# 小论文初稿

摘要

引言

行人重识别（Person Re-Identification, ReID）是智能监控系统的关键技术之一，是指利用计算机视觉技术在指定视频或图片序列中对特定行人进行识别的技术，旨在解决不同摄像头之间的行人检索问题。近年来，ReID获得了相当大的关注，取得了不错的进展。然而可见光摄像头在弱光或者无光（例如夜晚）的环境下，并不能捕获到有效的行人信息。与可见光相机相比，红外（IR）相机可以在黑暗环境中从场景中捕获足够的信息，但是，由于可见光图像和红外图像之间存在巨大的差异，传统单模态的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间的行人重识别，研究人员提出了基于可见光——红外图像跨模态行人重识别问题（Visible-Infrared Person Re-Identification, VI ReID）。

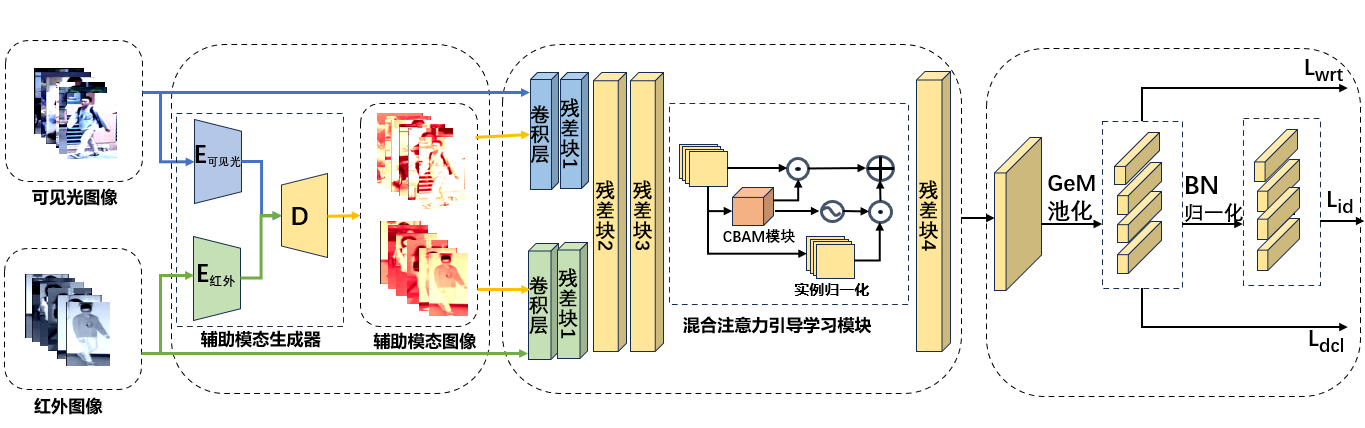
目前，解决跨模态行人重识别问题的方法可以分成以下两种方法：模态共享特征学习和模态特定信息补偿。模态共享特征学习主要通过特征对齐的方式，把红外图像和可见光图像的特征投影到一个统一的模态共享空间中，然后通过相关的损失函数进行度量学习，把不同模态、相同身份的行人特征距离拉近，同一模态不同身份的行人特征距离推远。2017年，Wu等人提出了一个可以在共享空间学习模态共享特征的深度零填充网络，并构建了第一个大规模的可见光-红外行人图像数据集SYSU-MM01。Ye等人提出了双向双约束top-ranking（BDTR）损失来缓解模态内和模态间差异，帮助模型提取出有鉴别性的行人特征信息。

此外，Ye等人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。Liu等人[22]首先从输入图像中提取模态共享特征，然后利用水平分割的方式来获得更具鉴别性的模态共享行人的局部特征。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来缓解不同模态间的差异。Wang等人提出了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产的红外伪图像进行匹配。Wang等人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进行相互转化来减缓模态间的差异。Li等人设计了一个轻量级网络，该网络可以将可见光图像生成一种新的辅助模态图像，以此来减缓可见光模态和红外模态间的模态差异。

相关方法

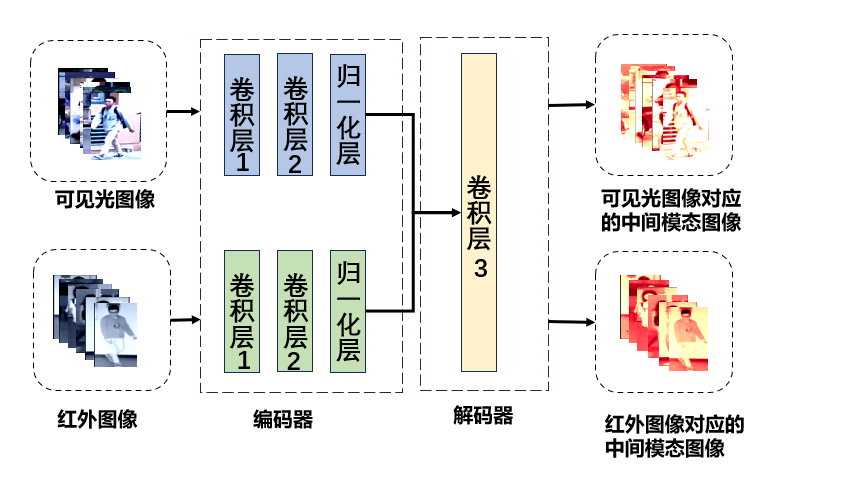
2.1网络整体架构

本文所提方法的网络结构如下图1所示。首先，将原始的可见光图像与红外图像输入中间模态生成器，以生成中间模态图像。然后，生成的中间模态图像分别和原始的可见光图像与红外图像一同输入至主干网络中提取模态共享特征。主干网络基于ResNet50进行改进，前面两层网络层的参数是独立的，用于提取模态特定特征，后面的网络层参数共享，用于提取模态共享特征。在第三个网络层后面嵌入混合注意力引导学习模块，有效地缓解模态间的差异，同时充分挖掘具有鉴别性的模态共享特征。将从主干网络提取的特征图水平划分为若干个部分，每个部分都被输入到分类器中以学习具有鉴别性的模态共享特征。最后，联合使用分布一致性损失函数、三元组损失函数、身份损失函数来对模型进行端到端训练。



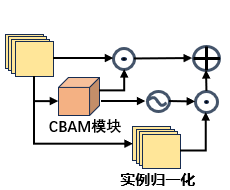
2.2辅助模态生成器

辅助模态生成器采用编码器-解码器架构，它可以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来生成辅助模态图像，生成的辅助模态图像可以有效地缓解可见光模态和红外模态间的模态差异。具体而言，辅助模态生成器是由两个参数独立的解码器和一个参数共享的解码器构成，首先将输入图片的尺寸调整为3x384x192，由于红外图像是单通道的，可见光图像是三通道的，因此需要把红外图像的通道数设置为3，和可见光图像的通道数保持一致。然后，把可见光图像和红外图像输入至对应的模态信息编码器中进行编码，在编码过程中，首先会通过一个卷积核为3x1x1的卷积层，将三通道图像转换为单通道图像，然后通过卷积核为1x1x1的卷积层，减少计算量，最后通过批归一化层进行数据归一化。然后将归一化后的数据输入至对应的解码器进行解码，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，以生成对应的辅助模态图像。



2.3混合注意力引导学习模块

本文设计了一个混合注意力引导学习模块，该模块具体结构如下图所示。该模块可以减缓模态间的差异，同时充分挖掘不同行人图像间的细微信息。

（补充一个CBAM结构图）

混合注意力引导学习模块使用了实例归一化来减缓模态间的差异，同时使用卷积注意力模块（Convolutional Block Attention Module，CBAM）来引导模型关注特征图中的有效信息，发掘不同行人图像之间的细微差异。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模块它能够引导模型关注特征图中较重要的通道信息和空间信息，模块的结构如下图所示。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：



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F，首先分别经过平均池化和最大池化得到最大池化特征和平均池化特征，然后经过一个共享的多层感知机MLP网络得到通道注意力特征图。然后，与输入特征相乘得到新特征，然后将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F‘，首先分别经过平均池化和最大池化得到最大池化特征和，然后将两个特征进行拼接，再输入至一个标准的卷积层进行卷积操作，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

本文联合使用分布一致性损失函数，身份损失函数和三元组损失函数对模型进行优化训练。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：



N是训练阶段每一批次的训练样本的总和。和是由可见光图像和红外图像分别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，它的计算公式如下所示：

，

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。在本发明中，四种模态间的三元组损失函数的表达式如下所示：



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：



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面所定义的公式相似。

3实验

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和RegDB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。图库集由测试模式决定，其中测试模式有两种，分别是全局搜索模式和室内搜索模式。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数据集包含412个行人身份，由一个可见光摄像头和一个红外摄像头拍摄收集，每个行人身份有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

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

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **方法** | **All search** | | | | **Indoor search** | | | |
| Rank-1 | Rank-10 | Rank-20 | mAP | Rank-1 | Rank-10 | Rank-20 | mAP |
| Zero-padding | 14.80 | 54.12 | 71.33 | 15.95 | 20.58 | 68.38 | 85.79 | 26.92 |
| HCML | 14.30 | 53.20 | 69.20 | 16.20 | 24.50 | 73.30 | 86.70 | 30.10 |
| BDTR | 17.0 | 55.40 | 72.00 | 19.70 | -- | -- | -- | -- |
| D2RL | 28.90 | 70.60 | 82.40 | 29.20 | -- | -- | -- | -- |
| AlignGAN | 42.40 | 85.00 | 93.70 | 40.70 | 45.90 | 87.60 | 94.40 | 54.30 |
| X-Modality | 49.92 | 89.79 | 95.96 | 50.73 | -- | -- | -- | -- |
| DDAG | 54.75 | 90.39 | 95.81 | 53.02 | 61.02 | 94.06 | 98.41 | 67.98 |
| LADN | 53.60 | 90.92 | 96.30 | 52.00 | 57.00 | 92.80 | 98.30 | 63.20 |
| AGW | 47.50 | 84.39 | 92.14 | 47.65 | 54.17 | 91.14 | 95.98 | 62.97 |
| NFS | 56.91 | 91.34 | 96.52 | 55.45 | 62.79 | 96.53 | 99.07 | 69.79 |
| CM-NAS | 61.99 | 92.87 | 97.25 | 60.02 | 67.01 | 97.02 | 99.32 | 72.95 |
| FMCnet | 66.34 | -- | -- | 62.51 | 68.15 | -- | -- | 74.09 |
| TSME | 64.23 | 95.19 | 98.73 | 61.21 | 64.80 | 96.92 | 99.31 | 71.53 |
| MID | 60.27 | 92.90 | -- | 59.40 | 64.86 | 96.12 | -- | 70.12 |
| **本文方法** | 67.08 | 92.74 | 97.08 | 63.93 | 72.42 | 97.92 | 99.73 | 77.47 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Method** | **Visible to Infrared** | | | | **Infrared to Visible** | | | |
| Rank-1 | Rank-10 | Rank-20 | mAP | Rank-1 | Rank-10 | Rank-20 | mAP |
| Zero-padding | 17.80 | 34.20 | 44.40 | 18.90 | 16.60 | 34.70 | 44.30 | 17.80 |
| HCML | 24.40 | 47.50 | 56.80 | 20.80 | 21.70 | 45.00 | 55.60 | 22.20 |
| BDTR | 33.47 | 58.42 | 67.52 | 31.83 | 68.06 | 85.15 | 90.31 | 61.80 |
| D2RL | 43.40 | 66.10 | 76.30 | 44.10 | -- | -- | -- | -- |
| AlignGAN | 57.90 | -- | -- | 53.60 | 56.30 | -- | -- | 53.40 |
| X-Modality | -- | -- | -- | -- | 62.21 | 83.13 | 91.72 | 60.18 |
| DDAG | 69.34 | 86.19 | 91.49 | 63.46 | 68.06 | 85.15 | 90.31 | 61.80 |
| LADN | 75.70 | -- | -- | 72.90 | 75.30 | -- | -- | 73.00 |
| AGW | 70.05 | -- | -- | 66.37 | -- | -- | -- | -- |
| NFS | 80.54 | 91.96 | 95.07 | 72.10 | 77.95 | 90.45 | 93.62 | 69.79 |
| CM-NAS | 82.80 | 95.10 | 97.70 | 79.30 | 81.70 | 94.10 | 96.90 | 77.60 |
| FMCnet | 89.12 | -- | -- | 84.43 | 88.38 | -- | -- | 83.86 |
| TSME | 87.35 | 97.10 | 98.90 | 76.94 | 86.41 | 96.39 | 98.20 | 75.70 |
| MID | 87.45 | 95.73 | -- | 84.85 | 84.29 | 93.44 | -- | 81.41 |
| **ours** | 91.84 | 97.82 | 99.47 | 84.62 | 87.48 | 95.39 | 97.72 | 80.47 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Method** | **All search** | | | | **Indoor search** | | | |
| Rank-1 | Rank-10 | Rank-20 | mAP | Rank-1 | Rank-10 | Rank-20 | mAP |
| Baseline | 58.66 | 90.55 | 96.89 | 53.76 | 61.63 | 94.64 | 98.34 | 67.66 |
| Baseline+MAML | 62.40 | 90.82 | 96.06 | 58.44 | 64.09 | 94.57 | 97.19 | 70.79 |
| **Baseline+MAML+AMG** | 67.08 | 92.74 | 97.08 | 63.93 | 72.42 | 97.92 | 99.73 | 77.47 |