



(12) 发明专利申请

(10) 申请公布号 CN 117314808 A

(43) 申请公布日 2023. 12. 29

(21) 申请号 202311218609.1

(22) 申请日 2023.09.20

(71) 申请人 南京信息工程大学

地址 210044 江苏省南京市江北新区宁六路219号

(72) 发明人 陆成 刘雪明

(74) 专利代理机构 南京经纬专利商标代理有限公司 32200

专利代理师 朱小兵

(51) Int. Cl.

G06T 5/50 (2006.01)

G06N 3/0455 (2023.01)

G06N 3/0464 (2023.01)

G06N 3/08 (2023.01)

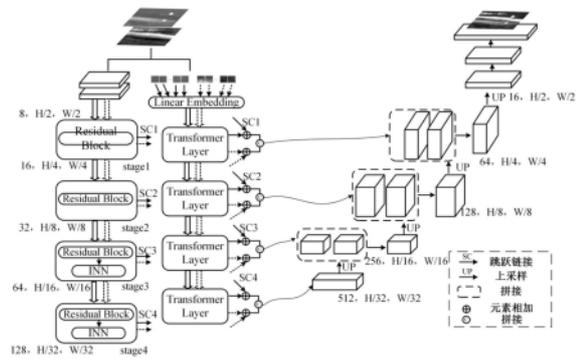
权利要求书4页 说明书11页 附图6页

(54) 发明名称

一种结合Transformer与CNN双编码器的红外与可见光图像融合方法

(57) 摘要

本发明公开了一种结合Transformer与CNN双编码器的红外与可见光图像融合方法,基于U-Net框架,构建结合Transformer与CNN双编码器的红外与可见光图像融合模型,其中基于U-Net框架的融合网络由双编码器和解码器组成,再利用训练集训练该模型;将红外图像、可见光图像输入到训练完成的结合Transformer与CNN双编码器的红外与可见光图像融合模型中,利用双编码器提取红外图像和可见光图像预融合的红外特征和可见光特征,并通过自上而下的方式进行特征表示;再利用解码器将特征表示映射到原始分辨率,通过自下而上的方式逐步融合图像特征,得到融合图像。本发明有效地将局部信息与全局信息相结合,提升融合图像质量,同时降低单一Transformer结构带来的计算复杂度。



1. 一种结合Transformer与CNN双编码器的红外与可见光图像融合方法,其特征在于,包括:

S1、基于U-Net框架,构建结合Transformer与CNN双编码器的红外与可见光图像融合模型,其中该模型由双编码器和解码器组成;

S2、选定数据集,对其进行处理后得到训练集,利用训练集训练步骤S1中的结合Transformer与CNN双编码器的红外与可见光图像融合模型;

S3、将红外图像、可见光图像输入到训练完成的结合Transformer与CNN双编码器的红外与可见光图像融合模型中,利用双编码器提取红外图像和可见光图像预融合的红外特征和可见光特征,并通过自上而下的方式进行特征表示;

S4、利用解码器将步骤S3中的特征表示映射到原始分辨率,通过自下而上的方式逐步融合图像特征,得到融合图像。

2. 根据权利要求1所述的结合Transformer与CNN双编码器的红外与可见光图像融合方法,其特征在于,步骤S1中,构建结合Transformer与CNN双编码器的红外与可见光图像融合模型包括以下内容:

S101、构建双编码器:双编码器包括局部细节特征提取编码器和全局信息提取编码器;

其中,局部细节特征提取编码器包括依次连接的卷积层和四个特征提取块,卷积层由卷积核为 4×4 、步长为2、填充为1的卷积以及批量归一化、LeakyReLU串行连接构成,卷积层用于下采样操作,并调整输入图像的通道数和尺寸;

第一、二特征提取块由残差模块构成,第三、四特征提取块由残差模块和可逆神经网络串行连接构成,四个特征提取块输入的通道数分别为16、32、64、128,输出的通道数分别为32、64、128、256;

全局信息提取编码器包括四个通过串行方式连接的Transformer Layer,其输入的通道数分别为16、32、64、128,输出的通道数分别为32、64、128、256;

S102、构建解码器:解码器包括四个特征融合模块和重建模块;每个特征融合模块包括上采样操作,拼接操作,依次连接的卷积核为 4×4 、步长为2、填充为1的反卷积,BatchNorm操作,激活函数ReLU;

重建模块包括依次连接的上采样操作,卷积核为 4×4 的卷积、填充为1的卷积,填充操作,双曲正切函数函数;

S103、双编码器通过自上而下方式逐步下采样进行特征提取,解码器通过自下而上的方式将双编码器提取的特征逐步上采样进行融合并重建,构成结合Transformer与CNN双编码器的红外与可见光图像融合模型。

3. 根据权利要求2所述的结合Transformer与CNN双编码器的红外与可见光图像融合方法,其特征在于,残差模块包括主路径和残差路径,主路径包含卷积核为 1×1 的卷积、BN、LeakyReLU激活函数、卷积核为 3×3 的卷积、卷积核为 4×4 的卷积;残差路径包括DSConv和卷积核为 4×4 的卷积;

可逆神经网络包括三个可逆残差模块,其包括卷积核为 1×1 的卷积、卷积核为 3×3 的深度可分离卷积、激活函数ReLU6。

4. 根据权利要求2所述的结合Transformer与CNN双编码器的红外与可见光图像融合方法,其特征在于,第一、二、四Transformer Layer包括两层混合注意力变换器结构,其中,第

一Transformer Layer的第一层包括通过串行方式连接的补丁嵌入操作、层归一化操作、W-MSA模块、局部-非局部注意力信息增强模块、层归一化操作和前馈神经网络,第二层包括通过串行方式连接的层归一化操作、SW-MSA模块、局部-非局部注意力信息增强模块、层归一化操作和前馈神经网络;

第二、四Transformer Layer的第一层包括通过串行方式连接的补丁合并操作、层归一化操作、W-MSA模块、局部-非局部注意力信息增强模块、层归一化操作和前馈神经网络,第二层包括通过串行方式连接的层归一化操作、SW-MSA模块、局部-非局部注意力信息增强模块、层归一化操作和前馈神经网络;

第三Transformer Layer包括六层混合注意力变换器结构,由第二Transformer Layer中的两层混合注意力变换器结构堆叠3次构成。

5. 根据权利要求4所述的结合Transformer与CNN双编码器的红外与可见光图像融合方法,其特征在于,局部-非局部注意力信息增强模块包括局部信息增强模块和非局部信息增强模块;其中局部信息增强模块包括拼接操作、全局平均池化、深度可分离卷积、Sigmoid函数、逐元素相乘操作和逐元素相乘操作;非局部信息增强模块包括LayerNorm操作和位置注意力模块。

6. 根据权利要求1所述的结合Transformer与CNN双编码器的红外与可见光图像融合方法,其特征在于,步骤S2中,训练结合Transformer与CNN双编码器的红外与可见光图像融合模型包括以下子步骤:

S201、从TNO数据集中选定m对图像作为数据集,将图像灰度值转为 $[-1, 1]$,利用 $h \times w$ 的窗口裁剪图像,步长设置为s,最终获得n对图像块作为训练集,其中,h表示图像的高,w表示图像的宽;

S202、采用结构相似性损失和对比损失监督结合Transformer与CNN双编码器的红外与可见光图像融合模型训练,获得最优网络参数,具体公式为:

$$L_{\text{total}} = \lambda_1 L_{\text{SSIM}} + \lambda_2 L_{\text{patchNCE}}$$

其中, L_{total} 表示损失函数, λ_1 、 λ_2 表示超参数, L_{SSIM} 表示结构相似性损失, L_{patchNCE} 表示对比损失;

S203、根据训练集和损失函数 L_{total} 对初始的结合Transformer与CNN双编码器的红外与可见光图像融合模型进行训练,训练过程使用的Adam优化器更新网络模型参数,得到训练完成的结合Transformer与CNN双编码器的红外与可见光图像融合模型。

7. 根据权利要求6所述的结合Transformer与CNN双编码器的红外与可见光图像融合方法,其特征在于,步骤S3中,利用双编码器提取特征包括以下子步骤:

S301、如步骤S201的操作内容对输入图像进行处理;

S302、将步骤S301中处理后的红外图像、可见光图像输入到双编码器,经过卷积层的下采样操作,调整该图像的通道数和尺寸;

S303、将步骤S302处理后的图像输入到四个特征提取块中,依次通过 1×1 的卷积、BN、LeakyReLU激活函数、卷积核为 3×3 的卷积、BN和卷积核为 4×4 的卷积操作后得到主路径中学习的卷积特征,再经过DSConv和卷积核为 4×4 的卷积卷积操作后得到补充信息特征,将卷积特征和补充信息特征相加后再经过LeakyReLU后得到信息增强特征;

S304、将步骤S303得到的信息增强特征输入到可逆神经网络中,将该特征 $F_{i,k}^S$ 在通道维

度上被均分成 $F_{I,k}^S[1:c]$ 和 $F_{I,k}^S[c+1:C]$ 两部分, 特征 $F_{I,k}^S[1:c]$ 经过第一个可逆残差模块得到的特征 $R_1(F_{I,k}^S[1:c])$ 与特征 $F_{I,k}^S[c+1:C]$ 逐元素相加得到特征 $F_{I,k+1}^S[c+1:C]$, 特征 $F_{I,k}^S[c+1:C]$ 经过第二个可逆残差模块与特征 $F_{I,k}^S[1:c]$ 逐元素相乘得到特征 $F_{I,k}^S[1:c] \otimes \exp(R_2(F_{I,k+1}^S[c+1:C]))$, 特征 $F_{I,k}^S[c+1:C]$ 经过第三个可逆残差模块得到特征 $(R_3(F_{I,k+1}^S[c+1:C]))$ 并与特征 $F_{I,k}^S[1:c] \otimes \exp(R_2(F_{I,k+1}^S[c+1:C]))$ 逐元素相加得到特征 $F_{I,k+1}^S[1:c]$, 将特征 $F_{I,k+1}^S[c+1:C]$ 和特征 $F_{I,k+1}^S[1:c]$ 在通道维度上拼接得到输出特征 $F_{I,k+1}^S$, 具体公式为:

$$F_{I,k+1}^S[c+1:C] = F_{I,k}^S[c+1:C] + R_1(F_{I,k}^S[1:c]);$$

$$F_{I,k+1}^S[1:c] = F_{I,k}^S[1:c] \otimes \exp(R_2(F_{I,k+1}^S[c+1:C])) + (R_3(F_{I,k+1}^S[c+1:C]));$$

$$F_{I,k+1}^S = \text{cat}\{F_{I,k+1}^S[1:c], F_{I,k+1}^S[c+1:C]\};$$

其中, C 表示完整通道数, c 表示完整通道数的一半, $[1:c]$ 表示当前特征通道数为 1 至 c , $[c+1:C]$ 表示当前特征通道数为 $c+1$ 至 C , S 表示 CNN 特征提取编码器, I 表示任意映射符, k 表示第 k 个可逆网络层数, R_n ($n=1, 2, 3$) 表示可逆残差模块, \exp 表示指数 e ;

最终输出红外局部细节特征和可见光局部细节特征;

S305、在第一 Transformer Layer 中, 红外图像和可见光图像的特征分别经过补丁嵌入操作后, 通过层归一化操作对该特征进行标准化, 再将其中输入到 W-MSA 模块, 按照设定的窗口大小对图像进行分块, 得到自注意力图, 再经过局部-非局部注意力信息增强模块, 得到局部-非局部注意力信息增强特征, 将该增强特征逐元素相加经得到注意力特征 F^1 , 经过层归一化操作操作后再经过 FFN 模块得到特征, 将该特征与注意力特征相加得到第一层 HAT 结构的输出特征 F^{1+1} ;

将输出特征 F^{1+1} 输入第二层 HAT 结构, 经过补丁嵌入操作后, 通过层归一化操作对特征进行标准化, 经过标准化的特征输入到 SW-MSA 模块, SW-MSA 模块按照设定的窗口大小对图像进行分块, 得到自注意力图, 自注意力图再经过局部-非局部注意力信息增强模块, 得到局部-非局部注意力信息增强特征, 将局部-非局部注意力信息增强特征逐元素相加经得到注意力特征 F^{1+1} , 注意力特征经过层归一化操作后经过 FFN 模块得到的特征与注意力特征相加得到第一 Transformer Layer 的红外全局信息特征和可见光全局信息特征, 具体公式为:

$$F^1 = \text{LG}(W\text{-MSA}(\text{LN}(F^{1-1}))) + F^{1-1};$$

$$F_1^1 = \text{FFN}(\text{LN}(F^1)) + F^1;$$

$$F^{1+1} = \text{LG}(SW\text{-MSA}(\text{LN}(F_1^1))) + F_1^1;$$

$$F_1^{1+1} = \text{FFN}(\text{LN}(F^{1+1})) + F^{1+1};$$

其中, F^{1-1} 表示红外图像和可见光图像的特征, F^1 表示第一层 HAT 输出的特征, F^{1+1} 表示红外全局信息特征和可见光全局信息特征, $\text{LG}()$ 表示局部-非局部信息增强操作, $\text{LN}()$ 表示层归一化操作, $W\text{-MSA}()$ 表示窗口的多头自注意力模块, $SW\text{-MSA}()$ 表示移位窗口的多头自

注意力模块操作；

S306、第二、三、四Transformer Layer重复步骤S305的内容,其中,第三Transformer Layer需重复三次后再输入第四Transformer Layer后输出红外与可见光的全局信息特征；

S307、将双编码器每个特征提取块与红外、可见光全局信息特征和红外、可见光局部细节特征对应模态逐元素相加得到四组预融合的红外特征和可见光特征。

8.根据权利要求7所述的结合Transformer与CNN双编码器的红外与可见光图像融合方法,其特征在于,步骤S305中,局部-非局部注意力信息增强模块提取特征的具体内容为：

输入特征分别经过全局平均池化、深度可分离卷积、Sigmoid操作得到局部信息增强图,再与原输入特征经过逐元素相乘操作得到局部信息增强特征;输入特征经过层归一化操作、位置注意力模块得到非局部信息增强特征,再与局部信息增强特征逐元素相加得到局部-非局部注意力信息增强特征。

9.根据权利要求1所述的结合Transformer与CNN双编码器的红外与可见光图像融合方法,其特征在于,步骤S4中,利用解码器得到融合图像包括以下子步骤：

S401、将第四组预融合的红外特征和可见光特征拼接后,经过上采样操作后与第三组预融合的红外特征和可见光特征在通道维度上拼接得到第三组重建特征,将第三组重建特征经过上采样操作后与第二组预融合的红外特征和可见光特征在通道维度上拼接得到第二组重建特征,将第二组重建特征经过上采样操作后与第三组预融合的红外特征和可见光特征在通道维度上拼接得到重建特征；

S402、将步骤S401中的重建特征经过上采样操作、填充操作、 4×4 卷积操作、Tanh操作后最终得到重建的融合图像。

一种结合Transformer与CNN双编码器的红外与可见光图像融合方法

技术领域

[0001] 本发明涉及图像处理技术领域,具体涉及一种结合Transformer与CNN双编码器的红外与可见光图像融合方法。

背景技术

[0002] 现有的图像融合方法主要分为两种:一种是传统融合方法,另一种是基于深度学习的图像融合方法。传统算法主要包括稀疏表示、多尺度变换、子空间和混合融合方法等,这些算法通常将图像分解为不同层级特征或依据图像特征建立稀疏矩阵,然后通过分解结果结合所设计的融合策略进行图像融合。虽然一定程度上能够保留源图像部分细节和突出目标,但源图像之间的模态差异使得传统方法手工设计越来越复杂,融合性能很难进一步提升,同时计算复杂度高,耗费时间长,并且容易忽略融合过程中的空间一致性,传统模型泛化性能较差。

[0003] 现有的基于深度学习的图像融合方法融合性能较好,网络模型具有更强的特征提取能力。目前主要的融合模型分为CNN(Convolution Neural Network,卷积神经网络)和GAN(Generative Adversarial Network,生成对抗网络)。基于CNN和GAN的图像融合方法通过卷积层分别对红外图像和可见光图像进行特征提取,一定程度上保留了红外像素幅度和可见光细节信息,但卷积核提取的特征仅在当前位置有较强的区域相关性,未考虑长距离上下文信息的提取能力,并且融合图像的细节纹理信息提取不充分,无法从全局角度增强信息融合。针对现有的红外与图像融合方法存在的问题,很多学者通过使用Transformer来实现信息远程依赖性,这主要得益于它的自注意力机制和全局特征提取能力,但仅通过Transformer来提取特征需要很大计算资源。

发明内容

[0004] 本发明所要解决的技术问题:提供一种结合Transformer与CNN双编码器的红外与可见光图像融合方法,将CNN与Transformer相结合,形成并行的双编码器结构,既能够关注局部上下文信息和全局信息,又能利用高效地计算优势训练网络模型,提升融合图像质量。

[0005] 本发明为解决以上技术问题而采用以下技术方案:

[0006] 本发明提出的一种结合Transformer与CNN双编码器的红外与可见光图像融合方法,包括:

[0007] S1、基于U-Net框架,构建结合Transformer与CNN双编码器的红外与可见光图像融合模型,其中该模型由双编码器和解码器组成。

[0008] S2、选定数据集,对其进行处理后得到训练集,利用训练集训练步骤S1中的结合Transformer与CNN双编码器的红外与可见光图像融合模型。

[0009] S3、将红外图像、可见光图像输入到训练完成的结合Transformer与CNN双编码器的红外与可见光图像融合模型中,利用双编码器提取红外图像和可见光图像预融合的红外

特征和可见光特征,并通过自上而下的方式进行特征表示。

[0010] S4、利用解码器将步骤S3中的特征表示映射到原始分辨率,通过自下而上的方式逐步融合图像特征,得到融合图像。

[0011] 进一步的,步骤S1中,构建结合Transformer与CNN双编码器的红外与可见光图像融合模型包括以下内容:

[0012] S101、构建双编码器:双编码器包括局部细节特征提取编码器和全局信息提取编码器。

[0013] 其中,局部细节特征提取编码器包括依次连接的卷积层和四个特征提取块,卷积层由卷积核为 4×4 、步长为2、填充为1的卷积以及批量归一化、LeakyReLU串行连接构成,卷积层用于下采样操作,并调整输入图像的通道数和尺寸。

[0014] 第一、二特征提取块由残差模块构成,第三、四特征提取块由残差模块和INN(Invertible Neural networks,可逆神经网络)串行连接构成,四个特征提取块输入的通道数分别为16、32、64、128,输出的通道数分别为32、64、128、256。

[0015] 可逆神经网络包括三个可逆残差模块,其包括卷积核为 1×1 的卷积、卷积核为 3×3 的深度可分离卷积、激活函数(ReLU6)。

[0016] 全局信息提取编码器包括四个通过串行方式连接的Transformer Layer,其输入的通道数分别为16、32、64、128,输出的通道数分别为32、64、128、256,第一、二、四Transformer Layer包括两层HAT(Hybrid Attention Transformer,混合注意力变换器)结构,其中,第一Transformer Layer的第一层包括通过串行方式连接的补丁嵌入操作、层归一化操作、W-MSA模块(Windows Multi-Head Self-Attention,窗口多头自注意力)、局部-非局部注意力信息增强模块、层归一化操作和前馈神经网络,第二层包括通过串行方式连接的层归一化操作、SW-MSA模块(Shift Windows Multi-Head Self-Attention,变换窗口多头自注意力)、局部-非局部注意力信息增强模块、层归一化操作和前馈神经网络。

[0017] 第二、四Transformer Layer的第一层包括通过串行方式连接的补丁合并操作、层归一化操作、W-MSA模块(Windows Multi-Head Self-Attention,窗口多头自注意力)、局部-非局部注意力信息增强模块、层归一化操作和前馈神经网络,第二层包括通过串行方式连接的层归一化操作、SW-MSA模块(Shift Windows Multi-Head Self-Attention,变换窗口多头自注意力)、局部-非局部注意力信息增强模块、层归一化操作和前馈神经网络。

[0018] 第三Transformer Layer包括六层混合注意力变换器结构,由第二Transformer Layer中的两层HAT(Hybrid Attention Transformer,混合注意力变换器)结构堆叠3次构成。

[0019] S102、构建解码器

[0020] 解码器包括四个特征融合模块和重建模块;每个特征融合模块包括上采样操作,拼接操作,依次连接的卷积核为 4×4 、步长为2、填充为1的反卷积,BatchNorm操作,激活函数ReLU。

[0021] 重建模块包括依次连接的上采样操作,卷积核为 4×4 的卷积、填充为1的卷积,填充操作,双曲正切函数函数(Tanh)。

[0022] S103、双编码器通过自上而下方式逐步下采样进行特征提取,解码器通过自下而上的方式将双编码器提取的特征逐步上采样进行融合并重建,构成结合Transformer与CNN

双编码器的红外与可见光图像融合模型。

[0023] 进一步的,步骤S1中,残差模块包括主路径和残差路径,主路径包含卷积核为 1×1 的卷积、BN、LeakyReLU激活函数、卷积核为 3×3 的卷积、卷积核为 4×4 的卷积;残差路径包括DSConv和卷积核为 4×4 的卷积。

[0024] 进一步的,步骤S1中,局部-非局部注意力信息增强模块包括局部信息增强模块和非局部信息增强模块;其中局部信息增强模块包括拼接操作、全局平均池化、深度可分离卷积、Sigmoid函数、逐元素相乘操作和逐元素相乘操作;非局部信息增强模块包括LayerNorm操作和位置注意力模块。

[0025] 进一步的,步骤S2中,训练结合Transformer与CNN双编码器的红外与可见光图像融合模型包括以下子步骤:

[0026] S201、从TNO数据集中选定m对图像作为数据集,将图像灰度值转为 $[-1, 1]$,利用 $h \times w$ 的窗口裁剪图像,步长设置为s,最终获得n对图像块作为训练集,其中,h表示图像的高,w表示图像的宽。

[0027] S202、采用结构相似性损失和对比损失监督结合Transformer与CNN双编码器的红外与可见光图像融合模型训练,获得最优网络参数,具体公式为:

$$[0028] \quad L_{\text{total}} = \lambda_1 L_{\text{SSIM}} + \lambda_2 L_{\text{patchNCE}}$$

[0029] 其中, L_{total} 表示损失函数, λ_1 、 λ_2 表示超参数, L_{SSIM} 表示结构相似性损失, L_{patchNCE} 表示对比损失。

[0030] S203、根据训练集和损失函数 L_{total} 对初始的结合Transformer与CNN双编码器的红外与可见光图像融合模型进行训练,训练过程使用的Adam优化器更新网络模型参数,得到训练完成的结合Transformer与CNN双编码器的红外与可见光图像融合模型。

[0031] 进一步的,步骤S3中,利用双编码器提取特征包括以下子步骤:

[0032] S301、将红外图像和可见光图像的灰度值转为 $[-1, 1]$,利用 $h \times w$ 的窗口裁剪图像,步长设置为s;

[0033] S302、将步骤S301中处理后的红外图像、可见光图像输入到双编码器,经过卷积层的下采样操作,调整该图像的通道数和尺寸;

[0034] S303、将步骤S302处理后的图像输入到四个特征提取块中,依次通过 1×1 的卷积、BN、LeakyReLU激活函数、卷积核为 3×3 的卷积、BN和卷积核为 4×4 的卷积操作后得到主路径中学习的卷积特征,再经过DSConv和卷积核为 4×4 的卷积卷积操作后得到补充信息特征,将卷积特征和补充信息特征相加后再经过LeakyReLU后得到信息增强特征;

[0035] S304、将步骤S303得到的信息增强特征输入到可逆神经网络中,将该特征 $F_{I,k}^S$ 在通道维度上被均分成 $F_{I,k}^S[1:c]$ 和 $F_{I,k}^S[c+1:C]$ 两部分,特征 $F_{I,k}^S[1:c]$ 经过第一个可逆残差模块得到的特征 $R_1(F_{I,k}^S[1:c])$ 与特征 $F_{I,k}^S[c+1:C]$ 逐元素相加得到特征 $F_{I,k+1}^S[c+1:C]$,特征 $F_{I,k}^S[c+1:C]$ 经过第二个可逆残差模块与特征 $F_{I,k}^S[1:c]$ 逐元素相乘得到特征 $F_{I,k}^S[1:c] \otimes \exp(R_2(F_{I,k+1}^S[c+1:C]))$,特征 $F_{I,k}^S[c+1:C]$ 经过第三个可逆残差模块得到特征 $(R_3(F_{I,k+1}^S[c+1:C]))$ 并与特征 $F_{I,k}^S[1:c] \otimes \exp(R_2(F_{I,k+1}^S[c+1:C]))$ 逐元素相加得到特征

$F_{I,k+1}^S[1:c]$,将特征 $F_{I,k+1}^S[c+1:C]$ 和特征 $F_{I,k+1}^S[1:c]$ 在通道维度上拼接得到输出特征 $F_{I,k+1}^S$,具体公式为:

$$[0036] \quad F_{I,k+1}^S[c+1:C] = F_{I,k}^S[c+1:C] + R_1(F_{I,k}^S[1:c]);$$

$$[0037] \quad F_{I,k+1}^S[1:c] = F_{I,k}^S[1:c] \otimes \exp(R_2(F_{I,k+1}^S[c+1:C])) + (R_3(F_{I,k+1}^S[c+1:C]));$$

$$[0038] \quad F_{I,k+1}^S = \text{cat}\{F_{I,k+1}^S[1:c], F_{I,k+1}^S[c+1:C]\};$$

[0039] 其中,C表示完整通道数,c表示完整通道数的一半,[1:c]表示当前特征通道数为1至c,[c+1:C]表示当前特征通道数为c+1至C,S表示CNN特征提取编码器,I表示任意映射符,k表示第k个可逆网络层数, R_n ($n=1,2,3$)表示可逆残差模块,exp表示指数e。

[0040] 最终输出红外局部细节特征和可见光局部细节特征。

[0041] 可逆神经网络是通过可逆性设计让输入和输出特征的相互生成来防止信息丢失,符合融合图像中保留高频特征的目标。

[0042] S305、在第一Transformer Layer中,红外图像和可见光图像的特征分别经过补丁嵌入操作后,通过层归一化操作对该特征进行标准化,再将其中输入到W-MSA模块,按照设定好的窗口大小对图像进行分块,得到自注意力图,再经过局部-非局部注意力信息增强模块,得到局部-非局部注意力信息增强特征,将该增强特征逐元素相加经得到注意力特征 F^1 ,经过层归一化操作操作后再经过FFN模块得到特征,将该特征与注意力特征相加得到第一层HAT结构的输出特征 F^{1+1} ;将输出特征 F^{1+1} 输入第二层HAT结构,经过补丁嵌入操作后,通过层归一化操作对特征进行标准化,经过标准化的特征输入到SW-MSA模块,SW-MSA模块按照设定好的窗口大小对图像进行分块,得到自注意力图,自注意力图再经过局部-非局部注意力信息增强模块,得到局部-非局部注意力信息增强特征,将局部-非局部注意力信息增强特征逐元素相加经得到注意力特征 F^{1+1} ,注意力特征经过层归一化操作后经过FFN模块得到的特征与注意力特征相加得到第一Transformer Layer的红外全局信息特征和可见光全局信息特征,具体公式为:

$$[0043] \quad F^1 = \text{LG}(W\text{-MSA}(\text{LN}(F^{1-1}))) + F^{1-1};$$

$$[0044] \quad F_1^1 = \text{FFN}(\text{LN}(F^1)) + F^1;$$

$$[0045] \quad F^{1+1} = \text{LG}(SW\text{-MSA}(\text{LN}(F_1^1))) + F_1^1;$$

$$[0046] \quad F_1^{1+1} = \text{FFN}(\text{LN}(F^{1+1})) + F^{1+1};$$

[0047] 其中, F^{1-1} 表示红外图像和可见光图像的特征; F^1 表示第一层HAT输出的特征; F^{1+1} 表示红外全局信息特征和可见光全局信息特征;LG(\cdot)表示局部-非局部信息增强操作;LN(\cdot)表示层归一化操作,用于对每层特征归一化;W-MSA(\cdot)表示窗口的多头自注意力模块;SW-MSA(\cdot)表示移位窗口的多头自注意力模块操作,通过两个不同的多头自注意力机制提高窗口之间的信息交互能力,从而提升模型泛化性能。

[0048] S306、第二、三、四Transformer Layer重复步骤S305的内容,其中,第三Transformer Layer需重复三次后再输入第四Transformer Layer后输出红外与可见光的全局信息特征;

[0049] S307、将双编码器每个特征提取块与红外、可见光全局信息特征和红外、可见光局部细节特征对应模态逐元素相加得到四组预融合的红外特征和可见光特征。

[0050] 进一步的,步骤S305中,局部-非局部注意力信息增强模块提取特征的具体内容为:

[0051] 输入特征分别经过全局平均池化、深度可分离卷积、Sigmoid操作得到局部信息增强图,再与原输入特征经过逐元素相乘操作得到局部信息增强特征,其次,输入特征经过层归一化操作、位置注意力模块得到非局部信息增强特征,再与局部信息增强特征逐元素相加得到局部-非局部注意力信息增强特征,该模块通过将局部信息归纳偏置引入补充学习非局部信息特征,提升模型抗干扰能力同时增强细节信息提取能力。

[0052] 进一步的,步骤S4中,利用解码器得到融合图像包括以下子步骤:

[0053] S401、将第四组预融合的红外特征和可见光特征拼接后,经过上采样操作后与第三组预融合的红外特征和可见光特征在通道维度上拼接得到第三组重建特征,将第三组重建特征经过上采样操作后与第二组预融合的红外特征和可见光特征在通道维度上拼接得到第二组重建特征,将第二组重建特征经过上采样操作后与第三组预融合的红外特征和可见光特征在通道维度上拼接得到重建特征。

[0054] S402、将步骤S401中的重建特征经过上采样操作、填充操作、 4×4 卷积操作、Tanh操作后最终得到重建的融合图像。

[0055] 本发明采用以上技术方案,与现有技术相比,其显著技术效果如下:

[0056] 1、本发明克服了基于卷积神经网络的图像融合方法和基于生成对抗网络的图像融合方法全局信息提取能力的局限性。

[0057] 2、本发明在U-net框架下,将Transformer嵌入到CNN中,形成并行的双编码器结构,克服了CNN在全局建模方面的不足,有效地将局部信息与全局信息相结合,提升融合图像质量,同时降低单一Transformer结构带来的计算复杂度。

[0058] 3、本发明在局部细节特征提取路径中引入可逆神经网络,通过可逆设计让输入和输出特征的相互生成来防止信息丢失,进一步提升高频局部细节特征提取能力。

[0059] 4、本发明开发了局部-非局部信息增强模块,通过将局部信息归纳偏置引入补充学习非局部信息特征,提升模型抗干扰能力同时增强细节信息提取能力。

[0060] 5、本发明融合效果相较于现有技术显著提升,模型抗干扰能力和泛化能力较好,可应用于多模态图像融合,对图像融合领域有很高应用价值。

附图说明

[0061] 图1为本发明的整体实施流程图。

[0062] 图2为局部细节特征提取路径中残差模块示意图。

[0063] 图3为局部细节特征提取路径中可逆神经网络示意图。

[0064] 图4为可逆神经网络中可逆残差模块示意图。

[0065] 图5为全局信息提取路径中混合注意力变换器示意图。

[0066] 图6为混合注意力变换器中局部-非局部注意力信息增强模块示意图。

[0067] 图7为解码器中图像重建模块示意图。

[0068] 图8为第一组图像(sandpath)的红外图像、可见光图像和各方法融合效果对比图。

[0069] 图9为第二组图像(Kaptein_1123)的红外图像、可见光图像和各方法融合效果对比图。

[0070] 图10为第三组图像(lake)的红外图像、可见光图像和各方法融合效果对比图。

具体实施方式

[0071] 下面将结合说明书附图和具体实施方式对本发明的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明的一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员所获得的所有其他实施例,都属于本发明保护的范围。

[0072] 本发明提出了一种结合Transformer与CNN双编码器的红外与可见光图像融合方法,如图1所示,包括以下步骤:

[0073] S1、基于U-Net框架,构建结合Transformer与CNN双编码器的红外与可见光图像融合模型,其中基于U-Net框架的融合网络由双编码器和解码器组成,具体内容为:

[0074] S101、构建双编码器:双编码器包括局部细节特征提取编码器和全局信息提取编码器。

[0075] 其中,局部细节特征提取编码器包括依次连接的卷积层和四个特征提取块,卷积层由卷积核为 4×4 、步长为2、填充为1的卷积以及批量归一化、LeakyReLU串行连接构成,卷积层用于下采样操作,并调整输入图像的通道数和尺寸。

[0076] 第二、二特征提取块由残差模块构成,第三、四特征提取块由残差模块和INN(Invertible Neural networks,可逆神经网络)串行连接构成,四个特征提取块输入的通道数分别为16、32、64、128,输出的通道数分别为32、64、128、256。

[0077] 如图2所示,残差模块包括主路径和残差路径,主路径包含卷积核为 1×1 的卷积、BN、LeakyReLU激活函数、卷积核为 3×3 的卷积、卷积核为 4×4 的卷积;残差路径包括DSConv和卷积核为 4×4 的卷积。

[0078] 如图3、4所示,可逆神经网络包括三个可逆残差模块,其包括卷积核为 1×1 的卷积、卷积核为 3×3 的深度可分离卷积、激活函数(ReLU6)。

[0079] 如图5所示,全局信息提取编码器包括四个通过串行方式连接的Transformer Layer,其输入的通道数分别为16、32、64、128,输出的通道数分别为32、64、128、256,第一、二、四Transformer Layer包括两层HAT(Hybrid Attention Transformer,混合注意力变换器)结构,其中,第一Transformer Layer的第一层包括通过串行方式连接的补丁嵌入操作、层归一化操作、W-MSA模块(Windows Multi-Head Self-Attention,窗口多头自注意力)、局部-非局部注意力信息增强模块、层归一化操作和前馈神经网络,第二层包括通过串行方式连接的层归一化操作、SW-MSA模块(Shift Windows Multi-Head Self-Attention,变换窗口多头自注意力)、局部-非局部注意力信息增强模块、层归一化操作和前馈神经网络。

[0080] 第二、四Transformer Layer的第一层包括通过串行方式连接的补丁合并操作、层归一化操作、W-MSA模块(Windows Multi-Head Self-Attention,窗口多头自注意力)、局部-非局部注意力信息增强模块、层归一化操作和前馈神经网络,第二层包括通过串行方式连接的层归一化操作、SW-MSA模块(Shift Windows Multi-Head Self-Attention,变换窗口多头自注意力)、局部-非局部注意力信息增强模块、层归一化操作和前馈神经网络。

[0081] 第三Transformer Layer包括六层混合注意力变换器结构,由第二Transformer Layer中的两层HAT(Hybrid Attention Transformer,混合注意力变换器)结构堆叠3次构成。

[0082] 其中,如图6所示,局部-非局部注意力信息增强模块包括局部信息增强模块和非局部信息增强模块;其中局部信息增强模块包括拼接操作、全局平均池化、深度可分离卷积、Sigmoid函数、逐元素相乘操作和逐元素相乘操作;非局部信息增强模块包括LayerNorm操作和位置注意力模块。

[0083] S102、构建解码器:

[0084] 解码器包括四个特征融合模块和重建模块;每个特征融合模块包括上采样操作,拼接操作,依次连接的卷积核为 4×4 、步长为2、填充为1的反卷积, BatchNorm操作, 激活函数ReLU。

[0085] 如图7所示,重建模块包括依次连接的上采样操作,卷积核为 4×4 的卷积、填充为1的卷积,填充操作,双曲正切函数函数(Tanh)。

[0086] S103、双编码器通过自上而下方式逐步下采样进行特征提取,解码器通过自下而上的方式将双编码器提取的特征逐步上采样进行融合并重建,构成结合Transformer与CNN双编码器的红外与可见光图像融合模型。

[0087] S2、选定数据集,对其进行处理后得到训练集,利用训练集训练步骤S1中的结合Transformer与CNN双编码器的红外与可见光图像融合模型,具体内容为:

[0088] S201、从TNO数据集中选定32对图像作为数据集,将图像灰度值转为 $[-1, 1]$,利用 128×128 的窗口裁剪图像,步长设置为32,最终获得6184对图像块作为训练集。

[0089] S202、采用结构相似性损失和对比损失监督结合Transformer与CNN双编码器的红外与可见光图像融合模型训练,获得最优网络参数,具体公式为:

$$[0090] \quad L_{\text{total}} = \lambda_1 L_{\text{SSIM}} + \lambda_2 L_{\text{patchNCE}}$$

[0091] 其中, L_{total} 表示损失函数, λ_1 、 λ_2 表示超参数, L_{SSIM} 表示结构相似性损失, L_{patchNCE} 表示对比损失。

[0092] 其中,结构相似性损失能够在亮度、对比度、结构上评价融合图像的质量,其结构相似度表示为:

$$[0093] \quad SSIM(x, y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

[0094] 其中, μ 表示像素均值; σ_{xy} 表示协方差; σ 表示方差; C_1 、 C_2 是为了避免 $\mu_x^2 + \mu_y^2$ 、 $\sigma_x^2 + \sigma_y^2$ 接近于零而设定的极小值,通常取 0.01^2 、 0.03^2 。

[0095] 结构相似性损失表示为:

$$[0096] \quad E(I | W) = \frac{1}{m \times n} \sum_{i=1}^{m \times n} P_i$$

$$[0097] \quad \text{Score}(I_f, I_y, I_x | W) = \begin{cases} SSIM(I_f, I_x | W), & \text{if } E(I_x | W) > E(I_y | W) \\ SSIM(I_f, I_y | W), & \text{if } E(I_x | W) \leq E(I_y | W) \end{cases}$$

$$[0098] \quad L_{\text{SSIM}} = 1 - \frac{1}{N} \sum_{W=1}^N \text{Score}(I_f, I_y, I_x | W)$$

[0099] 其中, I_x 、 I_y 和 I_f 分别表示可见光图像、红外图像和融合图像; W 表示滑动窗口,步幅

为1; P_i 表示像素i的值; m, n 表示滑动窗口大小, 本实施例设置为 16×16 。

[0100] 对比损失表示为:

$$[0101] \quad l(k, k^+, k^-) = -\log \left[\frac{\exp(k \cdot k^+ / \tau)}{\exp(k \cdot k^+ / \tau) + \sum_{n=1}^N \exp(k \cdot k^- / \tau)} \right]$$

$$[0102] \quad l_{\text{patchNCE}}^s = (D^s) = \begin{cases} l(d_{fx}^s, d_x^s, d_x^{S/s}) & \text{if } (d_{fx}^s \cdot d_x^s) > (d_{fy}^s \cdot d_y^s) \\ l(d_{fy}^s, d_y^s, d_y^{S/s}) & \text{if } (d_{fx}^s \cdot d_x^s) \leq (d_{fy}^s \cdot d_y^s) \end{cases}$$

$$[0103] \quad L_{\text{patchNCE}} = E_{z \sim Z} \sum_{s=1}^S l_{\text{patchNCE}}^s (D^s)$$

[0104] 其中, k 表示编码好的特征样本; k^+ 表示与 k 相似的正样本; k^- 表示与 k 不相似的负样本; τ 表示温度系数, 通常取 $\tau = 0.07$; S 表示图像特征层中采样的位置数, 其中 $s \in \{1, 2, \dots, S\}$; D^s 表示通道空间上任一特征序列。

[0105] 通过计算空间上某一区域与正负样本的相似度, 有效地保留融合图像和源图像最相似的部分。

[0106] S203、根据训练集和损失函数 L_{total} 对初始的结合Transformer与CNN双编码器的红外与可见光图像融合模型进行训练, 训练过程使用的Adam优化器更新网络模型参数, 学习率设置为0.01, Epoch设置为30, Batchsize设置为24, 得到训练完成的结合Transformer与CNN双编码器的红外与可见光图像融合模型。

[0107] S3、将红外图像、可见光图像输入到训练完成的结合Transformer与CNN双编码器的红外与可见光图像融合模型中, 利用双编码器提取红外图像和可见光图像预融合的红外特征和可见光特征, 并通过自上而下的方式进行特征表示, 具体内容为:

[0108] S301、将红外图像和可见光图像的灰度值转为 $[-1, 1]$, 利用 128×128 的窗口裁剪图像, 步长设置为32。

[0109] S302、将步骤S301中处理后的红外图像、可见光图像输入到双编码器, 经过卷积层的下采样操作, 调整该图像的通道数和尺寸。

[0110] S303、将步骤S302处理后的图像输入到四个特征提取块中, 依次通过 1×1 的卷积、BN、LeakyReLU激活函数、卷积核为 3×3 的卷积、BN和卷积核为 4×4 的卷积操作后得到主路径中学习的卷积特征, 再经过DSCov和卷积核为 4×4 的卷积卷积操作后得到补充信息特征, 将卷积特征和补充信息特征相加后再经过LeakyReLU后得到信息增强特征。

[0111] S304、将步骤S303得到的信息增强特征输入到可逆神经网络中, 将该特征 $F_{I,k}^S$ 在通道维度上被均分成 $F_{I,k}^S[1:c]$ 和 $F_{I,k}^S[c+1:C]$ 两部分, 特征 $F_{I,k}^S[1:c]$ 经过第一个可逆残差模块得到的特征 $R_1(F_{I,k}^S[1:c])$ 与特征 $F_{I,k}^S[c+1:C]$ 逐元素相加得到特征 $F_{I,k+1}^S[c+1:C]$, 特征 $F_{I,k}^S[c+1:C]$ 经过第二个可逆残差模块与特征 $F_{I,k}^S[1:c]$ 逐元素相乘得到特征 $F_{I,k}^S[1:c] \otimes \exp(R_2(F_{I,k+1}^S[c+1:C]))$, 特征 $F_{I,k}^S[c+1:C]$ 经过第三个可逆残差模块得到特征 $(R_3(F_{I,k+1}^S[c+1:C]))$ 并与特征

$F_{I,k}^S[1:c] \otimes \exp(R_2(F_{I,k+1}^S[c+1:C]))$ 逐元素相加得到特征 $F_{I,k+1}^S[1:c]$, 将特征 $F_{I,k+1}^S[c+1:C]$ 和特征 $F_{I,k+1}^S[1:c]$ 在通道维度上拼接得到输出特征 $F_{I,k+1}^S$, 具体公式为:

$$[0112] \quad F_{I,k+1}^S[c+1:C] = F_{I,k}^S[c+1:C] + R_1(F_{I,k}^S[1:c]);$$

$$[0113] \quad F_{I,k+1}^S[1:c] = F_{I,k}^S[1:c] \otimes \exp(R_2(F_{I,k+1}^S[c+1:C])) + (R_3(F_{I,k+1}^S[c+1:C]));$$

$$[0114] \quad F_{I,k+1}^S = \text{cat}\{F_{I,k+1}^S[1:c], F_{I,k+1}^S[c+1:C]\};$$

[0115] 其中, C表示完整通道数, c表示完整通道数的一半, [1:c]表示当前特征通道数为1至c, [c+1:C]表示当前特征通道数为c+1至C, S表示CNN特征提取编码器, I表示任意映射符, k表示第k个可逆网络层数, R_n ($n=1, 2, 3$)表示可逆残差模块, exp表示指数e。

[0116] 最终输出红外局部细节特征和可见光局部细节特征。

[0117] S305、在第一Transformer Layer中, 红外图像和可见光图像的特征分别经过补丁嵌入操作后, 通过层归一化操作对该特征进行标准化, 再将其中输入到W-MSA模块, 按照设定好的窗口大小对图像进行分块, 得到自注意力图, 再经过局部-非局部注意力信息增强模块, 得到局部-非局部注意力信息增强特征, 将该增强特征逐元素相加经得到注意力特征 F^1 , 经过层归一化操作操作后再经过FFN模块得到特征, 将该特征与注意力特征相加得到第一层HAT结构的输出特征 F^{1+1} ; 将输出特征 F^{1+1} 输入第二层HAT结构, 经过补丁嵌入操作后, 通过层归一化操作对特征进行标准化, 经过标准化的特征输入到SW-MSA模块, SW-MSA模块按照设定好的窗口大小对图像进行分块, 得到自注意力图, 自注意力图再经过局部-非局部注意力信息增强模块, 得到局部-非局部注意力信息增强特征, 将局部-非局部注意力信息增强特征逐元素相加经得到注意力特征 F^{1+1} , 注意力特征经过层归一化操作后经过FFN模块得到的特征与注意力特征相加得到第一Transformer Layer的红外全局信息特征和可见光全局信息特征, 具体公式为:

$$[0118] \quad F^1 = \text{LG}(W\text{-MSA}(\text{LN}(F^{1-1}))) + F^{1-1};$$

$$[0119] \quad F_1^1 = \text{FFN}(\text{LN}(F^1)) + F^1;$$

$$[0120] \quad F^{1+1} = \text{LG}(SW\text{-MSA}(\text{LN}(F_1^1))) + F_1^1;$$

$$[0121] \quad F_1^{1+1} = \text{FFN}(\text{LN}(F^{1+1})) + F^{1+1};$$

[0122] 其中, F^{1-1} 表示红外图像和可见光图像的特征; F^1 表示第一层HAT输出的特征; F^{1+1} 表示红外全局信息特征和可见光全局信息特征; $\text{LG}(\cdot)$ 表示局部-非局部信息增强操作; $\text{LN}(\cdot)$ 表示层归一化操作, 用于对每层特征归一化; $W\text{-MSA}(\cdot)$ 表示窗口的多头自注意力模块; $SW\text{-MSA}(\cdot)$ 表示移位窗口的多头自注意力模块操作, 通过两个不同的多头自注意力机制提高窗口之间的信息交互能力, 从而提升模型泛化性能。

[0123] 其中, 局部-非局部注意力信息增强模块提取特征的具体内容为: 输入特征分别经过全局平均池化、深度可分离卷积、Sigmoid操作得到局部信息增强图, 再与原输入特征经过逐元素相乘操作得到局部信息增强特征, 其次, 输入特征经过层归一化操作、位置注意力模块得到非局部信息增强特征, 再与局部信息增强特征逐元素相加得到局部-非局部注意力信息增强特征, 该模块通过将局部信息归纳偏置引入补充学习非局部信息特征, 提升模型抗干扰能力同时增强细节信息提取能力。

[0124] S306、第二、三、四Transformer Layer重复步骤S305的内容,其中,第三Transformer Layer需重复三次后再输入第四Transformer Layer后输出红外与可见光的全局信息特征。

[0125] S307、将双编码器每个特征提取块与红外、可见光全局信息特征和红外、可见光局部细节特征对应模态逐元素相加得到四组预融合的红外特征和可见光特征。

[0126] S4、利用解码器将步骤S3中的特征表示映射到原始分辨率,通过自下而上的方式逐步融合图像特征,得到融合图像,具体内容为:

[0127] S401、如图7所示,将第四组预融合的红外特征和可见光特征拼接后,经过上采样操作后与第三组预融合的红外特征和可见光特征在通道维度上拼接得到第三组重建特征,将第三组重建特征经过上采样操作后与第二组预融合的红外特征和可见光特征在通道维度上拼接得到第二组重建特征,将第二组重建特征经过上采样操作后与第三组预融合的红外特征和可见光特征在通道维度上拼接得到重建特征。

[0128] S402、将步骤S401中的重建特征经过上采样操作、填充操作、 4×4 卷积操作、Tanh操作后最终得到重建的融合图像。

[0129] 进一步地,为验证通过上述方法训练得到的结合Transformer与CNN双编码器的红外与可见光图像融合模型的图像融合效果,本实施例还对训练好的结合Transformer与CNN双编码器的红外与可见光图像融合模型进行验证。

[0130] 具体地,在测试阶段,从TNO数据集中选取了除训练集外的20组图像进行测试验证,对比方法选择7种典型方法,包括DenseFuse、U2Fusion、RFN-Nest、SEDRFuse、IFCNN、GANMcC、CSF。此外,定量评价指标采用信息熵(EN)、平均梯度(AG)、空间频率(SF)、互信息(MI)、标准差(SD)、视觉保真度(VIF)等6个指标,验证结果包括定性评价和定量评价两个方面。

[0131] 定性评价:图8、9、10给出了三组典型图像(sandpath,Kaptein_1123和lake)的定性比较结果图。通过对比,可以发现本发明的融合方法具有三个优势。首先,对于典型的红外目标,如图8、9中的人,本发明的融合结果比其他方法具更高亮度热辐射信息,更清晰的边缘轮廓。其次,融合结果可以保留可见光图像中的纹理细节和背景信息。例如图8的树和背景的天空亮度,图10的长椅,本发明的融合方法相比其他方法能够保留更清晰细节信息和更明显的背景信息。最后,本发明既能出色地捕捉接近红外图像的显著目标,同时最大限度地保留地可见光图像中的纹理细节和背景特征,融合图像更加清晰、自然。与源图像和其他融合结果相比,本发明方法可以更好地保留突出的目标特征和丰富的场景细节信息,本发明的融合图像方法细粒度信息提取更加充分,图像对比度高,更符合人眼视觉效果。

[0132] 定量评价:表1给出了TNO数据集的20组图像的客观比较结果。最优平均值和次优平均值分别用加粗和下划线标注。可以看出,本发明取得了EN、MI、SD、VIF、SF的最优平均值,指标AG为次优值,客观实验表明本发明方法比其他方法具有更好的融合性能。本发明所述的融合方法EN、MI取得做最优,这表明本发明融合图像高效地从源图像获取丰富的源图像信息,并有效地重建于融合图像,最大值SD表明本发明融合图像像素更分散,这表明本发明构造的混合变换器能够从局部和非局部两方面学习图像特征,提取图像细节信息,有效提高模型抗干扰能力。最大值VIF表明本发明的融合图像具有更好的视觉效果和更真实的人眼感受,这是由于本发明的融合方法采用Transformer嵌入到CNN中形成双编码器结构,

能够提取全局信息和局部信息,能够自适应融合不同模态特征。本发明所述的融合图像方法的SF也是最优,这说明本发明构造的可逆神经网络能够从不同模态的图像中提取高频语义信息,一定程度上防止输入和输出特征的相互生成来防止信息丢失。

[0133] 表1TN0数据集的20组图像的定量比较结果

[0134]	EN	AG	SF	MI	SD	VIF	
	DenseFuse	6.2880	2.3532	5.7371	1.8723	23.9162	0.4496
	U2Fusion	6.9814	5.2088	11.7202	1.7347	36.1467	<u>0.5437</u>
	RFN-Nest	7.0148	2.9111	6.2722	1.9224	38.8036	0.5177
	SEDRFuse	<u>7.1209</u>	4.0203	9.4567	2.0598	<u>41.4590</u>	0.5065
[0135]	IFCNN	6.7063	4.4162	10.9383	2.2720	32.3123	0.5247
	GANMcC	6.6821	2.4555	<u>12.2176</u>	<u>2.6669</u>	32.2925	0.4325
	CSF	6.9303	3.7744	8.7414	1.8938	37.64638	0.5259
	Ours	7.3182	<u>5.1952</u>	12.4650	3.0395	49.2826	0.6245

[0136] 以上所述仅为本发明的示例性实施例,并非因此限制本发明专利保护范围,凡是利用本发明说明书及附图内容所作的等效结构或等效流程变换,或直接或间接运用在其他相关的技术领域,均同理包括在本发明的专利保护范围内。

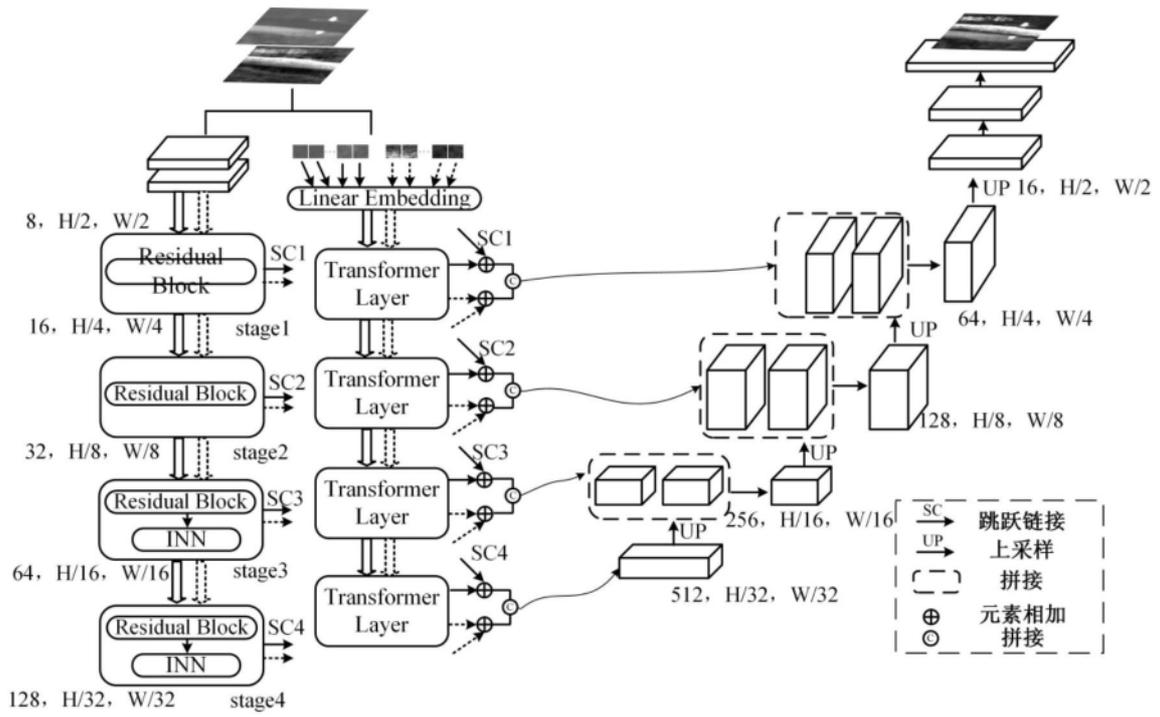


图1

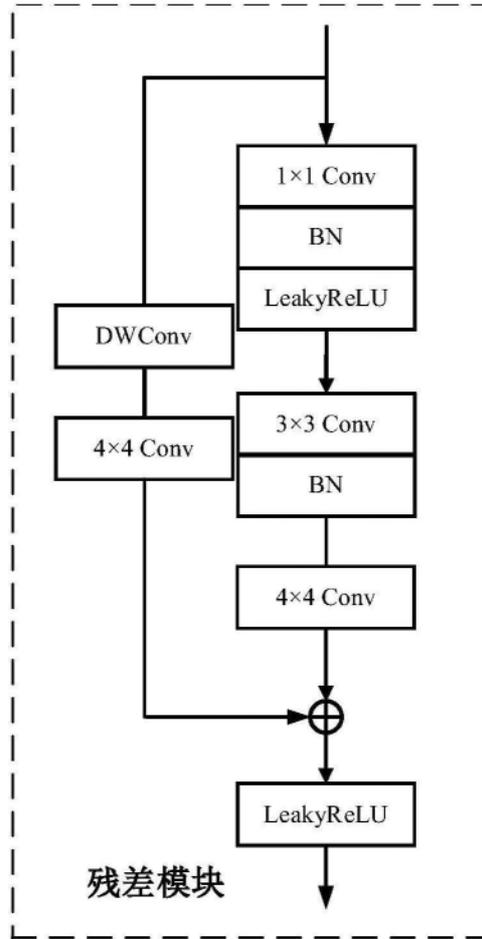


图2

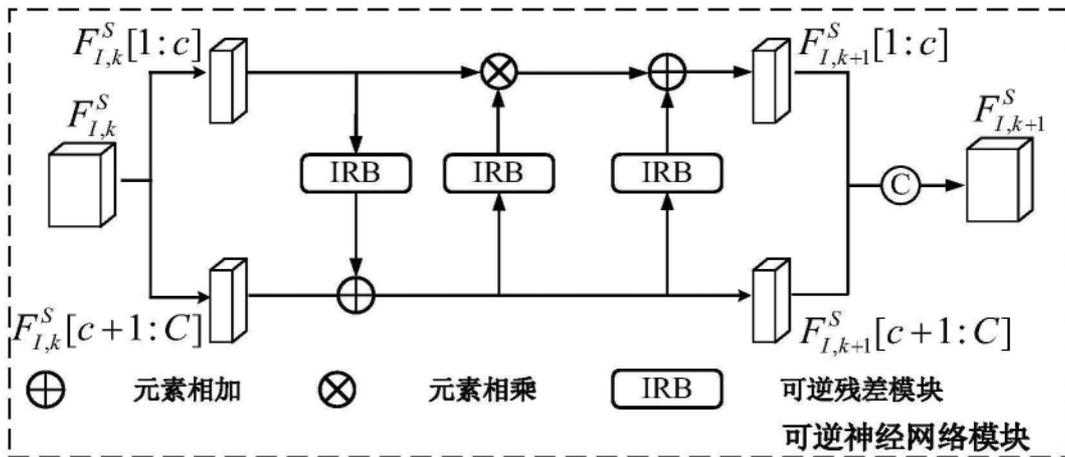


图3

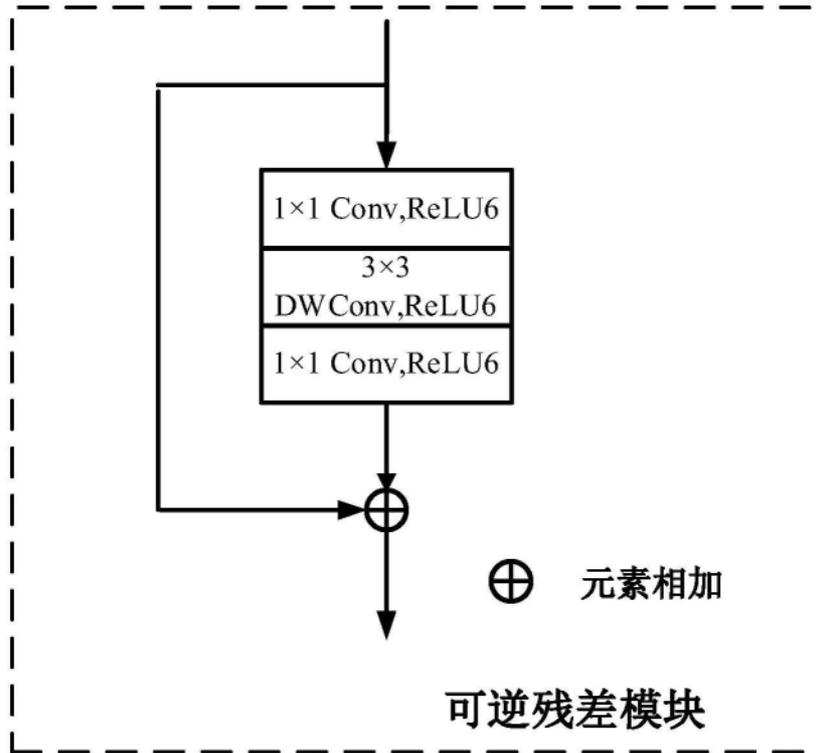


图4

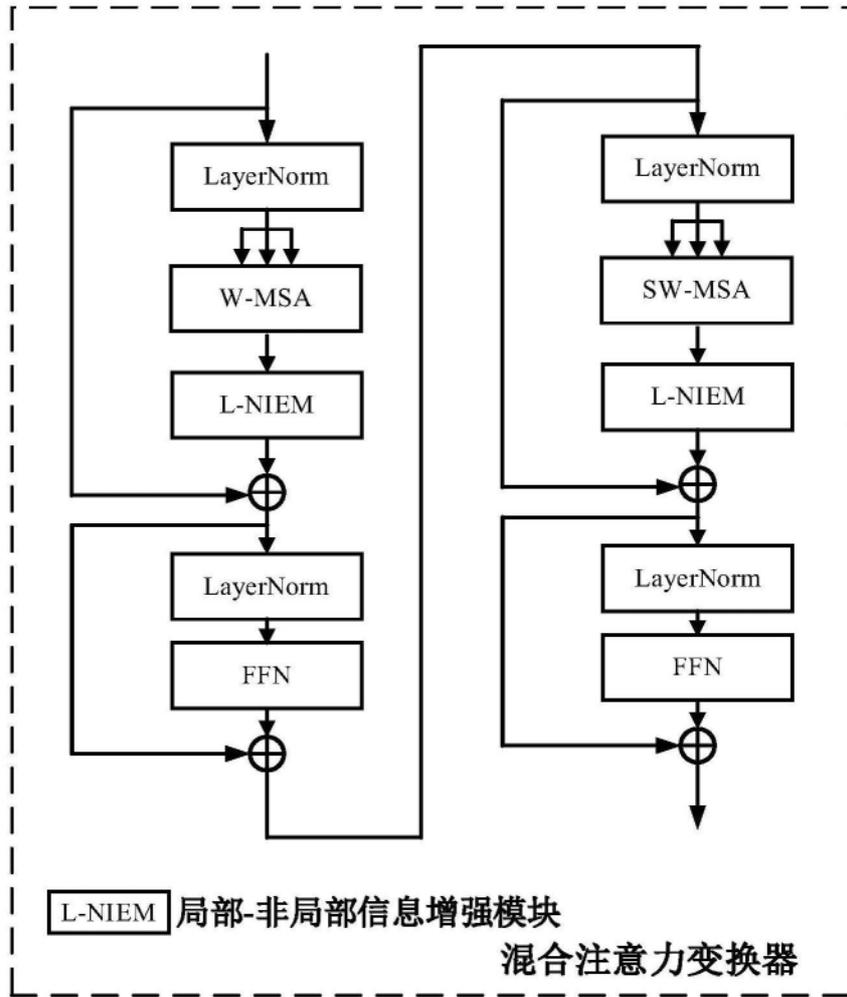


图5

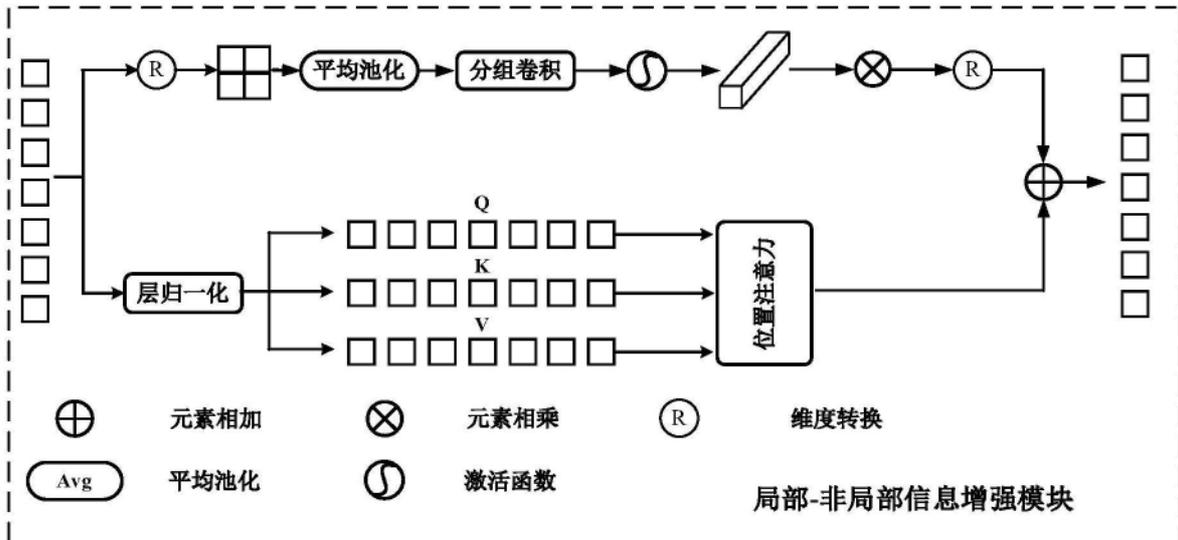


图6

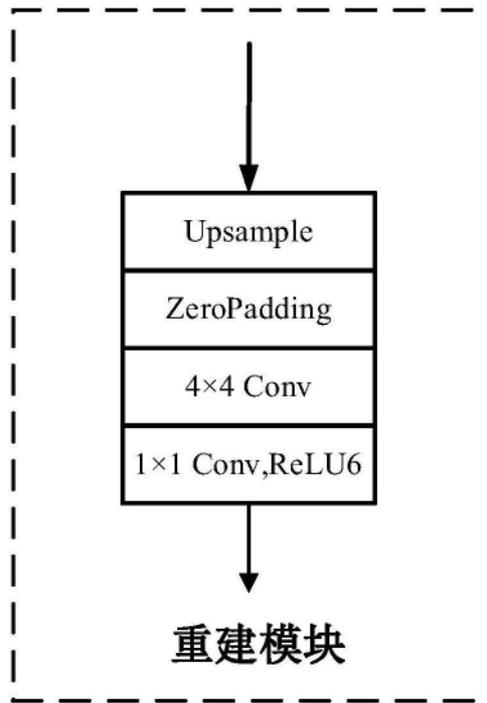


图7

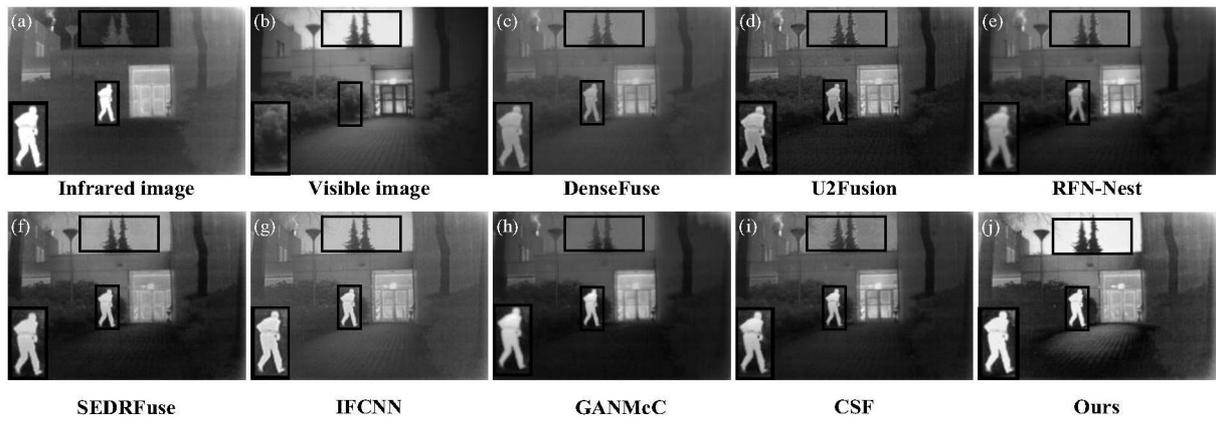


图8

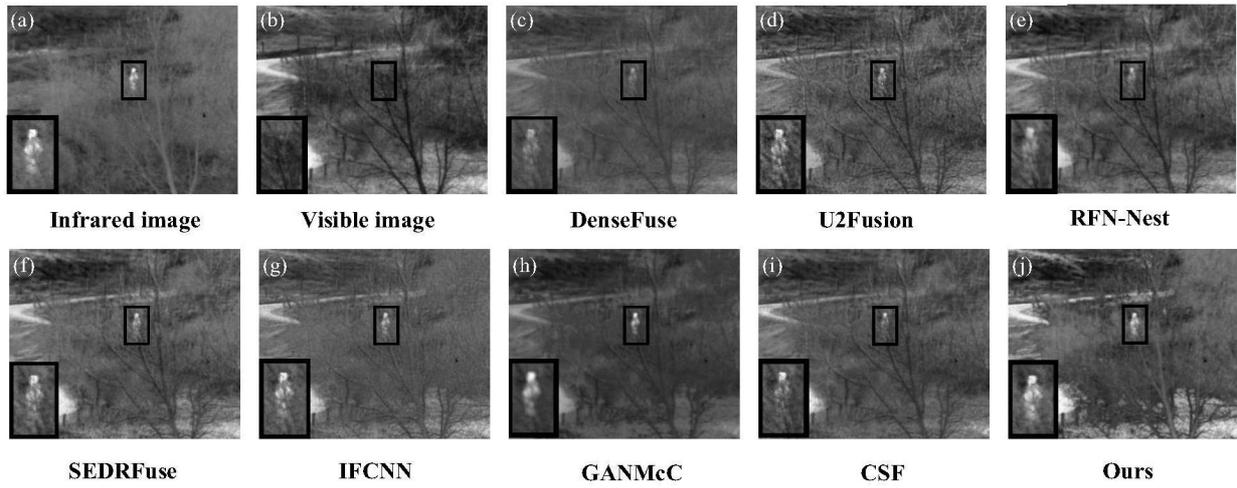


图9

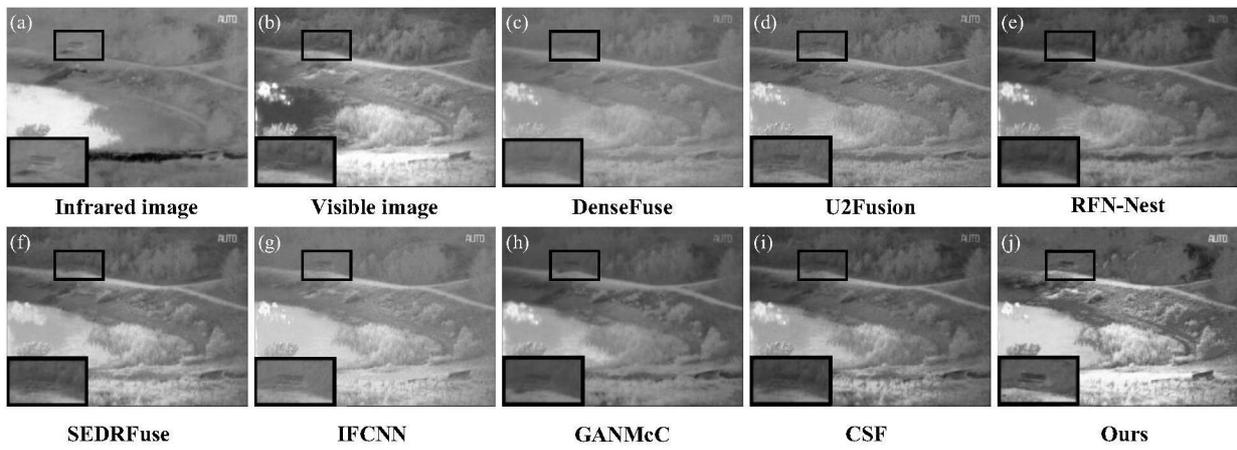


图10