

基于柔性加权原型对比学习的无监督可见光-红外行人重识别

赵涵硕¹, 马伊雯², 张艳霞¹, 王 珮³, 杨建伟⁴

(1. 河南开放大学 机电工程与智能制造学院, 河南 郑州 450046; 2. 郑州航空工业管理学院 计算机学院, 河南 郑州 450046;
3. 河南开放大学 建筑工程与智能建造学院, 河南 郑州 450046; 4. 南京邮电大学 自动化学院, 江苏 南京 210023)

摘要: 无监督可见光-红外行人重识别 (Unsupervised Visible-Infrared Person Re-identification, USVI-ReID) 是一项非常重要且具有挑战性的任务。USVI-ReID 的关键挑战是在不依赖任何跨模态标注的情况下有效地生成伪标签并建立跨模态对应。近年来, 通过聚类算法生成伪标签的方法在 USVI-ReID 中得到了越来越多的关注。然而, 以前的方法只是选择代表个体的单一类中心原型或按照一定的策略随机选择原型来建立跨模态对应。这不仅忽略了个体特征的多样性, 也没有考虑聚类过程中错误样本对模型训练的影响。为了解决这个问题, 本文提出了一种柔性加权原型对比学习 (Soft-Weight Prototype Contrastive Learning, SWPCL) 方法。该方法首先设计了一个柔性原型 (Soft Prototype, SP) 选择策略, 根据个体特征之间的相似度选择质心原型的最近邻样本作为柔性原型, 为模型提供丰富的正监督信息。为了进一步消除错误原型对模型训练的干扰, 提出了一种柔性加权 (Soft-Weight, SW) 策略, 定量地度量所选的柔性原型相对于当前质心原型的相关性, 将选择到的原型通过柔性加权的方式结合到对比学习中。最后, 引入了一种渐进式对比学习 (Progressive Contrastive Learning, PCL) 策略, 将模型的注意力逐渐转移到柔性原型上, 避免聚类退化。在 SYSU-MM01 和 RegDB 两个公共数据集上的大量实验证明了所提出的柔性加权原型对比学习方法的有效性。
关键词: 可见光-红外; 无监督学习; 行人重识别; 柔性加权; 对比学习

中图分类号: TP391

文献标识码: A

文章编号: 1001-8891(2025)05-0628-07

Soft-Weight Prototype Contrastive Learning for Unsupervised Visible-Infrared Person Re-Identification

ZHAO Hanshuo¹, MA Yiwen², ZHANG Yanxia¹, WANG Pei³, YANG Jianwei⁴

(1. School of Mechanical and Electrical Engineering, Henan Open University, Zhengzhou 450046, China;
2. School of Computer Science, Zhengzhou University of Aeronautics, Zhengzhou 450046, China;
3. School of Architecture and Intelligent Construction, Henan Open University, Zhengzhou 450046, China;
4. School of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

Abstract: Unsupervised visible-infrared person re-identification (USVI-ReID) is a highly important and challenging task. The key difficulty lies in effectively generating pseudo-labels and establishing cross-modality correspondences without relying on any annotations. Recently, generating pseudo-labels using clustering algorithms has attracted increasing attention in USVI-ReID. However, previous methods typically selected a single centroid prototype to represent an individual or randomly selected prototypes based on a fixed strategy for cross-modal correspondence. This approach not only overlooks the diversity of individual characteristics but also fails to account for the negative impact of incorrect samples on model training during clustering. To address these issues, we propose soft-weight prototype contrastive learning (SWPCL). This method first introduces a soft prototype (SP) selection strategy, which selects the nearest neighbor samples of the centroid prototype as the soft prototype based on the similarity between individual features, providing rich positive supervised information to the model. To further eliminate the interference of erroneous prototypes on model

收稿日期: 2024-05-23; 修订日期: 2024-06-27.

作者简介: 赵涵硕 (1995-), 男, 汉族, 河北衡水人, 硕士, 助教。研究方向: 毫米波雷达与机器视觉感知融合。E-mail: zhaohanshuo@yeah.net。

通信作者: 杨建伟 (1995-), 男, 汉族, 江苏南京人, 博士。研究方向: 计算机视觉。E-mail: yangjianwei0913@gmail.com。

基金项目: 河南省重点研发与推广专项 (科技攻关项目) (232102220003; 232102240088; 242102320029)。

training, a soft-weight (SW) strategy is proposed to quantitatively measure the correlation between each selected soft prototype and the corresponding centroid prototype. These prototypes are then integrated into contrastive learning through a soft-weighting mechanism. Finally, a progressive learning strategy is introduced to gradually shift the model's focus toward reliable soft prototypes, thereby avoiding clustering degradation. Extensive experiments on the public SYSU-MM01 and RegDB datasets demonstrate the effectiveness of the proposed SWPCL method.

Key words : visible-infrared, unsupervised learning, person re-identification, soft-weight, contrastive learning

0 引言

行人重识别 (Person Re-Identification, ReID) 是计算机视觉中的一个专门领域, 专注于跨不同摄像机视图或视频序列中的帧识别行人。想象这样一个场景: 一个行人被监控网络中的一台摄像机捕获, 然后移出画面, 之后重新出现在另一台摄像机的视野中。行人重识别旨在正确关联同一个行人的视图, 从而实现跨多个摄像机和位置的无缝跟踪和监控。它广泛应用于智能视频监控、智能安保等领域^[1-3]。然而, 在弱光或夜间环境中, 可见光图像可能不足以进行准确的行人识别^[4]。相比之下, 红外成像不依赖于可见光, 可以捕捉物体发出的热信号, 即使在低能见度情况下也能提供有价值的信息^[5]。随着红外成像技术的发展, 在光线不足的环境中从可见光模式切换到红外模式的智能监控摄像机已经普及, 并推动了监控系统的全天候可见光-红外行人重识别 (Visible-Infrared Person Re-Identification, VI-ReID) 的发展^[6-7]。

VI-ReID 的目标是在给定一种模态的图像时, 在一组可见/红外图像中识别到同一个人^[8-9]。近年来, VI-ReID 备受关注^[10]。许多专注于缓解模态之间巨大语义差距的 VI-ReID 工作已经取得了进展, 根据其不同的对齐方式, 可以将其分为两大类: 图像级对齐和特征级对齐。图像级对齐方法侧重于通过模态翻译减少跨模态差距。一些基于 GAN 的方法通过风格转换对图像进行跨模态对齐^[11-12]。然而, 生成的图像不可避免地含有噪声。因此, X 模态及其衍生方法通过引入中间模态来对齐跨模态图像^[13-15]。主流特征级对齐方法通过寻找模态共享特征空间来最小化跨模态差距^[16-19]。为此, Fang 等人^[20]提出了一种语义对齐和亲和推理框架, 该框架通过共享原型对齐潜在的部分语义特征, 从而允许在两种模态中聚集具有相同语义信息的潜在语义部分特征。Feng 等人^[21]通过正交分解将形状相关特征从形状擦除特征中分离出来, 提高了 VI-ReID 学习表征的多样性和识别性。然而, 这

些方法都需要标注良好的跨模态数据, 而标注数据集是一项耗时耗力的工程, 因此它们不太适用于真实场景中。由于这一限制, 本文试图研究 VI-ReID 的无监督解决方案。

对于无监督可见光-红外行人重识别 (Unsupervised VI-ReID, USVI-ReID), 一些无监督域适应 (Unsupervised Domain Adaptation) 方法使用标注良好的源域进行预训练, 然后将预训练好的模型迁移到目标域来解决 USVI-ReID 任务^[22-23], 但预训练阶段仍然依赖于标注数据。受 Dai 等人^[24]的启发, 一些真正的无监督方法采用对比学习来提高性能表现^[25-28]。这类方法利用聚类算法在齐次空间中产生监督信号, 生成伪标签来构造两种特定于模态的内存库, 一种用于存储可见光图像特征, 另一种用于存储红外图像特征。在训练过程中, 将内存中心作为质心原型, 最大限度地减小查询图像与质心原型之间的对比损失, 并基于相似度聚合相应的质心原型。然而, 伪标签往往是不准确和僵化的。Chen 等人^[29]利用文本信息提供更大的语义监控洞察, 以弥补伪标签的僵化。此外, 可靠的跨模态对应对于 USVI-ReID 至关重要, 因此 Wu 等人^[26]提出了一个渐进的图匹配框架来建立更可靠的跨模态对应。然而, 聚类中心主要呈现每个人的共同信息而缺乏不同个体之间的独特信息, 这导致基于质心原型的对比损失产生类别之间重叠的分类边界, 在遇到困难样本时出现模糊的跨模态对应。Shi 等人^[30]设计了一个硬原型对比学习来挖掘丰富而有意义的信息, 改善不同类别之间的边界。并且在每个聚类中随机抽取样本作为动态原型, 以满足特征的多样性。然而, 由于聚类操作不可避免地将错误样本聚类, 选择的困难原型或动态原型极有可能不属于当前类别, 从而会对模型的特征学习造成干扰。

针对上述 USVI-ReID 方法中存在的不足, 本文在现有研究基础上, 提出了一种柔性加权原型对比学习方法。首先, 设计了一个柔性原型 (Soft Prototype, SP) 选择策略来挖掘个体全面且正确的信

息,指导模型学习更高的鉴别性特征。与传统的对比学习方法不同,柔性原型与质心原型具有较高的相似性且包含自身独特的信息,能够提供更多正监督信号。在此基础上,又设计了一个柔性加权(Soft-Weight, SW)策略,定量地测量所选的柔性原型相对于当前质心原型的相关性,并将柔性加权原型融合到对比学习中,进一步消除错误原型对模型训练的干扰。由于训练初期的聚类结果是不可靠的,因此,最后引入渐进式对比学习(Progressive Contrastive Learning, PCL)策略,使模型将注意力逐

步转移到柔性原型上。

1 本文方法

本文提出的方法的框架如图1所示。首先利用双对比学习(Dual-Contrative Learning, DCL)框架学习模态内判别性,并通过联合模态内对比学习对其进行优化。接着将提出的柔性原型选择与柔性加权策略融入到对比损失中。最后通过渐进式对比学习策略优化网络模型。

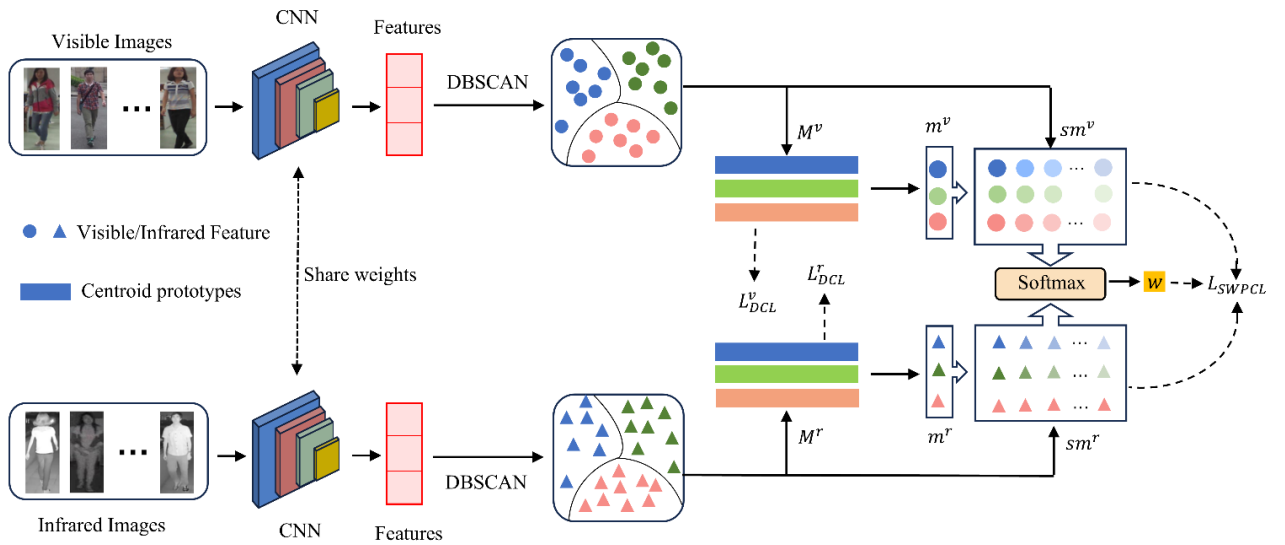


图1 本文方法的结构示意图

Fig.1 Framework of the proposed method

1.1 双对比学习

给定一个可见光-红外训练数据集 $T = \{T^v, T^r\}$, 其中 $T^v = \{x_i^v\}_{i=1}^{N_v}$ 表示 N_v 个可见光图像, $T^r = \{x_i^r\}_{i=1}^{N_r}$ 表示 N_r 个红外图像。利用双流主干网络 f_θ (θ 是网络参数) 分别提取行人的可见光图像的特征集 $F^v = \{f_i^v\}_{i=1}^{N_v}$ 和红外图像的特征集 $F^r = \{f_i^r\}_{i=1}^{N_r}$ 。大多数 USVI-ReID 方法通常使用聚类算法生成伪标签。遵循这一范式, 本文使用 DBSCAN 算法^[31]分别为可见光和红外图像生成伪标签:

$$Y^v = \text{DBSCAN}(F^v) \quad (1)$$

$$Y^r = \text{DBSCAN}(F^r) \quad (2)$$

在每次训练迭代开始时, 通过以下方式计算每个类的表征:

$$m_i^v = \frac{1}{|Y_i^v|} \sum_{f^v \in Y_i^v} f^v \quad (3)$$

$$m_i^r = \frac{1}{|Y_i^r|} \sum_{f^r \in Y_i^r} f^r \quad (4)$$

式中: $|Y_i^v|$ 和 $|Y_i^r|$ 表示第 i 个类的特征数量。将这些类别表征分别存储在可见光内存库 $M^v = \{m_i^v\}_{i=1}^{|Y^v|}$ 和红外内存库 $M^r = \{m_i^r\}_{i=1}^{|Y^r|}$ 中。在训练过程中, 通过动量更新策略更新两种模态的内存库:

$$m_i^v = \lambda m_i^v + (1 - \lambda) f^v, \quad f^v \in Y_i^v \quad (5)$$

$$m_i^r = \lambda m_i^r + (1 - \lambda) f^r, \quad f^r \in Y_i^r \quad (6)$$

式中: $\lambda \sim \text{Unif}(0, 1)$ 是动量更新速率。将内存库中的表征作为质心原型, 通过样本与质心原型的对比学习来优化特征提取器 f_θ , 计算如下:

$$L_{\text{DCL}}^v = -\log \frac{\exp(f^v \cdot m_i^v / \tau)}{\sum_{i=1}^{|Y^v|} \exp(f^v \cdot m_i^v / \tau)} \quad (7)$$

$$L_{\text{DCL}}^r = -\log \frac{\exp(f^r \cdot m_i^r / \tau)}{\sum_{i=1}^{|Y^r|} \exp(f^r \cdot m_i^r / \tau)} \quad (8)$$

$$L_{\text{DCL}} = L_{\text{DCL}}^v + L_{\text{DCL}}^r \quad (9)$$

式中： $|Y^v|$ 和 $|Y^r|$ 代表类别数量； τ 是用来控制数据分布的温度参数； $m_+^{v(r)}$ 代表与查询图像具有相同伪标签的质心原型。DCL 损失函数有助于模型学习模态内判别性和不变性特征。

1.2 柔性加权原型对比学习

为了确保原型有效地捕捉到个体的共性和差异性且减缓由于聚类算法引入的错误样本对模型的干扰，本文设计了一种新的柔性加权原型对比学习方法来提供对个体更加全面、精准的了解。其中，柔性原型选择策略利用网络提取的图像特征，根据特征之间的余弦相似度在每个类别中选择前 K 个最近邻样本作为柔性原型，具体如下：

$$(sm_i^v)_j = NN(m_i^v) = \arg \max_{f^v \in Y_i^v} (\cos(f^v, m_i^v)) \quad (10)$$

$$(sm_i^r)_j = NN(m_i^r) = \arg \max_{f^r \in Y_i^r} (\cos(f^r, m_i^r)) \quad (11)$$

$$L_{SWPCL}^v = -\log \frac{\sum_{i=1}^K w_i^v \cdot \exp(f^v \cdot (sm_+^v)_i / \tau)}{\sum_{i=1}^K w_i^v \cdot \exp(f^v \cdot (sm_+^v)_i / \tau) + \sum_{j=1}^{(|Y^v|-1)} \sum_{k=1}^K \exp(f^v \cdot (sm_j^v)_k / \tau)} \quad (14)$$

$$L_{SWPCL}^r = -\log \frac{\sum_{i=1}^K w_i^r \cdot \exp(f^r \cdot (sm_+^r)_i / \tau)}{\sum_{i=1}^K w_i^r \cdot \exp(f^r \cdot (sm_+^r)_i / \tau) + \sum_{j=1}^{(|Y^r|-1)} \sum_{k=1}^K \exp(f^r \cdot (sm_j^r)_k / \tau)} \quad (15)$$

$$L_{SWPCL} = L_{SWPCL}^v + L_{SWPCL}^r \quad (16)$$

SWPCL 促进灵活和适应性强的学习过程，旨在最大限度地减少样本与其各自的柔性原型之间的差异，而不是严格地将样本与固定的原型或随机选择的原型对齐。

1.3 渐进式对比学习

在初始训练阶段，模型学习到的特征通常不具备良好的判别性，在这个时期引入柔性样本可能会适得其反，将模型优化引向错误的方向。受 Shi 等人^[30]的启发，引入了渐进式对比学习，最终形成了整体损失函数：

$$L_{Total} = \begin{cases} L_{DCL}, & \text{epoch} \leq E_{DCL} \\ L_{SWPCL}, & \text{epoch} > E_{DCL} \end{cases} \quad (17)$$

式中： E_{DCL} 是一个超参数用来决定在 DCL 阶段训练的 epoch 次数。

2 实验

2.1 数据集和评估方法

在两个广泛使用的可见光-红外数据集 SYSU-

式中： $j=1, 2, \dots, K$ ，特征向量在计算前都进行了归一化。柔性原型选择策略不仅能够考虑到个体的全面性，为模型提供更多的正监督信息，而且选择到的原型有很大概率能够正确代表个体特征。同时，为了进一步消除错误原型对模型训练的干扰，本文提出了一种柔性加权策略，定量地测量所选的 K 个柔性原型相对于当前质心原型的相关性，具体如下：

$$w_j^v = \text{Softmax}[(sm_i^v)_j \times (m_i^v)^T] \quad (12)$$

$$w_j^r = \text{Softmax}[(sm_i^r)_j \times (m_i^r)^T] \quad (13)$$

将选择到的原型通过柔性加权的方式结合到对比损失函数中，得到柔性加权原型对比损失函数：

MM01^[32]和 RegDB^[33]上对所提出的方法进行了评估。SYSU-MM01 包含 22257 张可见光和 11909 张近红外图像，由室内和室外环境中的 4 台可见光相机和 2 台红外相机收集。RegDB 是一个较小的数据集，由两个对齐的相机（一个可见光和一个红外）收集，它由 412 个不同身份的行人组成，其中每个身份有 10 张可见光图像和 10 张红外图像。

在这两个数据集上采用累积匹配特征（Cumulative Matching Characteristics, CMC），即 Rank- k 准确率和平均精度均值（mAP）评估本文方法的性能。在 SYSU-MM01 上，有两种不同的测试设置，包括全搜索（All Search）和室内搜索（Indoor Search）模式。图库由可见光图像组成，查询由全搜索模式下的红外图像组成。对于室内搜索模式，丢弃由可见光的室外场景捕获的图像。RegDB 包含两个测试设置，包括可见模式到热模式（Visible to thermal）和热模式到可见模式（Thermal to Visible），并将训练集和测试集随机分割 10 次，报告总体平均结果。

本文方法使用 PyTorch 在 NVIDIA GeForce RTX 4050 上实现，采用 AGW^[34]中的非局部模块增强网

络, 该网络利用在 ImageNet^[35]预训练的 ResNet50^[36]作为特征提取器来输出 2048 维特征。在每个训练 Epoch 中, 随机抽样 16 个行人 ID, 每个 ID 抽取 16 个样本。所有行人图像都被调整为 288×144。使用 Adam 优化器对模型进行训练。权重衰减设置为 5e−4, 初始学习率为 3.5e−4, 每 20 个 Epoch 衰减到之前值的 1/10。前 50 个 Epoch 通过与质心原型的对比损失进行训练, 然后再用柔性加权原型对比损失训练整个网络 50 个 Epoch。在每个训练 Epoch 中, 使用 DBSCAN 对每个模态内的图像进行聚类。DBSCAN 的最大距离在 SYSU 上设置为 0.6, 在 RegDB 上设置为 0.3^[25]。在聚类过程中, 两个数据集的最少聚类数量设置为 4。内存库中的特征更新速率 λ 为 0.1, 柔性原型数量 K 为 3, 温度因子 τ 为 0.05。

2.2 与其它方法的比较

为了全面评价本文的方法, 将其与 13 种监督式 VI-ReID (SVI-ReID) 方法和 6 种无监督式 VI-ReID (USVI-ReID) 方法进行了比较。在 SYSU-MM01 与 RegDB 数据集上的对比结果见表 1。

如表 1 所示, 与最先进的 USVI-ReID 方法相比, 本文的方法具有更优越的性能。例如在 SYSU-MM01

数据集的 All Search 设置下, 本文方法的 Rank-1 和 mAP 的性能指标分别为 66.4%和 60.1%。在其余同等实验设置下, Rank-1 和 mAP 也获得了最好的性能。表 1 中提到的现有 USVI-ReID 方法, 虽然在行人重识别任务上取得了很大的进步, 但忽视了个体信息的多样性以及聚类误差带来的原型选择的错误性, 这将会给模型训练引入分歧。因此, 渐进式柔性原型对比学习方法不仅能最大概率地选择到具有全面和正确信息的样本作为原型, 而且对原型的柔性度量也能减缓错误样本对模型训练的干扰。该方法可以指导模型学习更多的鲁棒性和判别性特征。

与 SVI-ReID 方法的比较, 本文提出的方法优于其中一些最近提出的方法, 如 (Dynamic dual-attentive aggregation, DDAG^[17]), (Attention Generalized mean pooling with Weighted triplet loss, AGW^[34]) 和 (Neural Feature Search, NFS^[37])。这表明与完全依赖数据标注的 SVI-ReID 方法相比, 本文方法同样具有巨大的竞争力。这种现象主要得益于柔性度量的原型在对比学习中发挥的作用。必须承认, 由于缺乏带注释的跨模态对应, 与有监督的方法相比, 无监督方法仍有很大的改进空间。

表 1 SVI-ReID 和 USVI-ReID 方法在 SYSU-MM01 和 RegDB 数据集上结果的比较 (所有方法均采用 Rank-1(%)和 mAP(%)进行度量)

Table 1 Comparisons with state-of-the-art methods on SYSU-MM01 and RegDB, including SVI-ReID and USVI-ReID methods (All methods are measured by Rank-1 (%) and mAP (%))

%									
Settings		SYSU-MM01				RegDB			
		All Search		Indoor Search		Visable to Theamal		Thermal to Visable	
Type	Methods	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Ramk-1	mAP
SVI-ReID	DDAG ^[17]	54.8	53.0	61.0	68.0	69.4	63.5	68.1	61.8
	AGW ^[34]	47.5	47.7	54.2	63.0	70.1	66.4	70.5	65.9
	NFS ^[37]	56.9	55.5	62.8	69.8	80.5	72.1	78.0	69.8
	LbA ^[9]	55.4	54.1	58.5	66.3	74.2	67.6	72.4	65.5
	CAJ ^[10]	69.9	66.9	76.3	80.4	85.0	79.1	84.8	77.8
	MPANet ^[18]	70.6	68.2	76.7	81.0	83.7	80.9	82.8	80.7
	DART ^[8]	68.7	66.3	72.5	78.2	83.6	75.7	82.0	73.8
	FMCNet ^[4]	66.3	62.5	68.2	74.1	89.1	84.4	88.4	83.9
	MAUM ^[16]	71.7	68.8	77.0	81.9	87.9	85.1	87.0	84.3
	DEEN ^[7]	74.7	71.8	80.3	83.3	91.1	85.1	89.5	83.4
	SGIEL ^[21]	77.1	72.3	82.1	83.0	92.2	86.6	91.1	85.2
	PartMix ^[38]	77.8	74.6	81.5	84.4	85.7	82.3	84.9	82.5
	FDNM ^[39]	77.8	75.1	87.3	89.1	95.5	90.0	94.0	88.7
USVI-ReID	OTLA ^[23]	29.9	27.1	29.8	38.8	32.9	29.7	32.1	28.6
	ADCA ^[25]	45.5	42.7	50.6	59.1	67.2	64.1	68.5	63.8
	CCLNet ^[29]	54.0	50.2	56.7	65.1	69.9	65.5	70.2	66.7
	PGM ^[26]	57.3	51.8	56.2	62.7	69.5	65.4	69.9	65.2
	GUR ^[27]	61.0	57.0	64.2	69.5	73.9	70.2	75.0	69.9
	PCLMP ^[30]	64.4	58.7	69.5	74.4	84.3	80.7	82.7	78.4
	Ours	66.4	60.1	70.5	75.6	85.7	82.1	84.2	79.8

Note: The bold fonts represent the best in each category

2.3 消融实验

在本节中，对 SYSU-MM01 数据集进行了全搜索和室内搜索两种模式的消融研究，以展示本文方法中每个组件的有效性，结果如表 2 所示。使用 PGM^[26]作为 Baseline。当在 Baseline 上添加柔性样本 SP 时，Rank-1 和 mAP 的性能提升明显。原因在于使用柔性原型选择策略可以有效地挖掘个体全面且正确的特征，有利于模型学习更多的判别特征。当再加上柔性加权（SW）策略进一步缓解了错误样本对模型训练的干扰，提升了模型的性能。最后，渐进式对比学习（PCL）策略的引入使模型的注意力从单个质心原型平稳地转移到多个柔性原型上。结果证实了渐进式对比学习在柔性原型的选择中发挥了有价值的作用。

表 2 本文方法在 SYSU-MM01 上的消融研究

Methods				All search		Indoor search	
Baseline	SP	SW	PCL	Rank-1	mAP	Rank-1	mAP
✓				57.2	51.7	56.2	62.7
✓	✓			64.5	57.9	65.8	70.6
✓	✓	✓		65.9	59.6	68.3	74.5
✓	✓	✓	✓	66.4	60.1	70.5	75.6

此外，柔性原型选择策略中，原型数量 K 是一个重要的参数。在 SYSU-MM01 数据集上对不同 K 值做消融实验来评估其对模型性能的影响，结果如图 2 所示。当 K 逐渐增大时，Rank-1 不断提升，直到 K 为 3 时效果最佳。这证明了 SP 策略为模型提供了更全面、正确的信息。当 K 继续增大，Rank-1 性能开始下降。其原因在于选择的柔性原型与质心原型的相似度越来越低，对模型的训练造成了一定的干扰。

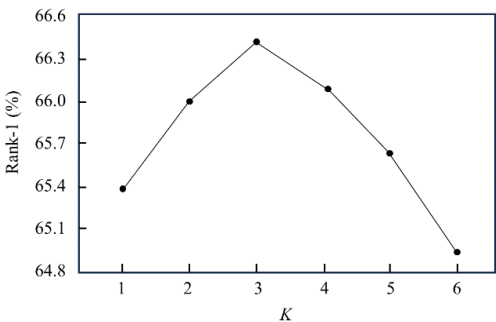


图 2 不同 K 值对模型性能的影响

Fig.2 Effects of different K values on model performance

3 结论

本文提出了一种基于柔性加权原型对比学习的无监督可见光-红外行人重识别方法。首先，设计了一个柔性原型选择策略来挖掘个体全面且正确的信息，指导模型学习更高的鉴别性特征。在此基础

上，又设计了一个柔性加权策略，定量地测量所选的柔性原型相对于当前质心原型的相关性，并将柔性加权原型融合到对比学习中，进一步消除错误原型对模型训练的干扰。最后引入渐进式对比学习策略，使模型将注意力逐步转移到柔性原型上。在 SYSU-MM01 和 RegDB 公共数据集上的大量实验证明了所提出的柔性加权原型对比学习方法的有效性。

参考文献：

[1] GE W, PAN C, WU A, et al. Cross-camera feature prediction for intra-camera supervised person re-identification across distant scenes[C]//*Proceedings of the 29th ACM International Conference on Multimedia*, 2021: 3644-3653.

[2] 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*, 2019: 6112-6121.

[3] GUO J, YUAN Y, HUANG L, et al. Beyond human parts: Dual part-aligned representations for person re-identification[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 3642-3651.

[4] ZHANG Q, LAI C, LIU J, et al. Fmcnet: Feature-level modality compensation for visible-infrared person re-identification[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 7349-7358.

[5] 杨瑞宇, 吕浩, 龚晓霞, 等. 红外成像系统中自动对焦功能优化方法研究[J]. *红外技术*, 2020, 42(10): 940-946.

YANG Ruiyu, LYV Hao, GONG Xiaoxia, et al. Optimizing method of autofocusing function in infrared imaging system[J]. *Infrared Technology*, 2020, 42(10): 940-946.

[6] YANG B, CHEN J, YE M. Top-k visual tokens transformer: Selecting tokens for visible-infrared person re-identification[C]//*ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023: 1-5.

[7] ZHANG Y, WANG H. Diverse embedding expansion network and low-light cross-modality benchmark for visible-infrared person re-identification[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 2153-2162.

[8] YANG M, HUANG Z, HU P, et al. Learning with twin noisy labels for visible-infrared person re-identification[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 14308-14317.

[9] Park H, Lee S, Lee J, et al. Learning by aligning: Visible-infrared person re-identification using cross-modal correspondences[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 12046-12055.

[10] YE M, RUAN W, DU B, et al. Channel augmented joint learning for visible-infrared recognition[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 13567-13576.

[11] WANG G, ZHANG T, CHENG J, et al. RGB-infrared cross-modality

- person re-identification via joint pixel and feature alignment[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 3623-3632.
- [12] YANG Y, ZHANG T, CHENG J, et al. Cross-modality paired-images generation and augmentation for RGB-infrared person re-identification[J]. *Neural Networks*, 2020, **128**: 294-304.
- [13] LI D, WEI X, HONG X, et al. Infrared-visible cross-modal person re-identification with an x modality[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, **34**(4): 4610-4617.
- [14] ZHANG Y, YAN Y, LU Y, et al. Towards a unified middle modality learning for visible-infrared person re-identification[C]//*Proceedings of the 29th ACM International Conference on Multimedia*, 2021: 788-796.
- [15] WEI Z, YANG X, WANG N, et al. Syncretic modality collaborative learning for visible infrared person re-identification[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 225-234.
- [16] LIU J, SUN Y, ZHU F, et al. Learning memory-augmented unidirectional metrics for cross-modality person re-identification[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 19366-19375.
- [17] YE M, SHEN J, J. Crandall D, et al. Dynamic dual-attentive aggregation learning for visible-infrared person re-identification[C]//*Computer Vision-ECCV*, 2020: 229-247.
- [18] WU Q, DAI P, CHEN J, et al. Discover cross-modality nuances for visible-infrared person re-identification[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 4330-4339.
- [19] SUN H, LIU J, ZHANG Z, et al. Not all pixels are matched: Dense contrastive learning for cross-modality person re-identification[C]//*Proceedings of the 30th ACM International Conference on Multimedia*, 2022: 5333-5341.
- [20] FANG X, YANG Y, FU Y. Visible-infrared person re-identification via semantic alignment and affinity inference[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023: 11270-11279.
- [21] FENG J, WU A, ZHENG W S. Shape-erased feature learning for visible-infrared person re-identification[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 22752-22761.
- [22] LIANG W, WANG G, LAI J, et al. Homogeneous-to-heterogeneous: Unsupervised learning for RGB-infrared person re-identification[J]. *IEEE Transactions on Image Processing*, 2021, **30**: 6392-6407.
- [23] WANG J, ZHANG Z, CHEN M, et al. Optimal transport for label-efficient visible-infrared person re-identification[C]//*European Conference on Computer Vision*, 2022: 93-109.
- [24] DAI Z, WANG G, YUAN W, et al. Cluster contrast for unsupervised person re-identification[C]//*Proceedings of the Asian Conference on Computer Vision*, 2022: 1142-1160.
- [25] 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*, 2022: 2843-2851.
- [26] WU Z, YE M. Unsupervised visible-infrared person re-identification via progressive graph matching and alternate learning[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 9548-9558.
- [27] YANG B, CHEN J, YE M. Towards grand unified representation learning for unsupervised visible-infrared person re-identification[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023: 11069-11079.
- [28] YANG B, CHEN J, CHEN C, et al. Dual Consistency-Constrained Learning for Unsupervised Visible-Infrared Person Re-Identification[J]. *IEEE Transactions on Information Forensics and Security*, 2023(19): 1767-1779.
- [29] CHEN Z, ZHANG Z, TAN X, et al. Unveiling the power of clip in unsupervised visible-infrared person re-identification[C]//*Proceedings of the 31st ACM International Conference on Multimedia*, 2023: 3667-3675.
- [30] SHI J, YIN X, WANG Y, et al. Progressive Contrastive Learning with Multi-Prototype for Unsupervised Visible-Infrared Person Re-identification[J]. arXiv preprint arXiv:2402.19026, 2024.
- [31] Ester M, Kriegl H P, Sander J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]//*KDD*, 1996, **96**(34): 226-231.
- [32] WU A, ZHENG W S, YU H X, et al. RGB-infrared cross-modality person re-identification[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2017: 5380-5389.
- [33] 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.
- [34] YE M, SHEN J, LIN G, et al. Deep learning for person re-identification: A survey and outlook[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, **44**(6): 2872-2893.
- [35] DENG J, DONG W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//2009 *IEEE conference on Computer Vision and Pattern Recognition*. IEEE, 2009: 248-255.
- [36] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770-778.
- [37] CHEN Y, WAN L, LI Z, et al. Neural feature search for rgb-infrared person re-identification[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 587-597.
- [38] KIM M, KIM S, Park J, et al. Partmix: Regularization strategy to learn part discovery for visible-infrared person re-identification[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 18621-18632.
- [39] ZHANG Y, LU Y, YAN Y, et al. Frequency domain nuances mining for visible-infrared person re-identification[J]. arXiv preprint arXiv:2401.02162, 2024.