 不同方法的性能对比

Table 1 Comparison of the performance of different methods

方法	Market		CUHK		MSMT	
	R1	mAP	R1	mAP	R1	mAP
SSKD（He et al.，2021a）	86. 3	65. 6	45. 4	45. 8	47. 2	20. 0
RaMoE（Dai et al.，2021）	82. 0	56. 5	36. 6	35. 5	34. 1	13. 5
ISR（Dou et al.，2023）	87. 0	70. 5	36. 6	37. 8	56. 4	30. 3
SVIL（Lv et al.，2024）	86. 1	65. 0	44. 1	43. 2	44. 6	19. 9
ViT（Dosovitskiy et al.，2020）	74. 4	50. 5	28. 8	30. 1	33. 7	16. 7
本文方法	**90. 6**	**72. 7**	**48. 4**	**48. 0**	**57. 4**	**31. 3**

表2 本文方法的消融性实验结果

Table 2 Ablation study results of the proposed method

	Market		CUHK		MSMT	
	R1	mAP	R1	mAP	R1	mAP
ViT（Dosovitskiy et al.，2020）	74. 4	50. 5	28. 8	30. 1	33. 7	16. 7
ViT+掩码图像建模（学生自监督）	74. 9	50. 8	29. 5	31. 0	42. 0	18. 0
ViT+掩码图像建模	75. 3	51. 4	29. 5	31. 1	42. 5	18. 9
ViT+知识蒸馏（学生挖掘）	87. 7	68. 2	38. 0	40. 5	50. 3	24. 1
ViT+知识蒸馏	87. 9	70. 1	41. 7	43. 2	55. 0	29. 4
本文方法	**90. 6**	**72. 7**	**48. 4**	**48. 0**	**57. 4**	**31. 3**

块带来显著性能提升，尤其在复杂场景数据集MSMT17上实现了Rank-1准确率8.8%的增长，且各基准数据集均呈现稳定提升。此外，使用学生网络自监督重建带来了一定的性能损失，展示了教师网络蒸馏知识重建对泛化能力的提高。

引入知识蒸馏技术后，模型在所有数据集上都取得突破性提高。如：在Market1501数据集上仅使用知识蒸馏的方法Rank-1提升了13.5%，mAP提升了19.6%，充分验证了利用大规模预训练模型经过域外数据训练知识进行迁移的泛化价值。此外，使用教师挖掘困难样本等知识蒸馏方法的优化相对于学生自行挖掘样本，在各个数据集间也展示出一致的提升。

结合知识蒸馏方法与掩码图像建模方法时，模型在各个目标数据集上均达到最优性能，相比单个方法生效时取得的优势更大，这表明两个方法之间存在一定的协同效应。

3 结 语

本文提出了一种大规模预训练模型知识蒸馏与掩码图像建模的域泛化行人重识别方法。一方面，基于模型蒸馏引入域外训练数据，增强模型的泛化能力；另一方面，通过MIM技术深入挖掘域内数据的潜在关联性，提升特征表示的鲁棒性。知识蒸馏通过迁移预训练模型的语义知识，缓解了目标域数据稀缺的问题，使模型能够学习更具判别性的特征表示。而MIM模块通过自监督学习策略，促使模型更好地捕捉行人图像的局部细节与结构信息，从而提升跨域场景下的特征不变性。实验结果表明，这两种机制具有显著的互补效应，两者同时使用的策略显著优于基准模型，在跨域泛化性上展现出强大的优势。

参考文献：

冯展祥，朱荣，王玉娟，等，2020. 非可控环境行人再识别综述[J]. 中山大学学报（自然科学版中英文），59(3)：1-11.

郭迎春，冯放，阎刚，等，2022. 基于自适应融合网络的跨域行人重识别方法[J]. 自动化学报，48(11)：2744-2756.

叶钰，王正，梁超，等，2020. 多源数据行人重识别研究综述

［J］. 自动化学报，46(9)：01869-01884.

朱锦雷，李艳凤，陈后金，等，2023. 近邻优化跨域无监督行人重识别算法［J］. 中国图象图形学报，28(11)：3471-3484.

DAI Y，LI X，LIU J，et al，2021. Generalizable person re-identification with relevance-aware mixture of experts［C］// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville，TN，USA：16140-16149.

DEVLIN J，CHANG M W，LEE K，et al，2019. Bert：Pre-training of deep bidirectional transformers for language understanding［C］// The North American Chapter of the Association for Computational Linguistics：Human Language Technologies. Minneapolis，MN，USA：4171-4186.

DOSOVITSKIY A，BEYER L，KOLESNIKOV A，et al，2020. An image is worth 16x16 words：Transformers for image recognition at scale［EB/OL］. arXiv：2010.11929.

DOU Z，WANG Z，LI Y，et al，2023. Identity-seeking self-supervised representation learning for generalizable person re-identification［C］// IEEE/CVF International Conference on Computer Vision. Paris，France：15801-15812.

ERGASTI A，FONTANINI T，FERRARI C，et al，2024. Mars：Paying more attention to visual attributes for text-based person search［EB/OL］.arXiv.2407.04287.

FU D，CHEN D，BAO J，et al，2021. Unsupervised pre-training for person re-identification［C］//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville，TN，USA：14745-14754.

HE L，LIU W，LIANG J，et al，2021a. Semi-supervised domain generalizable person re-identification［EB/OL］. arXiv：2108.05045.

HE S，LUO H，WANG P，et al，2021b. Transreid：Transformer-based object re-identification［C］// IEEE/CVF International Conference on Computer Vision. Montreal，QC，Canada：14993-15002.

LI W，ZHAO R，XIAO T，et al，2014. DeepReID：Deep filter pairing neural network for person re-identification［C］//IEEE Conference on Computer Vision and Pattern Recognition. Columbus，OH，USA：152-159.

LV K，CHEN H，ZHAO C，et al，2024. Style variable and irrelevant learning for generalizable person re-identification［J］. ACM Trans Multimedia Comput Commun Appl，20(9)：1-22.

MA H，LI X，YUAN X，et al，2023. Two-phase self-supervised pretraining for object re-identification［J］. Knowl Based Syst，261：0110220.

RUSSAKOVSKY O，DENG J，SU H，et al，2015. ImageNet large scale visual recognition challenge［J］. Int J Comput Vis，115(3)：211-252.

WANG F，LIU H，2021. Understanding the behaviour of contrastive loss［C］// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville，TN，USA：2495-2504.

WEI L，ZHANG S，GAO W，et al，2018. Person transfer gan to bridge domain gap for person re-identification［C］// IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City，UT，USA：79-88.

YANG S，ZHOU Y，ZHENG Z，et al，2023. Towards unified text-based person retrieval：A large-scale multi-attribute and language search benchmark［C］// 31st ACM International Conference on Multimedia. Ottawa，ON，Canada：4492-4501.

ZHAO Y，WANG G，LUO C，et al，2021. Self-supervised visual representations learning by contrastive mask prediction［C］// IEEE/CVF International Conference on Computer Vision. Montreal，QC，Canada：10160-10169.

ZHENG L，SHEN L，TIAN L，et al，2015. Scalable person re-identification：A benchmark［C］//IEEE International Conference on Computer Vision. Santiago，Chile：1116-1124.

ZHENG Z，ZHENG L，YANG Y，2017. Unlabeled samples generated by gan improve the person re-identification baseline in vitro［C］//IEEE International Conference on Computer Vision.Venice，Italy：3754-3762.

（责任编辑　王海蓉）