

基于多粒度注意力的可见光 - 红外行人重识别模态特异性算法

王娅迪

同济大学电子与信息工程学院, 上海

收稿日期: 2022年12月13日; 录用日期: 2023年1月9日; 发布日期: 2023年1月18日

摘要

可见光 - 红外行人重识别(Visible-infrared Person Re-identification, VI Re-ID)旨在根据裁剪后的查询行人样本, 检索另一模态的画廊集中相同身份的行人样本。现有基于特征融合的方法主要采用双流特征提取模型来解决此任务, 但由于两种模态的数据来源存在差异, 不同模态特征提取分支存在特征粒度不一致的问题。为了解决这一问题, 本文提出一种多粒度空间注意力模块, 并进一步提出模态特异性双注意力模块, 使用分散的细粒度注意力提取包含更多细节信息的可见光样本注意力描述符, 使用集中的粗粒度注意力提取包含更多模糊信息的红外样本注意力描述符。实验证明融入多粒度方法的注意力模块有效的提高了双流特征提取模型在SYSU-MM01和RegDB数据集上的精度。

关键词

跨模态行人重识别, 多粒度, 注意力机制, 深度学习, 计算机视觉

Multi-Granularity Modality Specific Attention Algorithm for Visible-Infrared Person Re-Identification

Yadi Wang

College of Electronic and Information Engineering, Tongji University, Shanghai

Received: Dec. 13th, 2022; accepted: Jan. 9th, 2023; published: Jan. 18th, 2023

Abstract

Visible-infrared Person Re-identification aims to recognize pedestrian images with the same iden-

tity as query pedestrian image in the gallery set of another modality. Existing methods based on two-stream feature extraction network have been widely used to solve this task. However, due to the different data sources, the feature granularities in different modality feature extraction branches are inconsistent. To alleviate this issue, we propose a multi-granularity spatial attention model and further proposes a modality specific attention module, which use decentralized fine-grained attention to extract the attention descriptor of visible samples containing more detailed information, and use concentrated coarse-grained attention to extract the attention descriptor of infrared samples containing more fuzzy information. Extensive experiments show that our method greatly improves the performance of the end-to-end framework on two widely used visible-infrared Person Re-identification datasets, SYSU-MM01 and RegDB.

Keywords

Cross-Modality Person Re-Identification, Multi-Granularity, Attention Mechanism, Deep Learning, Computer Vision

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来,随着智慧城市,智能安防等理念的提出和社会安定的需要,各种智能监控设备,如摄像头,传感器等组建成的网络在各大城市的建设中成为不可或缺的一部分,为满足大量跨设备场景下的城市安防监控场景,行人重识别(Person Re-Identification, ReID)任务应运而生。可见光到红外的行人重识别是跨模态行人重识别任务的重要研究方向,目的是给定待查询行人,确定该行人是否在另一模态相机捕捉的裁剪后的行人图像中出现。可见光摄像机捕捉可捕捉 RGB 格式的图像,其中记录了波长范围分别为 630~680 nm、525~600 nm、450~515 nm 的红波、绿波和蓝波的光强信息;红外摄像机可捕捉 IR 图像,其中记录了波长范围为 760~1000 nm 的短波近红外线的光强信息。与传统行人重识别相比,可见光到红外的行人重识别能更好进行光照不足这一特定场景对下的行人的重识别任务,例如夜晚及黑暗的室内。

可见光相机可以在光照条件良好的条件下清晰地捕捉到行人的细节信息和颜色信息,但在弱光照甚至完全黑暗的环境下很难捕捉到有效的目标信息;而红外相机对外部光照变化不敏感,通常可以呈现出人体清晰的轮廓,但会丢失了人体的细节信息和颜色信息。对于两种模态数据的研究在多个领域都有所涉及,且取得了良好的成果,例如可见光和红外的人脸识别、可见光和红外的目标检测这两大领域。在这两个领域中,红外信息作为补充信息加入,可有效丰富信息维度,提升识别和检测的准确率,因此研究重点大多集中在如何结合两种模态的信息。而可见光到红外的行人重识别作为跨模态任务中,红外信息作为一类独立信息,且与可见光信息不对齐,因此研究重点大多集中在如何将两种模态的特征映射到同一特征空间。

可见光-红外的行人重识别中也有注意力机制的相关应用,但较少有研究关注到由于两种模态间粒度不一致导致其适用的注意力机制粒度不一致的问题。其成因主要包含三点:红外摄像机和可见光摄像机成像原理差异,存储形式的差异以及红外相机的过曝问题。首先,由于成像原理的差异,可见光图像和红外图像描绘同一个人时的语义可能完全不同。由于红外光的波长更长,在空气中的散射更多,会失去一些纹理细节,红外图之间皮肤的颜色、反射率和光照也更难分辨,因此红外图更平滑。这也意味着

图中相似的纹理和颜色可以表示实际拍摄时不同的语义。同时，由于红外摄像机更倾向于捕捉发热的物体，而不是非热物体，这种热敏感性就会导致语义损失，这也可以看作是一种背景噪声的滤波。其次，在信息存储时，由于可见光相机成像时会捕捉红、绿、蓝这三个不同通道的数据，因此其单个样本会存储三个通道图像，但红外摄像机仅捕捉单个通道的数据，因此其单个样本的信息也仅包含单个通道图像，这就导致两类样本间信息粒度不一致。最后，当前监控设备的红外成像技术大多采用主动红外摄像技术，该技术需要在摄像设备上安装红外补光模块，通过红外光在物体表面的漫反射，采集成像。但红外补光设备排列成一定角度，导致补光设备照射在距离摄像机较近的物体表面时会发生强反射，多个补光模块的反射叠加导致反射强度过大，最终导致画面成像时出现过曝现象，造成图像细节丢失。介于以上提到的三个原因，通常来说同一身份行人样本的可见光图像包含更多细粒度语义信息，而红外图像包含更多粗粒度语义信息，两种特征的粒度不一致。但常用的双流提取网络中，首先通过相同结构的两个分支单独提取两种模态的特征，再将两种模态特征映射到相同的特征空间，没有关注到映射过程中两种模态特征的粒度不一致问题。

本文采用结合了 PCB (Part-Based Convolutional Baseline) [1] 的双流特征提取网络作为基准模型，目标是针对行人 VI-ReID 任务两种模态特征的粒度不一致问题，对现有注意力模块进行改进。本文的主要工作和贡献如下：

1) 针对可见光模态包含更多细节特征的问题，提出一种多粒度空间注意力模块，在空间维度上对可见光模态分支采用多粒度注意力机制，以提取细粒度的注意力描述符。

2) 针对模态特征粒度不一致问题，针对空间注意力和通道注意力的整体框架，提出基于多粒度的模态特异性双注意力模块，级联聚合两种维度上的注意力描述。对可见光模态分支采用多特征块的细粒度注意力描述；对红外模态分支采用单特征的粗粒度注意力描述。

3) 实验证明我们的方法极大的提高了 VI-ReID 方法在 SYSU-MM01 和 RegDB 数据集上的性能。

2. 相关工作

常见的行人 ReID 方法可分为两类，基于表征学习的方法和基于度量学习的方法[2]。首先通过传统方法或鲁棒性强的 CNN 网络提取行人特征，之后前者将重识别视为多分类问题，将每个行人身份 ID 作为一个类别[3] [4]；后者旨在计算两个行人的相似度，判断两者是否为同一行人[5] [6]。行人 VI-ReID 是跨模态行人重识别中的子类，其研究重点集中于缩小从相同行人的不同模态数据中提取特征的差距。RGB 图像有三通道信息，而 IR 图像只有单通道信息；且由于成像波长范围不同，传统重识别中 RGB 图提取的颜色等特征，在 IR 图中缺失，因此传统重识别网络无法直接应用于可见光到红外的行人重识别任务。目前主流的可见光到红外的行人重识别方法可分为两大类，一类是基于特征融合的方法，该类方法致力于将两种模态提取到的特征映射到同一特征空间；另一类是基于模态转换的方法，该类方法通过图像生成技术将其中一种模态的特征进行转换，后续可采用同模态下的重识别算法进行处理。

2.1. 基于特征融合的方法

Wu 等人[7]首次采用深度零填充的单流网络，对 RGB 图和 IR 图的特征进行跨模态匹配。对 RGB 图，将三通道信息转化为灰度一通道信息，放置在通道一，零填充放置在通道二；对 IR 图，将数据直接放置在通道二，零填充放置在通道一，采用单流网络区分来自不同域的特征。cmGAN [8]则通过生成对抗网络缩小跨模态差异，采用生成器提取特征向量，并映射到相应行人身份，鉴别器区分不同模态，约束两种模态特征向量处于同一空间。

依赖于人体结构的特殊性，各个部件位置分布相对固定，因此部分学者关注图像局部特征，最常见

的方法是采用 PCB 模块[1], 将提取后的特征进行均匀的横向切分, 共通过全局平均池化处理后接入卷积层缩减特征维度[9]。DEF [10]对每个切分后的特征分别预测, 并用 JS 散度约束两种模态的特征向量。Liu 等人[11]对特征切分的数量进行了实验, 并探讨了在双流网络的不同位置共享参数权重对网络性能的影响。由于人体姿态、条件遮挡等问题, 横向切分后的身体无法准确对齐, Alignedreid [12]中通过动态规划算法实现两份图像的部分级对齐问题, 增强目标细粒度刻画准确度。

注意力机制也广泛应用于行人 VI-ReID 算法。如 Ye 等人提出的动态双注意力聚合学习网络 DDAG [13], 不仅可以通过模态内的部分加权注意力, 自适应匹配不同身体部位权重, 还引入跨模态图结构注意力, 来挖掘两种模态间的特征关系。EDFL [14]关注到 ResNet 中不同卷积块所关注的特征区域不同, 高层网络会关注更多语义信息, 因此通过跳层连接将中间层特征融合, 以获得更丰富的特征。MPANet [15]中参考 SE-net 引入空间注意力以改善模态不一致性, 并在不同特征图上引入注意力机制以发现不同模式的细微差别。Huang 等人[16]为提取不同通道上的信息关系, 提出 PT3D, 将特征以空间维度分块, 并采用 3DCNN 进行卷积, 方法中提出的模态共享特征提取模块同时捕捉了模态特定特征和模态共享特征在通道维度上的关系。

2.2. 基于模态转换的方法

由于可见光模态图像和红外模态图像在通道数、成像方式上的差异, 一些学者通过图像生成算法, 将一种模态图像转换为另一种模态, 再采用通用重识别方法进行后续处理[17]。AlignGAN [18]首次采用 CycleGAN 将红外图转为 RGB 图, 并引入另一个 GAN 进行模态判别。D²RL 方法[19]通过自编码器进行风格分离, 通过 GAN 生成另一模态图像, 再采用特征提取网络缩小特征差异。Choi 等人在 Hi-CMD [20]中采用层次化方法解纠缠, 利用编码器分离出的模态特异属性和模态共享属性, 在从模态共享属性中分离出身份鉴别属性和身份无关属性, 将分离后的特征再纠缠以重构图像, 最终只采用身份鉴别信息判别身份。Wang 等人在 JSIA-ReID [21]提出对齐的图像对更利于判别, 因此采用编码器分别提取图像风格特征和内容特征, 并通过解纠缠后的另一模态风格特征和原始模态内容特征生成配对图像, 并在特征对齐模块分别进行集合对齐和实例对齐。但 GAN 只能实现 IR 图到 RGB 图一对一的生成, 事实上同一张 IR 图可能对应多种颜色的 RGB 图像, 同时生成的图像会破坏部分原有的信息并导致一定程度的噪声, 因此基于模态转换的方法训练难度较大。

3. 基于多粒度的模态特异性注意力模块

计算机视觉领域中, 对注意力机制的关注由来已久, 随着任务复杂性的不断加深, 如何在特征提取阶段提取到有辨别力的特征, 也成为各路学者争相研究的课题。注意力机制能够让模型关注到特征中更为重要的区域, 抑制背景等非重要区域的干扰, 对最终的任务完成有积极意义。本节首先介绍了本文采用的双流特征提取网络基线模型, 然后针对行人 VI-ReID 任务中的特征粒度不一致问题, 介绍改进后的模型。

3.1. 双流特征提取网络及空间通道注意力模块

双流特征提取网络是当前基于特征融合的 VI-ReID 方法中最常用的基线框架, 如图 1(a), 首先通过两个参数不同结构相同的特征提取模块提取两种模态的特异性特征, 再通过参数相同的网络结构和相应损失函数, 将不同特征空间下的特征约束到相同的特征空间。由于行人重识别任务的特殊性, 输入图像的相对方向大多是相同的, 即与身体结构相对应。身体结构是一个人的固有特征, 无论行人身份, 其均为人身体的不变信息。因此常用横向均匀切分的 PCB 模块提取图像的对应位置局部特征, 并采用广义平

均池化层将三维的特征块向量映射为一维的特征块向量，然后采用 1×1 的卷积层降低特征维数，最终提取到行人的全局和局部特征向量。

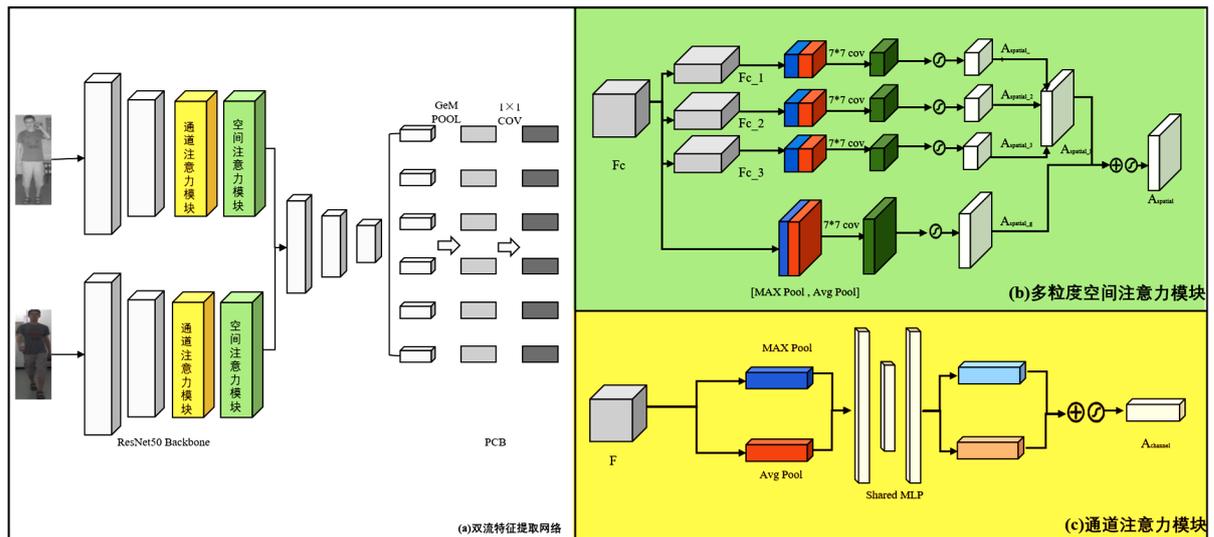


Figure 1. Diagram of Multi-granularity attention module structure
图 1. 多粒度注意力模块结构图

3.2. 多粒度空间注意力模块

卷积操作进行特征提取时，是通过卷积核在原始输入中不断滑动，最终计算出特征图的过程，其单个子过程为卷积核矩阵与覆盖的输入矩阵之间对应值相乘相加，得到特征矩阵上对应位置的单个值，最终将所有单个值进行组合，得到最终输出。输出特征矩阵上每个值的位置，由对应输入矩阵中相应值的位置决定。由于卷积操作中严格的映射关系，使得其输出特征上的表达是原始特征图上的平移，存在空间不变性。即存在于原始输入矩阵中左下角的特征，经过卷积操作后也位于输出矩阵的左下角。介于卷积操作在空间上的不变性，可将原始图像中空间上的划分简单映射至提取后特征块空间上的划分。行人 VI-ReID 中的样本图像主要是站立的行人，因此在模块的空间维度上对特征图进行横向划分，可映射至输入图像中人体部件的各个组成部分的特征。

由于可见光模态的特征相较于红外模态特征而言，具有更为丰富的特征信息，因此我们首先提出多粒度空间注意力(Multi-Granularity Spatial Attention, MGSA)模块，如图 1(b)。该模块将通道注意力模块的输出 F_c 作为输入，首先沿着空间维度将特征块横向切分为 I_l 个块，再分别对每个分块进行最大池化及平均池化，然后通过卷积网络降维及 sigmoid 激活函数后得到每个分块的空间注意力描述 $A_{spatial_l}$ ，最终所有分块的注意力描述符级联后的局部注意力描述符 $A_{spatial_l}$ 与原始全局注意力描述符 $A_{spatial_g}$ 进行对应位置的加和，最终通过 sigmoid 激活函数得到最终的空间注意力描述 $A_{spatial}$ ，如式(1)。

$$A_{spatial_g} = \delta \left(f^{7 \times 7} \left(\left[AvgPool(F_c); MaxPool(F_c) \right] \right) \right)$$

$$A_{spatial_l} = concat \left[\delta \left(f^{7 \times 7} \left(\left[AvgPool(F_{c_i}); MaxPool(F_{c_i}) \right] \right) \right) \right]$$

$$A_{spatial} = \delta \left(A_{spatial_g} + A_{spatial_l} \right) \quad (1)$$

其中 $f^{7 \times 7}$ 表示 7×7 的卷积层， δ 表示 sigmoid 激活函数。单粒度的全局注意力会使得模型更多关注在样

本整体有区分性的特征上, 本文提出的 MGSA 能迫使网络平等的关注每个分块区域内有辨别性的特征, 并重新分部他们之间的权重, 使得网络能更多注意到包含更多细粒度信息的可见光模态中的局部细节信息。

3.3. 多粒度模态特异性注意力框架

MGSA 仅在空间维度划分了不同粒度的注意力描述, 通道层面上仍将提取到的所有特征作为一个整体。因此本节基于能同时关注到通道维度和空间维度重要信息的通道空间混合注意力模块(Convolutional Block Attention Module, CBAM) [22], 针对 VI-ReID 任务提出模态特异性双注意力模块(Modality Specific Attention Module, MSAM), 在通道和空间维度针对不同模态分支进行不同的粒度划分, 如图 2。

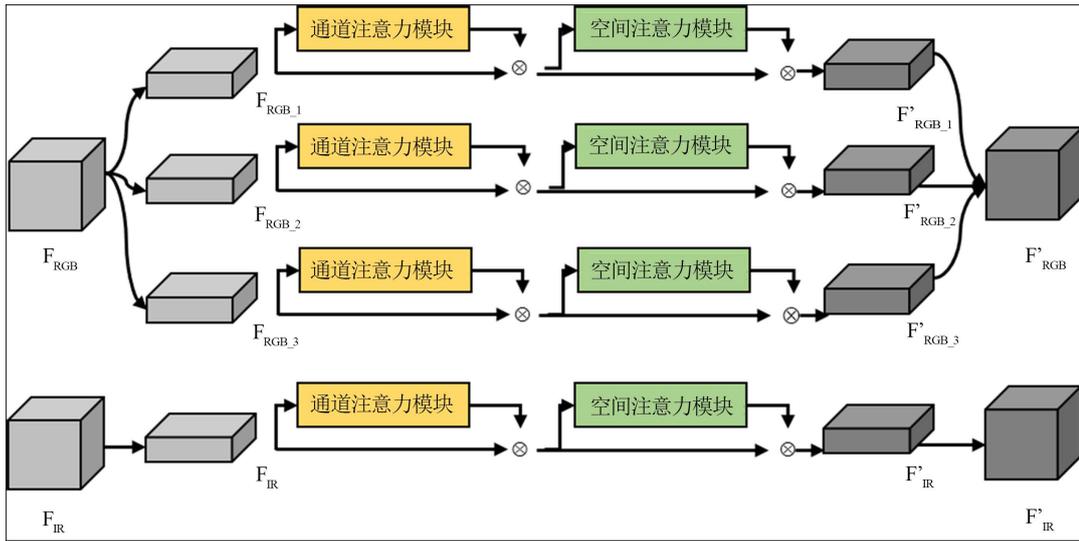


Figure 2. Diagram of modality specific attention module structure
图 2. MSAM 模块结构图

对于可见光模态的样本, 其属于多个粒度信息的整合, 包含更多细粒度信息, 因此在 MSAM 模块中将输入特征 F_{RGB} 横向划分为 I_2 个特征块, 得到包含 I_2 个元素的特征块集合 $\{F_{RGB_1}, F_{RGB_2}, \dots, F_{RGB_{I_2}}\}$, 以避免模型过于关注某些突出的特征, 忽略其他更为细节的特征。经过第 i 个子特征块通道注意力描述符 $A_{RGB_channel}$ 加权后得到 $F_{RGB_{i}}$, 然后通过空间注意力描述符 $A_{RGB_spatial}$ 加权后得到 F'_{RGB_i} , 最终所有特征块沿空间维度拼接后得到两种注意力模块级联增强后的特征输出 F'_{RGB} 。

$$\begin{aligned}
 F_{RGB_{i}} &= A_{RGB_channel} \otimes F_{RGB_i} \\
 F'_{RGB_i} &= A_{RGB_spatial} \otimes F_{RGB_{i}} \\
 F'_{RGB} &= \{F'_{RGB_1}, F'_{RGB_2}, \dots, F'_{RGB_{I_2}}\}
 \end{aligned} \tag{2}$$

其中 \otimes 表示向量对应位置点乘。相对而言红外模态的样本只包含单个粒度的信息, 同时由于红外相机容易过曝等原因, 导致红外样本所包含的特征更多为粗粒度特征, 因此直接采用单粒度的 CBAM 捕捉全局注意力描述符。其中的空间注意力模块均为单粒度分支, 如图 1(b), 首先通过并行的一个全局平均池化层及一个全局最大池化层对空间维度进行压缩, 提取到两个不同的空间上下文描述符, 其中平均池化对图中的每一个像素点都有反馈, 最大池化能反馈出较为突出的像素点。两个描述符经过共享参数的多层感知机(Multilayer Perceptron, MLP), 以保证对应位置的描述符形成相同的映射, 接着输出逐个元素求和的融合向量, 最后经过 sigmoid 激活函数, 将注意力权重映射到 0-1 区间。

4. 实验结果

4.1. 数据集

RegDB: RegDB 是由双摄像机(一个可见摄像机和一个热摄像机)系统构建, 每张可见光图像都有一一对应的红外图像, 最初提出是为了通过红外图像辅助可见光图像的重识别任务, 即多模态行人重识别任务, 因为红外图像可以有效减少噪声、光照变化、背景环境变化等对重识别任务带来的负面影响, 现在被广泛应用于跨模态行人重识别任务。整个数据集包含 412 个行人, 每个行人包含 10 张可见光相机拍摄的 RGB 图像和 10 张红外相机拍摄的 IR 图像, 每一个行人都有一个专属的身份类别标签, 其中包含 158 个男性和 254 个女性, 并且有 156 个行人从正面拍摄, 有 256 个行人从背面拍摄。整体来说相同身份行人在姿态上变化较小, 相同行人的 10 张图片中, 天气状况, 拍摄视角都是相同的, 检索难度较低。

SYSU-MM01: SYSU-MM01 是一个由 6 台摄像机收集的大规模数据集, 包括 4 台可见光摄像机和 2 台红外摄像机, 拍摄场地均在 SYSU 校园。其中相机 1、2 为可见光相机, 被放置在明亮的室内, 相机 3 为红外相机, 被放置在与 2 号相机相同房间的黑暗条件下; 相机 4、5 为可见光相机, 被放置在包含花园等场景的明亮的室外, 相机 6 为红外相机, 被放置在背景杂乱的室外通道。训练集包含 395 个行人, 其中可见光图像 22,258 张, 红外图像 11,909 张; 测试集包含 96 人, 其中 301 张红外图像作为查询图, 3803 张可见光图像作为画廊集。

4.2. 评价指标

本文采用 mAP (Mean Average Precision)、CMC (Cumulative Matching Characteristics)和 mINP (Mean Inverse Negative Penalty, mINP)作为衡量行人 VI-ReID 性能的标准指标。其中 mINP 用于衡量模型检索困难样本的能力。

4.3. 实验设置

本文基于 PyTorch 框架实现了双流特征提取网络以及提出的多粒度注意力模块, 并在 RTX 3090 显卡上运行实验。训练部分的网络参数采用 ImageNet 的预训练参数进行初始化。主干网络采用 Resnet50, 并将最后一个卷积块的步长由 2 改为 1, 以获得更细粒度的特征映射。训练阶段, 输入图像的尺寸调整为 288×144 , 再进行随机左右翻转的数据增强。训练过程采用随机梯度下降(SGD)优化器进行优化, 动量大小设置为 0.9, 衰减因子设置为 0.0005。模型训练共迭代了 60 个周期, 学习率设置为 0.1, 前 10 个周期采用预热策略, 在第 20、50 个周期时学习率依次衰减 10 倍。PCB 模块划分参数设置为 6。

4.4. 实验结果与分析

本文在两个公开数据集上对实验结果进行消融实验, 其中测试阶段仅采用余弦相似性进行距离度量, 未采用重排序算法。表 1 给出本文方法与基线方法[6]在检索精度上的对比结果。其中 baseline + MGSA 为在 baseline 中 ResNet50 的第二个 stage 与第三个 stage 之间, 对红外模态分支添加 CBAM, 对可见光模态分支添加空间注意力模块为 MGSA 的 CBAM 的实验结果, 从表中数据可以看到, MGSA 在基线网络的基础上有较大的提升。baseline + MSAM 的结果也展示出, 单独加入 MSAM 也能对模型产生正向的效果。同时加入 MGSA 和 MSAM, 即在 MSAM 的可见光模态分支中分别采用多分支的单粒度通道注意力描述符和多粒度空间注意力描述符的结构, 红外模态分支中分别采用单分支的单粒度通道注意力描述符和单粒度空间注意力描述符的结构, 能在半数指标上取得最优效果, 并在所有指标上优于基线。实验证明了 MGSA 和 MSAM 的优越性, 以及两模块间的相互促进作用, 后续本节在 SYSU-MM01 上进行了更多对比实验。

Table 1. Result of proposed method on SYSU-MM01 and RegDB
表 1. 标准本文方法在 SYSU-MM01 和 RegDB 上的实验结果

方法	SYSU-MM01			RegDB		
	top-1	mAP	mINP	top-1	mAP	mINP
baseline	59.14%	54.86%	38.22%	92.61%	87.95%	79.57%
Baseline + MGSA	61.60%	56.34%	45.37%	92.83%	89.24%	82.47%
Baseline + MSAM	60.35%	57.98%	40.71%	91.73%	88.01%	80.52%
Baseline + MGSA + MSAM	61.23%	57.96%	45.82%	92.31%	89.39%	82.94%

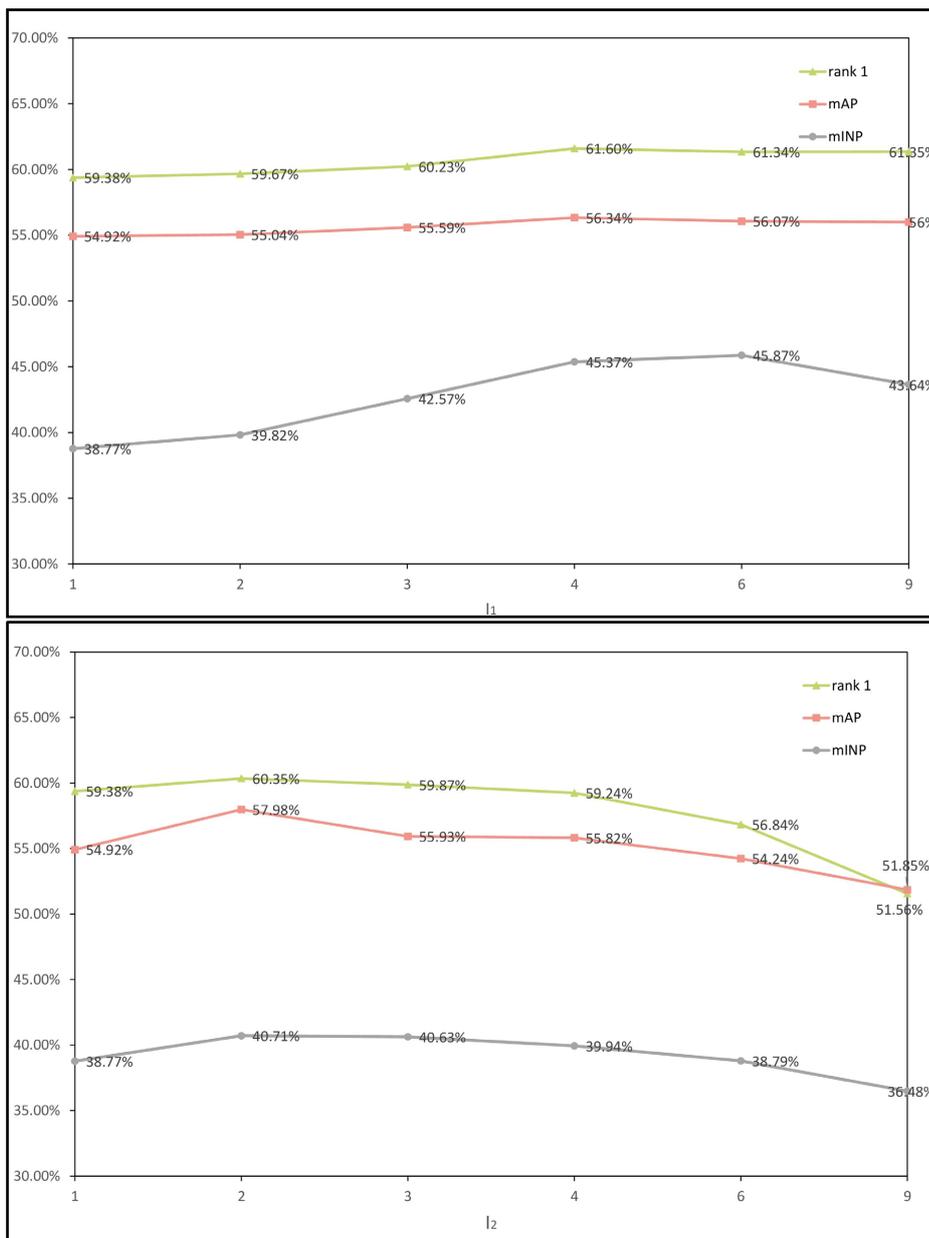


Figure 3. Parameter experiment results. The above image is experiment of parameter division I_1 in MGSA; the image below is experiment of parameter division I_2 in MSAM

图 3. 参数实验结果。上图为 MGSA 划分参数 I_1 实验；下图为 MSAM 划分参数 I_2 实验

MGSA 中划分参数 I_1 对模型性能的影响: 图 3 上图给出了 MGSA 中参数 I_1 的不同取值的对比图。特别的, 当 $I_1 = 1$ 时, MGSA 与 CBAM 中原始的空间注意力模块相同。最后, 我们可以从图中观察到, 当 $I_1 = 4$ 时, MGSA 效果最好, 相较于基线的 rank1 值和 mAP 值分别提升了 2.22% 和 1.42%。

MSAM 中划分参数 I_2 对模型性能的影响: 在 SYSU-MM01 数据集上, 图 3 下图给出了 MSAM 中参数 I_1 的不同取值的对比图。特别的, 当 $I_2 = 1$ 时, MSAM 与相当于在两个分支中插入权重不共享的 CBAM 模块。最后, 我们可以从图中观察到, 当 $I_2 = 2$ 时, MSAM 效果最好, 相较于基线的 rank1 值和 mAP 值分别提升了 0.97% 和 3.06%。

5. 结束语

本文针对现有行人 VI-ReID 任务中的模态特征粒度不一致问题, 在双流特征提取网络的基础模型中融入注意力机制和多粒度思想, 提出了一种多粒度的空间注意力提取模块, 并在此基础上提出了一种模态特异性的双注意力特征提取框架, 使得模型对包含更多细节信息的细粒度可见光样本, 通过分布更为均匀的特征描述符平等的关注每个特征块中有区分性的特征; 对包含更多模糊信息的粗粒度红外样本, 通过全局的特征描述符关注整体图像上有区分性的特征。本文方法有效的提高了算法在 SYSU-MM01 和 RegDB 数据集上的精度。

参考文献

- [1] Sun, Y., Zheng, L., Yang, Y., Tian, Q. and Wang, S. (2018) Beyond Part Models: Person Retrieval with Refined Part Pooling (and a Strong Convolutional Baseline). *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 8-14 September 2018, 480-496. https://doi.org/10.1007/978-3-030-01225-0_30
- [2] Zheng, L., Yang, Y. and Hauptmann, A.G. (2016) Person Re-Identification: Past, Present and Future.
- [3] Zheng, Z., Zheng, L. and Yang, Y. (2017) Unlabeled Samples Generated by Gan Improve the Person Re-Identification Baseline *in Vitro*. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 3754-3762. <https://doi.org/10.1109/ICCV.2017.405>
- [4] Qian, X., Fu, Y., Jiang, Y.G., Xiang, T. and Xue, X. (2017) Multi-Scale Deep Learning Architectures for Person Re-Identification. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 5399-5408. <https://doi.org/10.1109/ICCV.2017.577>
- [5] Moutafis, P., Leng, M. and Kakadiaris, I.A. (2016) An Overview and Empirical Comparison of Distance Metric Learning Methods. *IEEE Transactions on Cybernetics*, 47, 612-625. <https://doi.org/10.1109/TCYB.2016.2521767>
- [6] Deng, W., Zheng, L., Ye, Q., Kang, G., Yang, Y. and Jiao, J. (2018) Image-Image Domain Adaptation with Preserved Self-Similarity and Domain-Dissimilarity for Person Re-Identification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-22 June 2018, 994-1003. <https://doi.org/10.1109/CVPR.2018.00110>
- [7] Wu, A., Zheng, W. S., Yu, H. X., Gong, S. and Lai, J. (2017) RGB-Infrared Cross-Modality Person Re-Identification. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 5380-5389. <https://doi.org/10.1109/ICCV.2017.575>
- [8] Dai, P., Ji, R., Wang, H., Wu, Q. and Huang, Y. (2018) Cross-Modality Person Re-Identification with Generative Adversarial Training. *IJCAI*, 1, 6. <https://doi.org/10.24963/ijcai.2018/94>
- [9] Chen, Y., Wan, L., Li, Z., Jing, Q. and Sun, Z. (2021) Neural Feature Search for Rgb-Infrared Person Re-Identification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 587-597. <https://doi.org/10.1109/CVPR46437.2021.00065>
- [10] Hao, Y., Wang, N., Gao, X., Li, J. and Wang, X. (2019) Dual-Alignment Feature Embedding for Cross-Modality Person Re-Identification. *Proceedings of the 27th ACM International Conference on Multimedia*, Nice, 21-25 October 2019, 57-65. <https://doi.org/10.1145/3343031.3351006>
- [11] Liu, H., Tan, X. and Zhou, X. (2020) Parameter Sharing Exploration and Hetero-Center Triplet Loss for Visible-Thermal Person Re-Identification. *IEEE Transactions on Multimedia*, 23, 4414-4425. <https://doi.org/10.1109/TMM.2020.3042080>
- [12] Zhang, X., Luo, H., Fan, X., Xiang, W., Sun, Y., Xiao, Q., Sun, J., et al. (2017) Alignedreid: Surpassing Human-Level Performance in Person Re-Identification.

-
- [13] Ye, M., Shen, J., J Crandall, D., Shao, L. and Luo, J. (2020) Dynamic Dual-Attentive Aggregation Learning for Visible-Infrared Person Re-Identification. In: *European Conference on Computer Vision*, Springer, Cham, 229-247. https://doi.org/10.1007/978-3-030-58520-4_14
- [14] Liu, H., Cheng, J., Wang, W., Su, Y. and Bai, H. (2020) Enhancing the Discriminative Feature Learning for Visible-Thermal Cross-Modality Person Re-Identification. *Neurocomputing*, **398**, 11-19. <https://doi.org/10.1016/j.neucom.2020.01.089>
- [15] Wu, Q., Dai, P., Chen, J., Lin, C. W., Wu, Y., Huang, F., Ji, R., *et al.* (2021) Discover Cross-Modality Nuances for Visible-Infrared Person Re-Identification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 4330-4339. <https://doi.org/10.1109/CVPR46437.2021.00431>
- [16] Huang, N., Liu, J., Luo, Y., Zhang, Q. and Han, J. (2023) Exploring Modality-Shared Appearance Features and Modality-Invariant Relation Features for Cross-Modality Person Re-Identification. *Pattern Recognition*, **135**, Article ID: 109145. <https://doi.org/10.1016/j.patcog.2022.109145>
- [17] Fan, X., Luo, H., Zhang, C. and Jiang, W. (2020) Cross-Spectrum Dual-Subspace Pairing for RGB-Infrared Cross-Modality Person Re-Identification.
- [18] Wang, G.A., Zhang, T., Cheng, J., Liu, S., Yang, Y. and Hou, Z. (2019) RGB-Infrared Cross-Modality Person Re-Identification via Joint Pixel and Feature Alignment. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 3623-3632. <https://doi.org/10.1109/ICCV.2019.00372>
- [19] Wang, Z., Wang, Z., Zheng, Y., Chuang, Y.Y. and Satoh, S.I. (2019) Learning to Reduce Dual-Level Discrepancy for Infrared-Visible Person Re-Identification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 618-626. <https://doi.org/10.1109/CVPR.2019.00071>
- [20] Choi, S., Lee, S., Kim, Y., Kim, T. and Kim, C. (2020) Hi-CMD: Hierarchical Cross-Modality Disentanglement for Visible-Infrared Person Re-Identification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 10257-10266. <https://doi.org/10.1109/CVPR42600.2020.01027>
- [21] Wang, G. A., Zhang, T., Yang, Y., Cheng, J., Chang, J., Liang, X. and Hou, Z.G. (2020) Cross-Modality Paired-Images Generation for RGB-Infrared Person Re-Identification. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 12144-12151. <https://doi.org/10.1609/aaai.v34i07.6894>
- [22] Woo, S., Park, J., Lee, J.Y. and Kweon, I.S. (2018) Cbam: Convolutional Block Attention Module. *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, 8-14 September 2018, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1