Towards a Unified Middle Modality Learning for Visible-Infrared Person Re-Identification

统一的中间模态的多模态行人重识别

**摘要：**

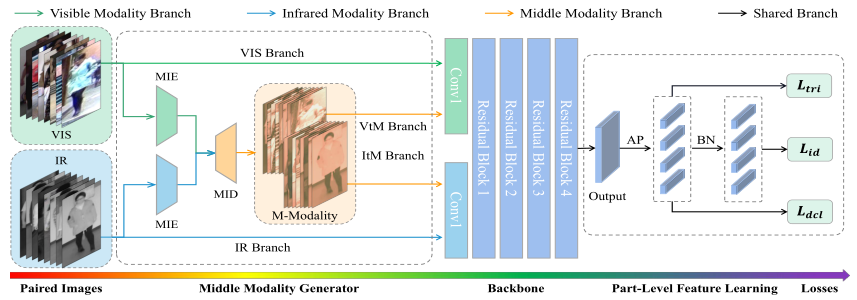
一些最先进的方法试图设计复杂的网络或利用生成网络来减轻模态差异，但是这些方法忽略可见图像和红外图像的模态之间的高度非线性关系。我们提出了一种非线性中间模态生成器（MMG），它有助于减少模态差异。我们的MMG可以有效地将VIS和IR图像投影到统一的中间模态图像（UMMI）空间中，以生成中间模态（M模态）图像。生成的M模态图像和原始图像被馈送到主干网络以减少模态差异。此外，为了将从UMMI空间中的VIS和IR图像生成的两种类型的M模态图像合并在一起，我们提出了分布一致性损失（DCL），以使生成的M模态图的模态分布尽可能一致。最后，我们提出了一种中间模态网络（MMN），以明确的方式进一步增强特征的辨别力和丰富性。

**1.主要贡献**

1. 我们提出了一种非线性中间模态生成器来生成中间模态图像，以辅助VI-ReID任务。中间模态生成器可以有效地将VIS和IR图像投影到统一的中间模态图像空间中。
2. 我们提出了一种有效的分布一致性损失，以使从VIS和IR图像获得的两种类型的中间模态图像在UMMI空间中的模态分布一致，这进一步提高了所提方法的性能。

**2.方法介绍**

**2.1整体框架**



MMN的输入是一对VIS-IR图像。VIS和IR图像被送到MMG模块中以生成中间模态（M模态）图像。生成的具有原始VIS和IR图像的M模态图像被馈送到双流ResNet50，其中每个流中的第一卷积块不同以捕获模态特定的低级别表示，而中间的深度卷积块被共享以学习模态共享的深层次的语义信息。并将特征图水平划分为几个部分，每个部分都被送入分类器以学习局部信息，将块数设置为4。将骨干网络中最后一个卷积块的步幅修改为1，以保留更多的空间信息。在平均池化（AP）层的卷积层之后，有一个批归一化（BN）层，该层是所有模态图像之间共享的参数，以使损失更容易收敛。最后，将BN层前后的特征输入到不同的损耗函数中，共同优化网络。

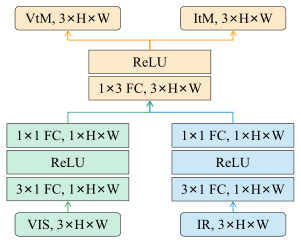
**2.2Middle Modality Generator（中间模态生成器MMG）**

分别表示来自数据集的VIS和IR图像对。所有输入图像的大小都调整为3×384×192，因此我们可以将红外图像视为三通道图像。事实上，它是一个只有一个通道的灰色图像。

MMG包括模态信息编码器（MIE）和模态信息解码器（MID），前者使用两个不共享参数的编码器对不同的模态信息进行编码，后者使用两个共享参数的解码器将VIS和IR图像投影到UMMI空间中。

**2.2.1Modality Information Encoder模态信息编码器**

由于IR图像有一个通道包含IR光的信息，而VIS图像有三个通道包含VIS光的颜色信息，因此我们首先在通道级别对齐VIS图像和IR图像。

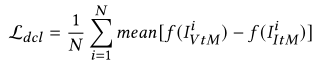


通过以上操作，我们获得了VIS和IR图像的特征，并在通道级执行VIS图像和IR图像之间的对齐。

**2.2.2 Modality Information Decoder模态信息解码器**

MID用于将编码为一个通道的数据投影到统一的三通道图像空间中。生成的M模态图像具有与VIS图像和IR图像相同的标签。最后，将M模态、VIS和IR模态图像一起馈入主干网络，以帮助VI ReID任务减少模态差异

**2.3 Distribution Consistency Loss (DCL)** **分布式一致性损失**

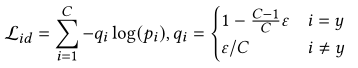


N代表一个批次的和图片数量

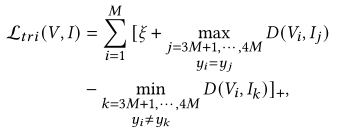
* 1. **Multi-Loss Optimization多种损失的优化**

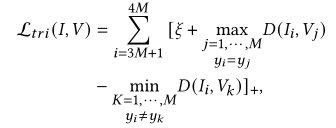
生成的M模态图像与原始VIS和IR图像一起馈送到双流ResNet50主干网络中，以帮助网络的优化。在MMN中，除了DCL之外，我们还将标签平滑、交叉熵损失和三元组损失结合起来，共同优化MMN。

标签平滑的交叉熵损失可以防止估计模型过度拟合：



使用M模态图像以及原始的VIS和IR图像来辅助MMN的训练，因此我们每个批次的图像数量为4M，0-M为图像，M-2M为图像，2M-3M为图像，3M-4M为图像。





其他模态三元组损失与此相似，最终的跨模态三元组损失函数计算如下：



总的损失函数如下：

