**技术交底书格式**

|  |  |  |  |
| --- | --- | --- | --- |
| 专利名称 | 一种基于多特征学习的双流网络跨模态行人重识别算法 | 所属技术领域 | 信息科学 |

1. 本发明要解决的技术问题是什么？

现有的跨模态行人重识别算法选择补充模态信息或直接将特征映射到共同特征空间，主要关注于缓解模态差异，而忽略细微差别的影响，这不利于模型学习到具有辨别性的身份特征。本发明使用一种基于多特征学习的跨模态行人重识别算法，可以以全局特征和局部特征结合的方式增强特征提取的信息容量，弥补仅使用一种特征造成的不足，增强特征的辨别性，提高跨模态行人重识别的精度。

1. 详细介绍技术背景,并描述已有的与本发明最相近似的实现方案。

（包括两部分：背景技术及现有技术方案，应详细介绍，以不需再去看文献即可领会该技术内容为准，如果现有技术出自专利、期刊、书籍，则提供出处）

行人重识别是图像检索的一个子问题，主要是实现给定一个摄像机捕获的行人图像，从其他摄像设备捕获的图像库中搜索该行人的其他图像的问题。行人重识别可以有效帮助刑侦人员在监控系统中定位目标人物，从而及时应对潜在危险。目前所研究的行人重识别问题都是单模态条件下的，即用于训练和测试的数据集行人图像都是在白天拍摄到的可见光RGB图像，在现实生活中不法分子不可能只在白天出没，刑侦人员很可能需要在光线昏暗的环境中寻找目标人物，而在夜晚或者光线很差的时候，RGB图像所包含的信息会很少。红外的环境适应性优于可见光，红外摄像机可以在光线昏暗的条件下捕获到具有行人信息的图像。因此，研究基于可见光——红外图像跨模态行人重识别问题，可以使得两种模态的行人图像能够相互匹配，充分利用可见光和红外摄像头拍摄到的监控数据，达到24小时监控目标。

跨模态行人重识别问题主要存在两种对齐困难，一种是由人体姿态、摄像机视角不同、遮挡等问题引起的空间不对齐，一种是由不同成像原理引起的模态不对齐。现有的解决跨模态行人重识别问题的方法有三种，分别是基于表征学习的方法、基于度量学习的方法以及模态互转的方法。基于表征学习的方法主要设计合适的神经网络结构，来提取出两种模态图像的共同特征以及具有辨别性的身份特征，减小模态间存在的差异性。基于度量学习的方法主要设计合理的度量损失函数，学习一个映射空间，使得在空间中不同模态同一身份的图像特征距离尽可能小，不同身份的图像特征距离尽可能大。基于模态互转的方法主要研究如何讲两种模态的图像转换成同一模态图像，减少两种模态之间的差异。但现有方法大多选择仅提取全部特征或局部特征的信息来辨别行人图像身份。

如图１所示，该方法提取可见光图像和红外图像的局部特征，通过身份损失函数、分布损失函数和相关损失函数优化模型，以提取出更具有辨别性和鲁棒性的行人特征。但这种仅使用单一特征信息的方法使得模型对图像的信息提取不够充分，可能导致模型对一些信息过分依赖，影响模型从图像库中检测目标行人的准确性。

3、现有技术的缺点是什么？针对这些缺点，说明本发明的目的。

（客观评价，现有技术的缺点是针对于本发明的优点来说的，本发明不能解决的缺点不必写；基于本发明能解决的问题写出发明的目的。）

可见光图像和红外图像不仅模态间有区别，同一模态内因行人姿态、监控视角、光照强度等变化，使得同一模态同一身份的图像之间也存在诸多差异。现有是方法大多选择仅提取全部特征或局部特征的信息来辨别行人图像身份，但这会使得模型对图像的信息提取不够充分，可能导致模型对一些信息过分依赖，影响模型从图像库中检测目标行人的准确性。

为了充分利用图像信息，提高模型的泛化能力，本发明旨在以全局特征和局部特征结合的方式增强特征提取的信息容量，弥补仅使用一种特征造成的不足，增强特征的辨别性，提高跨模态行人重识别的精度。

4、本发明技术方案的基本内容。

如图２所示，本发明采用基于注意力思想的BoTNet为骨干网络的双流网络提取全局特征，BoTNet最后卷积模块将卷积核为3×3的卷积层更换为多头自注意力层(MHSA)，增强网络架构的全局信息聚合能力。第一层卷积和卷积模块参数不共享，用于提取特定模态的特征。后面三个卷积模块及分类器C参数共享，用于提取两种模态之间的共同特征，该模型中的分类器包含一个全连接层和一个批归一化层。对网络提取出的全局特征做水平划分，得到K个局部特征。用于局部特征的卷积层、可见光分类器C1及红外分类器C2参数不共享，使用身份损失和五元组损失来指导模型学习具有鲁棒性和辨别性的行人特征。

5、本发明技术方案的详细阐述。

（本部分为专利申请最重要的部分，需要详细提供，专利必须是一个技术方案，应该阐述发明目的是通过什么技术手段来实现的，不能只有原理，也不能只做功能介绍；因此发明中每一功能的实现都要有相应的技术实现方案；所有英文缩写都应有中文注释；必须结合流程图、原理框图、电路图、时序图等附图进行说明**，**每个图都应有对应的文字详细的描述，以别人不看附图即可明白技术方案为准；同时附图中的关键词或方框图中的注释都尽量用中文；方法专利都应该提供一个流程图，并提供相关的系统装置。）

本发明的工作原理：将可见光图像和红外图像分别输入到双流网络的两个分支里，分别提取出每个图像的全局特征，对每个批次的全局特征计算身份损失和困难五元组损失来训练模型提取两个模态共有的信息，缩小模态间的差异。对全局特征做水平划分，得到K个局部特征，通过对局部特征计算身份损失，使得模型能够注意到每个图像里的细微差别，进而提取出同个模态下不同身份图像之间具有辨别性的信息。流程如图3所示，具体步骤如下：

1. 构建双流网络，选取BoTNet为骨干网络，定义网络仅在后面三个卷积模块及分类器C参数共享，后面用于局部特征的卷积层、可见光分类器C1及红外分类器C2参数不共享。
2. 对用于训练的数据集做数据增强，对预处理好的训练数据集做随机批量采样，定义每一批次里，从两个模态里选取N个行人的P张图像，即一个批次里输入到模型的共有2NP张图像。
3. 对模型提取出的三维张量T应用全局平均池化得到图像对应的全局特征。定义为第i个可见光图像的全局特征，为第i个红外图像的全局特征。
4. 联合每个批次中可见光图像和红外图像的全局特征，计算困难五元组损失，困难五元组损失由困难全局三元组损失和困难跨模态三元组损失组成。困难全局三元组损失计算如下：

其中和均为行人身份为i图像的全局特征，为行人身份不为i图像的全局特征。为批次中行人身份为i的第a张图像与其他行人身份图像中的最小欧式距离， 为批次中行人身份为i的第a张图像与其他该行人身份图像中的最大欧式距离。为定义的间隔距离，即图像之间最大欧式距离和最小欧式距离之间应该相差。中的\*若大于0则输出\*，若小于等于0则输出为0。

其中为行人身份为i图像的全局特征，为行人身份为i的另一个模态图像全局特征，为行人身份不为i并与不属于同一模态图像的全局特征。为批次中行人身份为i的第a张图像与另一个模态里其他行人身份图像中的最小欧式距离， 为批次中行人身份为i的第a张图像与另一个模态里其他该行人身份图像中的最大欧式距离。

困难五元组损失最终为：

1. 使用分类器C对每个批次图像的全局特征进行身份分类，计算交叉熵损失作为身份损失：

其中为第i张可见光图像的标签（即行人身份），为分类器对第i张可见光图像分类的概率结果。为第i张红外图像的标签（即行人身份），为分类器对第i张红外图像分类的概率结果。

1. 将模型提取出的三维张量T水平分割为K个水平部分，对每个部分应用全局平均池化得到K个列向量， 采用1×1卷积层对列向量进行降维得到局部特征。定义为第i个可见光图像的第j个局部特征。
2. 使用特定模态的分类器C1和C2对可见光图像和红外图像对应的局部特征进行身份分类，计算交叉熵损失作为特定身份损失。

其中为第i张可见光图像的标签（即行人身份），为分类器对第i张可见光图像的第j个局部特征分类的概率结果。为第i张红外图像的标签（即行人身份），为分类器对第i张红外图像的第j个局部特征分类的概率结果。

1. 设置模型整体损失函数L为：
2. 根据损失函数对设计的模型参数进行优化。

6、本发明的关键点和欲保护点是什么？

（发明内容部分提供的是为完成一定功能的完整技术方案，本部分是提炼出技术方案的关键创新点，列出1、2、3...，以提醒代理人注意，便于专利代理人撰写权利要求书。）

1)设计一个新的双流网络来更好地提取不同模态图像间具有鲁棒性和辨别性的行人特征。使用基于注意力思想的BoTNet为骨干网络，相较于使用Resnet50网络，减少了网络参数，并增强网络架构的全局信息聚合能力。模型只选取后三个卷积模块及分类器C参数共享，使得其能在提取特定于身份特征的同时提取到两个模态下的共同特征。

2)使用局部特征和全局特征计算身份损失，考虑细微差别，训练模型在提取两个模态下共同特征的同时提取特定身份的辨别性特征。

7、与第2条所属的最好的现有技术相比，本发明有何优点？

（效果一定要结合发明内容的技术方案来描述，做到有理有据；也可以对应本发明所要解决的技术问题来描述，一定是采用本发明技术方案带来的效果；效果可以是降低成本，提高了效率等。）

优点是：在减少模型网络参数、不添加改变原始数据操作的情况下，结合全局特征和局部特征增强特征提取的信息容量，弥补仅使用一种特征造成的不足，增强特征的辨别性，提高跨模态行人重识别的精度。

8、本发明是否经过实验、模拟、使用而证明可行，结果如何？

经过试验，证明可行。在减少模型网络参数、不添加改变原始数据操作的情况下，模型在图像库中检索目标图像的mAP（平均精度均值）得到提高，能够很好地解决跨模态行人重识别任务。

9、本发明的变更设计（替代方案）及其它用途：

（如果有，请尽量详细写明，内容的提供可以扩大专利的保护范围，防止他人绕过本技术去实现同样的发明目的；“替代方案”可以是部分结构、器件、方法步骤的替代，也可以是完整技术方案的替代。）

此算法还可用于一些跨模态人脸识别，跨模态目标检索等领域。

10、附图及说明

每幅图都应有相应的附图说明

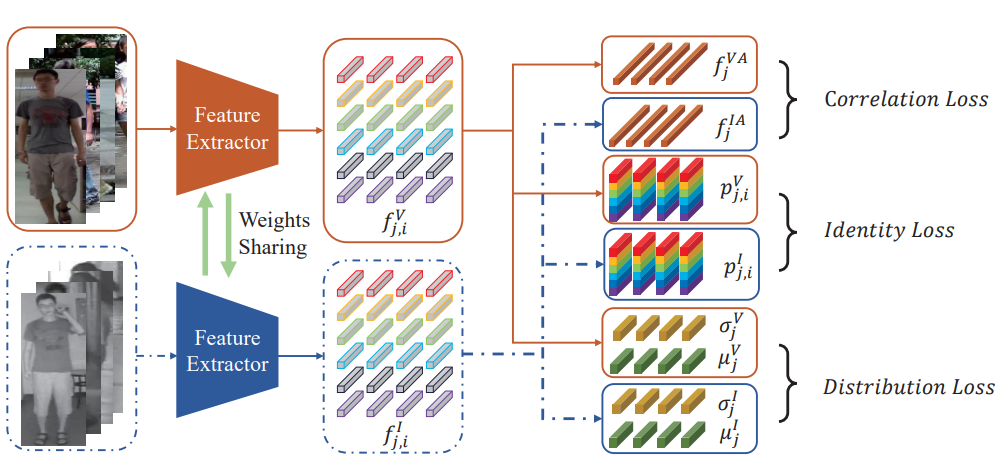


图1 现有跨模态行人重识别网络架构

图1：采用权重共享的ResNet50作为骨干模型，将可见光图像和红外图像同时传进模型中，得到相对于可见光图像和红外图像的局部特征，通过对局部特征联合求身份损失，约束局部特征之间的分布，约束可见光批处理图像和红外批处理图像聚合特征之间的Pearson相关系数，来训练模型提取模态不变特征。

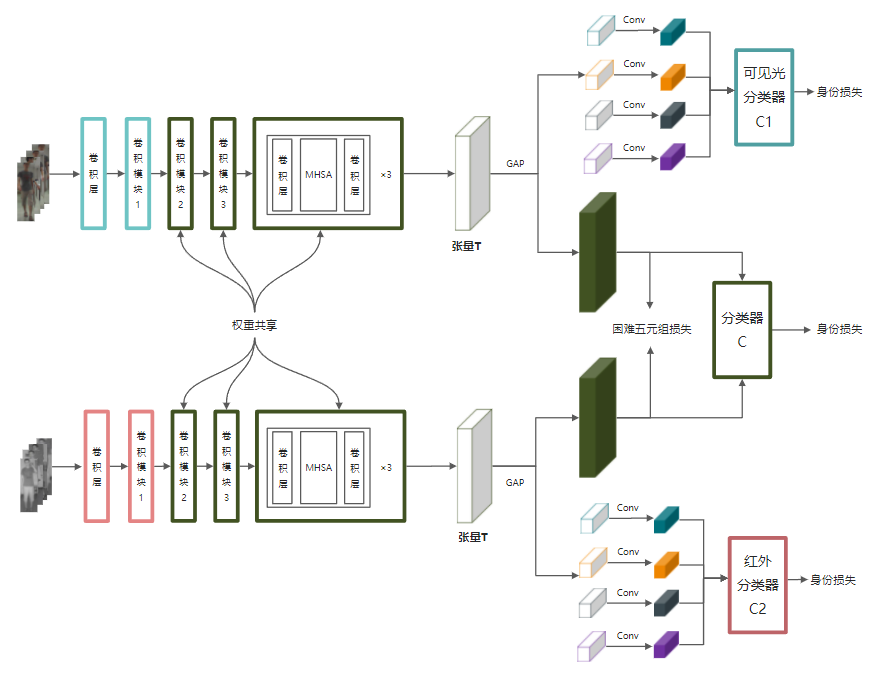


图2 本发明的跨模态行人重识别的模型架构

图2：采用基于注意力思想的BoTNet为骨干网络的双流网络提取全局特征，BoTNet最后卷积模块将卷积核为3×3的卷积层更换为多头自注意力层(MHSA)，增强网络架构的全局信息聚合能力。第一层卷积和卷积模块参数不共享，用于提取特定模态的特征。后面三个卷积模块及分类器C参数共享，用于提取两种模态之间的共同特征，该模型中的分类器包含一个全连接层和一个归一化层。对网络提取出的全局特征做水平划分，得到K个局部特征。用于局部特征的卷积层、可见光分类器C1及红外分类器C2参数不共享，使用身份损失和五元组损失来指导模型学习具有鲁棒性和辨别性的行人特征。

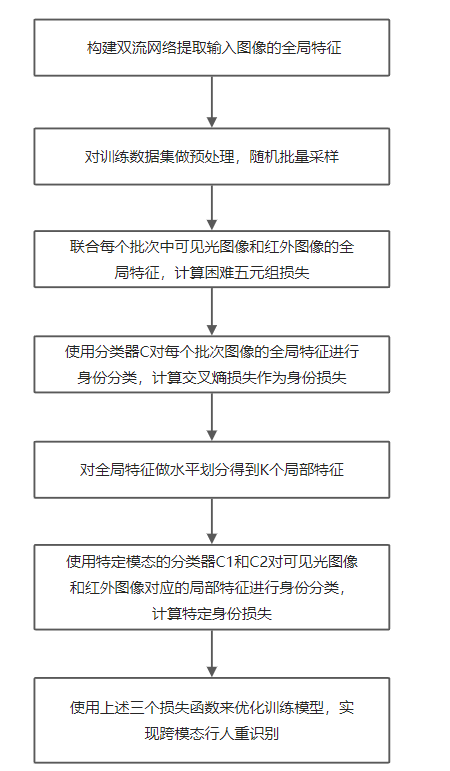


图3 本发明的跨模态行人重识别的流程图

图3： 构建双流网络提取输入图像的全局特征，对训练数据集做预处理，随机批量采样。联合每个批次中可见光图像和红外图像的全局特征，计算困难五元组损失。使用分类器C对每个批次图像的全局特征进行身份分类，计算交叉熵损失作为身份损失。对全局特征做水平划分得到K个局部特征。使用特定模态的分类器C1和C2对可见光图像和红外图像对应的局部特征进行身份分类，计算特定身份损失。使用上述三个损失函数来优化训练模型，实现跨模态行人重识别。

写技术交底书需注意：

1.英文缩写有中文译文，避免使用英文单词。

2.全文对同一事务的叫法应统一，避免出现一种东西多种叫法。

3.专利法规定：

1）专利必须是一个技术方案，应该阐述发明目的是通过什么技术方案来实现的，不能只有原理，也不能只做功能介绍；

2）专利必须充分公开，以本领域技术人员不需付出创造性劳动即可实现为准