**技术交底书格式**

|  |  |  |  |
| --- | --- | --- | --- |
| 专利名称 | 一种基于辅助模态和混合注意力的跨模态行人重识别方法 | 所属技术领域 | 图像检索 |

1. 本发明要解决的技术问题是什么？

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，现在所面临的主要挑战之一是可见光模态和红外模态之间存在巨大的差异。现有方法主要通过生成对抗网络进行模态互转或直接提取原始图像的模态共享特征来缓解模态差异。但是，模态互转需要复杂的生成网络和判别网络，并且生成的伪图像与真实图像之间仍然存在差距，容易引入噪声。同时，由于可见光图像和红外图像之间的存在巨大的模态差异，直接提取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。本发明设计了一种基于辅助模态和混合注意力的跨模态行人重识别方法，能有效地缓解模态差异，同时挖掘细微的、更具鉴别性的行人图像特征信息。

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（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受到许多关注，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有效的图像信息。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）。

跨模态行人重识别主要面临模态内行人图像的差异和模态间行人图像的差异所带来的困难和挑战。模态内行人图像的差异主要是由于不同的行人姿态、背景以及摄像头视角等因素造成的，例如待查询图像行人的姿势是正面，而行人图像候选库中可能是侧面、背面，甚至是被障碍物遮挡的行人图像。模态间行人图像的差异主要是由于可见光摄像头和红外摄像头不同的成像原理造成的，例如可见图像通常有三个通道，包含丰富的视觉信息，如形状、位置、颜色和纹理。而红外图像具有一个通道，主要包含轮廓和位置信息。

根据解决方式的不同，现有的跨模态行人重识别方法可以分成以下两种方法：模态共享特征学习和模态互转。

模态共享特征学习主要通过特征投影的方式，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。

模态互转主要使用生成对抗网络来生成伪图像，例如通过可见光图像来生成红外伪图像，实现了从一种模态转变成另一种模态，从而进行两种模态间的相互转换，把跨模态问题转变为单模态问题来解决。

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。）

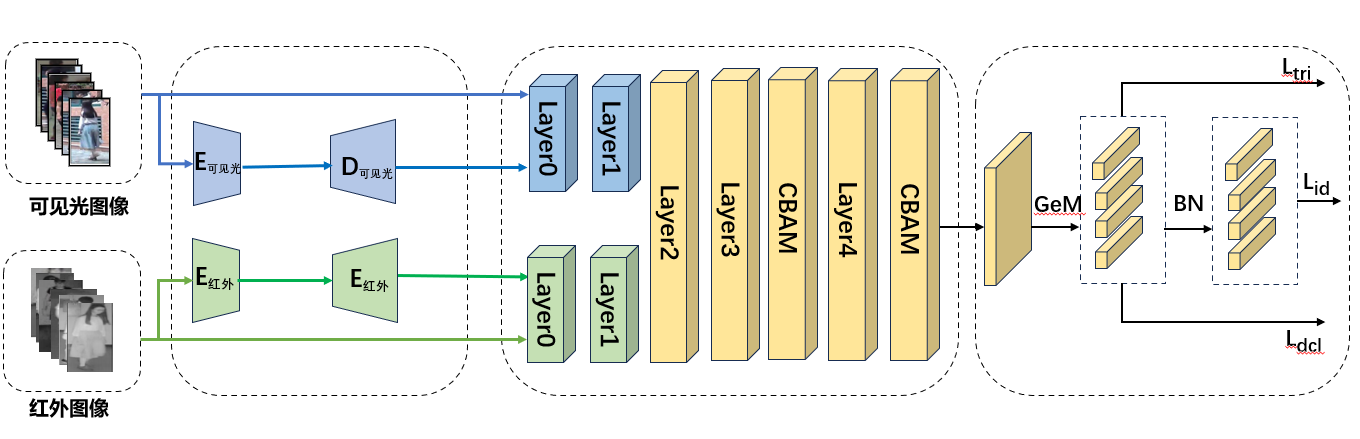
跨模态行人重识别面临的主要挑战之一是不同模态间存在巨大的差异。现有方法中，模态互转需要复杂的生成网络和判别网络，并且生成的伪图像与真实图像之间仍然存在差距，容易引入噪声。同时，由于可见光图像和红外图像之间的模态存在巨大的差异，直接提取模态共享特征十分困难，而且不能充分挖掘细微的、更具鉴别性的特征信息。

为了能有效地缓解模态差异，同时挖掘细微的、更具鉴别性的行人图像特征信息，本发明使用辅助模态来缓解模态差异，同时使用混合注意力机制来引导模型发现不同行人图像间的细微差异。

1. 本发明技术方案的基本内容。

本发明设计一种基于辅助模态和混合注意力的跨模态行人重识别方法，能有效地缓解模态差异，同时挖掘细微的、更具鉴别性的行人图像特征信息。

该方法的网络结构如下图所示。首先，将可见光图像与红外图像输入辅助模态生成器，以生成辅助模态图像。然后，生成的辅助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。其中，主干网络前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。在主干网络中嵌入CBAM（卷积注意力机制）模块，该模块属于一种混合注意力模块，它联合使用通道注意力和空间注意力，能帮助模型排除背景噪声的干扰，关注行人图像间的细微差异，挖掘更具鉴别性的特征信息。主干网络输出的特征图会被水平划分成若干个部分，进行局部特征的学习。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优化。



5、本发明技术方案的详细阐述。

本发明的工作流程如图所示，具体步骤：

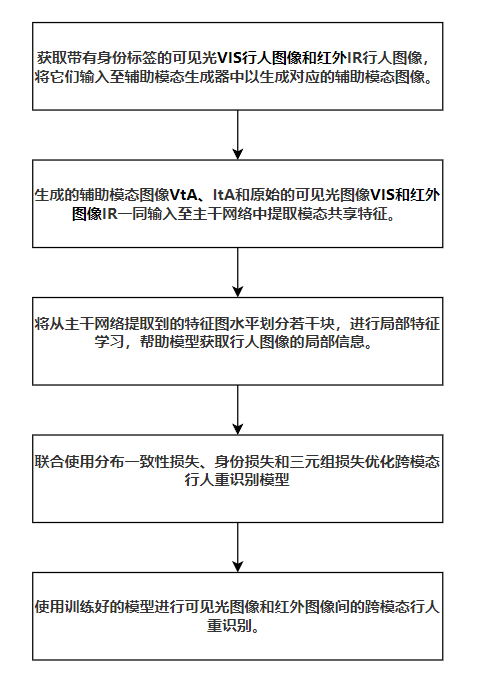
1.获取带有身份标签的可见光VIS行人图像和红外IR行人图像，将它们输入至辅助模态生成器中以生成对应的辅助模态图像。

2.生成的辅助模态图像VtA、ItA和原始的可见光图像VIS和红外图像IR一同输入至主干网络中提取模态共享特征。

3.将从主干网络提取到的特征图水平划分若干块，进行局部特征学习，帮助模型获取行人图像的局部信息。

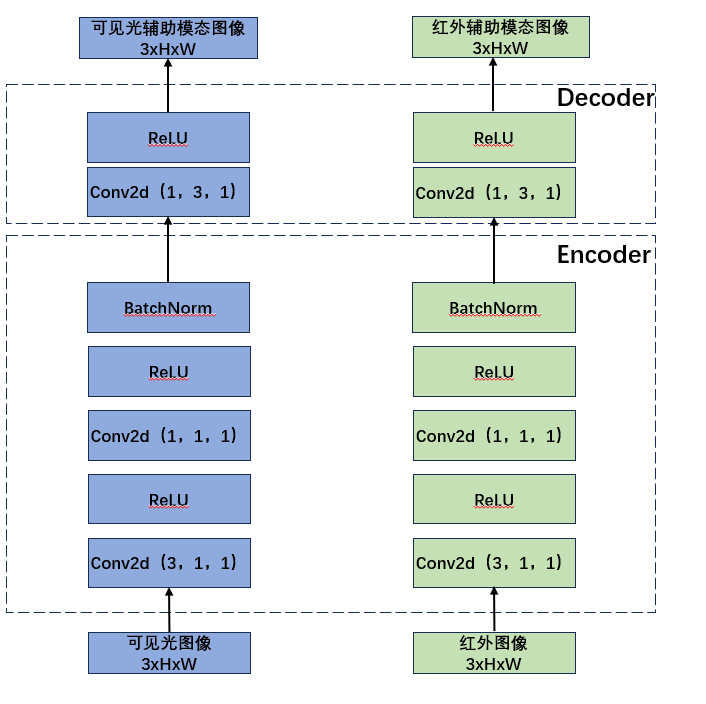
4.联合使用分布一致性损失、身份损失和三元组损失优化跨模态行人重识别模型

5.使用训练好的模型进行可见光图像和红外图像间的跨模态行人重识别。

****

**步骤1**所述辅助模态生成器结果如图所示，采用编码器-解码器架构，由两个编码器和两个解码器组成。生成辅助图像的过程为：原始的可见光图像和红外图像输入至辅助模态生成器中，可见光图像经过可见光模态信息编码器进行编码后输入至可见光模态解码器，解码器会将可见光图像投影到一个合适的辅助模态空间，以生成可见光的辅助模态图像。红外图像同理，经过红外模态信息编码器进行编码后输入至红外模态解码器以生成红外图像的辅助模态图像。

具体来说，首先对输入图像进行预处理，将输入尺寸调整为3×384×192，由于可见光图像和红外图像的通道数不同，因此将红外图像的通道数设置成3通道，与可见光图像的通道数保持一致。然后两种模态的图像分别输入至对应的模态信息编码器进行编码。在编码过程中，首先会通过一个卷积核尺寸大小为3\*1\*1的卷积层，将三通道图像转换为单通道图像，然后通过卷积核尺寸大小为1\*1\*1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尺寸大小为1\*1\*1的卷积层转换为三通道图像，以生成对应的辅助模态图像。



**步骤2**所述的主干网络基于ResNet50进行改进（该网络是目前主流的神经网络，它的源代码将网络结构分为了五层layer0、layer1、layer2、layer3、layer4，其实每一层layer里面的东西就是卷积、池化、残差连接等相关操作，这里通常概括为第零卷积块、第一卷积块、第二卷积块、第三卷积块、第四卷积块）。在此基础上，设计了一个双流网络（两条支路：一条支路处理是可见光图像+对应的辅助模态图像数据，对应图中蓝色部分；第二条支路是红外图像+对应的辅助模态图像数据，对应绿色部分）。具体而言，layer0和layer1是第零、一卷积块，分别处理对应支路的数据，在处理过程中，蓝色部分和绿色的layer0、1的参数是不同的，用于分别提取（可见光+辅助图像数据与红外图像+辅助图像数据这两种数据）特定的特征信息。在第零卷积块和第一卷积块提取完特定的特征信息后，layer2、3、4称为第二、第三、第四卷积块，它们参数是共享的（包括其中嵌入的CBAM模块，即图中主干网络的黄色部分都是参数共享的），目的是要提取两个支路数数据中共享的模态特征信息。前两个网络层参数独立，用以提取模态特定特征。后面的网络层参数共享，用以提取模态共享特征。其中，在网络层3和网络层4后面嵌入的CBAM（卷积注意力机制）模块，该模块采用了混合注意力机制，先后集成了通道注意力机制模块和空间注意力机制模块，能帮助模型排除背景噪声的干扰，关注行人图像间的细微差异，挖掘更具鉴别性的特征信息。CBAM模块的计算公式所下所示：



表示特征图F经过通道注意力模块后的输出，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。其定义如下所示：



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。

表示特征图F’经过空间注意力模块后的输出，空间注意力机制可以帮助模型了解哪些区域比较重要，其定义如下所示：



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。

**步骤3**所述的局部特征学习具体流程为：首先对主干网络提取的特征图进行广义平均池化，然后将其平均匀切分成四块，进行局部特征学习，这样可以帮助模型关注图像细节信息，减少背景噪声的干扰。广义均值池化，可以帮助模型获取图片的细粒度信息。它的定义如下所示：



表示池化层的输出，表示池化层的输入。是一个超参数，在反向传播中不断优化。当=1时，GeM广义均值池化近似于全局平均池化；当时，GeM广义均值池化近似于全局最大池化。

**步骤4**中提到的联合损失公式如下所示：



其中，是分布一致性损失,用于优化辅助模态生成器所生成的图像，缓解模态模态差异。它的定义如下所示：



N是训练阶段每一批次的训练样本的总和。和表示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。

其中，是交叉熵损失函数，它的定义如下所示：

，

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定义如下所示：



可见光图像和红外图像两种模态间的三元组损失定义如下：



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。

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)设计一个辅助模态生成器来生成可见光模态和红外模态所对应的辅助图像，有效地缓解了两种模态间的差异。辅助模态生成器能促进主干网络的优化过程，主干网络的优化也能促进辅助模态生成器生成更好的辅助模态图像。

（2）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。在主干网络中，前两个网络层参数独立，用以提取模态特定特征。后面的网络层参数共享，用以提取模态共享特征。在主干网络嵌入嵌入的CBAM（卷积注意力机制）模块，该模块使用混合注意力，联合通道注意力和空间注意力，能帮助模型排除背景噪声的干扰，关注行人图像间的细微差异，挖掘更具鉴别性的特征信息。然后对主干网络提取的特征图进行广义平均池化，然后将其水平均匀切分成四块，进行局部特征学习，帮助模型关注图像细节信息。

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：

1.相比于使用生成对抗网络进行模态互转，本发明所设计的辅助模态生成器是一个轻量级的生成器，减少了计算量，同时引入的辅助模态图像能有效地减缓两种模态图像间的模态差异。

2.大部分现有方法主要通过特征投影的方式来解决跨模态行人重识别问题，本发明设计的双流网络有效地缓解了模态差异的同时，能发现不同行人图像间的细微差异，提取出更具鲁棒性和鉴别性的特征信息，提高模型的性能。

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。模型在测试中的mAP（平均精度均值）和rank-k值得到明显地提高，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

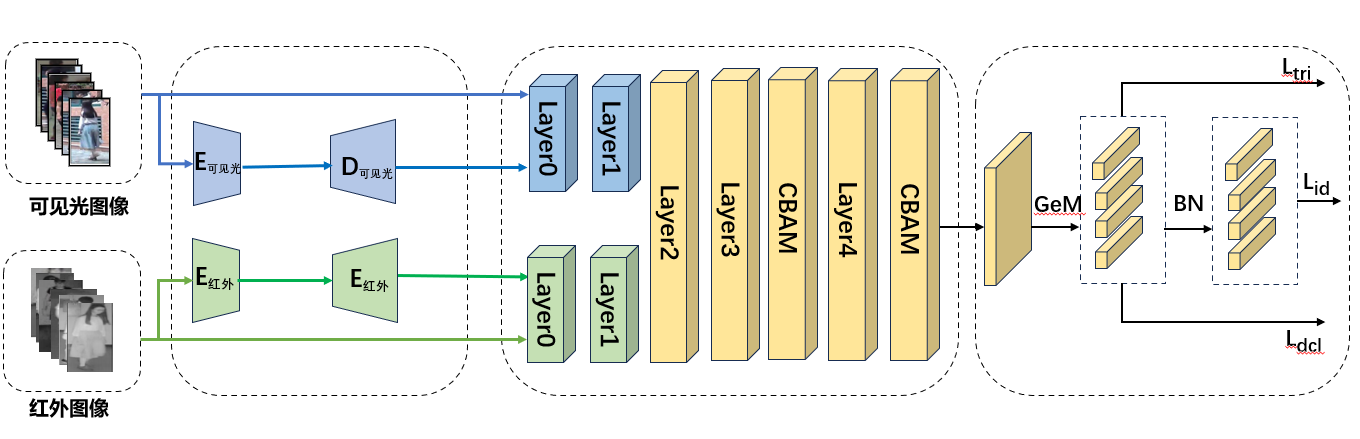


图1 本发明的模型网络结构图

图1: 本发明设计一种基于辅助模态和混合注意力的跨模态行人重识别方法，能有效地缓解模态差异，同时挖掘细微的、更具鉴别性的特征信息。

该方法的网络结构如下图所示。首先，将可见光图像与红外图像输入辅助模态生成器，以生成辅助模态图像。然后，生成的辅助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。其中，主干网络前两层网络层的参数是独立的，用于提取底层的模态特定特征，后面的网络层参数共享，用于提取模态共享特征。在主干网络中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。主干网络输出的特征图会被水平划分成若干个部分，进行局部特征的学习。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优化。

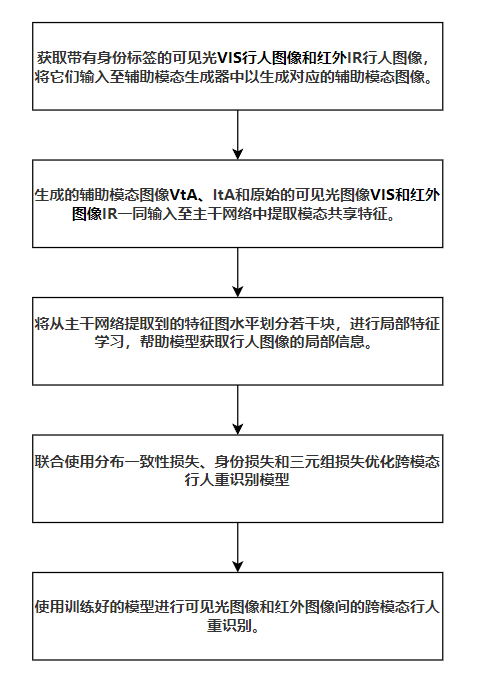
****

图2本发明的工作流程图

本发明的工作流程如图所示，具体步骤：

1.获取带有身份标签的可见光VIS行人图像和红外IR行人图像，将它们输入至辅助模态生成器中以生成对应的辅助模态图像。

2.生成的辅助模态图像VtA、ItA和原始的可见光图像VIS和红外图像IR一同输入至主干网络中提取模态共享特征。

3.将从主干网络提取到的特征图水平划分若干块，进行局部特征学习，帮助模型获取行人图像的局部信息。

4.联合使用分布一致性损失、身份损失和三元组损失优化跨模态行人重识别模型

5.使用训练好的模型进行可见光图像和红外图像间的跨模态行人重识别。

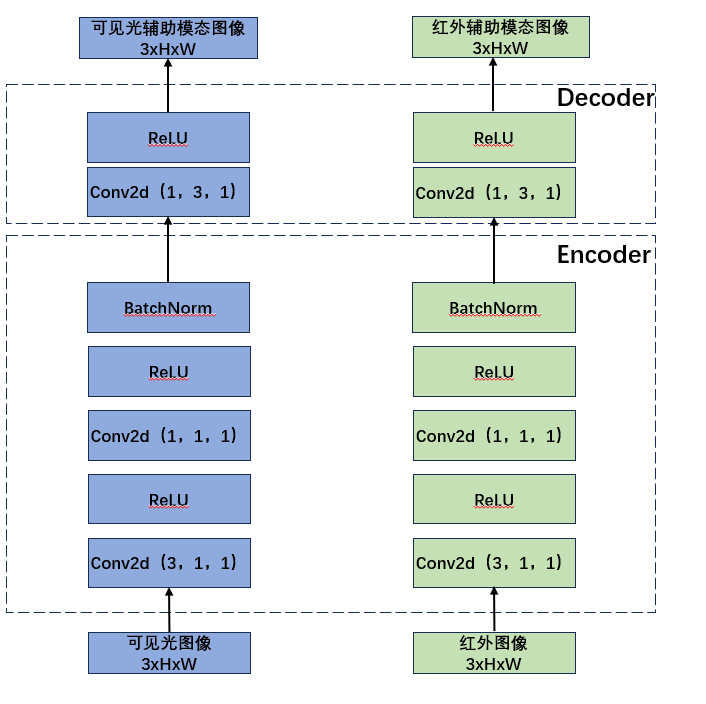


图3 辅助模态生成器

辅助模态生成器结果如图所示，采用编码器-解码器架构，由两个编码器和两个解码器组成。生成辅助图像的过程为：原始的可见光图像和红外图像输入至辅助模态生成器中，可见光图像经过可见光模态信息编码器进行编码后输入至可见光模态解码器，解码器会将可见光图像投影到一个合适的辅助模态空间，以生成可见光的辅助模态图像。红外图像同理，经过红外模态信息编码器进行编码后输入至红外模态解码器以生成红外图像的辅助模态图像。

具体来说，首先对输入图像进行预处理，将输入尺寸调整为3×384×192，由于可见光图像和红外图像的通道数不同，因此将红外图像的通道数设置成3通道，与可见光图像的通道数保持一致。然后两种模态的图像分别输入至对应的模态信息编码器进行编码。在编码过程中，首先会通过一个卷积核为3\*1\*1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1\*1\*1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1\*1\*1的卷积层转换为三通道图像，以生成对应的辅助模态图像。

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