

多粒度跨模态行人重识别算法研究

熊炜^{1,2*}, 乐玲¹, 周蕾¹, 张开¹, 李利荣¹

¹湖北工业大学电气与电子工程学院, 湖北 武汉 430068;

²美国南卡罗来纳大学计算机科学与工程系, 南卡罗来纳州 哥伦比亚 29201

摘要 针对当前跨模态行人重识别算法大多聚类能力不强、且难以提取高效判别性特征的问题,提出了一种多粒度跨模态行人重识别算法。首先,在骨干网络 Resnet50 中加入非局部注意力机制模块,关注长距离像素之间的关系,保留细节信息;其次,采用多分支网络提取不同细粒度特征信息,增强模型的判别性特征提取能力;最后,联合基于样本的三元组损失和基于中心的三元组损失监督训练,加速模型收敛。所提算法在 SYSU-MM01 数据集的全搜索模式下 Rank-1 和 mean average precision 分别达到 62.83% 和 58.10%,在 RegDB 数据集的可见光到红外模式下 Rank-1 和 mAP 分别达到 87.78% 和 76.22%。

关键词 光计算; 跨模态; 行人重识别; 注意力机制; 多粒度; 三元组损失

中图分类号 TP183

文献标志码 A

DOI: 10.3788/LOP202259.2220001

Multi-Granularity and Cross-Modality Pedestrian Re-Identification Algorithm

Xiong Wei^{1,2*}, Yue Ling¹, Zhou Lei¹, Zhang Kai¹, Li Lirong¹

¹School of Electrical & Electronic Engineering, Hubei University of Technology, Wuhan 430068, Hubei, China;

²Dept. of Computer Science & Engineering, University of South Carolina, Columbia 29201, SC, USA

Abstract Most current cross-modal pedestrian re-identification algorithms lack clustering ability and make it difficult to extract high-efficiency discriminative features; therefore, this paper proposes a multi-granular cross-modal pedestrian re-identification algorithm. First, the nonlocal attention mechanism module is added to the backbone network Resnet50 to focus on the relationship between long-distance pixels and retain detailed information. Second, a multi-branch network is used to extract fine-grained feature information to improve the distinguishing feature extraction ability of the model. Finally, the sample- and center-based triple losses are combined to supervise the training process, which achieves the purpose of accelerating the convergence of the model. The proposed method achieves Rank-1 and mean average precision of 62.83% and 58.10%, respectively, in the full search mode of the SYSU-MM01 dataset. In the visible-to-infrared mode of the RegDB dataset, Rank-1 and mAP reach 87.78% and 76.22%, respectively.

Key words optics in computing; cross-modality; pedestrian re-identification; attention mechanism; multi-granularity; triple loss

1 引言

行人重识别是根据查询图在图库中正确匹配到该行人其他图像的过程,现有的行人重识别算法,如罗浩等^[1]、张涛等^[2]、李聪等^[3]、卢健等^[4]侧重于单模态的研究,也就是 RGB-RGB 的检索问题,然而这些模型针对的仅是白天光照条件充足的摄像头获取的图像。日常生活中,普通摄像头在恶劣的光照环境下无法捕获具有

判别性特征的行人图像,带有红外功能的摄像头却能在夜晚捕获缺乏行人细节特征的近红外(IR)图像。因此,如何将具有三通道的 RGB 图和单通道的近红外图像进行匹配,在行人重识别研究中具有重大研究意义。

跨模态的行人重识别任务中难点在于如何减小 RGB 图和近红外图像的模态差异。RGB 图含有丰富的行人语义和颜色信息,而近红外图却没有,模态间的差异使得传统的单模态行人重识别方法无法在跨模态

收稿日期: 2021-08-23; 修回日期: 2021-09-09; 录用日期: 2021-09-28

基金项目: 国家自然科学基金(61571182, 61601177)、湖北省自然科学基金(2019CFB530)、湖北省科技厅重大专项(2019ZYYD020)、国家留学基金(201808420418)

通信作者: *xw@mail.hbut.edu.cn

行人重识别中取得理想的效果。

为解决上述问题,Nguyen^[5]等提出了第 1 个 RGB-Infrared 数据集,并提出了一种深度零填充(Zero-padding)的方法来训练单流网络,使之能够自动演化出特定领域的节点来进行跨模态匹配,并采用交叉熵损失(CE)监督训练过程。Ye等^[6]提出了双流网络提取模态差异性全局特征,同时将两个不同模态的特征转换为同一空间进行度量学习。然而全局特征对背景信息较为敏感,无法解决模态差异性,因此Zhu等^[7]利用部分特征对齐不同模态的相同部位特征,并提出异质中心损失(HC)监督训练过程。Dai等^[8]、Yang等^[9]等则采用生成对抗网络(GAN)技术来生成跨模态的行人图像,以减少图像和特征层面的跨模态差异。Li等^[10]生成了介于两种模态间的一种中间模态,称之为 X 模态,使得 3 种模态更容易匹配。

当前跨模态行人重识别方法,难以提取高效的行人区分特征,且难以保证类内特征的紧凑性。针对上述问题,本文提出了一种全新的多粒度跨模态行人重识别网络。首先采用非局部注意力机制(Non_local attention)^[11]建立间隔一定距离像素之间的联系;其次针对共享模态差异性特征提取能力不足的问题,采用

多粒度网络进行提取,实现不同级别的特征提取;最后采用异质中心三元组损失^[12]监督学习具有类间区分性和类内紧凑性的高质量行人特征。

2 所提算法原理

所提算法框图如图 1 所示。采用 Resnet50 作为骨干网络,前两层特定模态模块提取差异性特征,后三层共享模态模块提取共享特征,在共享网络中加入非局部注意力模块,获得所有位置的特征加权和作为输出响应,增强网络对高质量特征的提取能力。通过不同池化方式构建 3 个分支关注不同粒度的特征。细粒度特征 Branch 1 将特征图水平切分为 6 块,采用局部权重加权达到筛选出行人辨别性特征的目的,后与细粒度 Branch 2 实现特征融合;一方面可以减少噪声的影响,另一方面可以稳定并增强训练过程;细粒度全局分支 Branch 2 关注全局特征的学习;粗粒度全局分支 Branch 3 经细粒度全局分支 Branch 2 监督,后通过批量归一化(BN)层实现损失函数的同步收敛,约束特征。使用基于中心的三元组损失监督训练过程。通过样本中心代替样本,在降低计算量的同时,大幅度拉近类内距离,聚类能力更好。

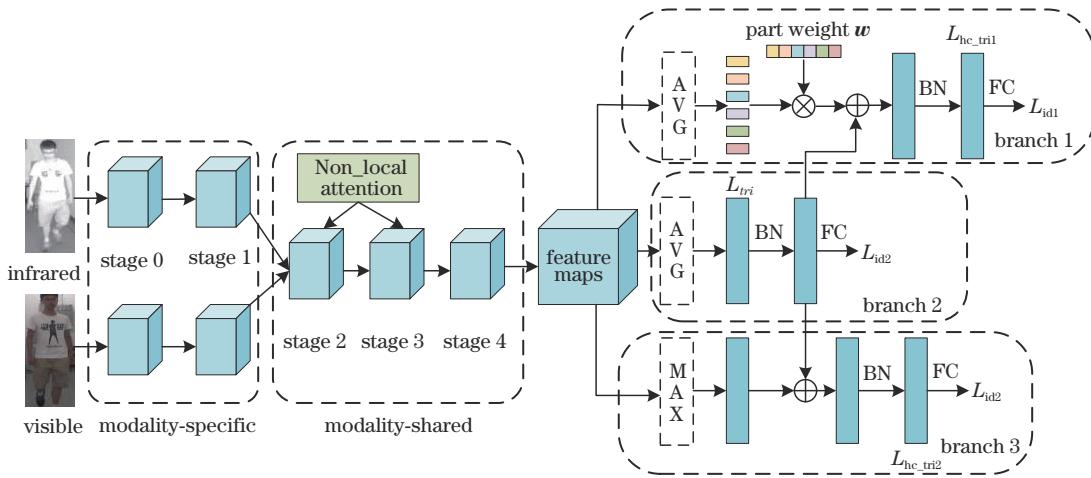


图 1 所提算法框图

Fig. 1 Block diagram of proposed algorithm

2.1 非局部注意力模块

卷积操作只关注特征图的局部区域,本实验组则采用非局部注意力模块来捕获长距离依赖,以此建立图像上两个有一定距离的像素之间的联系,图 2 为非局部注意力模块结构图。输入特征图 X 维度为 (N, C, H, W) ,依次代表批次大小、通道数、高和宽,该输入分别经过 3 个 1×1 卷积,得到 3 个输出大小为 $(N, C/2, H, W)$ 的特征分支,该操作通过降低通道数,减少计算量。由图 2 中 θ 和 Φ 两个分支得到的特征图的维度分别为 $(N, C/2, H \times W)$ 、 $(N, H \times W, C/2)$,所得两个特征图经矩阵相乘和归一化操作后与 g 分支再进行矩阵相乘,得到维度为 $(N, C/2, H, W)$ 的特征图,

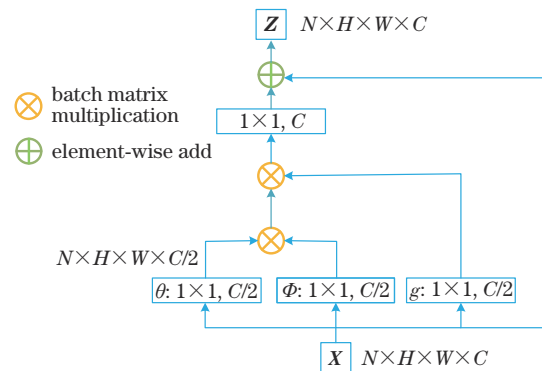


图 2 非局部注意力模块

Fig. 2 Non_local attention module

并通过 1×1 卷积升维,最后再与输入残差模块进行逐点相加得到输出 Z 。本实验组在骨干网络 stage 2 中增加 2 个非局部注意力模块,在 stage 3 中增加 3 个。具体可描述为

$$Z_i = W_z * \phi(X_i) + X_i, \quad (1)$$

式中: W_z 是待学习的权重矩阵; $\phi(\cdot)$ 是非局部操作; $+X_i$ 是残差学习策略; $*$ 表示矩阵点乘。

2.2 细粒度局部权重分支

受文献[13-14]启发,本实验组将图 1 中 Branch 1 的特征图经池化后水平切分成 6 块,随后通过一个自适应的权重向量 $w = (w_1, \dots, w_g, \dots, w_6)^T$ 来解决不同分块的模态差异性,最后通过相加与 Branch 2 中的残差结构融合,聚合注意力集中的特征来增强辨别能力。

$$f_{\text{Branch 1}}^k = \text{BN}(f_{\text{Branch 2}}^k) + \sum_{g=1}^6 w_g f_{\text{Branch 1}}^{k,g}, \quad (2)$$

式中: $f_{\text{Branch 1}}^k$ 为 Branch 1 的第 k 张特征图的特征向量; $\text{BN}(\cdot)$ 是批量归一化; w_g 是每一块的特征权重; $f_{\text{Branch 1}}^{k,g}$ 是第 k 张特征图的第 g 块特征。

2.3 损失函数

2.3.1 跨模态样本三元组损失

跨模态行人重识别任务中,在每次迭代中挑选 P 个行人,每个行人随机挑选 K 张可见光图像和 K 张红外图像,每次迭代总共获得 $2 \times P \times K$ 张行人图像。三元组损失从训练集中随机选取一个锚点(Anchor),再随机选取一个同类的正样本和不同类的负样本,由此构成一个三元组。通过难例样本挖掘^[15]从 $2 \times P \times K$ 张图像中找到最难的正样本和最难负本来计算难例三元组损失。常用跨模态行人重样本三元组损失的表达式为

$$L_{\text{tri}} = \sum_{i=1}^P \sum_{a=1}^{2K} \left[m + \max_{p=1, \dots, 2K} \|x_a^i - x_p^i\|_2 - \min_{\substack{n=1, \dots, 2K \\ j \neq i}} \|x_a^i - x_n^j\|_2 \right]_+, \quad (3)$$

式中: x_a^i 、 x_p^i 、 x_n^j 分别代表第 i 个行人锚点、正样本、负样本;折叶损失 $[x]_+ = \max(x, 0)$; $\|\cdot\|_2$ 代表欧氏距离; m 代表三元组损失的边缘值(Margin)。

2.3.2 跨模态中心三元组损失

三元组损失通过计算锚点到其他样本的距离来计算损失,中心损失使用样本的中心代表样本,通过样本中心的距离计算中心损失。所采用的异质中心三元组损失综合三元组损失和中心损失,通过锚点中心到其他样本中心的距离来计算度量损失。在一个批次中,各模态的中心距离公式为

$$\begin{cases} c_v^i = \frac{1}{K} \sum_{j=1}^K v_j^i \\ c_t^i = \frac{1}{K} \sum_{j=1}^K t_j^i \end{cases}, \quad (4)$$

式中: v_j^i 是第 i 个行人的第 j 张可见光图像; t_j^i 是第 i 个行人的第 j 张近红外图像; c_v^i 代表第 i 个行人的可见光图像中心; c_t^i 代表第 i 个行人的近红外图像中心。

度量学习的目的是缩小类内距离、扩大类间距离,结合中心距离计算公式,所用异质中心三元组损失可描述为

$$L_{\text{hc,tri}} = \sum_{i=1}^P \left[m_c + \|c_v^i - c_t^i\|_2 - \min_{\substack{n \in \{v, t\} \\ j \neq i}} \|c_v^i - c_n^j\|_2 \right]_+ + \sum_{i=1}^P \left[m_c + \|c_t^i - c_v^i\|_2 - \min_{\substack{n \in \{v, t\} \\ j \neq i}} \|c_t^i - c_n^j\|_2 \right]_+. \quad (5)$$

异质中心三元组损失一方面可以减小难例样本挖掘的计算量,另一方面还可以确保类内特征的紧凑性和类间特征的可区分性,在跨模态任务中效果较佳。

在梯度下降过程中,输入的是样本特征向量,输出是网络参数 $W = \{W_k\}$ 。首先通过正向传播,计算出 3 个样本中心 c_v^i 、 c_t^i 、 c_n^j 的值,其中 c_n^j 是属于第 j 个行人的可见光或红外图像,然后通过反向传播计算出 3 个样本中心关于网络参数 W_k 的偏导,再根据链式求导法则,有

$$\frac{\partial L_{\text{hc,tri}}}{\partial W_k} = \frac{\partial L_{\text{hc,tri}}}{\partial c_v^i} \frac{\partial c_v^i}{\partial W_k} + \frac{\partial L_{\text{hc,tri}}}{\partial c_t^i} \frac{\partial c_t^i}{\partial W_k} + \frac{\partial L_{\text{hc,tri}}}{\partial c_n^j} \frac{\partial c_n^j}{\partial W_k}. \quad (6)$$

根据式(6),更新网络参数的可描述为

$$W_k^{t+1} = W_k^t - \lambda_t \frac{\partial L_{\text{hc,tri}}}{\partial W_k}, \quad (7)$$

式中: λ_t 是学习率; W_k^t 是权重的初始值; W_k^{t+1} 是更新后的权值。梯度下降过程中,多次重复式(7),直至损失收敛。

2.3.3 分类损失

表征学习采用带有标签平滑的交叉熵损失防止模型过拟合:

$$\begin{cases} L_{\text{id}} = \sum_{i=1}^N -q_i \log p_i \\ q_i = \begin{cases} 1 - \frac{N-1}{N} \cdot \xi, & y = i, \\ \frac{\xi}{N}, & y \neq i \end{cases} \end{cases}, \quad (8)$$

式中: y 为图像的真实标签; p_i 为预测值; N 为行人数目; ξ 为容错率,以 $1 - \xi$ 作为真实标签进行训练,实验中设置 $\xi = 0.1$ 。

2.3.4 总体损失

这 3 个特征分支分别采用不同损失函数监督训

练,设置样本三元组损失的边缘值 $m = 0.3$,在 RegDB 数据集^[5]中设置中心三元组损失边缘值 $m_c = 0.3$,SYSU-MM01^[16]中设置 $m_c = 0.8$,参数设置参考基准线(Baseline)DGTL^[17]。总体损失的表达式为

$$L_{\text{total}} = L_{\text{id}} + L_{\text{tri}} + L_{\text{hc_tri}}, \quad (9)$$

$$L_{\text{id}} = L_{\text{id1}} + L_{\text{id2}}, \quad (10)$$

$$L_{\text{hc_tri}} = L_{\text{hc_tri1}} + L_{\text{hc_tri2}} \quad (11)$$

3 实验结果与分析

3.1 数据集和评价指标

在 RegDB 和 SYSU-MM01 两个公开的数据集上进行了实验。RegDB 数据集包含 412 个行人 ID, 总共包含 4120 张 RGB 图像和 4120 张 IR 图像, 该数据集图像源于 2 个摄像头, 包括 1 个在白天工作的普通摄像头和 1 个在夜晚工作的 IR 摄像头。训练时, 随机挑选 206 个行人 ID, 总计包含 RGB、IR 图像各 2060 张; 测试

时, 随机挑选 206 个行人 ID, 总计包含 RGB、IR 图像各 2060 张。测试模式分为 RGB-IR 和 IR-RGB, 本实验组采取两种模式。

SYSU-MM01 数据集包含 491 个行人 ID, 总共包含 287628 张 RGB 图像和 15793 张 IR 图像, 如图 3 所示。该数据集图像源于 6 个摄像头, 包括 4 个在白天工作的普通摄像头和 2 个在夜晚工作的 IR 摄像头, 3 个摄像头安置在室内, 3 个安置在室外。训练时, 挑选 395 个行人 ID, 总计包含 22258 张 RGB 图像和 11909 张 IR 图像; 测试时, 挑选剩余的 96 个行人 ID, 其中查询图集包含 3803 张 IR 图像, 图库集仅包含可见光 RGB 图像。测试模式分为两种: 一种是室内检索(Indoor-search)模式, 仅包含 3 个室内摄像头数据; 一种是全检索(All-search)模式, 包含 6 个摄像头数据。全检索模式更加接近于真实场景的行人重识别, 难度也更大, 本实验组选用全搜索模式和室内搜索, 并设置了 Single-shot。

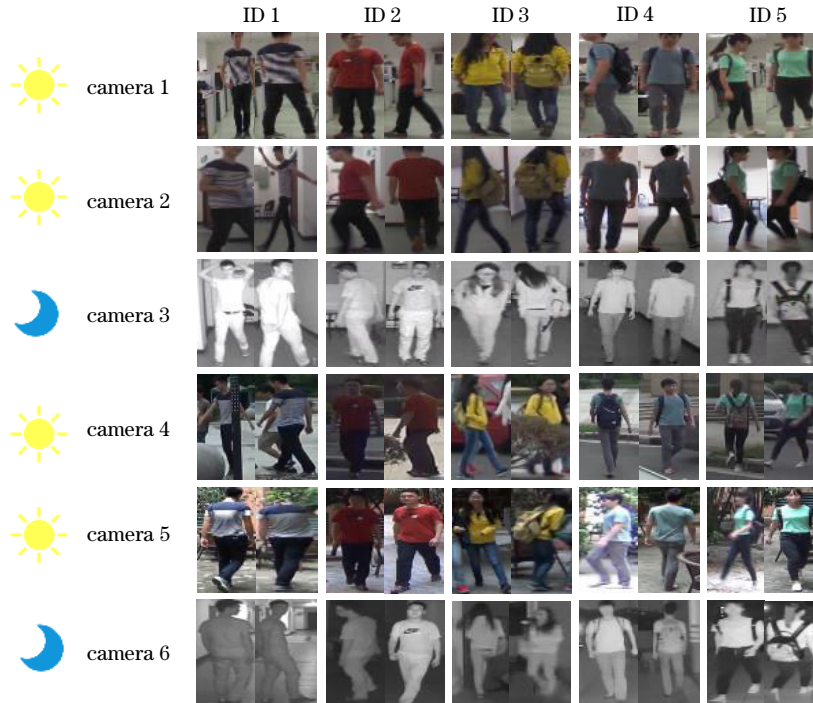


图 3 SYSU-MM01 数据集

Fig. 3 SYSU-MM01 dataset

在测试阶段, 采用 Rank-1、mean average precision (mAP)、mean inverse negative penalty (mINP) 作为评价指标。Rank-1 识别率代表了识别率排第 1 的图像是该行人的准确率, mAP 将行人重识别看作检索任务, 评价指标计算方式参考文献^[5], mINP 体现了检测出所有正确匹配项的工作量, 3 个评价指标数值越高, 代表算法效果越好。此外所有行人特征测试前都经了 L2 正则化。

3.2 实验过程

实验环境如下: 显卡 NVIDIA GeForce RTX 2070 (显存 8 GB)、CPU 内存 16 GB; 在 Ubuntu18.04 的系统上采用 Python3.6 的 Pytorch 深度学习框架。

训练时, 最后 1 层卷积层的跨度设置为 1, 输入行人图像调整为 288×144 , 通过镜像翻转和裁剪等操作进行数据增强。根据采样策略, 设置 $P=4$ 、 $K=8$, 总共训练 60 次, 采用随机梯度下降优化器 (SGD), 动量设置为 0.9, 采用预热的学习策略, 学习率设置如下:

$$R_{ir}(t) = \begin{cases} 0.1 \times \frac{t+1}{10}, & 0 \leq t < 10 \\ 0.1, & 10 \leq t < 20 \\ 0.01, & 20 \leq t < 50 \\ 0.001, & 50 \leq t \end{cases} \quad (12)$$

3.3 实验结果

3.3.1 消融实验

本小节在基准线的基础上依次增加注意力机制模块和 Branch 1 的细粒度局部权重模块来验证各个模块的有效性。消融实验在 SYSU-MM01 数据集上进行, 设置全搜索和室内搜索两种模式, 且均为 Single-shot, 结果如表 1 所示。从表 1 可以看出: 与基准线相比, 仅添加注意力机制 Non_local 模块时, 全搜索模式下 Rank-1 和 mAP 上涨了 3.82 个百分点和 1.9 个百分点, 室内搜索模式下分别上涨了 4.37 个百分点和 3.02 个百分点; 在此基础上继续添加 Branch 1 后, 全搜索模式下 Rank-1 和 mAP 又分别上涨了 1.67 个百分点、1.07 个百分点, 室内搜索模式下上涨了 1.41 个百分点、1.25 个百分点。当两个模块共同使用时, 在全搜索模式下, 评价指标较基准线分别上涨了 5.49 个百分点和 2.97 个百分点; 在室内搜索模式下, 评价指标较基准

表 1 消融实验

Table 1 Ablation experiment unit: %

Method	All Search		Indoor Search	
	Rank-1	mAP	Rank-1	mAP
Baseline	57.34	55.13	63.11	69.20
Baseline+Non_local	61.16	57.03	67.48	72.22
Baseline+Non_local+Branch 1	62.83	58.10	68.89	73.47

表 2 SYSU-MM01 数据集上比较结果

Table 2 Comparison results on SYSU-MM01 dataset

unit: %

Method	All Search			Indoor Search		
	Rank-1	mAP	mINP	Rank-1	mAP	mINP
One-stream ^[18] (ICCV2017)	12.04	13.67		16.94	22.95	
Two-stream ^[18] (ICCV2017)	11.65	12.85		15.60	21.49	
Zero-padding ^[18] (ICCV2017)	14.80	15.95		20.58	26.92	
HCML ^[6] (AAAI2018)	14.32	16.16		24.52	30.08	
cmGAN ^[8] (IJCAI2018)	26.97	31.49		31.63	42.19	
BDTR ^[19] (TIFS2019)	27.32	27.32		31.92	41.86	
D ² RL ^[20] (CVPR2019)	28.90	29.20				
MAC ^[21] (TIP2020)	33.26	36.22		36.43	37.03	
AlignGAN ^[22] (ICCV2019)	42.40	40.70		45.90	54.30	
Xmodal ^[10] (AAAI2020)	49.92	50.73				
DDAG ^[23] (ECCV2020)	54.75	53.02		61.20	67.98	
AGW ^[13] (TPAMI2021)	47.50	47.65	35.30	54.17	62.97	59.23
DF ² AM ^[14] (Arxiv2021)	56.93	55.10		66.39	71.52	
DGTL ^[17] (ISPL2021)	57.34	55.13		63.11	69.20	
Proposed method	62.83	58.10	45.62	68.89	73.47	68.06

线分别上涨了 5.78 个百分点和 4.27 个百分点。实验结果表明, 所提每个模块在最后的实验结果中都有重要作用。

3.3.2 结果比对

将所提方法与当前主流的跨模态行人重识别方法进行了比对, 包括 One-stream^[18]、Two-Stream^[18]、Zero-padding^[18]、HCML^[6]、cmGAN^[8]、BDTR^[19]、D²RL^[20]、MAC^[21]、AlignGAN^[22]、Xmodal^[10]、DDAG^[23]、AGW^[13]和 DF²AM^[14]。表 2 和表 3 分别为在数据集 SYSU-MM01 和 RegDB 上的比较结果。

从表 2 可以看出: 在 SYSU-MM01 数据集中所提方法在多个评价指标上均体现了优越性, 在全搜索模式下所提方法较基准线在 Rank-1 和 mAP 上分别上涨 5.49 个百分点和 2.97 个百分点, mINP 较 AGW 上涨了 10.32 个百分点; 在室内搜索模式下所提方法较基准线在 Rank-1 和 mAP 上分别涨了 5.78 个百分点和 4.27 个百分点, mINP 较 AGW 上涨了 8.83 个百分点。DDAG 采用的骨干网络只关注了行人全局特征, 缺乏对多粒度以及局部特征的考量, 导致识别率不高。所提方法采用的多粒度网络, 提取多粒度辨别性特征, 增强了对细节信息的关注。AGW 采用基于样本的三元组损失监督训练, 该损失函数计算的是锚点到其他所有样本的距离, 这个强力的约束同时也将异常的三元组包括在内, 导致效果不够理想。而所提方法的中心三元组损失函数采用样本间中心距离代替样本间的距离, 弱化了该约束, 在大大减少计算量的同时, 实现了样本间的聚类。

从表 3 可以看出: 在可见光到红外的模式下, 所提方法较基准线在 Rank-1 和 mAP 上分别涨了 3.86 个百

表 3 RegDB 数据集上比较结果
Table 3 Comparison results on RegDB dataset

unit: %

Method	Visible to Infrared			Infrared to Visible		
	Rank-1	mAP	mINP	Rank-1	mAP	mINP
HCML ^[6] (AAAI2018)	24.44	20.08		21.70	22.24	
Zero-padding ^[18] (ICCV2017)	17.75	18.90		16.63	17.82	
BDTR ^[19] (TIFS2019)	33.56	32.76		32.92	31.96	
D ² RL ^[20] (CVPR2019)	43.40	44.10				
MAC ^[21] (TIP2020)	36.43	37.03		36.20	36.63	
AlignGAN ^[22] (ICCV2019)	57.90	53.60		56.30	53.40	
Xmodal ^[10] (AAAI2020)	62.21	60.18				
DDAG ^[23] (ECCV2020)	69.34	63.46		68.06	61.80	
AGW ^[13] (TPAMI2021)	70.05	66.37	50.19			
DF ² AM ^[14] (Arxiv2021)	73.06	67.81		70.49	63.85	
DGTL ^[17] (ISPL2021)	83.92	73.78		81.59	71.65	
Proposed method	87.78	76.22	58.48	84.97	73.81	

分点和 2.44 个百分点;在红外到可见光的模式下,所提方法较基准线在 Rank-1 和 mAP 上分别涨了 3.38 个百分点和 2.16 个百分点。

4 结 论

提出了一种多粒度的跨模态行人重识别算法,在骨干网络中加入注意力机制,提取高质量共享模态信息,随后通过不同池化方式将特征图分成 3 个不同的分支,提取不同细粒度的辨别性特征,最后通过三元组损失和异质中心三元组损失监督学习,加速模型收敛,提高模型精度。实验结果表明,所提算法较为优越。

参 考 文 献

- [1] 罗浩,姜伟,范星,等.基于深度学习的行人重识别研究进展[J].自动化学报,2019,45(11):2032-2049.
Luo H, Jiang W, Fan X, et al. A survey on deep learning based person re-identification[J]. Acta Automatica Sinica, 2019, 45(11): 2032-2049.
- [2] 张涛,易争明,李璇,等.一种基于全局特征的行人重识别改进算法[J].激光与光电子学进展,2020,57(24):241503.
Zhang T, Yi Z M, Li X, et al. Improved algorithm for person re-identification based on global features[J]. Laser & Optoelectronics Progress, 2020, 57(24): 241503.
- [3] 李聪,蒋敏,孔军.基于多尺度注意力机制的多分支行人重识别算法[J].激光与光电子学进展,2020,57(20):201001.
Li C, Jiang M, Kong J. Multi-branch person re-identification based on multi-scale attention[J]. Laser & Optoelectronics Progress, 2020, 57(20): 201001.
- [4] 卢健,陈旭,罗毛欣,等.深度学习行人再识别研究综述[J].激光与光电子学进展,2020,57(16):160003.
Lu J, Chen X, Luo M X, et al. Person re-identification research via deep learning[J]. Laser & Optoelectronics Progress, 2020, 57(16): 160003.
- [5] 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.

- [6] Ye M, Lan X, Li J, et al. Hierarchical discriminative learning for visible thermal person re-identification[C]// Proceedings of the 32nd AAAI Conference on Artificial Intelligence, AAAI 2018, February 2-7, 2018, New Orleans, LA, United states. Virginia: AIAA Press, 2018: 7501-7508.
- [7] Zhu Y X, Yang Z, Wang L, et al. Hetero-Center loss for cross-modality person re-identification[J]. Neurocomputing, 2020, 386: 97-109.
- [8] Dai P, Ji R, Wang H, et al. Cross-modality person re-identification with generative adversarial training[C]// Proceedings of the 27th International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden. [S.l.: s.n.], 2018: 677-683.
- [9] Yang J H, Ruan D Y, Huang J W, et al. An embedding cost learning framework using GAN[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 839-851.
- [10] Li D G, Wei X, Hong X P, et al. Infrared-visible cross-modal person re-identification with an X modality[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(04): 4610-4617.
- [11] Wang X, Girshick R, Gupta A, et al. Non-local neural networks[C]//Proceedings of the 31st Meeting of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2018, June 18-22, 2018, Salt Lake City, UT, United states. New York: IEEE Press, 2018: 7794-7803.
- [12] Liu H J, Tan X H, Zhou X C. Parameter sharing exploration and hetero-center triplet loss for visible-thermal person re-identification[J]. IEEE Transactions on Multimedia, 2021, 23: 4414-4425.
- [13] Ye M, Shen J B, Lin G J, et al. Deep learning for person re-identification: a survey and outlook[EB/OL]. (2020-01-13)[2021-02-06]. <https://arxiv.org/abs/2001.04193>.

- [14] Yin J H, Ma Z Y, Xie J Y, et al. DF²AM: dual-level feature fusion and affinity modeling for RGB-infrared cross-modality person re-identification[EB/OL]. (2021-04-01)[2021-08-07]. <https://arxiv.org/abs/2104.00226>.
- [15] Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification[EB/OL]. (2017-03-22)[2021-05-06]. <https://arxiv.org/abs/1703.07737>.
- [16] 刘天瑜, 刘正熙. 跨模态行人重识别研究综述[J]. 现代计算机, 2021(07): 135-139.
Liu T Y, Liu Z X. Overview of cross modality person re-identification research[J]. Modern Computer, 2021(07): 135-139.
- [17] Liu H J, Chai Y X, Tan X H, et al. Strong but simple baseline with dual-granularity triplet loss for visible-thermal person re-identification[J]. IEEE Signal Processing Letters, 2021, 28: 653-657.
- [18] Wu A C, Zheng W S, Yu H X, et al. RGB-infrared cross-modality person re-identification[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 5390-5399.
- [19] Ye M, Lan X Y, Wang Z, et al. Bi-directional center-constrained top-ranking for visible thermal person re-identification[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 407-419.
- [20] Wang Z X, Wang Z, Zheng Y Q, et al. Learning to reduce dual-level discrepancy for infrared-visible person re-identification[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 618-626.
- [21] Ye M, Lan X Y, Leng Q M, et al. Cross-modality person re-identification via modality-aware collaborative ensemble learning[J]. IEEE Transactions on Image Processing, 2020, 29: 9387-9399.
- [22] Wang G A, Zhang T Z, Cheng J, et al. RGB-infrared cross-modality person re-identification via joint pixel and feature alignment[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 3622-3631.
- [23] Ye M, Shen J B, Crandall D J, et al. Dynamic dual-attentive aggregation learning for visible-infrared person re-identification[M]//Vedaldi A, Bischof H, Brox T, et al. Computer vision-ECCV 2020. Lecture notes in computer science. Cham: Springer, 2020, 12362: 229-247.