

文章编号: 1005-0523(2024)02-0087-08



基于双注意力机制的可见光-红外行人重识别

魏克铭¹, 韩星宇², 王辉², 范自柱^{1,2}

(1. 华东交通大学理学院, 江西 南昌 330013; 2. 华东交通大学江西省先进控制与优化重点实验室, 江西 南昌 330013)

摘要:【目的】由于可见光图像和红外图像之间的巨大模态差异,导致可见光-红外行人重识别是一项非常具有挑战性的图像检索问题。【方法】为了进一步减小两种模态之间的差异,重点关注行人信息,提出一种基于双注意力机制的网络结构用于可见光-红外行人重识别。一方面通过双注意力机制挖掘不同尺度的行人空间信息和增强局部特征的通道交互能力,另一方面利用全局分支和局部分支,学习多粒度的特征信息,使不同粒度信息可以相互补充,形成一个更具辨别性的特征。【结果】在两个公共数据集上的实验结果表明,该方法相较于基线有明显的提升,在RegDB数据集和SYSU-MM01数据集上均表现出理想的性能。【结论】该方法可为以后解决可见光-红外行人重识别的模态差异问题提供有效的参考。

关键词: 可见光-红外行人重识别; 注意力机制; 缓解模态差异

中图分类号: TP391; [U-9]

文献标志码: A

本文引用格式: 魏克铭, 韩星宇, 王辉, 等. 基于双注意力机制的可见光-红外行人重识别[J]. 华东交通大学学报, 2024, 41(2): 87-94.

Visible-Infrared Person Re-Identification Based on Dual Attention Mechanism

Wei Keming¹, Han Xingyu², Wang Hui², Fan Zizhu^{1,2}

(1. School of Science, East China Jiaotong University, Nanchang 330013, China; 2. Key Laboratory of Advanced Control and Optimization of Jiangxi Province, East China Jiaotong University, Nanchang 330013, China)

Abstract: 【Objective】Visible-infrared person re-identification is a very challenging image retrieval problem due to the huge modal difference between visible and infrared images. 【Method】In order to further reduce the difference between the two modalities and focus on pedestrian information, a network structure based on dual attention mechanism is proposed for visible-infrared person re-identification. On the one hand, through the dual attention mechanism to mine personal spatial information of different scales and enhance the channel interaction ability of local features. On the other hand, through learning multi-granular feature information through using global and local branches, different granular information can complement with each other to form a more discriminating feature. 【Result】Experimental results on two public datasets show that the proposed method has a significant improvement compared with the baseline, and shows ideal performance on both the RegDB dataset and the SYSU-MM01 dataset. 【Conclusion】The proposed method can provide an effective reference for solving the problem of modal difference of visible-infrared person re-identification in the future.

收稿日期: 2023-05-10

基金项目: 国家自然科学基金项目(61991401); 江西省自然科学基金项目(20224BAB212014)

Key words: visible-infrared person re-identification; attention mechanism; mitigate modal differences

Citation format: WEI K M, HAN X Y, WANG H, et al. Visible-infrared person re-identification based on dual attention mechanism[J]. Journal of East China Jiaotong University, 2024, 41(2): 87-94.

【研究意义】行人重识别主要任务是在多个不重叠的摄像机视图中匹配特定的人,在安全领域有着不可或缺的作用,近年来行人重识别一直受到广泛的关注。它的挑战主要集中在视图、姿态、光照、遮挡、背景变化等方面,为了解决这些问题,众多学者提取出了许多解决方法,取得不错的效果。这些方法主要集中在单模态的可见光行人重识别问题上,但在实际应用中,往往需要捕捉不同场景下的行人图像,特别是在夜晚光照极弱的情况下,可见光相机很难捕捉到有效的行人信息,因此可见光-红外行人重识别就引起了众多学者的注意。该领域主要研究可见光图像和红外图像之间的跨模态度量问题,以从不同模态的图像中匹配出相同的行人图像,目的是克服在复杂环境下传统行人重识别的局限性。红外图像相比于可见光图像,信息量更少、视觉效果模糊、分辨率差、对比度低,巨大的模态差异,导致很难提取到有效的特征信息,常规的单模态行人重识别也不能够发挥同等的效用。

【研究进展】不同模态间的巨大差异导致可见光-红外行人重识别非常具有挑战性,针对模态差异,众多学者提出了一系列解决方法。为缓解模态差异,一些方法通过设计模态生成器^[1-4],从而生成中间模态或实现模态转换,例如 Zhang 等^[2]提出特征级模态补偿网络,直接从其他模态的现有模态共享特征中生成缺失模态特定特征,但由于红外到可见光变换的不稳定性,导致生成的模态难以优化,而且不可避免地会引入噪声数据。另外一些方法采用单流、双流或多流网络^[5-10],通过设计不同的损失函数、注意力机制等提取不同模态共享特征,例如 Wu 等^[10]提出联合模态和模式对齐网络,以发现不同模态的细微差别,减轻模态差异,并且提出了一个中心聚类损失函数,进一步约束增强学习效果。然而,基于这些学习方法训练通用的网络模型,缺乏对特异性模态信息的关注度,导致关键信息丢失。

【关键问题】为了避免噪声数据的影响,充分利用有价值的行人信息,减小模态间的差异,从以下几个方面出发来解决此问题:**【创新特色】**首先,为了减小背景、光线等噪声数据的影响,受到交叉注意力和空洞卷积的启发,本文提出多尺度交叉注意力机制(multi-scale cross attention, MCA),同时利用不同尺度的空洞卷积和最大池化,扩大模块的感受野,关注更多的边缘信息。其次考虑到不同通道之间的信息交互和不同层次行人特征之间的差异性,本文提出局部通道交互注意力机制(part channel-interaction attention, PCA),在兼顾局部特征的同时,增强不同通道间的特征交互能力。最后,考虑到数据集规模有限,缺乏多样性,如果仅学习全局特征,容易导致信息丢失,而不同粒度的特征可以更有效的提取行人信息,因此本文结合局部特征和全局特征,共同优化网络模型。本文的模型在两个公开数据集 RegDB 和 SYSU-MM01 上均取得最优的识别效果。

1 方法

为了减小模态差异、背景噪声影响、增强模型的鲁棒性,基于随机通道交换^[9]提出了一种兼顾局部与全局特征的双注意力机制的网络结构用于可见光-红外行人重识别。这一部分主要介绍网络结构的模型框架,其整体结构如图 1 所示。主要包括以下几个部分组成:① 由 ResNet50 组成的双流骨干网络;② 多尺度交叉注意力机制(MCA);③ 局部通道交互注意力机制(PCA);④ 全局特征分支及其对应损失函数;⑤ 局部特征分支及其对应损失函数。在测试阶段,使用全局特征分支的输出结果进行预测,局部特征分支仅在训练过程中发挥效用。

1.1 模型框架

双流网络是用于可见光红外行人重识别特征提取的典型方法,而且它的有效性在众多文献中得到了有力的证明。本文利用在 ImageNet 上预训练的 ResNet50 作为骨干网络提取特征,为保证不同模

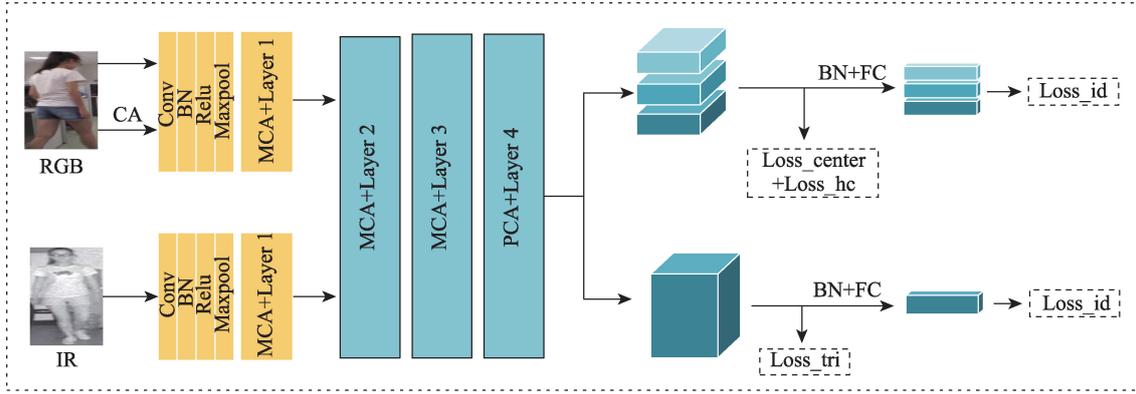


图1 模型框架

Fig. 1 Model framework

态特征的特异性,网络在第二个残差块前不共享参数;为避免噪声的影响,本文同时设计两个注意力模块,注意力模块在保持特征图身份识别能力的同时,减轻模态差异以及背景噪声的影响。其次,为了同时获得全局特征和局部特征,学习到细微的、具有鉴别性的特征,本文结合PCB^[11]特征分块机制,旨在学习不同行人图像之间的细微差别,以达到更好的效果。

1.2 多尺度交叉注意力机制(MCA)

受到交叉注意力和多尺度特征融合^[12]的启发,本文提出了MCA。考虑到最大池化可以加强网络对显著性区域的关注度,去除背景冗余信息,但容易丢失空间分辨率,因此引入空洞卷积弥补池化的不足。同时考虑到利用不同尺度最大池化和空洞卷积可以扩大感受野,关注行人的边缘特征,获取多尺度的上下文信息,从而增强像素级的表征能力。具体流程如图2左侧所示,图2右侧展示了MCA中最大池化模块和空洞卷积模块的具体结构。

给定特征 $x \in R^{C \times W \times H}$,该模块首先在 x 上应用两个具有 1×1 滤波器的卷积层,分别生成两个特征映射 Q 和 K ,其中 $\{Q, K\} \in R^{C' \times W \times H}$ 。 C' 为通道数,由于降维,通道数小于 C 。然后 Q 和 K 分别经过一个多尺度最大池化块,再经过 1×1 滤波器的卷积层得到 Q_1 和 K_1 ,用数学公式表示如下

$$Q_1 = \text{conv}_2^1(\max(\text{conv}_1^1(Q))) \quad (1)$$

$$K_1 = \text{conv}_2^2(\max(\text{conv}_1^2(K))) \quad (2)$$

另外给定的输入特征 $x \in R^{C \times W \times H}$ 分别经过不同大小滤波器的空洞卷积块得到 V ,而后输入 1×1 滤波器的卷积块和平均池化块得到 V_1 为

$$V_1 = \text{avg}(\text{conv}(\text{Dilated}(x))) \quad (3)$$

Q_1 和 K_1 经过仿射变换后,与 V_1 进行聚合变换,最终与输入特征 x 求和得

$$\text{output} = x + \text{Agg}(\text{Aff}(Q_1, K_1), V_1) \quad (4)$$

式中:Aff为仿射变换;Agg为聚合变换。

1.3 局部通道交互注意力机制(PCA)

不同通道的特征图受到的关注度理应是不同的,

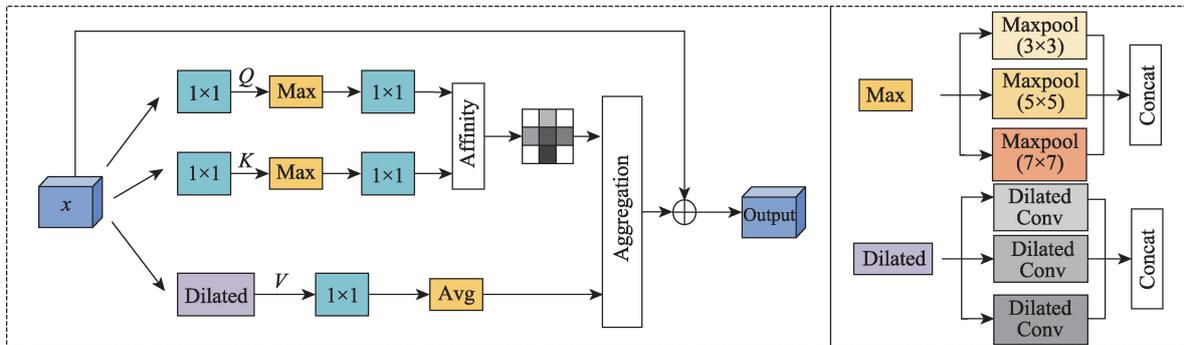


图2 MCA
Fig. 2 MCA

且特征图的不同层次也是如此,因此为了增强不同层次之间的通道交互能力,引入了PCA,如图3所示。首先给定一个输入特征 $x \in R^{C \times W \times H}$, x 在水平方向被均匀地分割成若干块,得到 $x_i, i=1,2,3$, 然后分别经过池化和 1×1 卷积块,进行缩放变换,随后对拼接的特征利用正切激活函数激活,具体过程用公式表示如下

$$x'_i = \max(\text{relu}(\text{conv}_1^i(x_i))), i=1,2,3 \quad (5)$$

$$x'_i = \tanh(\text{avg}(\text{conv}_1^i(x'_i))), i=1,2,3 \quad (6)$$

$$x' = \text{concat}(x'_1, x'_2, x'_3) \quad (7)$$

$$\text{output} = x + \tanh(x') \quad (8)$$

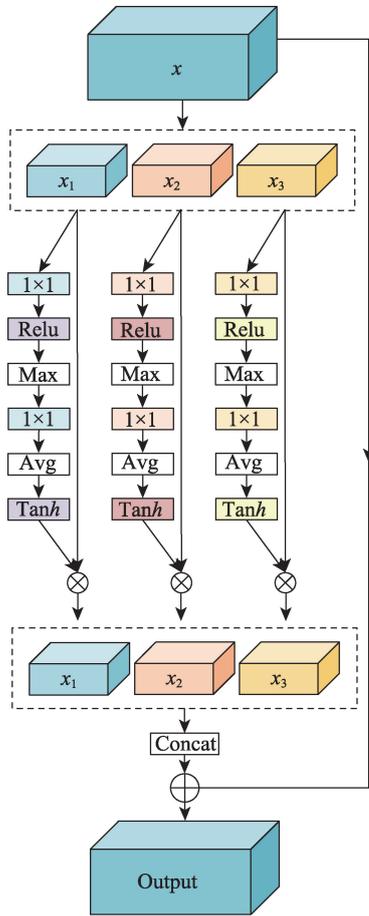


图3 PCA
Fig. 3 PCA

1.4 损失函数

本节主要介绍本文的框架中使用的损失函数,包括交叉熵损失、三元损失、聚类中心损失和中心三元损失。利用交叉熵损失和三元损失结合起来监督全局特征,利用交叉熵损失、聚类中心损失和中心三元损失作为局部分支的学习目标。

1) 交叉熵损失函数。交叉熵损失的目标是提取特定行人身份的信息进行分类。此方法被广泛应用于行人重识别任务中,以促进模型对样本进行有效的分类。在本文中,依旧采用交叉熵损失分别优化全局特征和局部特征,以捕获每个行人不同模态的身份鉴别信息。交叉熵损失的表达式如下

$$L_{id} = - \sum_{i=1}^N \log \frac{e^{w_{y_i}^T x_i}}{\sum_{k=1}^D e^{w_k^T x_i}} \quad (9)$$

式中: w_k 为第 k 类的权重向量; y_i 为特征 x_i 的真实身份标签; N 为批次大小; D 为训练集中的类数。

2) 三元损失函数。对于全局特征,利用三元损失优化不同模态下不同行人图像的特征,它可以拉近不同模态相同身份的行人特征间的距离,扩大不同身份的行人间的距离,本文沿用Ye等^[9]提出的三元损失,公式定义如下

$$L_{tri} = \sum_{i=1}^P \sum_{a=1}^K [\rho + \max \|x_a^i - x_p^i\|_2 - \min_{i \neq j} \|x_a^i - x_n^j\|_2]_+ \quad (10)$$

式中: K 为模态数量; P 为行人的数量; x_a 为锚点样本; x_p 为正样本对; x_n 为负样本对; ρ 为阈值参数,用以约束正负样本间的距离。

3) 中心三元损失函数。三元损失通过锚点与所有其他样本的比较计算损失。但由于图像本身存在的一些噪声,造成局部特征可能与全局特征有很大的差异,如果存在一些异常值,可能会过于严格地约束成对距离,三元损失将不能很好地优化类内类间距离。因此,利用中心三元损失函数优化局部特征的类内与类间距离,采用每个身份的中心作为身份代理,将锚点与所有其他样本的比较替换为锚点中心与所有其他中心的比较,具体计算公式如下

$$L_{center} = \sum_{i=1}^P [\rho + \|c_v^i - c_t^i\|_2 - \min_{i \neq j} \|c_v^i - c_n^j\|_2]_+ + \sum_{i=1}^P [\rho + \|c_t^i - c_v^i\|_2 - \min_{i \neq j} \|c_t^i - c_n^j\|_2]_+ + \sum_{i=1}^P [\rho + \|c_m^i - c_t^i\|_2 - \min_{i \neq j} \|c_m^i - c_n^j\|_2]_+ \quad (11)$$

式中: c_v^i 为可见光图像特征的聚类中心; c_t^i 为红外图像的聚类中心; c_m^i 为随机通道交换后得到的模态的聚类中心; c_n^i 为其他模态聚类中心。

4) 聚类中心损失函数。聚类中心损失通过惩罚不同模态分布的中心,优化不同模态的类内相似

度,公式定义如下

$$L_{hc} = \sum_{i=1}^P \|c_v^i - c_t^i\|_2 \quad (12)$$

式中: c_v^i 和 c_t^i 分别表示可见光和红外图像的聚类中心。

5) 均方差损失函数。为进一步缩小相同身份不同模态的行人图像之间的距离,简单地应用均方差损失进行约束,公式如下

$$L_{mse} = \sum_{i=1}^P \|x_v^i - x_t^i\|_2 \quad (13)$$

式中: P 为行人的数量。

综上所述,损失函数分为全局损失和局部损失,全局损失函数定义如下

$$L_{global} = L_{id} + L_{tri} + L_{mse} \quad (14)$$

局部损失函数定义如下

$$L_{local} = \lambda_1 L_{id} + \lambda_2 L_{center} + \lambda_3 L_{hc} \quad (15)$$

总损失函数定义如下

$$L = L_{global} + \alpha L_{local} \quad (16)$$

式中: α , λ_1 , λ_2 , λ_3 均为超参数,用以平衡各个损失函数之间的权重。

2 实验

2.1 数据集

为了评估本文提出方法的有效性,在两个公开的数据集(SYSU-MM01 和 RegDB)上做了充分的实验,数据集参数如下。

SYSU-MM01 数据集由 4 个可见光相机和 2 个红外摄像机在室内和室外拍摄而成,涉及 491 个身份。其中训练集包括 22 258 张可见光图像和 11 909 张红外图像,涉及 395 个身份。测试集包含 3 803 张用于被检索红外图像和 301 张用于检索的可见光图像,共 96 个身份。此数据集包含全局搜索和室内搜索两种测试模式。

RegDB 数据集由一对对齐的可见光和热成像相机拍摄而成,包括 412 个身份的 4 120 张图片,每个身份对应 10 张可见光图像和 10 张热成像图像。此数据集被随机划分为两部分,206 个身份用于训练,其余 206 个身份用于测试。训练和测试均需基于数据集的随机划分重复进行 10 次实验。

2.2 参数设置

采用双注意力机制增强的双流网络,引入 PCB 模块,以 ResNet50 作为骨干提取特征,共享后 3 个

残差块的参数。采用随机梯度下降(SGD)优化器进行训练。训练阶段,所有的可见光和红外图像的大小调整为 288×144 ,通过随机通道交换、擦除、翻转增强数据集。初始学习率设置为 0.1,在前 10 个训练周期采用预热策略,在第 20 个训练周期衰减为 0.01,在第 50 个训练周期衰减为 0.001。训练周期总数设置为 100。在每一个训练批次中,随机抽取 4 个行人,其中每个行人分别抽取 4 张可见光图像和 4 张红外图像,共 32 张行人图像。超参数的取值区间为 $[0, 1]$,根据实验结果的优劣,不断微调参数值,以取得更好的实验效果,最终总损失函数 L 的参数值在 RegDB 数据集上分别设置为 $\alpha=1$, $\lambda_1=1$, $\lambda_2=0.6$, $\lambda_3=0.6$; 在 SYSU-MM01 数据集上设置 $\alpha=1$, $\lambda_1=0.5$, $\lambda_2=0.1$, $\lambda_3=2$ 。

2.3 对比现有方法

对比了近几年来提出的可见光-红外行人重识别方法,表 1 和表 2 分别展示了在 RegDB 数据集和 SYSU-MM01 数据集上与不同方法比较的结果。在 RegDB 数据集中,可见光到红外模式下达到了 95.22% 的 $Rank1$, 87.70% 的 mAP 和 74.48% 的 $mINP$; 红外到可见光模式下达到了 93.67% 的 $Rank1$, 86.43% 的 mAP 和 71.68% 的 $mINP$ 。在 SYSU-MM01 数据集中,全局搜索模式下达到了 74.18% 的 $Rank1$, 70.04% 的 mAP 和 56.97% 的 $mINP$; 室内搜索模式下达到了 79.69% 的 $Rank1$, 83.08% 的 mAP 和 79.65% 的 $mINP$ 。基于所有这些评估和比较的结果,可以确认本文方法的优越性及有效性。

2.4 消融实验

以 RegDB 数据集为例,评估模型的有效性。

1) 不同模块的效果。首先从模型中删除两个注意力模块、PCB 模块及其对应的损失函数,以此作为基线方法 Base 进行比较。P 表示结合局部特征及其对应的损失函数, MCA 表示采用注意力模块 MCA, PCA 表示采用注意力模块 PCA。具体结果如表 3 所示,从中可以清晰地看出,在采用 MCA 时效果比 Base 增加 2.77%, 比 Base+P 增加 1.02%; 在采用 PCA 时效果比 Base 增加 2.72%, 比 Base+P 增加 0.68%; 同时加上两个注意力机制时效果比 Base 增加 4.56%, 比 Base+P 增加 1.87%。

2) 利用全局特征进行预测的有效性。为了验证仅使用全局特征分支进行预测的效果,在此以 RegDB 数据集为例,对比利用全局特征和局部特征

表1 在RegDB数据集上与现有的方法作对比

Tab.1 Comparison with existing methods on the RegDB dataset

%

Method	Visible to infrared					Infrared to visible				
	Rank1	Rank10	Rank20	mAP	mINP	Rank1	Rank10	Rank20	mAP	mINP
Zero-Pad ^[5]	17.75	34.21	44.35	18.90	-	16.63	34.68	44.25	17.82	-
HSME ^[6]	50.85	73.36	81.66	47.00	-	50.15	72.40	81.07	46.16	-
AlignGan ^[11]	57.90	-	-	53.60	-	56.30	-	-	53.40	-
X-Modal ^[3]	62.21	83.13	91.72	60.18	-	-	-	-	-	-
DDAG ^[7]	69.34	86.19	91.49	63.46	49.24	68.06	85.15	90.31	61.80	48.62
AGW ^[8]	70.05	86.21	91.55	66.37	50.19	70.49	87.21	91.84	65.90	51.24
MCLNet ^[13]	80.31	-	-	79.83	-	73.43	-	-	69.49	-
CAJ ^[9]	85.03	95.49	97.54	79.14	65.33	84.75	95.33	97.51	77.82	61.56
MPANet ^[10]	83.70	-	-	80.90	-	82.80	-	-	80.70	-
GLMC ^[14]	91.84	97.86	98.98	81.42	-	91.12	97.86	98.69	81.06	-
MMN ^[4]	91.60	97.70	98.90	84.10	-	87.50	96.00	98.10	80.50	-
MID ^[15]	87.45	95.73	-	84.85	-	84.29	93.44	-	81.41	-
SPOT ^[16]	80.35	93.48	96.44	72.46	56.19	79.37	92.79	96.01	72.26	56.06
CMT ^[17]	95.17	-	-	87.30	-	91.97	-	-	84.46	-
MAUM ^[18]	87.87	-	-	85.09	-	86.95	-	-	84.34	-
Ours	95.22	98.54	99.17	87.70	74.48	93.67	98.35	98.83	86.43	71.68

表2 在SYSU-MM01数据集上与现有方法作对比

Tab.2 Comparison with existing methods on the SYSU-MM01 dataset

%

Method	All search					Indoor search				
	Rank1	Rank10	Rank20	mAP	mINP	Rank1	Rank10	Rank20	mAP	mINP
Zero-Pad ^[5]	14.80	54.12	71.33	15.95	-	20.58	68.38	85.79	26.92	-
HSME ^[6]	20.68	32.74	77.95	23.12	-	-	-	-	-	-
AlignGan ^[11]	42.40	85.00	93.70	40.70	-	45.90	87.60	94.40	54.30	-
X-Modal ^[3]	49.90	89.80	96.00	50.70	-	-	-	-	-	-
DDAG ^[7]	54.75	90.39	95.81	53.02	39.62	61.02	94.06	98.41	67.98	62.61
AGW ^[8]	47.50	84.39	92.14	47.65	35.30	54.17	91.14	95.98	62.97	59.23
MCLNet ^[13]	65.40	93.33	97.14	61.98	-	72.56	96.88	99.20	76.58	-
CAJ ^[9]	69.88	95.71	98.46	66.89	53.61	76.26	97.88	99.49	80.37	76.69
MPANet ^[10]	70.58	96.21	98.80	68.24	-	76.74	98.21	99.57	80.95	-
GLMC ^[14]	64.37	93.90	97.53	63.43	-	67.35	98.10	99.77	74.02	-
MMN ^[4]	70.61	96.20	99.00	66.90	-	76.20	97.20	99.30	79.60	-
MID ^[15]	60.27	92.90	-	59.40	-	64.86	96.12	-	70.12	-
SPOT ^[16]	65.34	92.73	97.04	62.25	48.86	69.42	96.22	99.12	74.63	70.48
CMT ^[17]	71.88	96.45	98.87	68.57	-	76.90	97.68	99.64	79.91	-
MAUM ^[18]	71.68	-	-	68.79	-	76.97	-	-	81.94	-
Ours	74.18	96.58	99.20	70.04	56.97	79.69	98.12	99.73	83.08	79.65

的预测结果,其中Local表示仅利用局部特征进行预测,Global表示仅利用全局特征进行预测,Global+Local表示联合使用全局特征和局部特征进行预测。具体结果如表4所示,从中可以观察到利用全局特征分支进行预测的效果最好。

3) 层次划分的效果。不同的层次划分数量决定了行人局部特征通道交互的粒度。以RegDB数据集为例,图4展示了不同层次划分的效果,其中横轴Part表示分块的数量。为了保证Part的有效性,避免出现垂直方向上特征无法均分导致信息丢失

表3 各组成部分对模型性能的影响
Tab.3 The impact of each component on model performance %

Method	Visible to infrared		
	Rank1	mAP	mINP
Base	87.72	79.14	64.09
Base+MCA	90.49	81.27	66.15
Base+PCA	90.44	82.34	68.16
Base+P	93.35	84.86	70.23
Base+MCA+PCA	92.28	82.40	67.10
Base+P+MCA	94.37	85.92	72.81
Base+P+PCA	94.03	85.96	72.33
Base+P+MCA+PCA	95.22	87.70	74.48

表4 不同特征进行预测的影响
Tab.4 The impact of predictions by different features %

Method	Visible to infrared		
	Rank1	mAP	mINP
Global	95.22	87.70	74.48
Local	94.39	87.09	74.78
Global+Local	94.73	87.51	74.45

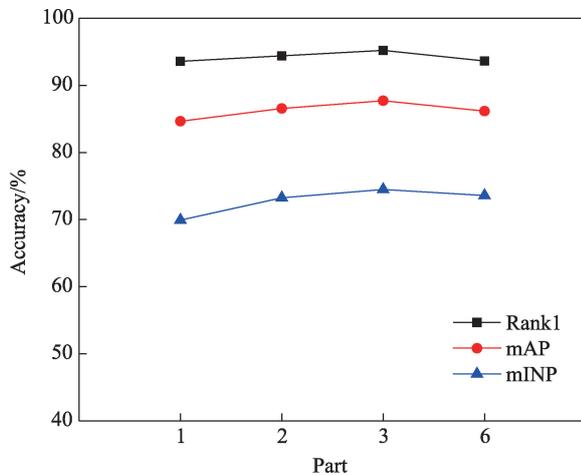


图4 不同层次划分的效果
Fig. 4 Effects divided by different levels

的情况,所以 Part 在此可以取 1, 2, 3, 6, 其中 Part=1 表示不做分块。可以观察到 Part=3 是分层提取局部行人特征的最佳设置。

3 结论

基于注意力机制,本文提出了一种端到端可见光-红外行人重识别模型。重点针对可见光图像和红外图像之间的模态差异,提出两种注意力机制,能够有效的提取判别性行人特征,主要结论如下。

1) 提出多尺度交叉注意力机制 MCA, 结合不同尺度的最大池化和空洞卷积, 扩大感受野, 获取多尺度的上下文信息。

2) 提出局部通道交互注意力机制 PCA, 增强了局部特征的通道交互能力, 对不同背景和遮挡等噪声具有更强的鲁棒性。

3) 通过设计网络结构, 结合全局特征和局部特征, 在 RegDB 数据集和 SYSU-MM01 数据集上均取得最优的效果。

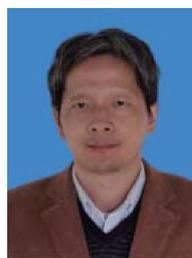
参考文献:

- [1] WANG G, ZHANG T, CHENG J, et al. RGB-infrared cross-modality person re-identification via joint pixel and feature alignment[C]//Seoul: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019.
- [2] ZHANG Q, LAI C, LIU J, et al. Fmcnet: Feature-level modality compensation for visible- infrared person re-identification[C]//New Orleans: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.
- [3] LI D, WEI X, HONG X, et al. Infrared-visible cross-modal person re-identification with an X modality[C]//New York: Proceedings of the AAAI Conference on Artificial Intelligence, 2020.
- [4] ZHANG Y, YAN Y, LU Y, et al. Towards a unified middle modality learning for visible-infrared person re-identification[C]//Chengdu: Proceedings of the 29th ACM International Conference on Multimedia, 2021.
- [5] WU A, ZHENG W S, GONG S, et al. RGB-IR person re-identification by cross- modality similarity preservation [J]. International Journal of Computer Vision, 2020, 128: 1765-1785.
- [6] HAO Y, WANG N, LI J, et al. HSME: Hypersphere manifold embedding for visible thermal person re-identification[C]//Honolulu: Proceedings of the AAAI Conference on Artificial Intelligence, 2019.
- [7] YE M, SHEN J B, CRANDALL D J, et al. Dynamic dual-attentive aggregation learning for visible-infrared person re- identification[C]//Glasgow: European Conference on Computer Vision - ECCV 2020, 2020.
- [8] YE M, SHEN J B, 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.
- [9] YE M, RUAN W, DU B, et al. Channel augmented joint learning for visible-infrared recognition[C]//Montreal: Pro-

- ceedings of the IEEE/CVF International Conference on Computer Vision, 2021.
- [10] WU Q, DAI P, CHEN J, et al. Discover cross-modality nuances for visible-infrared person re-identification[C]//Nashville: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
- [11] SUN Y, ZHENG L, YANG Y, et al. Beyond part models: Person retrieval with refined part pooling[C]//Munich: Proceedings of the European Conference on Computer Vision (ECCV), 2018.
- [12] 张泓, 范自柱, 石林瑞, 等. 一种基于多尺度特征融合的人头计数检测方法研究[J]. 华东交通大学学报, 2021, 38(2): 115-121.
- ZHANG H, FAN Z Z, SHI L R, et al. A head detection method based on multi-scale feature fusion[J]. Journal of East China Jiaotong University, 2021, 38(2): 115-121.
- [13] HAO X, ZHAO S, YE M, et al. Cross-modality person re-identification via modality confusion and center aggregation[C]//Montreal: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021.
- [14] ZHANG L Y, DU G D, LIU F, et al. Global-local multiple granularity learning for cross-modality visible-infrared person reidentification[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021: 34138719.
- [15] HUANG Z, LIU J, LI L, et al. Modality-adaptive mixup and invariant decomposition for RGB-infrared person re-identification[C]//Vancouver: Proceedings of the AAAI Conference on Artificial Intelligence, 2022.
- [16] CHEN C, YE M, QI M, et al. Structure-aware positional transformer for visible-infrared person re-identification[J]. IEEE Transactions on Image Processing, 2022, 31: 2352-2364.
- [17] JIANG K, ZHANG T, LIU X, et al. Cross-modality transformer for visible-infrared person re-identification[C]//Tel-Aviv: European Conference on Computer Vision, 2022.
- [18] LIU J, SUN Y, ZHU F, et al. Learning memory-augmented unidirectional metrics for cross-modality person re-identification[C]//New Orleans: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.



第一作者: 魏克铭(1998—), 男, 硕士研究生, 研究方向为深度学习、模式识别。E-mail: wkmqyr@163.com。



通信作者: 范自柱(1975—), 男, 博士, 教授, 博士生导师, 研究方向为模式识别、机器学习。E-mail: zzfan3@163.com。

(责任编辑: 吴海燕)