

一种结构化双注意力混合通道增强的跨模态行人重识别方法

庄建军^{*①②} 庄宇辰^①

^①(南京信息工程大学电子与信息工程学院 南京 210044)

^②(南京信息工程大学-中大医院智慧医疗研究院 南京 210044)

摘要: 在目前跨模态行人重识别技术的研究中,大部分现有的方法会通过单模态原始可见光图像或者对抗生成图像的局部共享特征来降低跨模态差异,导致在红外图像判别中由于底层特征信息丢失而缺乏稳定的识别准确率。为了解决该问题,该文提出一种结构化双注意力可交换混合随机通道增强的特征融合跨模态行人重识别方法,利用通道增强后的可视图像作为第三模态,通过图像通道可交换随机混合增强(I-CSA)模块对可见光图像进行单通道和三通道随机混合增强抽取,从而突出行人的姿态结构细节,在学习过程中减少模态间差异。结构化联合注意力特征融合(SAFF)模块在注重模态间行人姿态结构关系的前提下,为跨模态表征学习提供更丰富的监督,增强了模态变化中共享特征的鲁棒性。在SYSU-MM01数据集全搜索模式单摄设置下Rank-1和mAP分别达到71.2%和68.1%,优于同类前沿方法。

关键词: 行人重识别; 跨模态; 混合通道增强; 联合注意力; 特征融合

中图分类号: TN911.73; TP391.4

文献标识码: A

文章编号: 1009-5896(2024)02-0518-09

DOI: 10.11999/JEIT230614

A Cross-modal Person Re-identification Method Based on Hybrid Channel Augmentation with Structured Dual Attention

ZHUANG Jianjun^{①②} ZHUANG Yuchen^①

^①(School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China)

^②(Institute for AI in Medicine, School of Artificial Intelligence, Nanjing University of Information Science and Technology, Nanjing 210044, China)

Abstract: In the current research on cross-modal person re-identification technology, most existing methods reduce cross-modal differences by using single modal original visible light images or locally shared features of adversarially generated images, resulting in a lack of stable recognition accuracy in infrared image discrimination due to the loss of feature information. In order to solve this problem, A cross-modal person re-identification method based on swappable hybrid random channel augmentation with structured dual attention is proposed. The visual image after channel enhancement is used as the third mode, and the single channel and three channels random hybrid enhancement extraction of visible image is performed through the Image Channel Swappable random mix Augmentation (I-CSA) module, so as to highlight the structural details of pedestrian posture, Reduce modal differences in learning. The Structured joint Attention Feature Fusion (SAFF) module provides richer supervision for cross-modal Feature learning, and enhances the robustness of shared features in modal changes, under the premise of focusing on the structural relationship of pedestrian attitudes between modes. Under the single shot setting of full search mode in the SYSU-MM01 dataset, Rank-1 and mAP reached 71.2% and 68.1%, respectively, surpassing similar cutting-edge methods.

Key words: Person Re-identification; Cross-modal; Hybrid channel enhancement; Joint attention; Feature fusion

收稿日期: 2023-06-21; 改回日期: 2023-11-03; 网络出版: 2023-11-17

*通信作者: 庄建军 jjzhuang@njust.edu.cn

基金项目: 国家重点研发计划(2021YFE0105500), 国家自然科学基金(62171228), 江苏高校“青蓝工程”

Foundation Items: The National Key Research and Development Program (2021YFE0105500), The National Natural Science Foundation of China (62171228), Jiangsu Qinglan Project

1 引言

行人重识别技术(Person Re-identification)在目前迅速发展的智慧城市监控系统的建设中有着至关重要的作用。目前传统的作用于可见光图像下的行人重识别已经获得了很好的实验结果, Huang等人^[1]为解决在现实各种场景下特征丢失和图像退化问题提出了一种记忆重放策略, 通过训练退化编码器提取图像集的特征, 降低退化不变性学习和下游任务训练之间的严重性遗忘。Yang^[2]针对无监督行人重识别领域中背景干扰问题将前一历元的距离矩阵积分到当前历元中来生成新的距离矩阵提高动态学习过程的稳定性。Xuan^[3]等人设计了实例和相机风格归一化模块(Instance and Camera Style Normalization, ICSN), 通过自适应地学习实例和批处理归一化的组合, 减轻了相机内的变化以增强对域差异的鲁棒性。

但是在夜间或者光照不足的环境中, 监控系统难以通过单模态行人重识别来提取有效的特征信息, 实现跨模态重识别成为当前热门的研究方向。Dai等人^[4]提出一种端到端交叉模式生成对抗网络(conformer-based metric GAN, cmGAN), 将类间模糊性降到最小。在此基础上, Wang等人^[5]创新性提出了像素对齐的方法, 利用对齐生成对抗网络(Alignment Generative Adversarial Network, AlignGAN)更好地克服了交叉模态变化。Lu等人^[6]采用项目对抗策略, 确保共享和特定特征之间不包含相互关联信息。

在对抗学习的过程中, 提取双模态中特征间的关系, 适当分配局部特征权重成为目前前沿的跨模态行人重识别方法之一。Li等人^[7]提出了一种反事实干预特征转移方法(Counterfactual Intervention Feature Transfer, CIFT), 利用不平衡场景模拟来减少模态平衡差距。Wang等人^[8]设计了多粒度共享特征融合网络(Multi-Scale Feature Fusion module, MSFF), 实现完全的端到端, 提取更有效的共享特征。Liang等人^[9]提出了跨模态转换器(Cross-Modality Transformer, CMT), 通过模态级对齐模块补偿模态特定信息的缺失和实例级对齐模块库实例与同类查询实例对齐实现更鲁棒的模态补偿。

可见光行人重识别任务中加入注意力机制有很好的提升效果, Xu等人^[10]提出多样化局部注意力网络(Diversified Local Attention Network, DLAN), 倚靠空间注意力网络定位使各局部分支聚焦于非重叠的人体部位以提高精度。Jia等人^[11]通过基于遮挡的师生空间半注意力分区方法促进了

学生注意力与嘈杂的教师监督之间的合理契合。而在跨模态行人重识别的研究中, 红外检测行人结果中会有很强的背景杂波, 少有引入注意力机制, Ye等人^[12]提出了一种动态双注意聚合学习方法(Dynamic Dual-Attentive AGgregation Learning, DDAG), 通过模态间特征关系挖掘来提取有区别的部分特征信息增强特征表示。

为了最大限度地保留可见光图像的纹理结构, 降低噪声影响, 强调两模态下低层次特征, 避免在学习中发生对颜色的依赖性较强的情况, 本文提出了一种通过结构化双注意力混合随机通道增强联合学习跨模态行人重识别方法, 由图像通道可交换随机混合增强模块(Image Channel Swappable random mix Augmentation, I-CSA)和结构化联合注意力特征融合模块(Structured joint Attention Feature Fusion, SAFF)组成, 采用了新的通道混合随机性增强策略(Random Hybrid Channel Augmentation, RHCA)保证了对颜色变化的鲁棒性, 从而提高注意力模块的效率和精确度。

2 本文方法

2.1 网络框架

这里将原始的跨模态行人重识别数据集表示为 $Y = \{U^m, U^r\}$, $U^m = \{x_i^m | i = 1, 2, \dots, k^m\}$ 表示可见光图像的训练集, 其中的每个图像定义为 $x_i^m = \{x_i^R, x_i^G, x_i^B\}$, x_i^R, x_i^G, x_i^B 为图像颜色三通道的分量, $U^r = \{x_i^r | i = 1, 2, \dots, k^r\}$ 代表红外图像的训练集, 其中 x_i^r 表示单一过饱和灰度通道的红外图像。

结构化双注意力可交换混合随机通道增强的特征聚合方法模型的总结构框架如图1所示, 采用了以ResNet50^[13]作为双流特征提取主干网络, 将浅层卷积池化层和之后的4个残差块组定义为stage0~4。

2.2 图像通道可交换随机混合增强模块

本文设计图像通道可交换随机混合增强模块(I-CSA)是为了优化可见光模态下的特征 $f^m(x_i^{R,G,B})$ 和红外模态下的特征 $f^r(x_j^r)$ 间的关系, 可以表示为

$$L = \sum \ell(f^m(x_i^{R,G,B}), f^r(x_j^r), y_i, y_j) \quad (1)$$

其中, y_i 和 y_j 表示的是每个图像中带注释的训练身份标签, 而 $\ell(\cdot)$ 代表的是使用的损失函数。

2.2.1 通道混合随机性增强策略

通道混合随机性增强策略(Random Hybrid Channel Augmentation, RHCA)借助此前对可见光图像的训练集表示, 通过随机选择一个通道代替另两个通道, 形成单通道新的训练集图像如图2所示。可以将可见光, 单色通道以及红外模态下的三模态特征关系表示出来

$$L = \sum \ell(f^m(\tilde{x}_i^m), f^r(x_j^r), y_i, y_j) \quad (2)$$

其中, $f^m(x_i^{R,G,B})$ 表示通道增强后可见图像模态下的特征, 通过对 $[0, 1, 2, 3]$ (0代表RGB三通道, 1代表R单通道, 2代表G单通道, 3代表B单通道) 中抽取随机数, 选择输出图像的类型。

2.2.2 三重态图像感知随机擦除数据增强策略

为了提高样本丰富度以及提高对图像噪声的鲁棒性, 将单通道图像作为第3模态, 提出了三重态图像感知随机擦除数据增强 (Triplet State Image-aware Random Erasing Data Augmentation, TRE-DA) 策略。随机擦除的效果如图3所示。由于三模态下的特征学习受颜色干扰, 将随机灰度变换和随机水平翻转作为该策略下的补充。

训练中, 随机擦除选择1块矩形区域为 I_e , 并用随机值擦除该区域像素, 当训练图像大小为 $S = W \times H$, 将随机擦除的矩形区域初始化为 S_e , 其中 S_e/S 在最小值 S_l 和最大值 S_h 之间, 擦除矩形区域的长宽比在设定的 r_l 和 r_h 之间被随机初始化, 将其设置为 r_e 。 I_e 的大小定义 $[W_e, H_e]$, 其中 $W_e = \sqrt{S_e/r_e}$, $H_e = \sqrt{S_e \times r_e}$, 然后随机初始化一个点 $Q = (x_e, y_e)$, 如果 $x_e + W_e \leq W$ 且 $y_e + H_e \leq H$ 则设置一个区域 $I_e = (x_e, y_e, x_e + W_e, y_e + H_e)$, 作为选定的矩形区域。否则, 重复上述过程, 知道有一个合适的 I_e 已选择。使用选定擦除区域 I_e , 在 I_e 中的每一个像素在 $[0, 255]$ 中独立分配随机值。

2.3 双流网络

为了充分利用跨模态异构信息优势, 提高行人

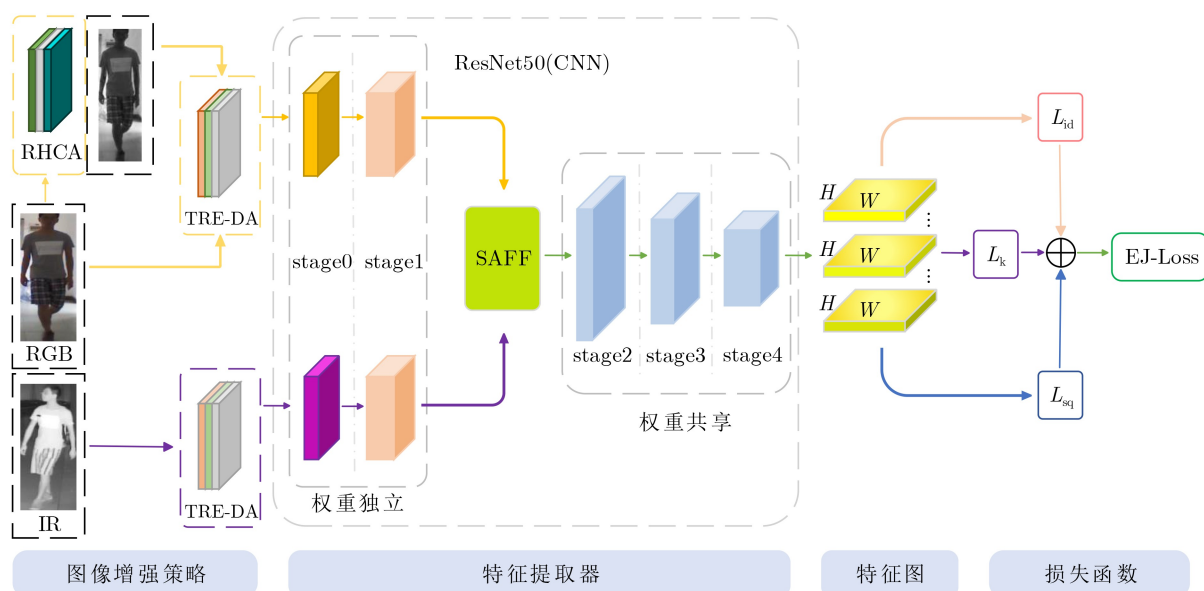


图1 本文模型总结构框架

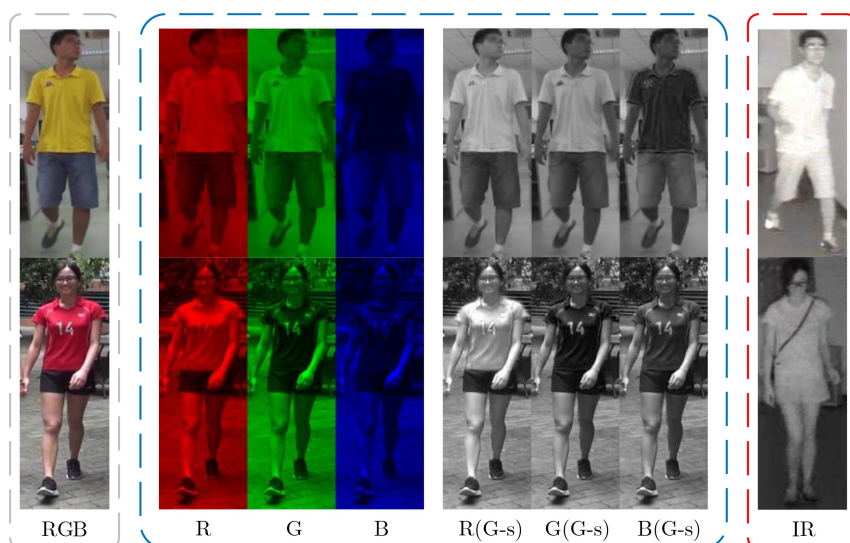


图2 RGB-IR及RHCA模块强化实例演示图

重识别的准确性和鲁棒性，本文采用双流特征提取主干网络结构来学习跨模态行人共享特征。

2.3.1 结构化联合注意力特征融合模块

在跨模态特征融合中，为了更好地利用可见光和红外图像间纹理结构的特征关联性，本文设计了结构化联合注意力特征融合(SAFF)模块，该模块分为两分支，其单一分支(Feature Aggregation Attention, FAA)分别对可见光图像和红外图像挖掘重要特征信息。模块结构图如图4所示。

将增强可视化化和红外图像输入的特征表示分别定义为 V_m 和 V_l ，而将他们经过FAA特征加权后的特征表示定义为 V'_m 和 V'_l ，通过加权平均后得到最终模块输出的特征表示 $V = \frac{1}{2}V'_m + \frac{1}{2}V'_l$ 。该模块单分支(FAA)如图5所示。

FAA结合了通道和空间注意力机制，在通道注意力机制中通过对ECA-net^[14]的学习，引入

GAP产生规格为 $1 \times 1 \times C$ 的特征图，通过聚合特征 $y \in R^C$ 不进行降维，通道注意力通过 $w = \sigma(Py)$ 进行学习， P 是通过SE块改进的一个 $C \times C$ 的参数矩阵，所以参数量巨大，为此引入拓展矩阵并且通过一个频带矩阵 P_k 来捕获局部跨通道交互关系，即

$$P_k = \begin{bmatrix} x^{1,1} & \dots & x^{1,k} & 0 & \dots & \dots & 0 \\ 0 & x^{2,2} & \dots & x^{2,k+1} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & x^{C,C-k+1} & x^{C,C} \end{bmatrix} \quad (3)$$

P_k 涉及了 k 个 C 参数，由此可以优化学习参数

$$w = \sigma(C1D_k(y)) \quad (4)$$

C1D表示1维卷积，该式由ECA块调用，由于只涉及 k 个参数，在复杂度降低的情况下即可获得SE-Var3的效果，提高了学习效率。

空间注意力机制中通过最大池化和平均池化，获得两个 $1 \times H \times W$ 的特征图，然后通过Concat操作对特征图拼接，通过 7×7 卷积变为单通道特征图，再经sigmoid函数获得 $1 \times H \times W$ 空间注意力机制后的特征图，通过与通道注意力机制后的特征图卷积后获得最终的特征表示。

2.4 损失函数

本文提出了增强型联合损失(EJ-Loss)，通过通道混合学习的策略来直接优化多模态间关系。具体的，增强型联合损失是由身份分类损失 L_{id} 和增强正则化三元组损失 L_{sq} 及交叉熵损失 L_k 组合的总损失。

本文介绍了一种改进的平方差方法，将向量对应维度差值求平方和，从而对样本对之间的差



图3 随机擦除数据增强实例演示图

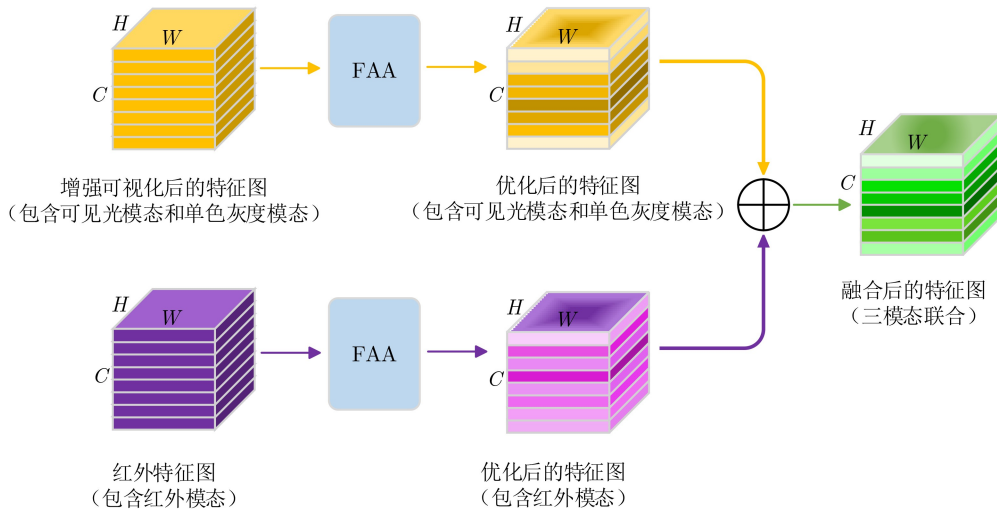


图4 SAFF模块结构图

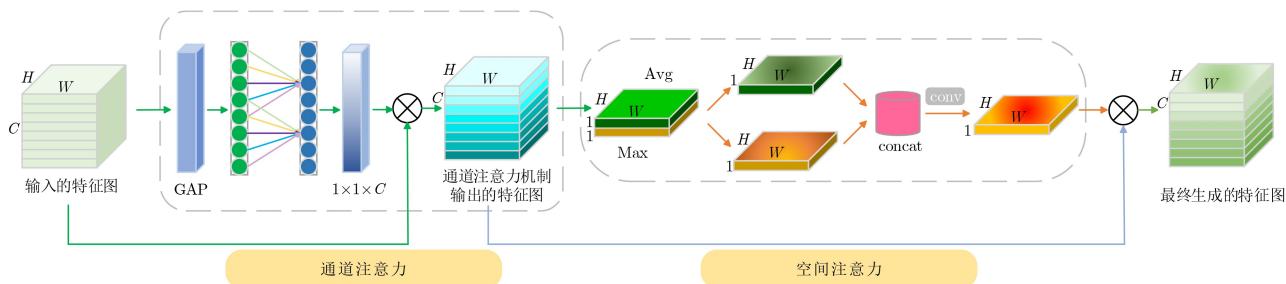


图5 SAFF模块单支FAA结构图

异更敏感,改进的增强正则化三元组损失公式表示为

$$L_{\text{sq}} = \frac{1}{N} \sum_{i=1}^N \log \left\{ 1 + \exp \left[\phi \left(\sum w_{ij}^p d_{ij}^p - \sum w_{ik}^p d_{ik}^p \right) \right] \right\} \quad (5)$$

其中 (i, j, k) 表示每个锚定样本每个训练批次中的三元组, d_{ij} 代表正样本对之间欧式距离, d_{ik} 代表负样本对之间欧式距离。在联合学习中, j 和 k 可以来自相同模态,也可以来自不同模态。

定义 k_i 为

$$k_i = \sum w_{ij}^p d_{ij}^p - \sum w_{ik}^p d_{ik}^p \quad (6)$$

这里的平方差表示为

$$\phi[k_i] = \begin{cases} k_i^2, & k_i > 0 \\ -k_i^2, & k_i < 0 \end{cases} \quad (7)$$

当处理难样本时,放大样本对间的差异,从而使网络学习到更好的表征。为了通过判别特征增强类间差异,本文加入交叉熵损失以实现该目的,总损失即增强型联合损失(EJ-Loss)表示为

$$L_{\text{EJ}} = L_{\text{id}} + L_{\text{sq}} + \gamma L_{\text{k}} \quad (8)$$

3 实验及结果分析

3.1 数据集和评价指标

本文跨模态行人重识别实验数据集包括两个公共数据集SYSU-MM01^[15]和RegDB^[16]。实验所用的评价指标包含了累计匹配特性曲线(cumulative matching characteristics)以及曲线上Rank准确率,查准均值mAP(mean Average Precision)。

3.2 实验设置

本文算法在PyTorch框架上呈现,训练评估均在NVIDIA 4070 GPU上进行,采用ResNet50双流主干网络,利用ImageNet^[17]预训练权重进行初始化加速收敛。将输入图像的尺寸均调整为144×288,采用随机梯度下降(SGD)优化器^[18]进行训练,动量参数为0.3,前10轮加入热身学习,学习率从0.01均

匀增加至0.1在此后保持不变,在第20轮和第50轮时学习率分别降低至当前学习率的1/10,在每次训练中分别随机抽取8个不同身份下的4张可见光图像和4张红外图像,将其8个身份作为一个批次。

3.3 对比实验结果分析

在SYSU-MM01数据集上分别在全搜索和室内搜索测试模式下通过单摄设置,使用本文方法与其他当前跨模态行人重识别先进算法的实验结果进行比较。如表1所示,相比于传统的通过对抗训练来降低类间歧义,增强模态相似性的方法如AlignGAN^[5],cm-SSFT^[6],本文的实验效果更好,说明SAFF模块能够避免对抗网络下不必要的噪声,保存了原图像中的纹理信息。与Liu等人^[19]在SFANET中设计的双向三约束损失相比,本文通过增强型联合损失(EJ-Loss)实现了更好的效果,说明本文提出的联合损失能够更有效地表达行人两模态间的关联性。相似地,使用双流网络来融合模态特征的方法如DDAG^[12],SFANET^[19],本文使用的SAFF模块更加注重提取浅层特征,在实验结果中也优于了这两种前沿的方法。

在RegDB数据集上使用本文方法与其他当前跨模态行人重识别先进算法的实验结果进行比较。如表2所示,由于RegDB数据集每个身份下的RGB图像和IR图像姿态是对应的,所以降低了跨模态行人重识别任务的难度,早期的如MAC^[24]结合模态特异身份分类器解决模态差异问题,虽然能够实现跨模态行人重识别任务,但是效果不佳。后来在研究通过优化损失函数来提高类内类间关系的方法,如LbA^[25]中设计的训练端到端网络的身份一致性和密集三联体损失和MCLNet^[26]引入的相机感知边际中心聚集损失,但是本文的方法在可见光图像查询远红外图像和远红外图像查找可见光图像下都取得了更好的效果。

3.4 消融实验结果分析

为了检验I-CSA模块,SAFF模块以及EJ-Loss的有效性,在全搜索模式下设置单摄,对SYSU-MM01进行如表3所示的消融实验,使用主干双流

表1 SYSU-MM01数据集在单摄设置下的实验对比结果(%)

方法	全搜索				室内搜索			
	单摄				单摄			
Method	Rank-1	Rank-10	Rank-20	mAP	Rank-1	Rank-10	Rank-20	mAP
AlignGAN ^[5]	42.4	85.0	93.7	40.7	35.9	87.6	94.4	54.3
AGW ^[20]	47.5	84.4	92.1	47.7	54.2	91.1	96.0	63.0
DDAG ^[12]	54.8	90.4	95.8	53.0	61.0	94.1	98.4	68.0
MID ^[21]	60.3	92.9	–	59.4	64.9	96.1	–	70.1
SFANET ^[19]	60.5	91.8	95.2	53.9	64.8	94.7	98.1	75.2
cm-SSFT ^[6]	61.6	89.2	93.9	63.2	70.5	94.9	97.7	72.6
SPOT ^[22]	65.3	92.7	97.0	62.3	69.4	96.2	99.1	74.6
MCLNet ^[26]	65.4	93.3	97.1	62.0	72.6	97.0	99.2	76.6
FMCNet ^[23]	66.3	–	–	62.5	68.2	–	–	74.1
本文方法	71.2	96.3	98.9	68.1	77.4	98.0	99.6	81.1

表2 RegDB数据集上的实验对比结果(%)

方法	可见光图像查询红外图像				红外图像查询可见光图像			
	Rank-1	Rank-10	Rank-20	mAP	Rank-1	Rank-10	Rank-20	mAP
MAC ^[24]	36.4	62.4	71.6	37.0	–	–	–	–
AlignGAN ^[5]	57.9	–	–	53.6	56.3	–	–	53.4
DDAG ^[12]	69.3	86.2	91.5	63.5	68.1	85.2	90.3	61.8
LbA ^[25]	74.2	–	–	67.6	72.4	–	–	65.5
MCLNet ^[26]	80.3	92.7	96.0	73.1	75.9	90.9	94.6	69.5
DCLNet ^[27]	81.2	–	–	74.3	78.0	–	–	70.6
本文方法	86.3	97.2	98.7	79.8	85.5	97.0	98.3	78.1

网络ResNet50和加权正则化三元组损失作为基线(Baseline), 通过加入RHCA子模块, 更加强调了行人姿态结构特征, 从图中可以看出加入该子模块之后Rank-1和mAP分别提高了8.6%和8.5%, 当加入完整I-CSA模块后, Rank-1和mAP分别提高了11.8%和10.7%, TRE-DA子模块引入随机擦除使模型对输入图像变化更加鲁棒, 减少过拟合风险。

在加入完整I-CSA模块基础上, 单独加入SAFF模块(嵌于stage1后)Rank-1和mAP分别提高了3.6%和2.6%, 通过FAA单支在前端残差块输出特征挖掘可见光和红外图像, 提高模型对行人姿态结构区域的关注度。若使用EJ-Loss作为损失函数, Rank-1和mAP分别提高了7.6%和5.0%, 这也证明了EJ-Loss在处理类内类间关系时拥有更出色的表现。在完整使用了本文方法, 即I-CSA模块, SAFF模块和EJ-Loss全部选用时达到了最佳的实验效果, Rank-1值达到71.2%, mAP达到68.1%。

3.5 SAFF模块嵌入位置效果研究

为了研究在ResNet50网络中的五层(stage0~4)后嵌入SAFF模块双注意力模块进行模态间特征

融合后的效果, 对SYSU-MM01和RegDB数据集分别对应做了相关实验如表4所示。SYSU-MM01数

表3 SYSU-MM01数据集上的消融实验(%)

RHCA	TRE-DA	SAFF	EJ-Loss	SYSU-MM01	
				Rank-1	mAP
–	–	–	–	48.8	46.6
✓	–	–	–	57.4	55.1
✓	✓	–	–	60.6	57.3
✓	✓	✓	–	64.2	60.9
✓	✓	–	✓	68.2	64.3
✓	✓	✓	✓	71.2	68.1

表4 SAFF模块嵌入位置研究结果(%)

SAFF	SYSU-MM01		RegDB	
	Rank-1	mAP	Rank-1	mAP
stage0	70.7	67.8	86.2	79.7
stage1	71.1	67.9	86.3	79.8
stage2	71.2	68.1	86.0	79.3
stage3	66.3	63.5	83.9	72.1
stage4	59.4	55.1	81.4	68.0

数据集在全搜索模式下设置单摄, RegDB数据集以可见光图像查询远红外图像作为评估标准。在图6中可以发现在底层级后加入SAFF模块的效果会优于在高层级后加的效果, 在底层级中包含重要的边缘和纹理结构特征, 并且拥有较小的感受野和参数量, 通过SAFF双注意力进行特征聚合后, 能更细粒度

地关注结构性特征以及提高局部特征的感知。在实验中发现在SYSU-MM01数据集上, 将SAFF模块放置于stage2后的实验结果最佳, 而在RegDB数据集上, 将SAFF模块放置于stage3后的实验结果最佳, 因此, 本文通过设计选择器, 通过数据集的选择灵活地嵌入SAFF模块。

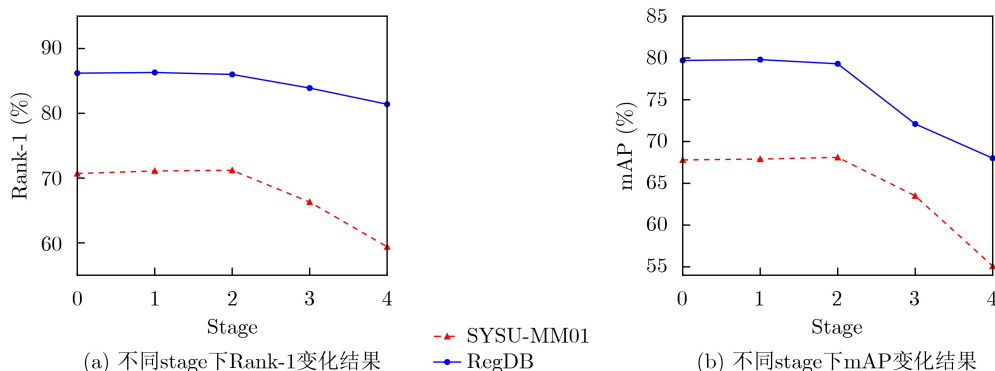


图6 不同stage下SAFF 模块嵌入效果对比

3.6 γ 参数的取值分析

在本研究中, 在SYSU-MM01数据集单摄设置下通过改变 γ 参数来控制交叉熵损失在整个增强型联合损失中的权重, 进行 γ 参数不同取值对训练性能的影响实验, γ 的取值区间为0以及0.1到0.9, 取值间隔为0.2, 实验结果如表5所示。

采用多尺度优化策略, 当 $\gamma = 0.47$ 时, 模型训练达到最优。但当 γ 大于0.5时, 发现由于当交叉熵损失权重过大时, 模型在训练过程中忽略跨模态下共享特征的学习导致过拟合现象发生, 模型性能大幅度下降。

4 结束语

本文提出了一种结构化双注意力可交换混合随机通道增强的特征聚合跨模态行人重识别方法, 由图像通道可交换随机混合增强模块、Resnet50主干双流网络、结构化联合双注意力特征融合模块以及

增强型联合损失构成。其中图像通道可交换随机混合增强模块主要通过可见光图像提取单通道图像并且与红外图像进行感知随机擦除数据增强, 缓解了模态偏差, 提高模型泛化能力。结构化联合双注意力特征融合模块主要利用新型注意力分支对可见光和红外图像进行重要特征的强调, 并通过特征融合增强模态间的联系。增强型联合损失通过身份分类损失和增强正则化三元组损失的组合, 提高了对难样本的敏感度, 具有出色的表征能力。在两个公共基准数据集上的实验中可以证明本文的方法在跨模态行人重识别任务相比于前沿的方法有着较好的表现效果。目前跨模态行人重识别依然有着很大的提升空间, 在模态差异的解决方法中大多更注重人体姿态结构特征的捕获, 如何解决姿态结构特征不稳定性带来的影响是解决实际应用问题的关键, 也是今后工作的一个重要探索方向。

参考文献

- [1] HUANG Yukun, FU Xueyang, LI Liang, *et al.* Learning degradation-invariant representation for robust real-world person Re-identification[J]. *International Journal of Computer Vision*, 2022, 130(11): 2770–2796. doi: [10.1007/s11263-022-01666-w](https://doi.org/10.1007/s11263-022-01666-w).
- [2] YANG Lei. Continuous epoch distance integration for unsupervised person Re-identification[C]. The 5th International Conference on Communications, Information System and Computer Engineering, Guangzhou, China, 2023: 464–469. doi: [10.1109/cisce58541.2023.10142496](https://doi.org/10.1109/cisce58541.2023.10142496).
- [3] XUAN Shiyu and ZHANG Shiliang. Intra-inter domain

表5 γ 参数不同取值下的训练实验结果(%)

γ	SYSU-MM01			
	Rank-1	Rank-10	Rank-20	mAP
0	66.7	94.6	97.1	64.7
0.1	68.4	95.2	97.8	65.9
0.3	69.9	95.8	98.5	67.3
0.5	70.8	96.2	98.7	67.9
0.7	59.3	91.1	95.4	57.7
0.9	55.4	89.0	94.2	53.1
0.47	71.2	96.3	98.9	68.1

- similarity for unsupervised person Re-identification[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 1. doi: [10.1109/tpami.2022.3163451](https://doi.org/10.1109/tpami.2022.3163451).
- [4] DAI Pingyang, JI Rongrong, WANG Haibin, *et al.* Cross-modality person Re-identification with generative adversarial training[C]. Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 2018: 677–683. doi: [10.24963/ijcai.2018/94](https://doi.org/10.24963/ijcai.2018/94).
- [5] WANG Guan'an, ZHANG Tianzhu, CHENG Jian, *et al.* RGB-infrared cross-modality person Re-identification via joint pixel and feature alignment[C]. The IEEE/CVF International Conference on Computer Vision, Seoul, Korea (South), 2019: 3622–3631. doi: [10.1109/ICCV.2019.00372](https://doi.org/10.1109/ICCV.2019.00372).
- [6] LU Yan, WU Yue, LIU Bin, *et al.* Cross-modality person Re-identification with shared-specific feature transfer[C]. The IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 13376–13386. doi: [10.1109/CVPR42600.2020.01339](https://doi.org/10.1109/CVPR42600.2020.01339).
- [7] LI Xulin, LU Yan, LIU Bin, *et al.* Counterfactual intervention feature transfer for visible-infrared person Re-identification[C]. 17th European Conference on Computer Vision, Tel Aviv, Israel, 2022: 381–398. doi: [10.1007/978-3-031-19809-0_22](https://doi.org/10.1007/978-3-031-19809-0_22).
- [8] 王凤随, 闫涛, 刘芙蓉, 等. 融合子空间共享特征的多尺度跨模态行人重识别方法[J]. 电子与信息学报, 2023, 45(1): 325–334. doi: [10.11999/JEIT211212](https://doi.org/10.11999/JEIT211212).
WANG Fengsui, YAN Tao, LIU Furong, *et al.* Multi-scale cross-modality person Re-identification method based on shared subspace features[J]. *Journal of Electronics & Information Technology*, 2023, 45(1): 325–334. doi: [10.11999/JEIT211212](https://doi.org/10.11999/JEIT211212).
- [9] LIANG Tengfei, JIN Yi, LIU Wu, *et al.* Cross-modality transformer with modality mining for visible-infrared person Re-identification[J]. *IEEE Transactions on Multimedia*, 2023: 1–13. doi: [10.1109/tmm.2023.3237155](https://doi.org/10.1109/tmm.2023.3237155).
- [10] 徐胜军, 刘求缘, 史亚, 等. 基于多样化局部注意力网络的行人重识别[J]. 电子与信息学报, 2022, 44(1): 211–220. doi: [10.11999/JEIT201003](https://doi.org/10.11999/JEIT201003).
XU Shengjun, LIU Qiuyuan, SHI Ya, *et al.* Person Re-identification based on diversified local attention network[J]. *Journal of Electronics & Information Technology*, 2022, 44(1): 211–220. doi: [10.11999/JEIT201003](https://doi.org/10.11999/JEIT201003).
- [11] JIA Mengxi, SUN Yifan, ZHAI Yunpeng, *et al.* Semi-attention partition for occluded person Re-identification[C]. The 37th AAAI Conference on Artificial Intelligence, Washington, USA, 2023: 998–1006. doi: [10.1609/aaai.v37i1.25180](https://doi.org/10.1609/aaai.v37i1.25180).
- [12] YE Mang, SHEN Jianbing, CRANDALL D J, *et al.* Dynamic dual-attentive aggregation learning for visible-infrared person Re-identification[C]. 16th European Conference on Computer Vision, Glasgow, UK, 2020: 229–247. doi: [10.1007/978-3-030-58520-4_14](https://doi.org/10.1007/978-3-030-58520-4_14).
- [13] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. The IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [14] WANG Qilong, WU Banggu, ZHU Pengfei, *et al.* ECA-Net: Efficient channel attention for deep convolutional neural networks[C]. The IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 11531–11539. doi: [10.1109/CVPR42600.2020.01155](https://doi.org/10.1109/CVPR42600.2020.01155).
- [15] WU Ancong, ZHENG Weishi, YU Hongxing, *et al.* RGB-infrared cross-modality person Re-identification[C]. The IEEE International Conference on Computer Vision, Venice, Italy, 2017: 5390–5399. doi: [10.1109/ICCV.2017.575](https://doi.org/10.1109/ICCV.2017.575).
- [16] 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. doi: [10.3390/s17030605](https://doi.org/10.3390/s17030605).
- [17] KRIZHEVSKY A, SUTSKEVER I, and HINTON G E. ImageNet classification with deep convolutional neural networks[J]. *Communications of the ACM*, 2017, 60(6): 84–90. doi: [10.1145/3065386](https://doi.org/10.1145/3065386).
- [18] SONG Shuang, CHAUDHURI K, and SARWATE A D. Stochastic gradient descent with differentially private updates[C]. Global Conference on Signal & Information Processing, Austin, USA, 2014: 245–248. doi: [10.1109/globalsip.2013.6736861](https://doi.org/10.1109/globalsip.2013.6736861).
- [19] LIU Haojie, MA Shun, XIA Daoxun, *et al.* SFANet: A spectrum-aware feature augmentation network for visible-infrared person reidentification[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, 34(4): 1958–1971. doi: [10.1109/tnnls.2021.3105702](https://doi.org/10.1109/tnnls.2021.3105702).
- [20] YE Mang, SHEN Jianbing, LIN Gaojie, *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. doi: [10.1109/TPAMI.2021.3054775](https://doi.org/10.1109/TPAMI.2021.3054775).
- [21] HUANG Zhipeng, LIU Jiawei, LI Liang, *et al.* Modality-adaptive mixup and invariant decomposition for RGB-infrared person Re-identification[C/OL]. The 36th AAAI Conference on Artificial Intelligence, 2022: 1034–1042. doi: [10.1609/aaai.v36i1.19987](https://doi.org/10.1609/aaai.v36i1.19987).
- [22] CHEN Cuiqun, YE Mang, QI Meibin, *et al.* Structure-aware positional transformer for visible-infrared person Re-identification[J]. *IEEE Transactions on Image Processing*, 2022, 31: 2352–2364. doi: [10.1109/tip.2022.3141868](https://doi.org/10.1109/tip.2022.3141868).
- [23] ZHANG Qiang, LAI Changzhou, LIU Jianan, *et al.* FMCNet: Feature-level modality compensation for visible-

- infrared person Re-identification[C]. The IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, USA, 2022: 7339–7348. doi: [10.1109/cvpr52688.2022.00720](https://doi.org/10.1109/cvpr52688.2022.00720).
- [24] YE Mang, LAN Xiangyuan, and LENG Qingming. Modality-aware collaborative learning for visible thermal person Re-identification[C]. The 27th ACM International Conference on Multimedia, Nice, France, 2019: 347–355. doi: [10.1145/3343031.3351043](https://doi.org/10.1145/3343031.3351043).
- [25] PARK H, LEE S, LEE J, *et al.* Learning by aligning: Visible-infrared person Re-identification using cross-modal correspondences[C]. The IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021: 12026–12035. doi: [10.1109/iccv48922.2021.01183](https://doi.org/10.1109/iccv48922.2021.01183).
- [26] HAO Xin, ZHAO Sanyuan, YE Mang, *et al.* Cross-modality person Re-identification via modality confusion and center aggregation[C]. The IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021: 16383–16392. doi: [10.1109/ICCV48922.2021.01609](https://doi.org/10.1109/ICCV48922.2021.01609).
- [27] SUN Hanzhe, LIU Jun, ZHANG Zhizhong, *et al.* Not all pixels are matched: Dense contrastive learning for cross-modality person Re-identification[C]. The 30th ACM International Conference on Multimedia, Lisbon, Portugal, 2022: 5333–5341. doi: [10.1145/3503161.3547970](https://doi.org/10.1145/3503161.3547970).

庄建军: 男, 教授, 研究方向为视频信号的智能处理.

庄宇辰: 男, 硕士生, 研究方向为计算机视觉、行人重识别.

责任编辑: 马秀强