# 小论文之摘要

跨模态行人重识别旨在匹配来自不同模态的相同身份的行人图像。现有方法主要通过模态特定信息补偿或直模态共享特征学习来缓解模态差异。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，由于可见光图像和红外图像之间存在模态差异，直接提取两种模态的模态共享特征十分困难，而且不能充分挖掘细微的、更具鉴别性的特征信息。本文提出了一种基于辅助模态和混合注意力引导学习模块的可见光-红外跨模态行人重识别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来生成可见光图像和红外图像对应的辅助模态图像，这样有助于缓解模态差异。同时，在主干网络中嵌入混合注意力引导学习模块，该模块主要使用了混合注意力机制和实例归一化，能够有效减缓模态差异，同时引导模型关注细微的、更具鉴别性的特征信息。最后，联合使用分布一致性损失函数、三元组损失函数、身份损失函数来对网络进行端到端训练。在SYSU-MM01和RegDB数据集的大量实验证明了该方法的有效性。

**摘要：**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以生成辅助模态图像，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，该模块使用混合注意力和实例归一化，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。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。在SYSU-MM01和RegDB数据集的大量实验证明了该方法的有效性。

One of the major challenges in cross-modal pedestrian re-identification is the presence of large inter-modal differences. Existing methods mainly mitigate modal differences through modal-specific information compensation or learning modal shared features. However, modality-specific information compensation requires complex computational effort and is prone to introduce noise. Meanwhile, existing methods generally extract modal shared features directly, and have not yet fully explored the subtle and more discriminative feature information. Therefore, this paper proposes a two-stream parameter sharing network based on auxiliary modality and hybrid attention to mitigate modal differences while mining subtle and more discriminative feature information. The network uses an auxiliary modality generator to project visible and infrared images into a unified auxiliary modality space to generate auxiliary modality images, which can effectively mitigate modal differences. Meanwhile, a hybrid attention-guided learning module is designed, 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, it is horizontally and uniformly sliced into blocks for local feature learning.

Translated with www.DeepL.com/Translator (free version)

Rank-1 and mAP reached 67.08% and 63.93% respectively in the global search mode of SYSU-MM01 dataset, which experimentally proved the effectiveness of the method.

Hello ChatGPT, I would like to request your help in correcting my research paper in the area of video captioning pedestrian re-identification. I have already written the paper, but I need assistance in improving the language, grammar, spelling, and punctuation. My paper is intended for submission to a scholarly journal, and I need it to be error-free and polished to the highest standards:

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 shared features across modalities. However, modality-specific information compensation often requires complex computational effort and can introduce noise. On the other hand, 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 an 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, utilizing hybrid attention and instance normalization to guide the model's focus towards 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 the SYSU-MM01 dataset, our method achieves a Rank-1 accuracy of 67.08% and a mean Average Precision (mAP) of 63.93%, demonstrating the superiority of our approach

以下是一篇学术论文的一段话，打磨文字以符合学术风格，提高拼写，语法，清晰度，简洁性和整体可读性。必要时重写整个句子。此外，在标记表格中列出所有修改并解释修改的原因：