

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-13

论文引用格式: Li Qihong, Ma Bingpeng. Role-aware hard sample modeling for unsupervised visible-infrared person re-identification[J/OL]. Journal of Image and Graphics, XXXX: 1-13. DOI: 10.11834/jig.260448. (李骑宏, 马丙鹏. 面向无监督可见光-红外行人重识别的角色感知难样本建模[J/OL]. 中国图象图形学报, XXXX: 1-13. DOI: 10.11834/jig.260448.) [DOI: 10.11834/jig.260448]

面向无监督可见光-红外行人重识别的角色感知难样本建模

李骑宏, 马丙鹏

中国科学院大学计算机科学与技术学院

摘要: 目的 无监督可见光-红外行人重识别(Unsupervised Visible-Infrared Person Re-identification, USVI-ReID)旨在无标注条件下实现跨模态行人匹配。然而,现有方法中的困难样本挖掘策略通常未考虑困难正负样本在对比学习中的不同优化作用,导致模型的类内紧凑性不足且类间分离性有限。针对上述问题,提出一种角色感知难样本建模方法,在嵌入空间中对困难正负样本进行差异化建模,以协同优化类内一致性与类间判别性。方法 首先,基于共享伪标签构建模态特定身份中心,并利用同一身份在不同模态中心之间的偏移生成困难正样本,使生成样本能够在保持身份一致性的同时包含更具挑战性的跨模态差异;其次,通过子簇划分与近邻簇搜索,挖掘不同身份间的局部相似关系,并沿近邻簇方向构造困难负样本,以增强模型对易混淆身份的判别能力;此外,受课程学习思想启发,引入渐进式训练策略,通过逐步调整困难样本生成幅度控制训练难度,提高模型优化稳定性。结果 在SYSU-MM01和RegDB两个公开可见光-红外行人重识别数据集上的实验结果表明,所提方法能够有效提升无监督跨模态特征学习性能。在SYSU-MM01数据集全搜索模式下,相比现有最优方法,Rank-1和mAP(mean average precision)分别提升5.5%和4.1%;在RegDB数据集的可见光-热成像搜索模式中,Rank-1和mAP分别达到93.0%和86.2%。结论 通过区分困难正负样本在对比学习中的优化角色,并针对性地构造跨模态变化样本和类别边界样本,所提方法能够同时增强跨模态一致性和身份判别能力,为无监督可见光-红外行人重识别提供了一种有效的难样本建模策略。

关键词: 计算机视觉;行人重识别;可见光-红外;无监督学习;难样本学习;对比学习

Role-aware hard sample modeling for unsupervised visible-infrared person re-identification

Li Qihong, Ma Bingpeng

School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: **Objective** Visible-infrared person re-identification (VI-ReID) aims to retrieve images of the same person across visible and infrared modalities captured by different cameras, and has attracted considerable attention due to its potential applications in intelligent surveillance and all-weather pedestrian retrieval. However, VI-ReID remains challenging because of the large modality discrepancy caused by different imaging mechanisms. Visible images usually contain rich color and texture information, whereas infrared images mainly describe thermal radiation and structural characteristics.

收稿日期: 2026-07-23;修回日期: 2026-08-23

基金项目: 国家自然科学基金项目(项目编号: 62276246)

Supported by: National Natural Science Foundation of China(62276246)

© 中国图象图形学报版权所有

resulting in significant distribution differences between the two modalities. Recently, deep learning-based VI-ReID methods have achieved remarkable progress by learning modality-invariant feature representations with identity annotations. Nevertheless, acquiring large-scale identity annotations for visible and infrared images is expensive and time-consuming, which limits their application in practical scenarios. Therefore, unsupervised visible-infrared person re-identification (USVI-ReID) has gradually become an important research topic. Existing USVI-ReID methods generally rely on clustering algorithms to generate pseudo labels and employ contrastive learning frameworks to optimize feature representations. Although these methods have achieved promising performance, the quality of learned representations is still affected by insufficient hard sample mining strategies. Most existing methods select hard samples according to feature distances and fail to consider the different optimization roles of hard positive and hard negative samples. In contrastive learning, hard positives should maintain identity consistency while introducing challenging cross-modality variations to reduce intra-class discrepancy, whereas hard negatives should emphasize visually similar but identity-different samples to improve inter-class discrimination. Therefore, directly adopting a unified hard sample selection strategy may introduce noisy positive samples or ineffective negative samples, which restricts the optimization ability of contrastive learning. To address this problem, this study aims to propose a role-aware hard sample modeling framework for improving unsupervised cross-modality representation learning. **Method** A role-aware hard sample modeling framework is proposed to construct hard positive and hard negative samples according to their different optimization objectives. First, shared pseudo labels are obtained through cross-modality clustering, and modality-specific identity centers are constructed for each identity cluster. Based on the feature shift between identity centers of different modalities, hard positive samples are generated along the cross-modality shift direction. Unlike conventional hard positive mining strategies that select samples with large distances from cluster centers, the proposed method introduces challenging modality variations while preserving identity consistency, thereby improving cross-modality feature alignment. Second, a hard negative modeling strategy based on local cluster relationships is proposed. Specifically, intra-identity features are divided into multiple sub-clusters, and neighboring clusters belonging to different identities are searched according to feature similarity. Hard negative samples are then generated along the direction of neighboring clusters. By introducing locally confusing identity information, the proposed strategy enables the model to focus on difficult inter-class relationships and learn more discriminative representations. Furthermore, inspired by curriculum learning, a progressive training strategy is introduced to gradually adjust the difficulty of generated hard samples. Instead of introducing excessively difficult samples at the beginning of training, the proposed strategy progressively increases the generation magnitude, allowing the model to learn from simple samples to challenging samples and improving optimization stability. **Result** Extensive experiments are conducted on two widely used USVI-ReID benchmark datasets, SYSU-MM01 and RegDB. SYSU-MM01 contains visible and infrared images collected from multiple indoor and outdoor cameras, with significant variations in illumination, viewpoint, and modality distribution. RegDB provides paired visible and infrared images and is widely used for evaluating cross-modality retrieval performance. Rank-1 accuracy and mean average precision (mAP) are adopted as evaluation metrics. Experimental results demonstrate that the proposed method achieves superior performance compared with existing state-of-the-art USVI-ReID methods. On the SYSU-MM01 dataset under the all-search protocol, the proposed method improves Rank-1 accuracy and mAP by 5.5 and 4.1 percentage points compared with the best existing method, respectively. On the RegDB dataset under the visible-to-thermal retrieval setting, the proposed method achieves 93.0% Rank-1 accuracy and 86.2% mAP. Ablation studies further demonstrate the effectiveness of hard positive modeling, hard negative modeling, and progressive training strategy. **Conclusion** The proposed method provides a new perspective for hard sample modeling in USVI-ReID by considering the different optimization roles of hard positive and hard negative samples. By constructing cross-modality shift-based hard positives and locally confusing hard negatives, the proposed framework effectively enhances cross-modality consistency and identity discrimination. Experimental results on two benchmark datasets verify the effectiveness of the proposed role-aware hard sample modeling strategy, providing a potential solution for robust unsupervised visible-infrared person re-identification.

Key words: computer vision; person re-identification; visible-infrared; unsupervised learning; hard sample learning; contrastive learning

论文引用格式:[DOI: 10. 11834/jig. 260448]

0 引言

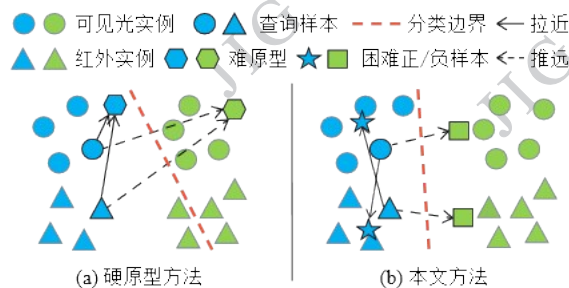
可见光-红外行人重识别(Visible-Infrared Person Re-Identification, VI-ReID)旨在跨可见光与红外模态匹配同一行人的图像。得益于其在夜间监控和全天候人员检索方面的潜在应用价值,该任务近年来受到了研究社区的广泛关注(Sun 等, 2024; Ye 等, 2020; Zheng 等, 2022)。尽管已有若干工作在VI-ReID任务中取得了令人鼓舞的成果(Cui 等, 2024; Wei and Yang, 2025; Zhang and Wang, 2023),但这些方法普遍依赖昂贵且耗时的跨模态人工标注数据,严重制约了其在真实场景中的部署与应用。针对这一瓶颈,无监督可见光-红外行人重识别(Unsupervised VI-ReID, USVI-ReID)逐渐成为研究热点(Liang 等, 2021; Yang 等, 2022),其核心目标是在无需人工身份标签的情况下学习模态不变的判别性特征。

早期研究主要通过引入辅助监督信息或迁移已有单模态知识,探索跨模态特征对齐方法。例如,Liang 等人(2021)提出从同构到异构的两阶段学习框架,通过利用可见光模态模型学习到的身份知识辅助跨模态特征对齐;Wang 等人(2022)进一步结合伪标签生成和最优传输策略,实现可见光与红外模态之间的标签迁移。然而,这类方法通常依赖额外标注数据或预训练模型,难以适用于真实无标注场景。

为降低对人工标注信息的依赖,近年来研究逐渐转向完全无监督可见光-红外行人重识别任务,通过挖掘目标数据自身的结构关系实现身份识别与跨模态特征学习。在该方向中,研究者主要围绕伪标签生成、跨模态对齐以及身份表示学习展开探索。Yang 等人(2022)提出增强双通道对比聚合学习方法,通过双路径对比学习和跨模态记忆聚合模块,实现不同模态之间的特征融合;Wu 等人(2023)进一步提出渐进式图匹配方法,通过构建跨模态图结构挖掘潜在身份对应关系,提升伪标签质量和跨模态身份关联能力。在此基础上,后续研究进一步关注身份信息的细粒度表达。Yang 等人(2024)提出浅层-深层协同学习框架,通过融合浅层细节特征与深层语义信息增强身份表示能力;Wang 等人(2025)提出

多令牌匹配框架,通过学习多个互补嵌入空间提升复杂场景下的细粒度身份匹配能力。

经过上述研究的发展,当前USVI-ReID方法通常基于聚类伪标签和对比学习框架进行特征优化。具体而言,这些方法首先利用聚类算法为每张训练图像分配伪身份标签,然后为每个身份构建模态特定的记忆库作为该身份的原型。在训练过程中,通过最小化查询图像特征与所属身份原型之间的对比损失,迫使模型学习到更具判别性的表示。然而,单个原型主要存储了行人的共性信息,难以刻画聚类内部的分布多样性,同时在类别边界区域缺乏细粒度约束。为弥补这一不足,部分研究引入硬原型对比学习(Pang 等, 2025; Shi 等, 2024; Zou 等, 2023),即将距离聚类中心最远的样本视为困难样本参与对比学习,以增强类内建模能力并强化类间分离。



((a) hard prototype mining method; (b) proposed method)

图1 硬原型方法与本文方法的对比示意图

Fig. 1 Comparison between hard prototype mining and the proposed method

然而,传统硬原型方法本质上采用一种距离驱动的困难性定义,即将样本相对于聚类中心的偏离程度直接作为其困难程度的衡量依据。该定义忽略了困难正样本与困难负样本在对比学习中具有不同的优化目标,导致所挖掘的困难样本可能与实际优化需求不一致。在正样本侧,距离聚类中心最远的样本(如图1(a)蓝色六边形)很可能源于错误聚类,或者受到极端姿态、严重遮挡和成像质量退化等非模态因素干扰。强制对齐此类样本会误导模型的优化方向,削弱跨模态表示的一致性。另一方面,在负样本侧,距离聚类中心最远的样本(如图1(a)绿色六边形)仅表示其偏离自身类别中心,并不一定靠近最具混淆性的邻近类别边界。通过拉远这类样本得到的分类边界往往难以反映真实的类间混淆关系,因此对提升类间判别性的作用有限。综上,仅依据

中心距离选取极端样本,实质上混淆了“特征偏离程度”与“困难样本对模型优化的实际价值”。

针对这一问题,本文提出一种角色感知的难样本建模框架。与传统方法使用统一距离准则衡量样本困难程度不同,本文根据困难正负样本在对比学习中的不同优化角色,分别界定其应当具备的结构属性,并据此进行针对性建模。其中,困难正样本应在尽可能保持身份语义一致的前提下刻画类内复杂变化,以提升特征的鲁棒性与泛化能力;困难负样本则聚焦于类间决策边界附近的难例区域,以增强模型对相似身份的判别能力。具体而言,在正样本侧,我们利用模态特定身份中心之间的跨模态偏移构造困难正样本。模态身份中心之间的偏移刻画了同一身份在不同模态下的整体变化趋势,因此沿该方向生成的样本能够在保持身份语义一致性的前提下引入有挑战性的跨模态变化,使模型学习到模态差异本身。在负样本侧,我们通过搜索样本的K近邻簇作为生成方向并构造困难负样本。K近邻簇刻画了样本在嵌入空间中的局部邻域关系,而邻域中彼此接近的样本往往更容易发生类别混淆。因此,沿该方向生成的负样本能够更接近类间分界区域,从而在训练过程中有效地强化类间区分能力。如图1(b)所示,本文方法能够有效实现跨模态身份对齐,并且优化不同身份之间的易混淆边界区域。此外,受到课程学习(Bengio 等,2009)思想的启发,我们引入一种渐进式训练策略,在训练过程中逐步增加难

样本的困难程度,从而在保持训练稳定性的同时提升特征表示的判别能力。

我们在 SYSU-MM01 (Wu 等,2017) 和 RegDB (Nguyen 等,2017) 两个基准数据集上对所提方法进行了系统评估。实验结果表明,该方法能够有效缓解现有困难样本挖掘策略中的偏差,并显著提升跨模态行人重识别的性能。

1 本文方法

1.1 方法概览

在无监督可见光-红外行人重识别任务中,给定无标签的可见光数据集 $D^v = \{\mathbf{x}_i^v\}_{i=1}^{N^v}$ 和红外数据集 $D^r = \{\mathbf{x}_i^r\}_{i=1}^{N^r}$, 式中 N^v 和 N^r 分别表示可见光和红外数据集的图片数目。我们的目标是训练一个深度神经网络 $f_\theta(\cdot)$ 将两种模态的图片映射到共享的特征空间中,以实现同一身份在可见光与红外图像之间的准确匹配。

如图2所示,本文提出一种基于角色感知的难样本建模框架(role-aware hard sample modeling, RAHSM)。该框架包含三个模块,分别是基于跨模态偏移的困难正样本构造模块(Cross-modality shift based Hard Positive sample, CHP)、基于K近邻簇的困难负样本构造模块(K-nearest neighbor cluster based Hard Negative sample, KHN)以及渐进式训练策略模块(Progressive Training Strategy, PTS)。

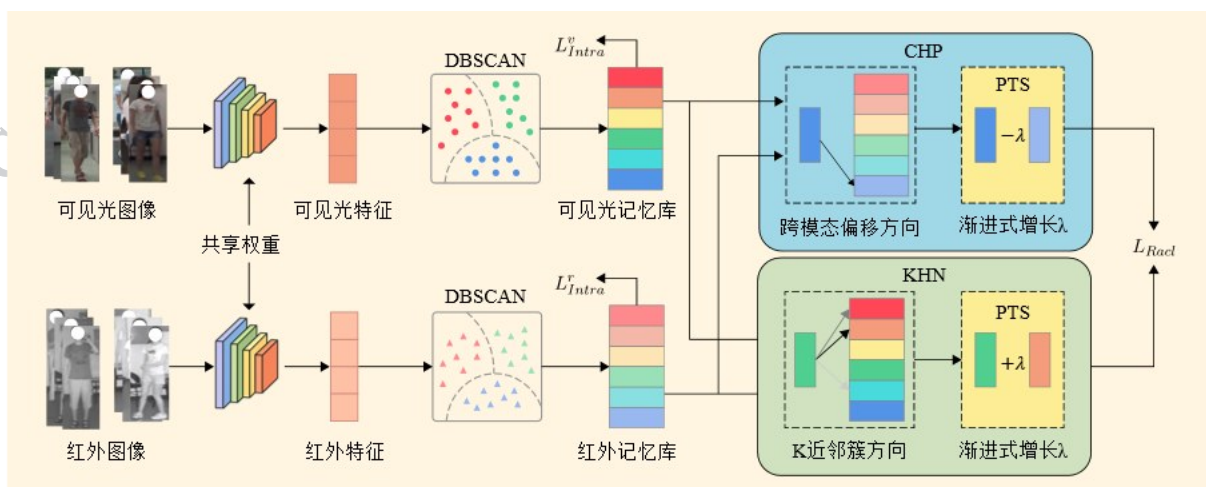


图2 角色感知难样本建模框架结构示意图

Fig. 2 Overview of the proposed role-aware hard sample modeling framework

在基础表征学习阶段,主干网络对输入图像提取特征,然后通过 DBSCAN (Ester 等, 1996) 聚类获得伪标签,并结合模态内对比学习获得初步判别能力。进入跨模态对齐阶段后,CHP 模块与 KHN 模块在嵌入空间中分别从不同角度构造困难样本:前者基于跨模态偏移刻画类内复杂变化的困难正样本,后者基于 K 近邻簇挖掘类间边界附近的困难负样本,从而提高模型类内紧凑性和类间判别性。同时,PTS 模块在训练过程中动态调节难样本构造的强度,以逐步提升样本难度,使模型在保持训练稳定性的同时增强特征表示能力。

1.2 基于跨模态偏移的困难正样本构造模块

在无监督跨模态对比学习中,困难正样本需要在保证身份一致的前提下引入更具有挑战性的跨模态差异,从而提升特征表示的跨模态对齐能力。对于每个跨模态共享身份 $y^s = k$,我们分别在可见光与红外模态下维护模态特定的身份中心,记为 p_k^{vs} 和 p_k^{rs} ,用于刻画同一身份不同模态下的共性特征。具体地,中心通过编码器特征的均值计算得到:

$$p_k^{vs} = \frac{1}{|C_k^{vs}|} \sum_{x_i^v \in C_k^{vs}} f_\theta(x_i^v) \# (1)$$

$$p_k^{rs} = \frac{1}{|C_k^{rs}|} \sum_{x_i^r \in C_k^{rs}} f_\theta(x_i^r) \# (2)$$

式中, C_k^{vs} 和 C_k^{rs} 分别表示共享身份 k 下的可见光和红外样本集合。

训练过程中,为了使 p_k^{vs} 和 p_k^{rs} 保持具备模态特定身份共性信息的性质,我们利用当前批次中属于同身份且来自对应模态的特征均值对其动量更新。

$$p_k^{vs} \leftarrow \mu p_k^{vs} + (1 - \mu) \frac{1}{|Q_k^v|} \sum_{x_i^v \in Q_k^v} f_\theta(x_i^v) \# (3)$$

$$p_k^{rs} \leftarrow \mu p_k^{rs} + (1 - \mu) \frac{1}{|Q_k^r|} \sum_{x_i^r \in Q_k^r} f_\theta(x_i^r) \# (4)$$

式中, μ 是用于控制更新速度的动量参数。 Q_k^v 表示本轮查询样本中身份 k 的可见光样本集合, $|Q_k^v|$ 为该集合的样本数。 Q_k^r 和 $|Q_k^r|$ 分别表示查询样本中身份 k 的红外模态下的对应集合和样本数。

最后,我们基于 p_k^{vs} , p_k^{rs} , 以及二者之间的跨模态偏移构造困难正样本,即沿着和另一个模态特定中心相反的方向进行外推:

$$\hat{p}_k^{rs} = p_k^{rs} - \lambda(p_k^{vs} - p_k^{rs}) \# (5)$$

$$\hat{p}_k^{vs} = p_k^{vs} - \lambda(p_k^{rs} - p_k^{vs}) \# (6)$$

式中, λ 控制偏移幅度。由于 p_k^{vs} 和 p_k^{rs} 分别为身份 k 在可见光与红外模态下的特征中心,二者之间的偏移主要反映该身份在不同模态下的整体变化趋势。进一步地,由式(5)和式(6)可得:

$$\frac{\hat{p}_k^{vs} + \hat{p}_k^{rs}}{2} = \frac{p_k^{vs} + p_k^{rs}}{2} \# (7)$$

$$\hat{p}_k^{vs} - \hat{p}_k^{rs} = (1 + 2\lambda)(p_k^{vs} - p_k^{rs}) \# (8)$$

上述关系表明,对称外推保持了两个模态特定中心的中点不变,同时沿原有跨模态偏移方向扩大了二者之间的差异。基于这种构造,生成样本仍具有模态特定的身份特征,但相较于原始模态中心具有更显著的模态差异。将其作为困难正样本进行对比优化,可以促使模型在更大的跨模态差异下保持对同一身份的正确匹配,从而提升跨模态对齐能力。

1.3 基于 K 近邻簇的困难负样本构造模块

在无监督跨模态对比学习中,困难负样本旨在刻画易混淆的类别关系,从而提升模型在类别边界的判别能力。嵌入空间中距离较近的不同身份通常具有较高的特征相似性,更容易发生错误分类,且二者相互邻接的区域也更可能形成类别混淆边界。因此,本文利用 K 近邻簇关系挖掘局部易混淆身份,并以此确定困难负样本的生成方向。

具体而言,我们基于跨模态共享伪标签 $y^s = k$,在每个身份内部对可见光与红外特征分别进行 K-means 聚类,得到该身份的子簇中心集合 $\mathcal{U}_k^s = \{u_{k,j}^s\}_{j=1}^{K_1}$, 式中 K_1 为子簇个数。对于任意子簇中心 $u_{k,j}^s$,在其他身份中搜索与其距离最近的 K_2 个近邻簇。以 $K_2=1$ 为例,记当前身份在对应模态下的中心为 p_k ,距离 $u_{k,j}^s$ 最近的其他身份簇中心为 p_j 。差向量 $(p_j - p_k)$ 表示由当前身份中心指向最易混淆的近邻身份中心。沿该方向移动子簇 $u_{k,j}^s$,可以减小其与近邻身份之间的特征差异,使生成样本向两个身份之间的局部类别边界靠近。困难负样本定义为:

$$\hat{u}_{k,j}^s = u_{k,j}^s + \lambda(p_j - p_k) \# (9)$$

式中, λ 控制外推幅度。基于这种构造,生成样本以当前身份的局部特征为基础,同时缩小了与易混淆身份之间的特征差异,因而更接近两个身份之间的局部类别边界。将其作为困难负样本进行对比优化,可以为模型提供更具判别性的训练信号,从而提升相似身份之间的区分能力。

1.4 渐进式训练策略

在难样本构造过程中,偏移幅度 λ 决定了生成样本的困难程度及其包含的上下文信息。受到课程学习思想的启发,我们认为难样本的困难程度应随训练过程逐步递进,而非在训练初期直接设定为高水平。基于这一动机,本文在难样本构造中引入渐进式训练策略,通过动态调节偏移幅度 λ 使生成样本的困难程度随训练过程自适应变化:

$$\lambda = \lambda_0 \cdot \log\left(\frac{e-1}{T} \cdot t + 1\right) \quad (10)$$

式中, t 表示当前迭代次数, T 表示总迭代次数, λ_0 是基础幅度参数。该策略使模型在训练初期侧重于学习结构稳定的低难度样本,以获得可靠的特征表示;随着训练的推进,逐步增大偏移幅度,使构造样本由低难度向高难度过渡,并引入更具挑战性的类内跨模态变化与类间易混淆样本。在这一由易到难的优化过程中,模型能够在保持训练稳定性的同时,有效提升特征表示的泛化能力和判别能力。

1.5 损失函数

在获得困难正样本与困难负样本的候选集合后,我们采用基于InfoNCE(Oord等,2018)形式构建角色感知的对比损失函数。具体而言,对于任意查询样本 $\mathbf{x}_q$,选取同身份的两个困难正样本中与其距离较远者进行对齐,选取其它身份的最近的困难负样本进行对比优化:

$$hm_i = \begin{cases} \arg \max_{\forall p \in \{\hat{p}_i^+, \hat{p}_i^-\}} \|f_\theta(\mathbf{x}_q) - p\|_2, & \text{if } y_q = k \\ \arg \min_{\forall p \in \{\hat{u}_i\}} \|f_\theta(\mathbf{x}_q) - p\|_2, & \text{if } y_q \neq k \end{cases} \quad (11)$$

式中, y_q 是 $\mathbf{x}_q$ 的共享身份标签。在此基础上:

$$L_{\text{Raci}} = -\log \frac{\exp\left(f_\theta(\mathbf{x}_q) \cdot \frac{hm_+}{\tau}\right)}{\sum_{i=1}^{N^+} \exp\left(f_\theta(\mathbf{x}_q) \cdot \frac{hm_i}{\tau}\right)} \quad (12)$$

式中, N^+ 是共享标签空间中的身份总数, hm_+ 表示选取的困难正样本, τ 则是温度超参数。在理想情况下,查询特征与同模态生成中心的距离较近,而与跨模态生成中心的距离较远。因此,式(11)的距离最大准则通常会从两个候选中选出跨模态生成中心作为困难正样本进行对齐。与传统方法从身份簇内的所有实例中直接选取最远样本不同,本文构造的困难正、负样本分别以模态特定身份中心和子簇中心为基础,而这些中心均由多个样本特征聚合得到,并

非直接采用某个单一实例。尽管异常样本仍可能干扰中心估计,但单个离群样本不会直接作为困难样本参与优化,其对模型训练的影响也相应减弱。

2 实验

2.1 数据集和评估方法

本文在两个广泛使用的可见光-红外数据集SYSU-MM01(Wu等,2017)和RegDB(Nguyen等,2017)上对所提方法进行测试。SYSU-MM01数据集包含由4个可见光摄像头和2个红外摄像头采集的491个身份,其中训练集由来自395个身份的22258张可见光图像和11909张红外图像组成。RegDB数据集包含由1个可见光摄像机和1个红外相机采集的412个身份,每个身份有10张可见光图片和10张红外图片。

本文使用的评价指标为可见光-红外行人重识别任务的标准评估指标(Sun等,2022),即累计匹配特性(Cumulative Matching Characteristics, CMC)、平均精度均值(mean Average Precision, mAP)以及平均逆负惩罚指标(mean Inverse Negative Penalty, mINP)进行评估。在SYSU-MM01数据集上,我们给出全搜索(All Search)和室内搜索(Indoor Search)两种模式的实验结果。RegDB则包含可见光到热成像(Visible to thermal)和热成像到可见光(Thermal to Visible)两种测试模式,并将训练集和测试集进行10次随机分割,报告平均结果。

2.2 实验环境及网络参数设置

本文方法使用PyTorch在两张NVIDIA GeForce RTX 3090显卡上实现。我们采用TransReID(He等,2021)的特征提取器作为骨干网络,并结合增强双对比学习策略(Yang等,2022)。在数据预处理阶段,输入图像被调整为 288×144 的分辨率,并且使用通道增强(Ye等,2021)、随机裁剪、水平翻转以及随机擦除进行数据增强。每个训练批次对可见光和红外模态分别抽取8个身份,每个身份包含16张图片。使用SGD优化器,初始学习率为 3.5×10^{-4} ,每隔20个epoch衰减为原来的1/10。模型总共训练50个周期,其中前30个周期为单模态训练,本文方法在后20个周期引入。在共享伪标签生成方面,本文参考SDCL(Yang等,2024)的跨模态排序关联策略,并采用仅基于深层特征的简化实现。具体而言,根据

表1 与主流方法在 SYSU-MM01 数据集上的性能比较

Table 1 Performance comparison with state-of-the-art methods on the SYSU-MM01 dataset

方法	全搜索模式			室内搜索模式		
	Rank-1	mAP	mINP	Rank-1	mAP	mINP
AGW [†] (Ye 等, 2022)	47.5	47.7	35.3	54.2	63.0	59.2
CAJ [†] (Ye 等, 2021)	69.9	66.9	54.6	76.3	80.4	76.8
FMCNet [†] (Zhang 等, 2022)	66.3	62.5	—	68.2	74.1	—
DEEN [†] (Zhang and Wang, 2023)	74.7	71.8	—	80.3	83.3	—
PartMix [†] (Kim 等, 2023)	77.8	74.6	—	81.5	84.4	—
RLE [†] (Zhang 等, 2024)	75.4	72.4	60.9	84.7	87.0	83.7
CSDN [†] (Yu 等, 2025)	76.7	73.0	—	84.5	86.8	—
H2H(Liang 等, 2021)	30.2	29.4	—	—	—	—
OTLA(Wang 等, 2022)	29.9	27.1	—	29.8	38.8	—
ADCA(Yang 等, 2022)	45.5	42.7	28.3	50.6	59.1	55.2
PGM(Wu and Ye, 2023)	57.3	51.8	34.5	56.2	62.7	58.1
GUR(Yang 等, 2023)	61.0	57.0	41.9	64.2	69.5	64.8
MMM(Shi 等, 2025)	61.6	57.9	—	64.4	70.4	—
PCLHD(Shi 等, 2024)	64.4	58.7	—	69.5	74.4	—
RPNR(Yin 等, 2024)	65.2	60.0	—	68.9	74.4	—
SDCL(Yang 等, 2024)	64.5	63.2	51.1	71.4	76.9	73.5
ASM(Pang 等, 2025)	65.1	63.4	51.3	71.1	76.9	73.7
RAHSM(Ours)	70.6	67.5	54.1	76.0	80.4	77.0

注: “—”表示对应论文中没有该结果, “†”表示为有监督方法; 加粗字体为每列无监督方法中的最优值。

图像特征的跨模态排序结果建立可见光簇与红外簇之间的身份对应关系, 并据此生成共享伪标签。进入跨模态训练阶段后, 每个训练周期开始时均重新生成伪标签, 因此共享身份对应关系会随网络特征的更新而动态调整。基础幅度参数 λ_0 设为 0.3, 和困难负样本构建相关参数 K_1 和 K_2 设置为 16 和 1, 温度超参数 τ 设置为 0.05。

2.3 与现有方法比较

为验证本文方法的有效性, 将其与 7 个监督可见光-红外行人重识别(SVI-ReID)及 10 个无监督可见光-红外行人重识别(USVI-ReID)方法进行对比, 结果如表 1 和表 2 所示。

与当前先进的 USVI-ReID 方法相比, 本文方法在核心指标上均取得领先。例如在 SYSU-MM01 的 All-Search 模式下, 本文比先进的同类方法 ASM 在 Rank-1 和 mAP 指标上分别提高了 5.5% 和 4.1%。在其它的检索设置下, 本文方法在 Rank-1 准确率与

mAP 上均达到最优或极具竞争力的水平。尽管表 1 中提及的现有 USVI-ReID 方法在该任务上已取得显著进展, 但忽视了难正负样本在训练中的不同作用, 因而难以同时兼顾跨模态对齐与类别判别。相比之下, 本文方法对难正负样本进行角色感知的差异化建模, 使模型能够获得更加有效的优化信号, 从而学习到更加鲁棒且具有判别性的特征表示。

在与 SVI-ReID 方法的对比中, 本文方法在无需人工标注的情况下已超过部分有监督方法, 表现出优异的竞争力。这说明在充分挖掘数据内在结构信息的条件下, 无监督方法同样具备逼近有监督性能的潜力。尽管如此, 由于缺乏精确的跨模态身份标注, 本文方法与当前最先进的有监督方法仍存在一定性能差距, 还有进一步提升空间。

2.4 消融实验

在本节中, 我们在 SYSU-MM01 数据集的全搜索模式和室内搜索模式下进行消融实验, 以展示我

们方法中各个组件的有效性。结果如表3所示。

表2 与主流方法在 RegDB 数据集上的性能比较
Table 2 Performance comparison with state-of-the-art methods on the RegDB dataset

方法	可见光-热成像			热成像-可见光		
	Rank-1	mAP	mINP	Rank-1	mAP	mINP
AGW [†] (Ye 等, 2022)	70.1	66.4	50.2	70.5	65.9	51.2
CAJ [†] (Ye 等, 2021)	85.0	79.1	65.3	84.8	77.8	61.6
FMCNet [†] (Zhang 等, 2022)	89.1	84.4	–	88.4	83.9	–
DEEN [†] (Zhang and Wang, 2023)	91.1	85.1	–	89.5	83.4	–
PartMix [†] (Kim 等, 2023)	85.7	82.3	–	84.9	82.5	–
RLE [†] (Zhang 等, 2024)	92.8	88.6	–	91.0	86.6	–
CSDN [†] (Yu 等, 2025)	95.4	87.7	–	92.3	85.5	–
H2H(Liang 等, 2021)	23.8	18.9	–	–	–	–
OTLA(Wang 等, 2022)	32.9	29.7	–	32.1	28.6	–
ADCA(Yang 等, 2022)	67.2	64.1	52.7	68.5	63.8	49.6
PGM(Wu and Ye, 2023)	69.5	65.4	–	69.9	65.2	–
GUR(Yang 等, 2023)	73.9	70.2	58.9	75.0	69.9	56.2
MMM(Shi 等, 2025)	89.7	80.5	–	85.8	77.0	–
PCLHD(Shi 等, 2024)	84.3	80.7	–	82.7	78.4	–
RPNR(Yin 等, 2024)	90.9	84.7	–	90.1	83.2	–
SDCL(Yang 等, 2024)	86.9	78.9	62.8	85.8	77.3	59.6
ASM(Pang 等, 2025)	88.2	79.7	63.1	86.9	79.2	60.0
RAHSM(Ours)	93.0	86.2	73.0	91.2	83.8	68.6

注：“–”表示对应论文中没有该结果，“†”表示为有监督方法；加粗字体为每列无监督方法中的最优值。

表3 不同模块组合对模型性能的影响
Table 3 Influence of different module combinations on model performance

方法配置				全搜索		室内搜索	
Base	CHP	KHN	PTS	Rank-1	mAP	Rank-1	mAP
└				59.52	58.61	66.71	73.05
└	└			68.47	65.53	73.51	78.28
└		└		65.50	64.76	72.27	77.65
└	└	└		69.30	66.68	74.37	79.27
└	└	└	└	70.63	67.46	76.04	80.35

在基线模型中,模型仅依赖基础对比学习进行优化,性能相对有限。引入基于跨模态偏移的困难正样本构造模块 CHP 后,Rank-1 明显提升,说明基于跨模态偏移生成的样本能够有效刻画类内跨模态变化,从而增强跨模态对齐能力。在此基础上引入

基于K近邻簇的困难负样本构造模块 KHN,mAP 指标显著提升。这说明基于K近邻簇构造的困难负样本能够有效刻画类间易混淆区域,从而提升模型的类间判别能力。最后,通过引入渐进式训练策略 PTS,模型性能进一步提升。该策略通过逐步增加

难样本的困难程度,在保证训练稳定性的同时增强模型对复杂分布的适应能力。综上,各模块分别从类内对齐、类间判别及训练稳定性三个方面协同优化,共同提升模型整体性能。

2.5 超参数分析

为分析关键超参数对模型性能的影响,本文在 SYSU-MM01 数据集全搜索模式下进行了超参数敏感性实验。具体而言,分别研究困难样本偏移幅度参数 λ_0 、子簇数量 K_1 以及近邻簇搜索数量 K_2 对模型性能的影响,实验结果如图3所示。

1) 偏移幅度参数 λ_0 : 该参数用于控制困难样本生成过程中的偏移强度。如图3(a)所示,随着 λ_0 增大,模型性能先提升后下降,并在 $\lambda_0 = 0.3$ 时取得最佳结果(Rank-1 为 70.63%, mAP 为 67.46%)。较小的 λ_0 无法充分挖掘困难样本信息,而过大的偏移会引入额外噪声,导致生成样本过分偏离真实分布。因此,本文选择 $\lambda_0 = 0.3$ 。

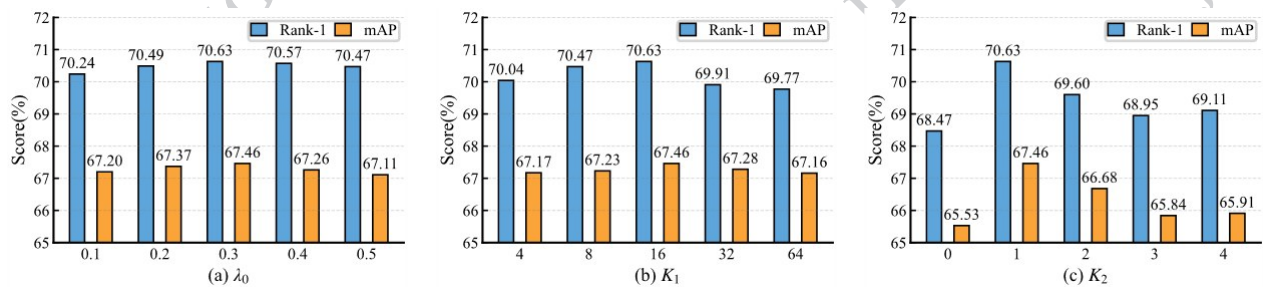
2) 子簇数量 K_1 : 该参数用于控制身份内部子簇划分粒度。如图3(b)所示,模型性能随着 K_1 增加呈现先升高后降低的趋势,并在 $K_1 = 16$ 时达到最佳性能。当 K_1 较小时,子簇数量不足,难以充分描述身份内部存在的多样性特征变化,导致特征分布建模不充分。随着 K_1 进一步增大,子簇划分过细,部分子簇包含的样本数量过少,容易受到噪声样本的波动影响,从而降低特征分布描述的稳定性。因此,本文设置 $K_1 = 16$ 。

3) 近邻簇搜索数量 K_2 : 该参数用于控制困难负样本生成时选择的跨模态近邻身份簇数量。如图3(c)所示,模型在 $K_2 = 1$ 时达到最佳性能。并且随着 K_2 增加,模型性能逐渐下降。我们推测这可能是由于行人身份特征在空间中的分布具有一定的流形结

构,在沿着簇间方向插值时,需要保证生成的特征仍然在合理的局部邻域内。当 K_2 较大时,参与生成的近邻身份簇距离当前身份簇更远,生成样本可能偏离原始特征流形,甚至无法保持有效的身份语义信息,从而引入噪声并降低模型性能。此外,由于可见光和红外模态之间存在较大的分布差异,跨模态簇间插值的不确定性会进一步增加。因此,本文选择 $K_2 = 1$ 作为默认设置。

2.6 偏移幅度的更新策略对比

为分析偏移幅度 λ 的调度方式对模型性能的影响,本文在 SYSU-MM01 数据集上对多种更新策略进行了对比,其中包括常数型 $\lambda = \lambda_0$ 、线性增长型 $\lambda = \frac{\lambda_0 \cdot t}{T}$ 、平方增长型 $\lambda = \frac{\lambda_0 \cdot t^2}{T^2}$ 以及本文采用的对数增长策略(见式(10))。式中, t 表示当前训练迭代, T 为总迭代次数,所有设置中统一取 $\lambda_0 = 0.3$ 。实验结果如表4所示。可以观察到,常数型策略性能较差,其在训练早期即引入固定强度的难样本扰动,容易影响特征空间的稳定收敛;平方增长策略在训练初期增长过于缓慢,难样本强度长期处于较低水平,难以提供有效的判别信号。线性增长策略在难度与稳定性之间取得一定平衡,因此表现较为稳定。相比之下,对数增长策略取得最佳性能。我们推测这主要得益于其“先快后缓”的增长特性:在训练初期能够较快提升难样本强度,使模型尽早接触具有判别性的样本;而在训练后期增长逐渐趋缓,有效避免偏移幅度增速过大导致样本越过类别边界。该特性使构造样本始终分布在可学习的判别区域附近,从而在训练稳定性与难度控制之间取得更优平衡。需要说明的是,对数增长策略仅作为满足渐进式难度调整需求的一种有效方式,并不意味着其对



((a) offset magnitude λ_0 ; (b) number of sub-clusters K_1 ; (c) number of neighboring clusters K_2)

图3 不同超参数设置对模型性能的影响

Fig. 3 Influence of different hyperparameter settings on model performance

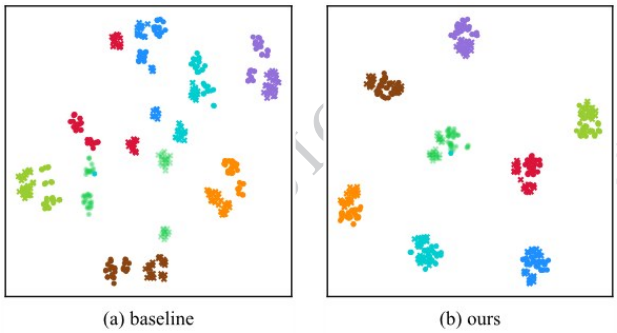
所有数据分布均具有普适优势。由于不同数据集存在模态差异、样本分布和伪标签质量差异,最优调度方式可能有所不同。因此,本文重点验证渐进式难度调节机制的有效性。

表4 不同偏移幅度更新策略的性能比较
Table 4 Performance comparison of different offset magnitude updating strategies

λ 更新类型	全搜索		室内搜索	
	Rank-1	mAP	Rank-1	mAP
常数型	69.30	66.68	74.37	79.27
平方型	69.91	66.52	73.73	78.23
线性型	70.27	67.13	75.42	79.81
对数型	70.63	67.46	76.04	80.35

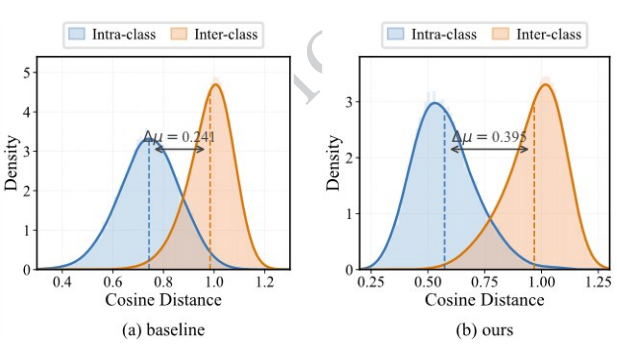
2.7 可视化分析

为进一步验证本文方法对跨模态特征学习的有效性,本文分别从t-SNE特征分布、跨模态样本对距离分布以及跨模态检索性能三个角度进行可视化分析。



((a) baseline; (b) ours)
图4 t-SNE特征分布可视化
Fig. 4 Visualization of t-SNE feature distributions

1)t-SNE可视化:为观察不同方法的特征分布情况,本文在SYSU-MM01数据集上通过t-SNE方法进行可视化分析,如图4所示。其中,不同颜色表示不同身份类别,不同形状表示不同模态。可以看出,基线方法学习到的特征仍存在明显的模态偏移,同一身份的跨模态样本分布较为分散,且部分身份之间距离较近。相比之下,本文方法能够使同一身份的跨模态样本更加聚集,同时扩大不同身份之间的距离,形成更加清晰的类别边界。这表明本文方法能够有效缓解模态差异并提升特征判别能力。



((a) baseline; (b) ours)
图5 正负跨模态匹配对的余弦距离分布可视化
Fig. 5 Visualization of cosine distance distributions for positive and negative cross-modal matching pairs

2)跨模态对距离分布可视化:为了进一步分析不同方法学习到的跨模态特征分布,本文统计了同类与异类跨模态样本对之间的余弦距离,并对其分布进行可视化,如图5所示。可以观察到,基线方法中Intra-class与Inter-class分布存在较明显的重叠,二者均值间隔 $\Delta\mu$ 仅为0.241,表明特征空间中的类别边界仍不够清晰。相比之下,本文方法有效拉近了同身份跨模态样本之间的距离,使Intra-class分布向低距离区域移动,其均值由0.743降低至0.574;同时,Inter-class与Intra-class之间的距离间隔扩大至0.395,使正负样本分布具有更明显的分离趋势。上述结果表明,本文方法能够优化跨模态特征空间结构,在增强类内紧凑性的同时提升类间判别能力。

3)跨模态检索结果分析:如图6所示,本文在SYSU-MM01数据集上对跨模态检索结果进行了可视化分析。图中展示了不同查询样本的Top-5检索结果,绿色/红色边框分别表示正确与错误匹配,各检索图上方标注余弦相似度。与基线模型相比,本文方法在光照差异、姿态变化和背景环境变化等典型场景下,Top-5内正确匹配数量更多且排序更为靠前,同时正确匹配的相似度整体更高,有着更强的身份匹配能力。

上述可视化结果表明,本文提出的方法能够同时改善特征空间中的类内紧凑性和类间分离性,并有效提升跨模态身份匹配的可靠性,进一步验证了角色感知难样本建模策略在无监督可见光-红外行人重识别任务中的有效性。



((a) query images; (b) baseline results; (c) results of the proposed method)

图6 SYSU-MM01数据集上的跨模态检索结果可视化

Fig. 6 Visualization of cross-modality retrieval results on the SYSU-MM01 dataset

3 结论

本文针对无监督可见光-红外行人重识别中困难样本挖掘策略与对比学习优化目标不一致的问题,提出了一种角色感知的难样本建模方法。通过在共享嵌入空间中区分困难正样本与困难负样本的优化作用,本文重新刻画了困难样本在类内对齐与类间分离中的功能,提升了特征学习的有效性。此外,本文还引入了渐进式训练策略,使模型在训练过程中由易到难逐步适应困难样本,在提升特征判别能力的同时保证优化稳定性。实验结果表明,该方法在SYSU-MM01和RegDB两个基准数据集上均取得了具有竞争力的性能,验证了角色感知的难样本建模策略在跨模态特征对齐与判别学习中的优势。

需要指出的是,本文方法依赖共享伪标签建立跨模态身份关联,因此其有效性会受到伪标签质量的影响。当不同真实身份被错误合并至同一伪标签簇时,模态特定身份中心会混入其他身份的特征,使中心间偏移难以准确反映同一身份的跨模态变化,进而影响困难正样本的身份一致性。同时,错误混入的样本还会干扰身份内部的子簇划分,使部分子簇中心包含其他身份的信息,进而影响近邻簇搜索

及困难负样本的生成方向。另一方面,当同一真实身份被错误拆分为多个伪标签簇时,不同样本子集将被视为不同身份,导致本应属于同一身份的特征参与困难负样本构造。因此,本文方法更适用于初始特征具有一定判别能力、伪标签质量较为良好的场景。

上述局限也反映了无监督可见光-红外行人重识别与有监督方法在身份不变性约束来源上的关键差异:有监督方法能够利用真实身份标签,明确约束同一身份在不同外观条件下保持一致,而无监督方法则需要依据聚类结果建立身份关联。在缺少真实身份先验的情况下,聚类结果容易受到光照、背景和服饰等身份无关因素的影响,并产生错误的身份关联。例如,同一身份可能因是否携带背包而被划分至不同聚类,使模型在后续优化中错误地将背包作为身份判别依据。因此,未来研究需要进一步探索如何减少伪标签噪声对模型优化的影响,并提升模型挖掘身份相关判别信息的能力,以缩小无监督方法与有监督方法之间的性能差距。

参考文献(References)

- Bengio Y, Louradour J, Collobert R and Weston J. 2009. Curriculum
© 中国图象图形学报版权所有

- learning//Proceedings of the 26th Annual International Conference on Machine Learning. Montreal Quebec Canada; ACM: 41-48 [DOI: 10.1145/1553374.1553380].
- Cui Z, Zhou J and Peng Y. 2024. DMA: Dual Modality-Aware Alignment for Visible-Infrared Person Re-Identification. *IEEE Transactions on Information Forensics and Security*, 19: 2696-2708 [DOI: 10.1109/TIFS.2024.3352408].
- Ester M, Kriegel H P, Sander J and Xu X. 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise//Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining. Portland, Oregon, USA: AAAI Press: 226-231.
- He S, Luo H, Wang P, Wang F, Li H and Jiang W. 2021. TransReID: Transformer-Based Object Re-Identification//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 15013-15022 [DOI: 10.1109/ICCV48922.2021.01474].
- Kim M, Kim S, Park J, Park S and Sohn K. 2023. PartMix: Regularization Strategy to Learn Part Discovery for Visible-Infrared Person Re-Identification//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, BC, Canada: IEEE: 18621-18632 [DOI: 10.1109/CVPR52729.2023.01786].
- Liang W, Wang G, Lai J and Xie X. 2021. Homogeneous-to-Heterogeneous: Unsupervised Learning for RGB-Infrared Person Re-Identification. *IEEE Transactions on Image Processing*, 30: 6392-6407 [DOI: 10.1109/TIP.2021.3092578].
- Nguyen D, Hong H, Kim K and Park K. 2017. Person Recognition System Based on a Combination of Body Images from Visible Light and Thermal Cameras. *Sensors*, 17 (3) : 605 [DOI: 10.3390/s17030605].
- Oord A V D, Li Y and Vinyals O. 2018. Representation learning with contrastive predictive coding[EB/OL]. [2026-07-22]. <https://arxiv.org/pdf/1807.03748.pdf>.
- Pang Z, Wang C, Zhao L and Wang J. 2025. Augmented and Softened Matching for Unsupervised Visible-Infrared Person Re-Identification//Proceedings of the IEEE/CVF International Conference on Computer Vision. Honolulu, USA: IEEE: 11100-11109 [DOI: 10.1109/ICCV51701.2025.01033].
- Shi J, Yin X, Chen Y, Zhang Y, Zhang Z, Xie Y, et al. 2025. Multi-memory Matching for Unsupervised Visible-Infrared Person Re-identification//Proceedings of the European Conference on Computer Vision. Cham, Switzerland: Springer Nature Switzerland: 456-474 [DOI: 10.1007/978-3-031-72649-1_26].
- Shi J, Yin X, Zhang Y, Zhang Z, Xie Y and Qu Y. 2024. Learning Commonality, Divergence and Variety for Unsupervised Visible-Infrared Person Re-Identification//Advances in Neural Information Processing Systems 37. Vancouver, Canada: Neural Information Processing Systems Foundation, Inc.: 99715-99734 [DOI: 10.52202/079017-3163].
- Sun H, Liu J, Zhang Z, Wang C, Qu Y, Xie Y, et al. 2022. Not All Pixels Are Matched: Dense Contrastive Learning for Cross-Modality Person Re-Identification//Proceedings of the 30th ACM International Conference on Multimedia. Lisboa, Portugal: ACM: 5333-5341 [DOI: 10.1145/3503161.3547970].
- Sun Y B, Wang R, Zhang Q and Lin R H. 2024. A cross-modality person re-identification method for visible-infrared images. *Journal of Beijing University of Aeronautics and Astronautics*, 50(6) : 2018-2025 (孙义博, 王蓉, 张琪, 林榕辉. 面向可见光-红外图像的跨模态行人再识别方法. 北京航空航天大学学报, 50(6) : 2018-2025 [DOI: 10.13700/j.bh.1001-5965.2022.0554]).
- Wang J, Zhang Z, Chen M, Zhang Y, Wang C, Sheng B, et al. 2022. Optimal Transport for Label-Efficient Visible-Infrared Person Re-Identification//Proceedings of the European Conference on Computer Vision. Cham, Switzerland: Springer Nature Switzerland: 93-109 [DOI: 10.1007/978-3-031-20053-3_6].
- Wang X, Liu L, Yang B, Ye M, Wang Z and Xu X. 2025. Token-Matcher: Diverse Tokens Matching for Unsupervised Visible-Infrared Person Re-Identification//Proceedings of the AAAI Conference on Artificial Intelligence. Philadelphia, USA: AAAI Press: 7934-7942 [DOI: 10.1609/aaai.v39i8.32855].
- Wei S and Yang W. 2025. Visible-infrared person re-identification algorithm integrating structural and visual features. *Journal of Image and Graphics*, 30(10) : 3335-3345 (魏思, 杨文璐. 2025. 融合结构与视觉特征的可见光-红外行人重识别. 中国图象图形学报, 30(10) : 3335-3345) [DOI: 10.11834/jig.240600].
- Wu A, Zheng W S, Yu H X, Gong S and Lai J. 2017. RGB-Infrared Cross-Modality Person Re-identification//IEEE/CVF International Conference on Computer Vision. Venice: IEEE: 5390-5399 [DOI: 10.1109/ICCV.2017.575].
- Wu Z and Ye M. 2023. Unsupervised Visible-Infrared Person Re-Identification via Progressive Graph Matching and Alternate Learning//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, BC, Canada: IEEE: 9548-9558 [DOI: 10.1109/CVPR52729.2023.00921].
- Yang B, Chen J and Ye M. 2023. Towards Grand Unified Representation Learning for Unsupervised Visible-Infrared Person Re-Identification//IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 11035-11045 [DOI: 10.1109/ICCV51070.2023.01016].
- Yang B, Chen J and Ye M. 2024. Shallow-Deep Collaborative Learning for Unsupervised Visible-Infrared Person Re-Identification//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE: 16870-16879 [DOI: 10.1109/CVPR52733.2024.01596].
- Yang B, Ye M, Chen J and Wu Z. 2022. Augmented Dual-Contrastive Aggregation Learning for Unsupervised Visible-Infrared Person Re-Identification//Proceedings of the 30th ACM International Conference on Multimedia. Lisboa, Portugal: ACM: 2843-2851 [DOI: 10.1145/3503161.3547970].

- 10.1145/3503161.3548198].
- Ye M, Lan X, Leng Q and Shen J. 2020. Cross-Modality Person Re-Identification via Modality-Aware Collaborative Ensemble Learning. *IEEE Transactions on Image Processing*, 29: 9387-9399 [DOI: 10.1109/TIP.2020.2998275].
- Ye M, Ruan W, Du B and Shou M Z. 2021. Channel Augmented Joint Learning for Visible-Infrared Recognition//*IEEE/CVF International Conference on Computer Vision*. Montreal, QC, Canada: IEEE: 13547-13556 [DOI: 10.1109/ICCV48922.2021.01331].
- Ye M, Shen J, Lin G, Xiang T, Shao L and Hoi S C H. 2022. Deep Learning for Person Re-Identification: A Survey and Outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44 (6): 2872-2893 [DOI: 10.1109/TPAMI.2021.3054775].
- Yin X, Shi J, Zhang Y, Lu Y, Zhang Z, Xie Y, et al. 2024. Robust Pseudo-label Learning with Neighbor Relation for Unsupervised Visible-Infrared Person Re-Identification//*Proceedings of the 32nd ACM International Conference on Multimedia*. Melbourne, Australia: ACM: 2242-2251 [DOI: 10.1145/3664647.3680951].
- Yu X, Dong N, Zhu L, Peng H and Tao D. 2025. CLIP-Driven Semantic Discovery Network for Visible-Infrared Person Re-Identification. *IEEE Transactions on Multimedia*, 27: 4137-4150 [DOI: 10.1109/TMM.2025.3535353].
- Zhang Q, Lai C, Liu J, Huang N and Han J. 2022. FMCNet: Feature-Level Modality Compensation for Visible-Infrared Person Re-Identification//*IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, LA, USA: IEEE: 7339-7348 [DOI: 10.1109/CVPR52688.2022.00720].
- Zhang Y and Wang H. 2023. Diverse Embedding Expansion Network and Low-Light Cross-Modality Benchmark for Visible-Infrared Person Re-identification//*IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, BC, Canada: IEEE: 2153-2162 [DOI: 10.1109/CVPR52729.2023.00214].
- Zhang Yukang, Han K, Dai P, Zhang Yan, Wu Y and Ji R. 2024. RLE: A Unified Perspective of Data Augmentation for Cross-Spectral Re-identification//*Advances in Neural Information Processing Systems 37*. Vancouver, Canada: Neural Information Processing Systems Foundation, Inc.: 126977-126996 [DOI: 10.52202/079017-4032].
- Zheng H, Zhong X, Huang W, Jiang K, Liu W and Wang Z. 2022. Visible-Infrared Person Re-Identification: A Comprehensive Survey and a New Setting. *Electronics*, 11 (3): 454 [DOI: 10.3390/electronics11030454].
- Zou C, Chen Z, Cui Z, Liu Y and Zhang C. 2023. Discrepant and Multi-instance Proxies for Unsupervised Person Re-identification//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Paris, France: IEEE: 11024-11034 [DOI: 10.1109/ICCV51070.2023.01015].

作者简介

李骑宏,男,硕士研究生,主要研究方向为行人重识别。E-mail: liqihong24@mailsucas.ac.cn

马丙鹏,通信作者,男,教授,主要研究方向为计算机视觉和模式识别。E-mail: bpma@ucas.ac.cn