# 小论文润色

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，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生成器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，以此生成相同模态的辅助模态图像。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，引入混合注意力引导学习模块，该模块利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，采用广义平均池化，然后水平均匀切分成若干块，进行局部特征学习。

最后，为了对网络进行端到端训练，联合使用分布一致性损失、三元组损失和身份损失。