

DOI: 10.20189/j.cnki.CN/61-1527/E.202506009

双重协同细化的无监督跨模态 行人重识别方法

张 聪, 王 涛*, 王 念, 姜 柯, 苏延召, 夏 菁, 李爱华
(火箭军工程大学, 西安 710025)

摘 要: 针对跨模态行人重识别过程中存在的噪声伪标签和模态间特征分布差异问题, 以双分支对比学习网络为框架, 提出一种基于双重协同细化的无监督跨模态行人重识别方法。首先, 以聚类簇内可靠实例的聚类中心为锚点, 设计了伪标签细化模块, 通过计算实例与锚点间的相似度来对伪标签进行优化, 从而提高伪标签的质量; 其次, 利用图匹配初步确定可见光与红外模态间的特征对应关系, 设计了跨模态双向对齐模块, 通过构建模态无关混合记忆字典来促使模型学习跨模态间的共性特征; 最后, 在伪标签和跨模态匹配双重细化的基础上, 设计了模态相关和模态无关的对比学习模块, 增强了模型学习判别性特征的能力。与次优方法多记忆匹配 (multi-memory matching, MMM) 相比, 所提方法在 SYSU-MM01 数据集的 All-search 和 Indoor-search 模式下平均均值精度 (mean average precision, mAP) 分别提高 1.4% 和 2.0%, 在 RegDB 数据集的 Visible2Infrared 和 Infrared2Visible 模式下 mAP 分别提高 1.8% 和 4.7%, 表明该方法具备良好的鲁棒性及跨模态检索能力。

关键词: 行人重识别; 无监督学习; 跨模态; 对比学习

中图分类号: TP391

文献标志码: A

开放科学标识码(OSID):

文章编号: 2097-2741(2025)06-0091-13



听语音
与作者互动
聊科研

Dual Collaborative Refinement-Based Unsupervised Cross-Modality Person Re-Identification

ZHANG Cong, WANG Tao*, WANG Nian, JIANG Ke,
SU Yanzhao, XIA Jing, LI Aihua
(Rocket Force University of Engineering, Xi'an 710025, China)

ABSTRACT: An unsupervised cross-modality person re-identification method based on dual collaborative refinement was proposed using a dual contrastive learning network framework. It was aimed to address the issues of noise pseudo labels and cross-modality feature distribution gaps in unsupervised person re-identification. First, a pseudo-label refinement module was designed, with the cluster centers of reliable instances serving as anchors. The module optimized the quality

收稿日期: 2025-07-03

修回日期: 2025-08-25

录用日期: 2025-09-16

第一作者: 张聪 (1987—), 男, 工程师, 博士研究生, 主要从事智能视频监控研究。E-mail: zhangc1022@126.com

***通信作者:** 王涛 (1977—), 男, 教授, 主要从事故障诊断、计算机视觉、图像处理等研究。E-mail: twang305@126.com

引用格式: 张聪, 王涛, 王念, 等. 双重协同细化的无监督跨模态行人重识别方法[J]. 火箭军工程大学学报, 2025, 39 (6): 91-103. ZHANG Cong, WANG Tao, WANG Nian, et al. Dual collaborative refinement-based unsupervised cross-modality person re-identification[J]. Journal of Rocket Force University of Engineering, 2025, 39 (6): 91-103.

of pseudo labels by calculating the similarity between instances and anchors. Second, a cross-modal bidirectional alignment module was designed using graph matching to preliminarily determine the feature correspondence between visible light and infrared modalities. It facilitated the model's learning of modality-invariant information by constructing modality independent mixed memory banks. Finally, based on the dual refinement of pseudo labeling and cross-modality matching, a modality-related and modality-independent contrastive learning module was designed to enhance the model's ability to learn discriminative features. Compared with the suboptimal multi-memory matching (MMM), the proposed method enhances the mean average precision (mAP) by 1.4% and 2.0% under the All-search and Indoor-search modes of the SYSU-MM01 dataset, respectively, and by 1.8% and 4.7% under the Visible2Infrared and Infrared2Visible modes of the RegDB dataset, respectively. The results demonstrate the excellent robustness and cross-modal retrieval capability of the proposed method.

KEYWORDS: person re-identification; unsupervised learning; cross-modality; contrastive learning

行人重识别 (person re-identification, ReID) 指从不同相机拍摄的各种行人图像数据集中检索特定行人, 通常侧重于在理想光照条件下拍摄的可见光图像^[1-2]。然而, 在光线不足或夜间环境下捕获的行人图像与可见光图像的模态域差异较大, 导致传统方法的检索能力明显下降^[3-4]。为解决这一问题, 特别是为了满足全天候智能监控的需要, 可见光-红外跨模态行人重识别 (visible-infrared person re-identification, VI-ReID)^[5-6]已经成为一个重要的研究方向。VI-ReID旨在给定一个红外/可见光模态下的行人图像, 在另一组可见光/红外模态图库中检索同一身份行人图像, 由于其在全天候智能监控系统和公共安全中的重要意义, 受到了业内的广泛关注。早期研究^[7-9]主要集中在需要大量标注信息的有监督方法, 但高昂的人工标注代价限制了其在现实应用场景中的可扩展性。

为解决大规模跨模态数据的标注问题, 一些学者提出了基于无监督学习的跨模态行人重识别 (unsupervised visible-infrared person re-identification, USL-VI-ReID) 方法^[10-11]。该方法通常以基于伪标签的对比学习方法^[12]为基线, 采用聚类算法^[13-14]作为训练样本预测伪标签, 在此基础上创建 2 个模态特定的记忆字典, 分别为可见光模态记忆字典和红外模态记忆字典。在模型训练过程中, 将记忆字典中存储的聚类中心作为原型, 最大限度地减少待查询图像特征和原型特征之间的对比损失。为进一步探索模态间的对应关系, 一些方法利用图匹配^[15]或最优传输^[16]来建立模态间的对应关系, 促进模型学习模态无关和身份相关的判别性特征, 并取得一定进展。例如, Wu 等^[17]量化

了可见光图片-红外图片的共享特征, 通过互学习来辅助融合可见光和红外模态中的互补信息; Pang 等^[18]提出一种跨模态层次聚类算法, 逐步将不同模态间的正样本融入到同一聚类簇中; Ye 等^[19]设计了一种实例级的跨模态匹配策略, 从而构建更加可靠的跨模态对应关系; Pang 等^[10]利用对抗生成网络构建中间模态, 促进模型学习模态无关信息。然而, 由于聚类算法的局限性, 预测伪标签中往往存在被错误预测的噪声伪标签, 从而降低了聚类中心原型的准确度。另外, 由于不同模态间的差异较大, 这种类间的对应关系很可能是不可靠的。尤其是在训练初期, 当模型表征能力不足时, 不可靠的伪标签和不准确的模态间对应关系将会导致模型学习错误特征信息。因此, 提高伪标签质量以及建立可靠的跨模态间对应关系是 USL-VI-ReID 方法研究的关键。

为了解决上述问题, 本文提出一种基于双重协同细化 (dual collaborative refinements, DCR) 的 USL-VI-ReID 方法。针对现有方法容易忽视噪声伪标签与全局伪标签之间的关系, 设计了一个伪标签细化模块, 根据噪声伪标签与全局伪标签的相似度来对伪标签进行细化, 从而提高伪标签的质量; 同时, 针对现有跨模态匹配方法仅在聚类级构建对应关系而忽视了特征层面的细粒度信息, 提出一个跨模态双向对齐模块, 通过双向约束构建 2 个模态无关混合记忆字典, 从而促进模型学习可靠的模态无关特征, 并通过模态相关和模态无关的对比学习模块来提高模型学习模态内和模态间共性特征的能力; 最后, 在 SYSU-MM01 与 RegDB 数据集上验证了该方法的有效性。

1 基于 DCR 的 USL-VI-ReID 方法

给定未标注的跨模态行人重识别数据集 D 为

$$D = \{D^v, D^r\} \quad (1)$$

其中, 可见光模态 D^v 与红外模态 D^r 分别为

$$D^v = \{x_1^v, x_2^v, \dots, x_{N^v}^v\} \quad (2)$$

$$D^r = \{x_1^r, x_2^r, \dots, x_{N^r}^r\} \quad (3)$$

式中: x_i^v 、 x_j^r 分别为可见光模态和红外模态的第 i 个和第 j 个实例; N^v 、 N^r 分别为可见光和红外模态的样本数量。

本文将训练过程分为 2 个阶段。在第一阶段, 采用双分支对比学习 (dual contrastive learning, DCL) 网络框架提取各模态图片特征, 聚类得到可见光模态和红外模态的伪标签集分别为

$$Y^v = \{y_1^v, y_2^v, \dots, y_{N^v}^v\} \quad (4)$$

$$Y^r = \{y_1^r, y_2^r, \dots, y_{N^r}^r\} \quad (5)$$

式中: y_i^v 、 y_j^r 分别为实例 x_i^v 、 x_j^r 的伪标签。模型以伪标签作为监督信号, 分别学习可见光模态和红外模态特征, 使模型具备一定表征能力。为了提高模型学习模态无关特征能力, 本文在第二阶段提出双重协同细化方法, 主要包括伪标签细化 (pseudo-label refinement, PR)、跨模态双向对齐 (cross-modality bidirectional alignment, CBA) 以及模态相关和模态无关的对比学习 (modality-specific and modality-invariant contrastive learning, MMCL) 3 个模块, 方法框架如图 1 所示。其中, 不同颜色表示不同行人。

如图 1 所示, 本文采用双分支网络分别对提取到的可见光和红外模态特征进行聚类, 并通过伪标签细化模块对错误预测的伪标签进行矫正; 然后, 分别基于可见光和红外模态的聚类原型来对模态相关的可见光和红外记忆字典进行初始化, 同时基于图匹配建立聚类级模态间的对应关系 V2R 和 R2V, 并初始化模态无关的混合记忆字典; 最后, 基于模态相关和模态无关记忆字典, 采用对比学习来优化模型。

1.1 伪标签细化

由于聚类算法的局限性, 预测伪标签不可避免地包含噪声。本文设计了伪标签细化模块, 以可靠的聚类原型为锚点, 通过实例特征与聚类原型的相似度来矫正错误预测的伪标签, 其原理示意图如图 2 所示。

首先, 计算每个聚类簇中实例之间的余弦相似度, 即

$$S_{i,j}^t = d(\phi(x_i^t), \phi(x_j^t)) \quad (6)$$

式中: t 表示模态, 取 v (可见光模态) 或 r (红外模态); $d(\cdot)$ 为余弦相似度函数; $\phi(\cdot)$ 为模型函数。

对于第 i 个样本, 该聚类簇中的其余样本与该样本的相似度分数 U_i^t 为

$$U_i^t = \sum_{j=1}^{n-1} \text{sign}(S_{i,j}^t - \sigma), \quad i = 1, 2, \dots, n \quad (7)$$

式中: $\text{sign}(\cdot)$ 为符号函数; σ 为阈值参数; n 为该聚类簇内的实例个数。由于被正确聚类的实例应该具有较高的相似度, 因此伪标签细化模块将相

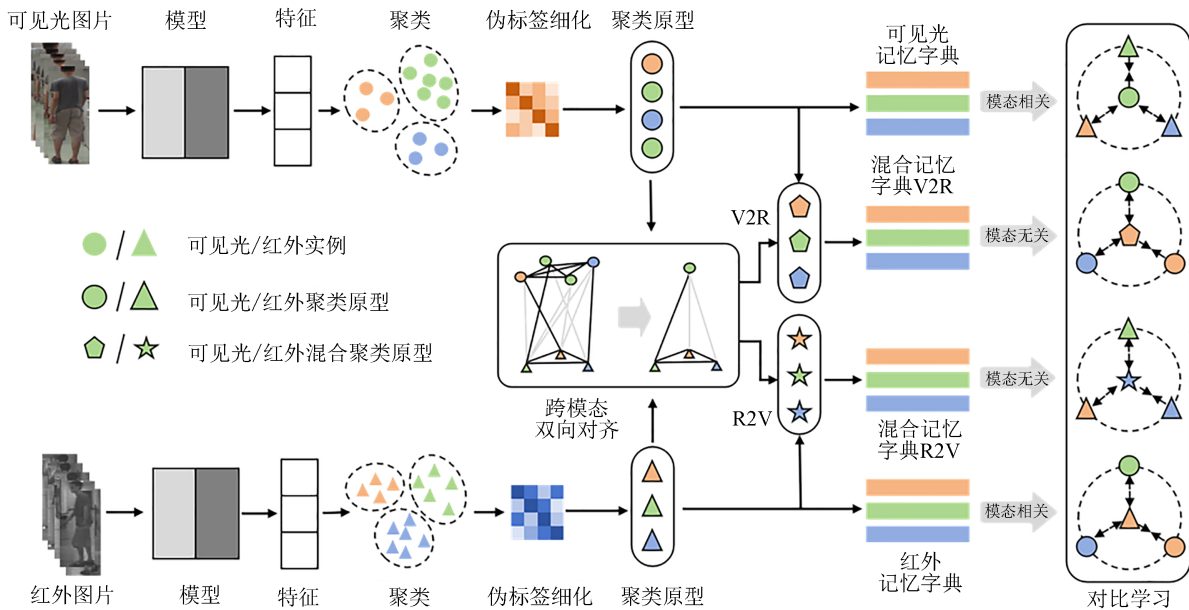


图 1 双重协同细化方法框架

Fig. 1 Framework of dual collaborative refinement

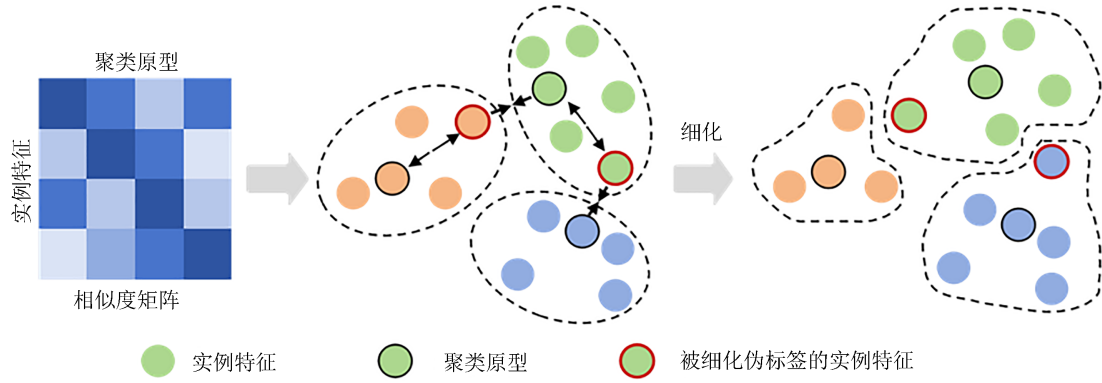


图 2 伪标签细化模块

Fig. 2 Pseudo label refinement module

似度较高的 K 个样本实例视为可靠样本, 即

$$e = \arg \max_K U_i^t \quad (8)$$

式中: e 为该 K 个实例的索引。

然后, 本文以可靠实例特征的均值作为该聚类簇的锚点, 即

$$l^t = \frac{1}{K} \sum_{i \in e} \phi(x_i^t) \quad (9)$$

同理, 可以得到所有聚类簇的可靠锚点集合 $L^t = \{l_1^t, l_2^t, \dots, l_{N_c}^t\}$ 。其中, N_c 为可见光或红外模态聚类簇的个数。

最后, 计算每个实例与所有锚点的余弦相似度并得到相似度分数, 将相似度分数最高的锚点所对应的类别索引作为该实例细化后的伪标签, 数学表达式为

$$\hat{y}_i^t = \arg \max_p U_i^t, P = 1, 2, \dots, N_c^t \quad (10)$$

经细化后的伪标签可以优化记忆字典中原型特征的准确度, 从而提高模型学习判别性特征的能力。

1.2 跨模态双向对齐

不同模态之间的信息交互是影响模型性能的关键因素。然而, 由于红外图像所包含的信息比可见光图像少, 在高维特征空间中可能被限制在很小的子空间内, 导致大多数红外聚类原型只能与少数特定的可见光聚类原型进行匹配, 可能使模型在训练过程中丢失一些聚类原型信息。因此, 现有方法提出的单向跨模态匹配方法在无监督条件下的映射关系并不可靠。为了解决这一问题, 本文提出跨模态双向对齐模块, 其原理示意如图 3 所示。

由图 3 可见, 通过图匹配建立可见光与红外聚类原型之间的跨模态对应关系 R2V 和 V2R, 然后, 将可见光聚类原型特征与其对应的红外聚类

原型特征进行加权融合, 并初始化混合记忆字典 M^{r2v} 。同理, 以红外聚类原型特征与其对应的可见光聚类原型特征加权融合来初始化混合记忆字典 M^{v2r} 。跨模态双向对齐模块为每个模态建立一个图, 以聚类中心原型为节点, 则可见光模态图 G^v 的节点集合 V 和红外模态图 G^r 的节点集合 R 分别表示为

$$V = \{m_i^v \mid i = 1, 2, \dots, N_c^v\} \quad (11)$$

$$R = \{m_j^r \mid j = 1, 2, \dots, N_c^r\} \quad (12)$$

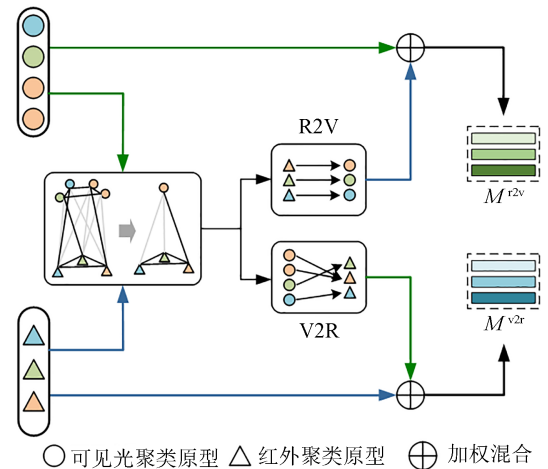


图 3 跨模态双向对齐模块

Fig. 3 Cross-modality bidirectional alignment module

图匹配的目标函数可表示为

$$\begin{cases} G(w) = \arg \min_w C^T w \\ \text{s.t. } \forall i \in V, \forall j \in R: w_{ij} \in \{0, 1\} \\ \forall i \in V: \sum_{j \in R} w_{ij} \leq 1 \\ \forall j \in R: \sum_{i \in V} w_{ij} = 1 \end{cases} \quad (13)$$

式中: 矩阵 w 中的元素 w_{ij} 表示 m_i^v 和 m_j^r 是否匹配的指标。当 $w_{ij} = 1$ 时表示节点匹配, 否则表示不匹

配。成本矩阵 $\mathbf{C}$ 可表示为

$$C_{ij} = \frac{1}{\exp [d(m_i^v, m_j^r)]} \quad (14)$$

跨模态对齐的目标是在最小成本的前提下, 构建不同模态间每个节点的对应关系, 可以使用匈牙利算法求得。

跨模态双向对齐模块将可见光模态的节点和红外模态的节点进行一对一匹配。在执行一次图匹配之后, G^v 中仍可能存在未找到对应关系的剩余节点, 而 G^r 中所有节点均已匹配。针对这一问题, 模块利用剩余节点动态构建一个新图, 与 G^r 再次执行图匹配。重复执行这一过程, 直至所有节点完成匹配。

针对基于图匹配得到的对应关系 R2V 和 V2R, 跨模态双向对齐模块进一步设计了 2 个模态无关的混合记忆字典 M^{v2v} 和 M^{v2r} , 分别由加权融合的聚类中心原型进行初始化, 即

$$m_i^{v2v} = \eta \times m_i^v + (1 - \eta) \times m_{r2v,i}^r \quad (15)$$

$$m_j^{v2r} = \eta \times m_j^r + (1 - \eta) \times m_{v2r,j}^v \quad (16)$$

式中: η 为平衡可见光和红外聚类中心原型融合比例的超参数; $m_{r2v,i}^r$ 表示与可见光聚类中心原型 m_i^v 匹配的红外聚类中心原型; $m_{v2r,j}^v$ 表示与红外聚类中心原型 m_j^r 对齐的可见光聚类中心原型。2 个模态无关记忆字典的维度不同, 且含有的模态信息也不相同。跨模态双向对齐能够从特征层面融合模态间对应的特征, 细化跨模态匹配的可靠性, 从而提高了模型学习模态无关特征的能力。

1.3 模态相关和模态无关的对比学习

本文基于伪标签细化和跨模态双向对齐分别得到 2 个模态相关记忆字典 M^v 和 M^r 以及 2 个模态无关记忆字典 M^{v2v} 和 M^{v2r} , 并构建了模态相关和模态无关的对比学习框架。

对于待查询的可见光实例 x_q^v 和红外实例 x_q^r , 模态相关的对比损失可以表示为

$$\mathcal{L}_{ms} = \mathcal{L}_v + \mathcal{L}_r \quad (17)$$

$$\mathcal{L}_v = -\log \frac{\exp [\phi(x_q^v) \cdot m^{v+} / \tau]}{\sum_{i=1}^{N_v^+} \exp [\phi(x_q^v) \cdot m_i^v / \tau]} \quad (18)$$

$$\mathcal{L}_r = -\log \frac{\exp [\phi(x_q^r) \cdot m^{r+} / \tau]}{\sum_{j=1}^{N_r^+} \exp [\phi(x_q^r) \cdot m_j^r / \tau]} \quad (19)$$

式中: m^{v+} 和 m^{r+} 分别为待查询实例在 2 个模态相关记忆字典中的正对; τ 为温度因子。模态无关的对比损失可以表示为

$$\mathcal{L}_{mi} = \mathcal{L}_v^{v2v} + \mathcal{L}_r^{v2r} \quad (20)$$

$$\mathcal{L}_v^{v2v} = -\log \frac{\exp [\phi(x_q^v) \cdot m^{v2v+} / \tau]}{\sum_{i=1}^{N_v^+} \exp [\phi(x_q^v) \cdot m_i^{v2v} / \tau]} \quad (21)$$

$$\mathcal{L}_r^{v2r} = -\log \frac{\exp [\phi(x_q^r) \cdot m^{v2r+} / \tau]}{\sum_{j=1}^{N_r^+} \exp [\phi(x_q^r) \cdot m_j^{v2r} / \tau]} \quad (22)$$

式中: m^{v2v+} 和 m^{v2r+} 分别为待查询实例在 2 个模态无关记忆字典中的正对。模态相关和模态无关对比学习的总目标损失函数为

$$\mathcal{L} = \mathcal{L}_{ms} + \lambda \mathcal{L}_{mi} \quad (23)$$

式中: λ 为平衡系数。模态相关和模态无关记忆字典均以动量方式更新, 即

$$m_i \leftarrow \mu m_i + (1 - \mu) \phi(x_q) \quad (24)$$

式中: μ 为动量更新系数; x_q 为待查询实例。

2 实验与分析

2.1 数据集与评价指标

本文采用广泛使用的跨模态行人重识别数据集 SYSU-MM01^[20] 和 RegDB^[21] 评估 DCR 方法。模型性能评价指标采用平均均值精度 (mean average precision, mAP)、累积匹配特征 (cumulative matching characteristic, CMC) 以及度量平均反向负惩罚 (mean inverse negative penalty, mINP)。其中, mAP 综合了所有相关目标的检索精度, 衡量整体检索效果; CMC 选取 Rank-1、Rank-10 和 Rank-20 作为具体指标, Rank-1 体现模型检索时首次命中目标的能力, 反映精准度, Rank-10 和 Rank-20 体现前 10、20 次检索中包含目标的概率, 反映召回的全面性; mINP 则聚焦于困难样本的检索性能, 体现模型对复杂实例的区分能力。

2.2 实验设置

本文提出的 DCR 方法以在 ImageNet^[22] 上预训练的 ResNet50 作为骨干网络。所有图像被调整为 288 像素 $\times$ 144 像素, 并且使用了随机翻转、随机剪裁、随机擦除和通道增强^[23] 等作为数据增强手段。训练阶段的采样批量大小设置为 128, 采用 Adam optimizer 优化网络, 权重衰减率为 0.000 5。初始学习率设为 0.000 35, 之后每经过 20 个训练周期降为原来的 1/10。在第一个训练阶段的 50 个周期内, 采用 DCL 框架训练模型来学习模态特定信息。在第二个训练阶段的 50 个周期内, 采用本文提出的 DCR 方法训练模型。记忆字典的动量更

新系数 μ 设为 0.1, 融合比例参数 η 设置为 0.5, 阈值参数 σ 设置为 0.5, 平衡系数 λ 设置为 0.4。本文采用 Py-torch 深度学习平台实现 DCR 方法的构建和运行, 使用 4 个 GeForce RTX 3090 GPU 进行训练, 使用 1 个 GeForce RTX 3090 GPU 进行测试。

2.3 与相关方法性能比较

实验进一步将本文所提出的 DCR 方法与多个无监督跨模态行人重识别方法进行比较, 包括 ADCA (augmented dual contrastive aggregation)^[12]、DOTLA (dual optimal transport label assignment)^[24]、MBCCM (many-to-many bilateral cross-modality cluster matching)^[16]、CCLNet (CLIP-enhanced contrastive learning network)^[25]、PGM (progressive graph matching)^[15]、GUR (grand unified representation)^[26]、MMM (multi-memory matching)^[27]。表 1 和表 2 分别展示了 DCR 方法与上述几种方法在 SYSU-MM01 和 RegDB 数据集上的对比结果。

其中, 最优结果用粗体标出, 次优结果用下划线标出。对比方法的结果均取自其原始论文, 部分未列出的结果因原文未提供而缺失。

由表 1 和表 2 可以看出, 在 SYSU-MM01 和 RegDB 数据集上, DCR 方法的绝大多数性能评估指标均优于其他几种对比方法, 体现出更强的跨模态行人重识别能力, 这得益于其在模态差异消解与特征判别性增强等方面的有效设计, 能更稳健地应对不同场景下的跨模态检索任务。值得注意的是, 表中所列 MMM 方法的多数性能评估指标均为次优, 且在 RegDB 数据集上的 Visible2Infrared 模式下, MMM 方法的 Rank-1 指标显著高于 DCR 方法。这是由于 RegDB 数据集由同一摄像机在相同时间内同步采集 2 种模态行人图像, 模态间纹理特征更为相似。因此, MMM 方法在 Visible2Infrared 模式下, 能较好捕获待检索样本的纹理信息与可见光模态的颜色信息, 在单样本精准

表 1 在 SYSU-MM01 数据集上的对比实验结果

Table 1 Results of comparative experiments on the SYSU-MM01 dataset

%

方法	SYSU-MM01									
	All-search					Indoor-search				
	Rank-1	Rank-10	Rank-20	mAP	mINP	Rank-1	Rank-10	Rank-20	mAP	mINP
ADCA	45.5	85.3	93.2	42.7	28.3	50.6	90.0	94.2	59.1	55.2
DOTLA	50.4	89.0	96.0	47.4	—	53.5	92.2	97.8	61.7	—
MBCCM	53.1	—	—	48.2	32.5	55.2	—	—	62.0	57.1
CCLNet	54.0	88.8	95.0	50.2	—	56.7	91.1	<u>98.2</u>	65.1	—
PGM	57.3	92.5	97.2	51.8	35.0	56.2	90.2	95.4	62.7	58.1
GUR	61.0	—	—	57.0	<u>41.9</u>	64.2	—	—	69.5	<u>64.8</u>
MMM	<u>61.6</u>	<u>93.3</u>	<u>98.0</u>	<u>57.9</u>	—	<u>64.4</u>	<u>95.0</u>	<u>98.2</u>	<u>70.4</u>	—
DCR	62.4	93.4	98.6	59.3	44.9	66.4	95.3	98.5	72.4	68.4

表 2 在 RegDB 数据集上的对比实验结果

Table 2 Results of comparative experiments on the RegDB dataset

%

方法	RegDB									
	Visible2Infrared					Infrared2Visible				
	Rank-1	Rank-10	Rank-20	mAP	mINP	Rank-1	Rank-10	Rank-20	mAP	mINP
ADCA	67.2	82.0	87.4	64.1	52.7	68.5	83.2	88.0	63.8	49.6
DOTLA	85.6	<u>94.1</u>	<u>95.5</u>	76.7	61.6	82.9	<u>92.3</u>	<u>94.9</u>	75.0	58.6
MBCCM	83.8	—	—	77.9	<u>65.0</u>	82.8	—	—	76.7	<u>61.7</u>
CCLNet	69.9	—	—	65.5	—	70.2	—	—	66.7	—
PGM	69.5	—	—	65.4	—	69.9	—	—	65.2	—
GUR	73.9	—	—	70.2	58.9	75.0	—	—	69.9	56.2
MMM	89.7	—	—	<u>80.5</u>	—	<u>85.8</u>	—	—	<u>77.0</u>	—
DCR	<u>87.0</u>	95.2	97.3	82.3	71.9	86.8	95.2	97.1	81.7	68.5

匹配方面表现突出,而在 Infrared2Visible 模式下,由于待检索样本缺失关键颜色信息,致使其检索性能受到限制。

2.4 消融实验分析

本节通过消融实验进一步验证 DCR 方法中 3 个关键模块的有效性。实验以 DCL 作为基线模型 (Baseline),在 SYSU-MM01 数据集的 All-search 模式下进行对比分析,结果如表 3 所示。

从表 3 中可以看出,本文所提出的 3 个模块均在不同程度上提升了模型的跨模态行人重识别性能。具体而言,引入 PR 模块后,模型的 mAP 和 Rank-1 较 Baseline 分别提升了 3.7% 和 5.3%;进一步加入 CBA 模块后, mAP 和 Rank-1 较 Baseline 分别提升了 17.8% 和 22.8%;当最终引入 MMCL 模块后,模型性能达到最优, mAP 和 Rank-1 较 Baseline 分别提升了 23.4% 和 27.3%。实验结果表明:引入 CBA 模块后模型性能提升显著,表明在无监督跨模态行人重识别任务中,有效缓解模态间差异对算法性能提升具有关键作用。上述消融实验充分验证了本文所提出 3 个模块的有效性及其协同优化能力。

为了评估 PR 模块对聚类预测伪标签质量的影响,本文参考文献 [28] 和 [29],选取标准化互信息 (normalized mutual information, NMI)、调整兰德指数 (adjusted Rand index, ARI)、福勒克斯-马洛克斯指数 (Fowlkes-Mallows index, FMI)、调整互信息 (adjusted mutual info, AMI) 评分和 V 度量 (V-measure) 评分作为评价指标,系统对比了 PGM^[15]伪标签细化方法与 DCR 伪标签细化方法的性能,实验结果如图 4 所示。其中,“可见光”与“红外”分别指可见光与红外模态下的伪标签与真实标签之间的评价指标得分,得分越高表明伪标签与真实标签之间的匹配度越好,聚类预测伪标签的质量越高。

从评价指标的分数变化曲线可以看出,PGM

方法的聚类质量随着训练迭代步数逐步提升,但在第 20 个训练周期后趋于稳定,整体提升幅度有限。相比之下,本文所提 DCR 方法引入 PR 模块后,模型在训练初期受表征能力有限的影响,各项聚类评价指标得分并不理想,但随着迭代训练,评价指标得分显著提升。实验结果表明:PR 模块通过度量聚类簇内的样本与全局聚类原型锚点之间的关系,有效提高了聚类预测伪标签的质量。

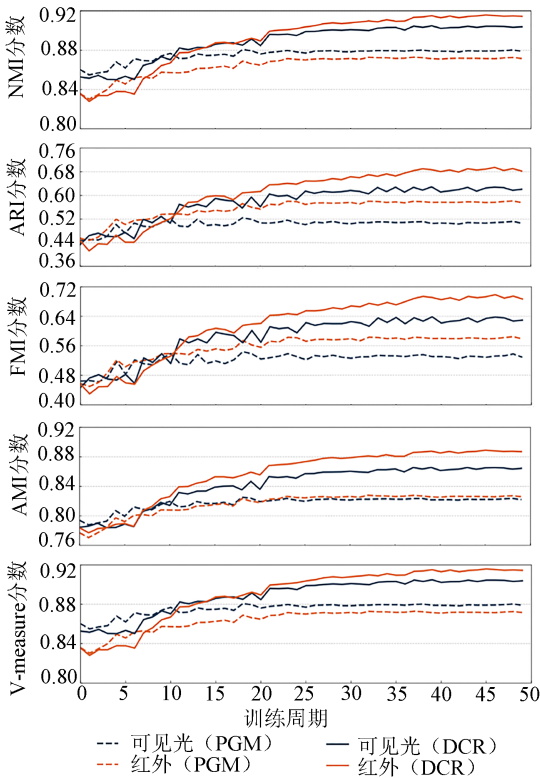


图 4 聚类预测伪标签质量分析
Fig. 4 Quality analysis of pseudo labels in clustering predictions

为进一步分析 CBA 模块对跨模态匹配性能的影响,绘制得到使用不同匹配策略的模型在训练过程中的跨模态匹配准确率随迭代周期的变化曲线,结果如图 5 所示。其中,虚线表示采用跨模态单向对齐 (cross-modality unidirectional alignment,

表 3 在 SYSU-MM01 数据集的消融实验结果
Table 3 Results of ablation experiments on the SYSU-MM01 dataset %

序号	模块				SYSU-MM01				
	Baseline	PR	CBA	MMCL	Rank-1	Rank-10	Rank-20	mAP	mINP
1	√				35.1	78.4	85.0	35.9	23.5
2	√	√			40.4	80.1	89.5	39.6	26.5
3	√	√	√		57.9	92.0	96.9	53.7	38.1
4	√	√	√	√	62.4	93.4	98.6	59.3	44.9

CUA) 模块的模型匹配准确率变化曲线。

由图 5 可知, 在训练初期约 10 个周期内, 由于使用双向匹配策略的 CBA 模块引入了更加严格的对齐约束, 模型仅学习可靠实例对的共性特征, 导致模型初期匹配准确率略低于单向匹配策略。然而, 随着训练迭代, 使用双向匹配策略的 CBA 模块能够有效促进模型学习模态无关特征, 使跨模态匹配准确率显著优于单向匹配策略。该可视化结果验证了 CBA 模块在提升跨模态语义一致性与匹配可靠性方面的有效性。

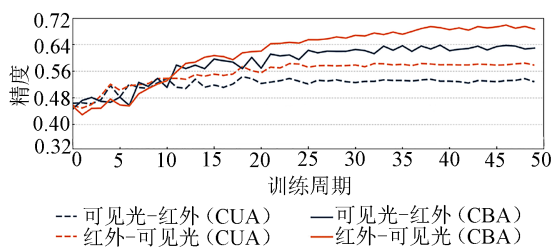


图 5 跨模态匹配准确率曲线

Fig. 5 Accuracy curves of cross-modality matching

2.5 参数分析

本文对 DCR 方法中的关键参数 η 、 λ 、 σ 和 μ 进行了系统分析, 以评估其对模型性能的影响。

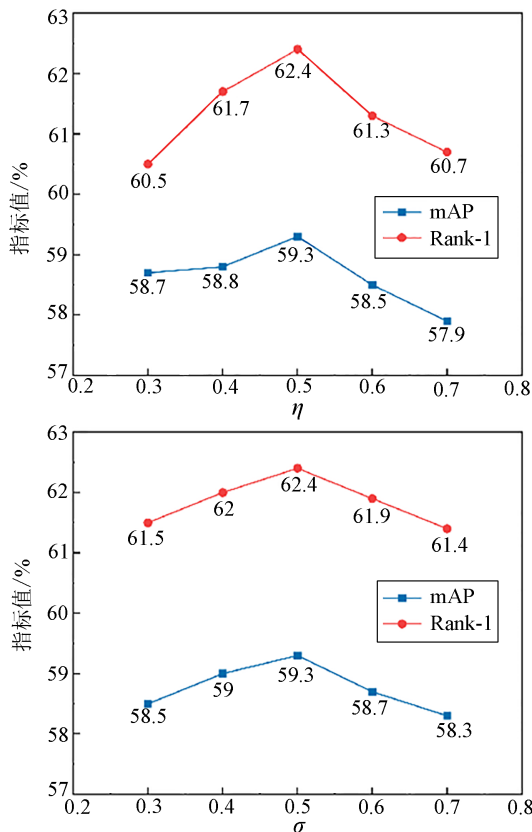


图 6 关键参数分析

Fig. 6 Analysis of key parameters

实验在 SYSU-MM01 数据集的 All-search 模式下进行, 通过调整参数并观察模型 mAP 和 Rank-1 的变化情况来确定最优参数配置。图 6 分别展示了这 4 个关键参数对模型性能的影响。其中, 默认参数 $\eta = 0.5$, $\lambda = 0.4$, $\sigma = 0.5$, $\mu = 0.1$, 采用控制变量法研究每个参数对模型性能的影响。

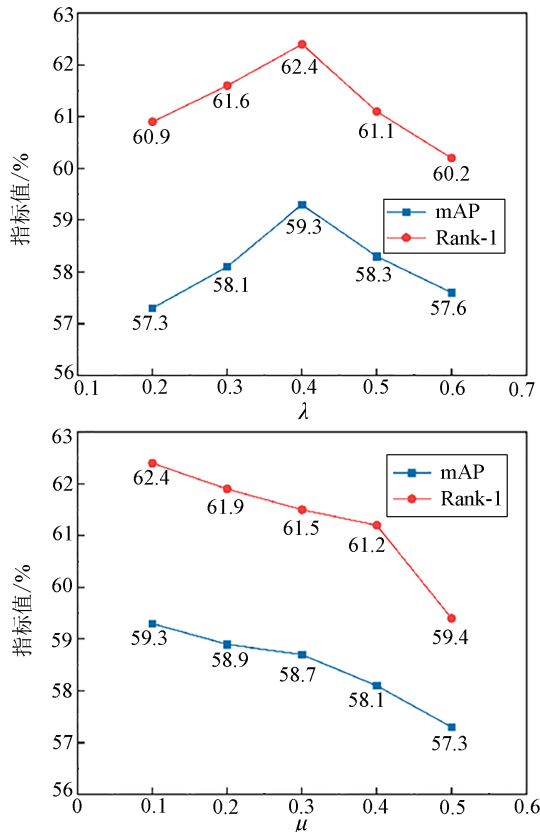
从图 6 中可以看到, 随着 η 、 λ 和 σ 值的增加, 模型的 mAP 和 Rank-1 先增大后减小, 在 η 、 λ 、 σ 值分别取 0.5、0.4、0.5 时, 模型性能达到最佳。此外, 模型的 mAP 和 Rank-1 随着 μ 值的增加而逐渐减小, 表明过大的 μ 值会显著降低模型性能。

2.6 可视化分析

为了定性评估 DCR 方法的性能, 本文进行了 3 组可视化实验, 包括行人特征空间分布可视化、行人特征分布可视化和行人检索排序可视化。

1) 特征空间分布可视化。实验随机选取 7 个不同身份的行人, 利用 t-SNE 方法将可见光模态和红外模态的特征分布合并聚类, 并将其嵌入到二维空间, 可视化结果如图 7 所示。其中, 不同颜色代表不同行人身份, 不同形状代表不同模态。

从图 7 中可以看出, 与 Baseline 相比, DCR 方



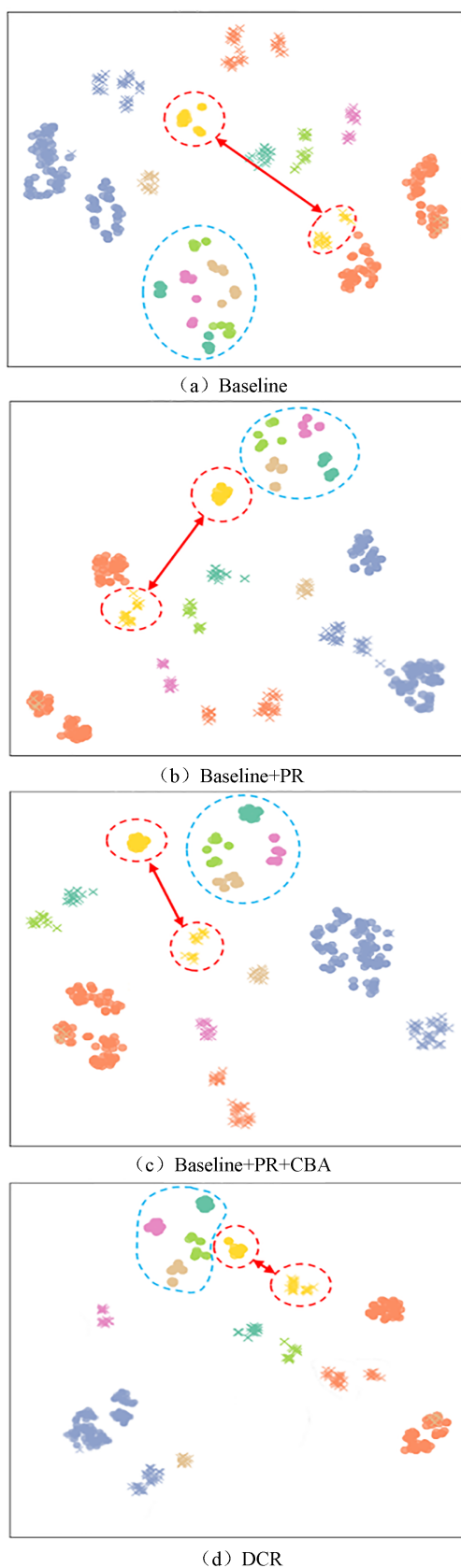
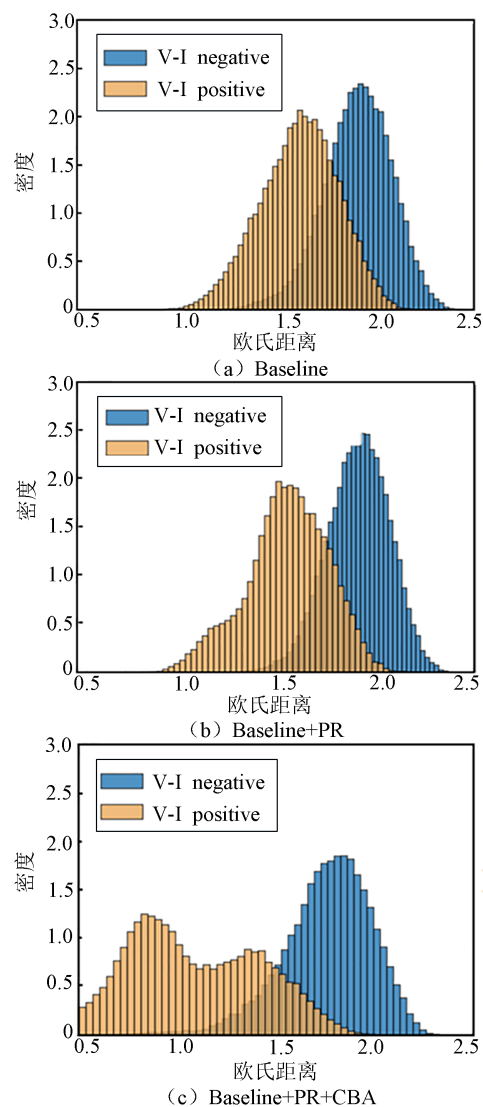


图 7 随机选择的 7 个行人身份的 t-SNE 可视化对比
Fig. 7 t-SNE visualization comparison of seven randomly selected pedestrian identities

法的 3 个模块逐渐将同一模态下行人的类内特征聚合在一起, 使类内特征分布更加紧凑、类间特征区分明显 (见蓝色虚线圈)。同时, 不同模态间的行人类内特征分布逐渐拉近 (见红色虚线圈), 表明模型学习同一身份行人的模态间共性特征能力被逐渐加强。该可视化对比结果表明: 本文所提方法通过建立可靠的跨模态对应关系, 既强化了同一模态内的类间区分度, 又显著提高了模型学习模态无关特征能力。

2) 特征分布可视化。为了进一步定性展示本文所提方法的跨模态匹配性能, 将随机选取的跨模态正、负样本对的特征分布可视化, 结果如图 8 所示。

由图 8 (a) 可得, Baseline 的可见光-红外正对 (V-I positive) 的欧氏距离主要集中在 1.7 附近, 密度约为 2.2; 可见光-红外负对 (V-I negative)



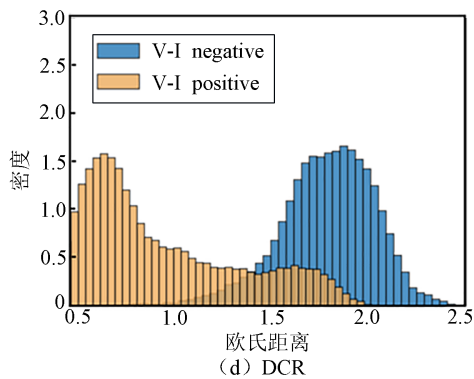


图 8 跨模态正/负对的特征分布可视化对比

Fig. 8 Visualization comparison of feature distribution of cross-modality positive and negative pairs

tive) 的欧式距离集中在 2.0 附近, 密度约为 2.4, 二者分布的重叠部分较多, 区分度有限。加入 PR 和 CBA 模块后, V-I positive 的欧氏距离分布向更小值区域移动, 主要集中在 0.9~1.4, 密度约为 1.1; V-I negative 的欧氏距离集中在 1.8, 密度约

为 2.0, 二者之间的重叠区域大幅减少。当引入全部模块时, V-I positive 的欧氏距离趋向于 0.7, 密度约为 1.5; V-I negative 的欧氏距离集中在 1.8~2.4, 密度约为 1.6, 正对与负对的分布几乎完全分离。由此可见, 本文所提的 3 个模块使可见光-红外正对的收敛性逐渐紧凑, 同时与可见光-红外负对的区分度显著提高, 证明了 DCR 方法的 3 个模块均能不同程度提升模型学习模态无关特征的能力, 从而改善模型的跨模态检索性能。

3) 行人检索排序可视化。实验从数据集中随机抽取一个红外模态行人图片, 并分别列出了不同模型给出的图库中与其最相似的 5 个可见光模态行人图片, 如图 9 所示。其中, 绿色底框表示正确检索的行人图片, 红色底框表示错误检索的行人图片。从检索排序结果中可以看到, Baseline 错误检索了 4 个行人身份, 而本文提出的 3 个模块均不同程度提高了行人检索排序精度。对比 Baseline 的



图 9 跨模态行人检索排序可视化对比

Fig. 9 Visualization comparison of cross-modality pedestrian retrieval and sorting results

检索排序结果, DCR 方法在一定程度上解决了因模态差异而导致的跨模态特征分布不一致的问题, 且对行人的姿态差异和光照差异更具鲁棒性, 从而提高了模型跨模态检索性能。

3 结论

本文为解决无监督跨模态行人重识别过程中存在的噪声伪标签干扰与模态间特征分布差异等核心问题, 以双分支对比学习网络为基础框架, 设计 PR、CBA 及 MMCL 核心模块, 提出 DCR 方法, 并通过在 SYSU-MM01 与 RegDB 数据集上的对比实验、消融实验及可视化分析, 验证了该方法的有效性, 主要结论如下。

1) PR 模块可显著提升伪标签质量。该模块以聚类簇内可靠实例特征均值为锚点, 通过计算实例与锚点间的余弦相似度来矫正错误伪标签, 相较于 PGM 方法, 其在训练后期能够显著提升聚类质量, NMI、ARI 和 FMI 等多个评价指标得分更高, 能有效减少噪声伪标签对模型训练的干扰, 为后续特征学习提供更可靠的监督信号。

2) CBA 模块可有效缓解模态差异, 提升跨模态匹配可靠性。该模块通过图匹配建立可见光与红外模态的双向对应关系, 构建模态无关混合记忆字典, 虽然在训练初期因严格对齐约束而导致匹配准确率略低于单向匹配策略, 但随着训练迭代, 跨模态匹配准确率得到显著提高, 且使不同模态间同一身份行人的类内特征分布更接近, 为模型学习模态无关共性特征提供关键支撑。

3) MMCL 模块可增强模型判别性特征学习能力。该模块结合模态相关记忆字典与模态无关记忆字典构建对比损失, 在 SYSU-MM01 和 RegDB 数据集上的性能达到最佳, 多数指标优于 ADCA、DOTLA 和 MMM 等现有主流方法。

然而, 本文 DCR 方法仍存在一定的局限性: 一是模型训练需依赖 4 个 GeForce RTX 3090 GPUs, 计算资源消耗较高, 在低资源场景下适用性受限; 二是在 RegDB 数据集的 Visible2Infrared 模式下, Rank-1 指标得分明显低于 MMM 方法, 表明本文方法对纹理特征相似的跨模态样本精准匹配能力仍有提升空间。未来, 将进一步探索模型轻量化方案, 强化纹理特征学习, 以增强无监督跨模态行人重识别模型在实际应用中的鲁棒性。

参考文献

- [1] 李擎, 胡伟阳, 李江昀, 等. 基于深度学习的行人重识别方法综述[J]. 工程科学学报, 2022, 44(5): 920-932.
LI Q, HU W Y, LI J Y, et al. A survey of person re-identification based on deep learning[J]. Chinese Journal of Engineering, 2022, 44(5): 920-932.
- [2] 钱亚萍, 王凤随, 熊磊. 基于局部细化多分支与全局特征共享的无监督行人重识别方法[J]. 电子测量与仪器学报, 2023, 37(1): 106-115.
QIAN Y P, WANG F S, XIONG L. Unsupervised person re-identification method based on local refinement multi-branch and global feature sharing[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(1): 106-115.
- [3] 范慧杰, 郁航, 赵颖畅, 等. 可见光红外跨模态行人重识别方法综述[J]. 信息与控制, 2025, 54(1): 50-65.
FAN H J, YU H, ZHAO Y C, et al. Review of visual-infrared cross-modal person re-identification methods[J]. Information and Control, 2025, 54(1): 50-65.
- [4] 刘章平. 面向单模态和跨模态场景的无监督行人重识别方法研究[D]. 合肥: 中国科学技术大学, 2024: 5-7.
LIU Z P. Unsupervised person re-identification methods for single and cross modality scenarios[D]. Hefei: University of Science and Technology of China, 2024: 5-7.
- [5] YE M, SHEN J B, LIN G J, et al. Deep learning for person re-identification: a survey and outlook[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(6): 2872-2893.
- [6] TAN X H, CHAI Y X, CHEN F L, et al. A fourier-based semantic augmentation for visible-thermal person re-identification[J]. IEEE Signal Processing Letters, 2022, 29: 1684-1688.
- [7] 赵三元, 阿琪, 高宇. 基于跨模态近邻损失的可见-红外行人重识别[J]. 北京航空航天大学学报, 2024, 50(2): 433-441.
ZHAO S Y, A Q, GAO Y. Cross-modality nearest neighbor loss for visible-infrared person re-identification[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(2): 433-441.
- [8] LIU M, SUN Y Q, WANG X P, et al. Pose-guided modality-invariant feature alignment for visi-

- ble-infrared object re-identification[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 5017610.
- [9] ZHANG D M, ZHANG Z Z, JU Y, et al. Dual mutual learning for cross-modality person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(8): 5361-5373.
- [10] PANG Z Q, WANG C Y, PAN H H, et al. MIMR: modality-invariance modeling and refinement for unsupervised visible-infrared person re-identification[J]. Knowledge-Based Systems, 2024, 285: 111350.
- [11] 孙锐, 余益衡, 张磊, 等. 基于语义伪标签和双重特征存储库的无监督跨模态行人重识别[J]. 模式识别与人工智能, 2022, 35(10): 904-914.
- SUN R, YU Y H, ZHANG L, et al. Unsupervised cross-modality person re-identification based on semantic pseudo-label and dual feature memory banks[J]. Pattern Recognition and Artificial Intelligence, 2022, 35(10): 904-914.
- [12] YANG B, YE M, CHEN J, et al. Augmented dual-contrastive aggregation learning for unsupervised visible-infrared person re-identification[C]//Proceedings of the 30th ACM International Conference on Multimedia. New York, USA: ACM, 2022: 2843-2851.
- [13] 谢佳伟, 余志勇, 吕典, 等. 结合基于密度的噪声应用空间聚类 and 反距离加权算法的城市环境电磁频谱地图构建[J]. 火箭军工程大学学报, 2024, 38(4): 93-97.
- XIE J W, YU Z Y, LYU D, et al. Construction of electromagnetic spectrum map for urban environment based on DBSCAN-IDW algorithm[J]. Journal of Rocket Force University of Engineering, 2024, 38(4): 93-97.
- [14] 王念, 张炜, 崔智高, 等. 基于离散指示矩阵优化的多视角图聚类方法[J]. 火箭军工程大学学报, 2024, 38(3): 51-59.
- WANG N, ZHANG W, CUI Z G, et al. A graph-based multi-view clustering method based on discrete indicator matrix optimization[J]. Journal of Rocket Force University of Engineering, 2024, 38(3): 51-59.
- [15] WU Z S, YE M. Unsupervised visible-infrared person re-identification via progressive graph matching and alternate learning[C]//2023 IEEE / CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2023: 9548-9558.
- [16] CHENG D, HE L F, WANG N N, et al. Efficient bilateral cross-modality cluster matching for unsupervised visible-infrared person ReID[C]//Proceedings of the 31st ACM International Conference on Multimedia. New York, USA: ACM, 2023: 1325-1333.
- [17] WU A C, LIN C Z, ZHENG W S. Asymmetric mutual learning for unsupervised transferable visible-infrared re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(11): 10524-10537.
- [18] PANG Z Q, WANG C Y, ZHAO L L, et al. Cross-modality hierarchical clustering and refinement for unsupervised visible-infrared person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(4): 2706-2718.
- [19] YE M, WU Z S, DU B. Dual-level matching with outlier filtering for unsupervised visible-infrared person re-identification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(5): 3815-3829.
- [20] WU A C, ZHENG W S, YU H X, et al. RGB-infrared cross-modality person re-identification[C]//Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2017: 5390-5399.
- [21] NGUYEN D T, HONG H G, KIM K W, et al. Person recognition system based on a combination of body images from visible light and thermal cameras[J]. Sensors, 2017, 17(3): 605.
- [22] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database[C]//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2009: 248-255.
- [23] YE M, RUAN W J, DU B, et al. Channel augmented joint learning for visible-infrared recognition [C]//Proceedings of 2021 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2021: 13567-13576.
- [24] CHENG D, HUANG X J, WANG N N, et al. Unsupervised visible-infrared person ReID by collaborative learning with neighbor-guided label refinement[C]//Proceedings of the 31st ACM International Conference on Multimedia. New York, USA: ACM, 2023: 7085-7093.
- [25] CHEN Z, ZHANG Z Z, TAN X, et al. Unveil-

- ing the power of clip in unsupervised visible-infrared person re-identification[C]//Proceedings of the 31st ACM International Conference on Multimedia. New York, USA: ACM, 2023: 3667-3675.
- [26] YANG B, CHEN J, YE M, et al. Towards grand unified representation learning for unsupervised visible-infrared person re-identification[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2023: 11035-11045.
- [27] SHI J M, YIN X B, CHEN Y Y, et al. Multi-memory matching for unsupervised visible-infrared person re-identification[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2024: 456-474.
- [28] GE Y X, ZHU F, CHEN D P, et al. Self-paced contrastive learning with hybrid memory for domain adaptive object re-ID[J]. Advances in Neural Information Processing Systems, 2020, 33(1): 11309-11321.
- [29] ZHANG X Y, LI D D, WANG Z G, et al. Implicit sample extension for unsupervised person re-identification[C]//2022 IEEE / CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2022: 7359-7368.

(编辑: 刘建民)