# 辅助模态生成器介绍

**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\*3\*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。

模态特定的编码器部分[0029]

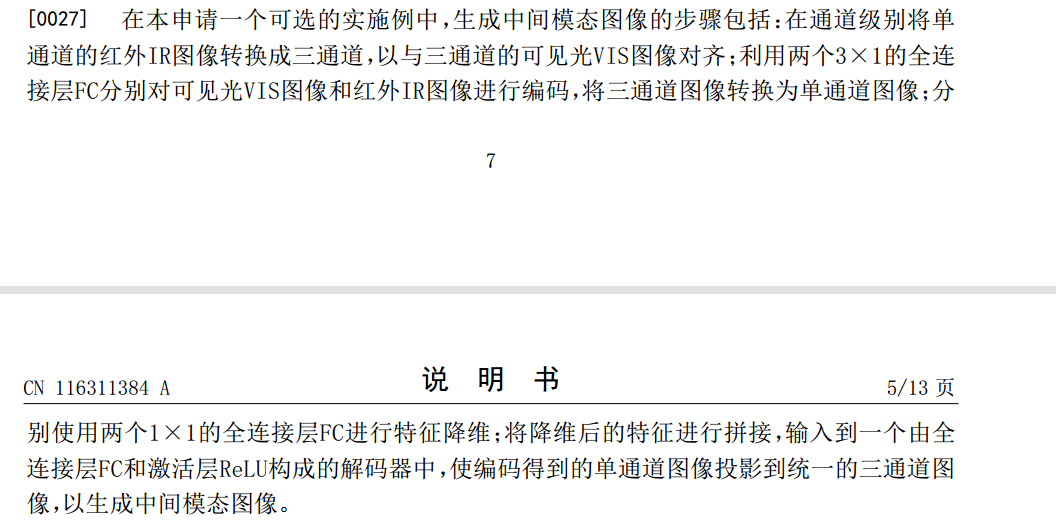
在跨模态行人重识别中，模态差异主要存在于白天的三通道彩色图像(VIS图像)与夜间的单通道灰度图像(NIR图像)之间，因此，本发明首先将这两种图像的差异在通道层面实现对齐，这里采用由两个3\*1全连接层(为方便，在输入网络时本发明将单通道NIR图像设置为三通道的灰度图像)组成的模态特定的编码器Evis和Enir来实现这个目的，从而得到两种1\*1的单通道图像；

[0030]2)模态特定的非线性激活部分

[0031]由于跨模态的两种图像之间存在着巨大的非线性跨模态差异，接下来本发明将上述两种1\*1的单通道图像输入到两个3\*1全连接层后跟ReLU激活函数中用于提升其特征表达能力；

[0032]3)模态共享的解码器部分[0033]然后上述两种通过ReLU激活函数的1\*1的单通道图像输入到一个模态共享的由1\*3全连接层组成的解码器中，从而将这两种模态的图像解码到一个共享的图像空间中，得到本发明发明所提出的中间模态图像(本发明中采用VtM和NtM表示这两种中间模态图像)。

[0034]最后，本发明将原始的VIS图像、NIR图像与中间模态图像一起输入到ResNet‑50网络中用于辅助原始图像进行学习，通过损失函数进行自监督的优化，实现降低模态差异的目的。



从上图可以看出，中间模式生成器由编码器和解码器组成。中间模态生成器将具有相同标签的两种模态的图像对作为输入。输入图像尺寸统一调整为3×384×192，其中单通道红外图像复制到三个通道，以确保与三通道可见光图像对齐。然后，两种模态的图像都输入编码器进行编码，首先通过Covn2d(3,1,1)将三通道图像转换为单通道图像，然后通过Covn2d(1,1,1)将减少计算量，最后通过BN层进行数据归一化。然后将归一化后的数据输入到解码器进行解码，将编码后的单通道图像通过Covn2d(1,3,1)转换为三通道图像，生成与标签相同的中间模态图像