**基于辅助模态和混合注意力的跨模态行人重识别**

梁靖贤，谭台哲

（广东工业大学计算机学院 广州 510000）

**摘要：**目前，跨模态行人重识别面临的主要挑战之一是模态间存在巨大差异。现有方法主要通过模态特定信息补偿或学习模态共享特征来缓解模态差异。但是，模态特定信息补偿需要复杂的计算量，且容易引入噪声。同时，现有方法一般直接提取模态共享特征，尚未充分挖掘细微的、更具鉴别性的特征信息。因此，本文提出了一个基于辅助模态和混合注意力的双流参数共享网络来减缓模态差异，同时挖掘细微的、更具鉴别性的特征信息。该网络使用辅助模态生成器把可见光图像和红外图像投影到统一的辅助模态空间来生成辅助模态图像，辅助模态图像能有效地缓解模态差异。同时，设计了一个混合注意力引导学习模块，该模块通过混合注意力和实例归一化来引导模型关注细微的、更具鉴别性的特征信息。在提取特征图后，将其水平均匀切分成若干块，进行局部特征学习。在SYSU-MM01数据集全局搜索模式下Rank-1和mAP分别达到了67.08%和63.93%，实验证明该方法的有效性。

**关键词：**跨模态；行人重识别；辅助模态；混合注意力

Cross-modality person re-identification model based on auxiliary modality and mix attention mechanism

Liang JingXian, TAN TaiZhe

(School of Computer, Guangdong University of Technology Guangzhou Guangzhou 510000)

**Abstract** One of the primary challenges in cross-modal pedestrian re-identification is dealing with significant inter-modal differences. Existing methods mainly address these differences by either compensating for modality-specific information or learning modality-shared feature. However, modality-specific information compensation requires complex computational effort and is prone to introduce noise. Meanwhile, existing methods typically extract modality-shared features directly, without fully exploring the subtle and more discriminative feature information. To overcome these limitations, we propose a two-stream parameter sharing network based on auxiliary modality and hybrid attention, aiming to mitigate modal differences while mining subtle and more discriminative feature information. Our network employs an auxiliary modality generator to project visible and infrared images into a unified auxiliary modality space, generating auxiliary modality images that effectively mitigate modal differences. Additionally, we design a hybrid attention-guided learning module, which uses hybrid attention and instance normalization to guide the model to focus on subtle and more discriminative feature information. After extracting the feature map, we horizontally and uniformly slice it into blocks for local feature learning. In the All-search mode of SYSU-MM01 our method achieves a Rank-1 accuracy of 67.08% and a mean Average Precision (mAP) of 63.93%, demonstrating the superiority of our method

**Key Words** Cross-Modality; Person Re-identification; Auxiliary Modal; Hybrid Attention

1 引 言

行人重识别（Person Re-Identification, ReID）是智能监控系统中的关键技术之一，用于在给定的视频或图像序列中对特定行人进行识别。其目标是解决不同摄像头之间的行人检索问题。近年来，ReID已经受到广泛关注并取得了不错的进展[1-3]。然而，可见光摄像头在弱光或无光的环境（例如夜晚）下，往往无法捕获到有效的行人信息。相比之下，红外相机可以在黑暗环境中从场景中获取足够的信息。但是，由于可见光图像和红外图像之间存在巨大的差异，传统的单模态行人重识别模型难以处理这两种不同模态图像之间的行人重识别任务。因此，为了更有效地进行两种模态图像之间的行人重识别，研究人员提出了基于可见光——红外图像的跨模态行人重识别问题（Visible-Infrared Person Re-Identification, VI ReID）。

目前，解决跨模态行人重识别问题的方法可以分为以下两种：模态共享特征学习和模态特定信息补偿。模态共享特征学习主要通过特征对齐的方式，将红外图像和可见光图像投影到一个统一的模态共享空间中，然后利用相关的损失函数进行度量学习。这样做可以将不同模态但相同身份的行人特征距离拉近，同时将同一模态但不同身份的行人特征距离推远。例如，2017年，Wu等人[4]提出了一种能够在共享空间中学习模态共享特征的深度零填充网络，并创建了第一个大规模的可见光-红外行人图像数据集SYSU-MM01。Ye等人[5]则提出了双向双约束top-ranking损失，用于缓解模态内和模态间的差异，从而帮助模型提取具有鉴别性的行人特征信息。此外，Ye等人[6]还提出了一种基于双流网络的分层跨模态匹配模型，对模态特定特征和模态共享特征进行联合度量优化。而Liu等人[7]首先从输入图像中提取模态共享特征，然后通过水平分割的方式获得更具鉴别性的模的局部特征。模态特定信息补偿主要使用生成对抗网络来生成伪图像，从现有的模态特定信息中生成缺失的模态特定信息，以缓解不同模态间的差异。例如，Wang等人[8]提出了AlignGAN，该网络首先通过原始的可见光图像生成红外伪图像，实现像素级的特征对齐，然后将原始的红外图像和生成的红外伪图像进行匹配。另外，Wang等人[9]提出了一种基于双流网络的双层差异减少学习策略（D2RL），该策略主要通过可见光图像和红外图像之间的相互转化来减缓模态间的差异。还有Li等人[10]设计了一个轻量级网络，该网络能够将可见光图像生成一种新的辅助模态图像，从而减缓可见光模态和红外模态之间的模态差异。

通过分析研究现有方法，发现模态特定信息补偿需要复杂的生成对抗网络，且生成的伪图像容易引入噪声，从而影响模型性能。同时，由于可见光图像和红外图像之间的模态差异巨大，直接提取模态共享特征十分困难，且无法充分挖掘细微的、更具鉴别性的特征信息。为了解决上述问题，本文设计了一个基于辅助模态和混合注意力的双流参数共享网络，旨在减缓两种模态之间的差异，同时发掘更具鉴别性和鲁棒性的特征信息。所提出的网络基于Resnet50[11]进行改进。首先，通过一个轻量级的辅助模态生成器，将可见光图像和红外图像投影到一个统一的辅助模态空间，以此生成辅助模态图像。然后，将这些辅助模态图像与原始图像一起输入主干网络，以提取特征信息，从而有效减缓模态之间的差异。在主干网络中嵌入混合注意力引导学习模块，该模块使用混合注意力机制和实例归一化，有助于排除背景噪声的干扰，引导模型关注图像中的细节信息，提取更具鉴别性的特征。在提取特征图后，将其水平均匀切分成若干块，进行局部特征学习。最后，联合使用分布一致性损失、三元组损失、身份损失来对网络进行端到端训练。

2 算法原理

2.1网络总体结构

本文的网络总体结构如下图1所示。该网络基于Resnet50进行改进，首先，将可见光图像与红外图像输入至辅助模态生成器，以生成辅助模态图像。然后，辅助模态图像和原始图像一同输入至主干网络提取模态共享特征。主干网络中的第一个卷积层和第一个残差块参数独立的，用于提取模态特定特征，后面三个残差块参数共享，用于提取模态共享特征。同时，在第三个残差块后嵌入混合注意力引导学习模块，该模块通过混合注意力机制和实例归一化来缓解模态间的差异，同时充分挖掘具有鉴别性的行人图像信息。在主干网络提取特征图后，将特征图水平划分为若干个部分，进行局部特征学习。最后，联合使用分布一致性损失、三元组损失、身份损失来对模型进行端到端训练。

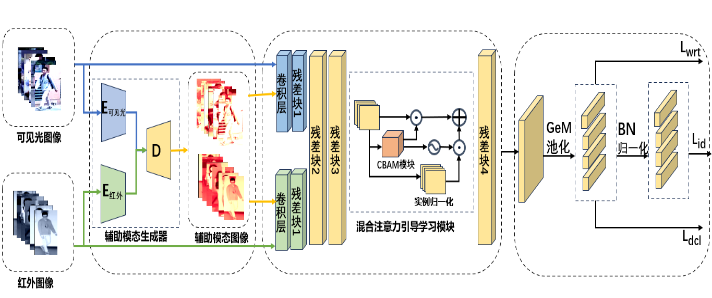


图1 网络总体架构图

2.2 辅助模态生成器

辅助模态生成器采用编码器-解码器架构，其主要功能是将可见光图像和红外图像映射到一个统一的辅助模态空间，以生成辅助模态图像。这样生成的辅助模态图像有效地缓解了可见光模态和红外模态之间的模态差异。具体而言，辅助模态生成器由两个参数独立的编码器和一个参数共享的解码器组成。首先，输入图像的尺寸被调整为3x384x192。由于红外图像是单通道的，而可见光图像是三通道的，因此需要将红外图像的通道数设置为3，与可见光图像的通道数保持一致。然后，可见光图像和红外图像分别输入对应的模态信息编码器进行编码。在编码过程中，首先通过一个卷积核为3x1x1的卷积层，将三通道图像转换为单通道图像。接着，通过卷积核为1x1x1的卷积层以减少计算量，并最后通过批归一化层进行数据归一化。随后，归一化后的数据输入至参数共享的解码器进行解码。解码器将编码后的单通道图像通过卷积核为1x1x1的卷积层转换为三通道图像，从而生成对应的辅助模态图像。辅助模态生成器能够将可见光图像和红外图像转换为辅助模态图像，从而缓解模态间的差异，为后续模态共享特征的提取提供帮助。

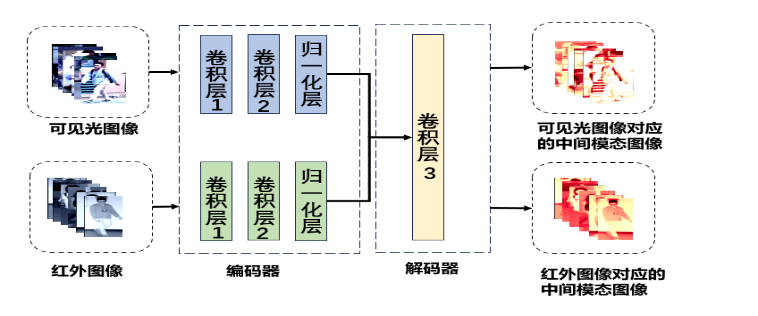


图2 辅助模态生成器结构图

2.3 混合注意力引导学习模块

本文中设计了一个混合注意力引导学习模块，该模块旨在有效减缓模态间的差异，并充分挖掘行人图像的细微信息。

混合注意力引导学习模块采用了实例归一化[13]来减缓模态间的差异，同时结合了卷积注意力模块（Convolutional Block Attention Module，CBAM）[12]来引导模型关注特征图中的有效信息混合注意力引导学习模块的计算公式如下所示：

其中，表示经过该模块输出的特征图，表示输入的特征，表示经过实例归一化后的输出，表示特征经过CBAM模块后得到混合注意力掩码。

具体而言，CBAM模块通过通道注意力和空间注意力引导模型关注重要的通道信息和空间信息。该模块先使用了通道注意力，再使用空间注意力。的计算公式如下所示：

其中，表示特征经过通道注意力模块后的得到的通道注意力掩码。通道注意力机制可以帮助模型了解哪个通道的内容比较重要。其计算公式如下所示：

表示Sigmoid激活函数。对于输入特征图，特征经过平均池化和最大池化后得到最大池化特征和平均池化特征，通过共享的多层感知机MLP网络得到通道注意力特征图。与输入特征相乘得到新特征，然后将进一步输入至空间注意力模块。

空间注意力机制可以帮助模型了解哪些区域比较重要，表示特征图经过空间注意力模块后的空间注意力掩码，其计算公式如下所示：

表示Sigmoid激活函数，表示卷积核尺寸为7x7的卷积运算。对于输入特征图，特征经过平均池化和最大池化后得到最大池化特征和，然后将两个特征拼接后进行卷积操作，最后经过激活函数输出得到空间注意力特征图。

实例归一化的计算公式如下所示：

其中，表示输入特征的第维，和表示计算每个维度的均值和标准差，用于避免分母为零。

2.4 损失函数

本文采用联合损失函数来优化训练模型，以扩大类间距离和提高类内不同模态特征相似度。联合损失函数包括分布一致性损失函数（）、交叉熵损失函数（）和三元组损失函数（）。联合损失函数的表达式如下所示：

其中，是分布一致性损失函数,用于优化辅助模态生成器，生成更好的辅助模态图像。它的计算公式如下所示：

其中，N是每一批次的训练样本总。。和分别表示由可见光图像和红外图像生成的辅助模态图像。表示辅助模态图像经过全连接层后的输出。表示取均值。

是交叉熵损失函数，帮助模型对不同身份的行人进行分类。它的计算公式如下所示：

其中，N表示训练过程中的行人身份个数，y表示图像的行人身份标签，表示预测是类别i的可能性，是一个常量，这里设置为0.1.

是三元组损失函数，它能帮助模型拉近不同模态相同身份的行人特征间的距离，推远相同模态不同身份的行人特征间的距离。四种模态间的三元组损失函数的表达式如下所示：



其中，可见光图像和红外图像两种模态间的三元组损失函数的计算公式如下：



其中，N表示每种模态图像在一批次的训练样本中的数量总和，表示欧式距离，是一个阈值常量，这里设置为0.3。其他模态间的三元组损失和上述定义的公式类似。

3 实验结果与分析

3.1 数据集和评价指标

本文所提出的方法在两个公开的可见光-红外行人重识别数据集SYSU-MM01[4]和RegDB[14]上进行实验。

SYSU-MM01数据集由四个可见光相机和两个红外相机拍摄收集，包含491个行人身份，共有287628张可见光图像和15792张近红外图像，涵盖室内和室外环境。训练集包含395个行人身份，共有19659张可见光图像和12792张红外图像；测试集包含 96 个行人身份，其中有3803张红外图像作为查询图像。测试时，图库集根据全局搜索模式和室内搜索模式进行划分，在全局搜索模式下，所有由可见光摄像头拍摄的图像构成图库集。在室内搜索模式下，仅使用两个室内可见光摄像头拍摄的图像构成图库集。

RegDB数据集则包含412个行人身份，每个身份有10张可见光图像和10张红外图像。数据集被随机分为两半，206个行人身份的图像用于训练，另外的206个行人身份的图像用于测试。测试有两种评估模式，分别是用可见光图像查询红外图像和用红外图像查询可见光图像。

在实验过程中，本文使用累积匹配特征(cumulative matching characteristics, CMC) 和平均精度(mean average precision, mAP)作为模型性能的评估指标。

3.2 实验细节

本文实验基于PyTorch环境，总共进行了80个Epoch的模型训练。输入图像的尺寸被调整为3×384×192，并在训练阶段采用随机水平翻转和随机擦除进行数据增强。在训练过程中，采用warm-up策略来平滑训练梯度，初始学习率设置为1×10−2，经过10个epoch线性增加到1×10−1，再在第20个epoch时将学习率衰减为1×10−2，最后在第60个epoch时进一步衰减至1×10−3。每个批次随机选择4个行人，每个行人分别选取4张可见光图像和4张红外图像进行训练。优化过程中采用了SGD优化器，动量参数设置为0.9，三元组损失函数的边距超参数设定为0.3，权重超参数设定为1，而分布一致性损失函数的权重超参数设定为0.5。

3.3 实验结果

3.3.1 方法比较

本文所提方法与主流的可见光-红外跨模态行人重识别方法进行对比实验，包括：Zero-padding[4]、BDTR[5]、HCML[6]、LADN[7]、AlignGAN[8]、D2RL[9]、X-Modality[10]、DDAG[15]、AGW[3]、NFS[16]、CM-NAS[17]、TSME[18]、FMCNet[19]、MID[20]。其中，基于模态共享特征提取的方法有：Zero-padding、HCML、BDTR、DDAG、LADN、AGW、NFS、CM-NAS；基于模态特定信息补充的方法有：BDTR、D2RL、AlignGAN、X-Modality、FMCNet、TSME、MID。表1是在SYSU-MM01数据集和RegDB数据集上的实验结果。

表1 在SYSU-MM01数据集和RegDB数据集上的对比实验结果

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Method** | **SYSU-MM01** | | | | | | | | **RegDB** | | | | | | | |
| **All search** | | | | **Indoor search** | | | | **Visible to Infrared** | | | | **Infrared to Visible** | | | |
| **R-1** | **R-10** | **R-20** | **mAP** | **R-1** | **R-10** | **R-20** | **mAP** | **R-1** | **R-10** | **R-20** | **mAP** | **R-1** | **R-10** | **R-20** | **mAP** |
| Zero-padding[4] | **14.80** | **54.12** | **71.33** | **15.95** | **20.58** | **68.38** | **85.79** | **26.92** | **17.80** | **34.20** | **44.40** | **18.90** | **16.60** | **34.70** | **44.30** | **17.80** |
| BDTR[5] | **17.0** | **55.40** | **72.00** | **19.70** | **--** | **--** | **--** | **--** | **33.47** | **58.42** | **67.52** | **31.83** | **68.06** | **85.15** | **90.31** | **61.80** |
| HCML[6] | **14.30** | **53.20** | **69.20** | **16.20** | **24.50** | **73.30** | **86.70** | **30.10** | **24.40** | **47.50** | **56.80** | **20.80** | **21.70** | **45.00** | **55.60** | **22.20** |
| D2RL[9] | **28.90** | **70.60** | **82.40** | **29.20** | **--** | **--** | **--** | **--** | **43.40** | **66.10** | **76.30** | **44.10** | **--** | **--** | **--** | **--** |
| AlignGAN[8] | **42.40** | **85.00** | **93.70** | **40.70** | **45.90** | **87.60** | **94.40** | **54.30** | **57.90** | **--** | **--** | **53.60** | **56.30** | **--** | **--** | **53.40** |
| X-Modality[10] | **49.92** | **89.79** | **95.96** | **50.73** | **--** | **--** | **--** | **--** | **--** | **--** | **--** | **--** | **62.21** | **83.13** | **91.72** | **60.18** |
| DDAG[15] | **54.75** | **90.39** | **95.81** | **53.02** | **61.02** | **94.06** | **98.41** | **67.98** | **69.34** | **86.19** | **91.49** | **63.46** | **68.06** | **85.15** | **90.31** | **61.80** |
| LADN[7] | **53.60** | **90.92** | **96.30** | **52.00** | **57.00** | **92.80** | **98.30** | **63.20** | **75.70** | **--** | **--** | **72.90** | **75.30** | **--** | **--** | **73.00** |
| AGW[3] | **47.50** | **84.39** | **92.14** | **47.65** | **54.17** | **91.14** | **95.98** | **62.97** | **70.05** | **--** | **--** | **66.37** | **--** | **--** | **--** | **--** |
| NFS[16] | **56.91** | **91.34** | **96.52** | **55.45** | **62.79** | **96.53** | **99.07** | **69.79** | **80.54** | **91.96** | **95.07** | **72.10** | **77.95** | **90.45** | **93.62** | **69.79** |
| CM-NAS[17] | **61.99** | **92.87** | **97.25** | **60.02** | **67.01** | **97.02** | **99.32** | **72.95** | **82.80** | **95.10** | **97.70** | **79.30** | **81.70** | **94.10** | **96.90** | **77.60** |
| FMCnet[19] | **66.34** | **--** | **--** | **62.51** | **68.15** | **--** | **--** | **74.09** | **89.12** | **--** | **--** | **84.43** | **88.38** | **--** | **--** | **83.86** |
| TSME[18] | **64.23** | **95.19** | **98.73** | **61.21** | **64.80** | **96.92** | **99.31** | **71.53** | **87.35** | **97.10** | **98.90** | **76.94** | **86.41** | **96.39** | **98.20** | **75.70** |
| MID[20] | **60.27** | **92.90** | **--** | **59.40** | **64.86** | **96.12** | **--** | **70.12** | **87.45** | **95.73** | **--** | **84.85** | **84.29** | **93.44** | **--** | **81.41** |
| **ours** | **67.08** | **92.74** | **97.08** | **63.93** | **72.42** | **97.92** | **99.73** | **77.47** | **89.66** | **96.61** | **98.45** | **83.43** | **85.19** | **94.61** | **97.31** | **79.44** |

根据表1的实验结果，可以得出本文所提出的方法在跨模态行人重识别任务中表现出不错的性能效果。具体而言，相较于基于模态共享特征提取的CM-NAS方法，在SYSU-MM01数据集的all-search模式下，本文方法的Rank-1指标提升了5.09%，mAP提升了3.91%。在in-door模式下，Rank-1指标提升了5.41%，mAP提升了4.52%。在RegDB数据集的可见光图像检索红外图像的测试模式下，本文方法的Rank-1指标提升了9.04%，mAP提升了5.32%。在红外图像检索可见光图像的测试模式下，本文方法的Rank-1指标提升了5.78%，mAP提升了2.87%。此外，与基于模态特定信息补偿的FMCnet方法相比，在SYSU-MM01数据集的in-door模式下，本文方法的Rank-1指标提升了4.27%，mAP提升了3.38%。在RegDB数据集的可见光图像检索红外图像的测试模式下，本文方法的Rank-1指标提升了2.72%。

3.3.2 消融实验

在SYSU-MM01数据集的all-search模式下，进行了对辅助模态生成器和混合注意力引导学习模块的消融实验，实验结果如表2所示。具体来看，在基线模型中，Rank-1和mAP值分别为58.66和53.76。加入混合注意力引导学习模块后，Rank-1和mAP值分别提高到了62.40和58.44。在此基础上再加入辅助模态生成器，Rank-1和mAP值进一步提高到了67.08和63.93。综合来看，辅助模态生成器和混合注意力引导学习模块的联合使用可以有效地提升跨模态行人重识别任务的性能，使得模型在图像匹配和检索方面表现更为优越。

表2 在SYSU-MM01数据集All search模式的消融实验结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Method** | **All search** | | | |
| R-1 | R-10 | R-20 | mAP |
| Baseline | 58.66 | 90.55 | 96.89 | 53.76 |
| Baseline+MAML | 62.40 | 90.82 | 96.06 | 58.44 |
| **Baseline+MAML+AMG** | 67.08 | 92.74 | 97.08 | 63.93 |

4 结束语

本文针对跨模态行人重识别问题，设计了一个基于辅助模态和混合注意力的双流参数共享网络。该网络在Resnet50的基础上进行改进，首先通过辅助模态生成器将可见光图像和红外图像投影到统一的辅助模态空间，生成辅助模态图像，并与原始图像一起输入主干网络进行特征提取。在主干网络嵌入混合注意力引导学习模块，该模块使用混合注意力机制和实例归一化，能有效减缓模态间差异，并挖掘细微的、更具鉴别性的特征信息。在主干网络提取特征图后，将特征图水平均匀切分成若干块，进行局部特征学习。最后，联合分布一致性损失、三元组损失、身份损失对网络进行优化。大量实验证明本文所提方法的有效性。

**参考文献**

1. 罗浩, 姜伟, 范星, 张思朋. 基于深度学习的行人重识别研究进展.自动化学报, 2019, 45(11): 2032−2049)
2. Z. Wang, Z. Wang, Y. Zheng, Y. Wu, W. Zeng, S. Satoh.Beyond intra-modality: A survey of heterogeneous person re-identification[C]//Proceedings of the International Joint Conference on Artificial Intelligence, 2020:4973–4980.
3. M. Ye, J. Shen, G. Lin, T. Xiang, L. Shao, and S. C. H. Hoi, “Deep learning for person re-identification: A survey and outlook,” IEEE Trans.Pattern Anal. Mach. Intell., vol. 44, no. 6, pp. 2872–2893, 2022.
4. Ancong Wu, Wei-Shi Zheng, Hong-Xing Y u, Shaogang Gong, and Jianhuang Lai. Rgb-infrared cross-modality person re-identification[C]//Proceedings of the IEEE International Conference on Computer Vision.2017: 5390–5399.
5. Mang Y e, Zheng Wang, Xiangyuan Lan, and Pong C. Y uen.Visible thermal person re-identification via dual-constrained top-ranking[C]// Proceedings of the International Joint Conference on Artificial Intelligence, 2018: 1092–1099
6. Mang Y e, Xiangyuan Lan, Jiawei Li, and Pong C. Y uen. Hierarchical discriminative learning for visible thermal person re-identification[C]// Proceedings of the AAAI Conference on Artificial Intelligence, 2018: 7501–7508
7. S. Liu, J. Zhang, Local alignment deep network for infrared-visible cross-modal person reidentification in 6G-enabled internet of things, IEEE Internet Things J. 8 (20) (2021) 15170–15179.
8. G. Wang, T. Zhang, J. Cheng, S. Liu, Y. Yang, Z. Hou, RGB-Infrared crossmodality person re-identification via joint pixel and feature alignment[C]//Proceedings of the IEEE International Conference on Computer Vision, 2019:622–3631.
9. Z. Wang, Z. Wang, Y. Zheng, Y.-Y. Chuang, S. Satoh, Learning to reduce duallevel discrepancy for infrared-visible person re-identification[C]//Proceedings of Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 618–626.
10. 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(04): 4610-4617.
11. He K, Zhang X, Ren S, et al Deep residual learning for image recognition[C]//Proceedings of the IEEE con- ference on computer vision and pattern recognition. 2016: 770-778.
12. Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. CBAM: convolutional block attention module. In ECCV, pages 3–19, 2018. 3, 8
13. Xingang Pan, Ping Luo, Jianping Shi, and Xiaoou Tang.Two at once: Enhancing learning and generalization capacities via ibn-net. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, ECCV. 4
14. Dat Tien Nguyen, Hyung Gil Hong, Ki-Wan Kim, and Kang Ryoung Park. Person recognition system based on a combination of body images from visible light and thermal cameras. Sensors, 17(3):605, 2017. 5
15. Ye M, Shen J, J Crandall D, et al. Dynamic dual-attentive aggregation learning for visible-infrared person re-identification[C]//European Conference on Computer Vision. Springer, Cham, 2020: 229-247.
16. Y . Chen, L. Wan, Z. Li, Q. Jing, and Z. Sun, Neural feature search for RGB-infrared person re-identification[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021:8503-8510
17. C. Fu, Y. Hu, X. Wu, H. Shi, T. Mei, and R. He, “CM-NAS: cross-modality neural architecture search for visible-infrared person reidentification,” in ICCV, 2021, pp. 11 803–11 812.
18. J. Liu, J. Wang, N. Huang, Q. Zhang, and J. Han, “Revisiting modality-specific feature compensation for visible-infrared person reidentification,” IEEE Trans. Circuits Syst. Video Technol., vol. 32, no. 10, pp. 7226–7240, 2022
19. Q. Zhang, C. Lai, J. Liu, N. Huang, and J. Han, “Fmcnet: Feature-level modality compensation for visible-infrared person re-identification,” in CVPR, 2022, pp. 7339–7348.
20. Z. Huang, J. Liu, L. Li, K. Zheng, and Z. Zha, “Modality-adaptive mixup and invariant decomposition for rgb-infrared person re-identification,” pp. 1034–1042, 2022.

作者简介：梁靖贤，男，硕士研究生，研究方向：跨模态行人重识别；谭台哲，男，博士，副教授，研究方向：图像处理与计算机视觉、机器学习与大数据处理、区块链技术。

作者联系方式

通信地址：广东省广州市番禺区大学城外环西路100号广东工业大学生活东区

邮政编码：510006

联系电话：13433203028

电子邮件：285988424@qq.com