第 58 卷 第 14 期/2021 年 7 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于双通路生成对抗网络的红外与可见光图像融合方法

侯春萍[1]，王霄聪[1]，夏晗[2]，杨阳[1*]

[1] 天津大学电气自动化与信息工程学院，天津 300072；

[2] 天津大学国际工程师学院，天津 300072

摘要 针对已有红外与可见光图像融合方法没有充分考虑不同模态间及相同模态内的信息差异，融合图像存在细节纹理信息丢失、对比度低等问题，提出了一种基于双通路生成对抗网络的红外与可见光图像融合方法。通过对抗方式训练生成网络和鉴别网络，并将训练的生成网络作为最终的图像融合模型。在融合模型中采用双通路分别对红外与可见光图像进行特征提取，以保留更多的跨模态信息。此外，为了加强模态内特征像素的全局依赖关系，引入自注意力机制增强输入特征，提高特征的细节丰富度。在 TNO 公开数据集上的实验结果表明，相比现有图像融合方法，本方法得到的融合图像对比度更高，细节纹理更丰富；且能良好地契合人类的视觉感知，在各类评估指标上均能达到较高水平。

关键词 图像处理；生成对抗网络；自注意力机制；双通路；红外图像；可见光图像

中图分类号 TP391.4 **文献标志码** A **doi**：10.3788/LOP202158.1410024

Infrared and Visible Image Fusion Method Based on Dual-Channel Generative Adversarial Network

Hou Chunping[1], Wang Xiaocong[1], Xia Han[2], Yang Yang[1*]

[1] *School of Electrical and Information Engineering, Tianjin University, Tianjin* 300072, *China*;

[2] *Tianjin International Engineering Institute, Tianjin University, Tianjin* 300072, *China*

Abstract Considering that the existing infrared and visible image fusion methods do not entirely consider the information differences between different modalities and within the same modalities and the fusion image exhibit problems such as loss of detailed texture information and low contrast, this paper proposes an infrared and visible image fusion method based on a dual-channel generative adversarial network. The generation and identification networks are trained through confrontation, and the trained generation network is used as the final image fusion model. The fusion model uses dual-channel to extract features from infrared and visible images to retain more cross-modal information. Furthermore, a self-attention mechanism is introduced to enhance the input features and improve feature-richness in the modal to strengthen the global dependence of feature pixels. The experimental results on the TNO public dataset show that compared with the existing image fusion methods, the fused image obtained by the method has a higher contrast and richer detailed texture than the image obtained using existing image fusion methods. The fused image can efficiently fit human visual perception and performs well across a range of evaluation metrics.

Key words image processing; generative adversarial network; self-attention mechanism; dual-channel; infrared image; visible image

OCIS codes 100.2000; 100.3010

收稿日期：2020-08-17；**修回日期**：2020-09-10；**录用日期**：2020-09-30

基金项目：重点国际(地区)合作研究(61520106002)、南方电网公司科技项目(ZBKJXM20170086)

通信作者：*yang_yang@tju.edu.cn

1 引 言

红外与可见光图像融合是图像融合领域的重要研究内容之一。红外图像能够捕获目标场景中的热辐射信息，而可见光图像的分辨率较高，包含丰富的场景结构和细节纹理信息[1]，将红外图像与可见光图像融合可以最大程度上保留二者的有效信息，实现信息互补[2-5]。根据实现过程可将红外与可见光图像融合方法分为传统方法和基于深度学习的方法[6]。其中，传统方法基于人工设计的融合规则，通过分解和变换方式获得融合图像[7]。而基于深度学习的融合方法能自适应地对输入图像进行融合，具有重要的研究价值。

传统的红外与可见光图像融合方法主要有基于小波变换的 Wavelet 方法[8]、基于多尺度几何变换的 CVT 方法[9]、基于多尺度奇异值分解的 MSVD 方法[10]等。这些方法将图像进行分解变换后，根据图像特性，按照一定的融合规则（如简单加权法、基于区域的最大能量法）进行融合。Li 等[11]在传统方法的基础上将手工特征提取改为用预训练好的 VGG（Visual geometry group）神经网络提取特征，然后根据红外与可见光图像同一位置点的像素值大小进行权重分配，完成融合，但该方法仍然没有解决传统方法繁琐的融合规则设计。深度学习在图像融合领域中可实现端到端的图像融合，其网络模型的输入是红外与可见光图像对，输出为融合图像。Ma 等[12]提出了一种基于生成对抗网络（GAN）的融合模型 FusionGAN，将图像的融合问题表示为生成模型的建模问题，完成了红外与可见光图像的融合任务。Li 等[13]提出了一种基于自编码器的深度融合方法 DenseFuse，其使用的 Dense Block 结构对多层特征重复利用，丰富了融合图像的特征。Zhang 等[14]将红外与可见光图像的融合问题表述为原始图像纹理和强度比例的保持问题，保证了融合图像中的信息保留效果。唐超影等[15]提出了一种低照度可见光与近红外图像的融合方法，同时进行降噪与融合处理，在实际工程中有重要的应用价值。目前基于深度学习的红外与可见光图像融合方法中，不同模态的图像数据均通过统一的神经网络结构进行特征编码，多模态图像共用一组网络权重，忽略了不同模态数据的统计特性差异[16]和训练网络参数时模态信息相互干扰的风险[17]。此外，利用卷积神经网络（CNN）提取特征时，得到的特征信息流以相同的权重向后流动，但对于多模态图像，全局信息并非同等重要。如果能从原始图像中获得一些特征内像素位置的长距离依赖先验知识，就能根据这些先验知识抑制无用信息，增强关键信息的表达，使最终的融合图像更利于被理解和描述。但现有网络结构受卷积层数的限制，获取的感受野难以对特征位置上的长距离依赖关系进行描述，特征对原始图像的表达欠缺，从而在视觉表达方面对融合结果产生一定的影响。

朱东涛[18]将双通路结构用于交通标志的识别中，在通路 1 上获取交通标志的全局特征，在通路 2 上获取交通标志核心区域的特征，提高了网络对交通标志识别的精度。Kamnitsas 等[19]设计了双通路网络，分别实现了局部细节特征和全局特征的提取，在脑损伤图像分割任务上的表现较好。针对上述问题，本文提出了一种基于双通路 GAN 的红外与可见光图像融合方法。首先，针对红外与可见光图像跨模态间的差异性，用设计的双通路结构在输入端分别提取红外和可见光图像的特征，使融合图像尽可能地保证原始图像的信息。其次，为了弥补卷积层数对特征位置上长距离依赖关系的限制，在两支信息流中提取关键信息进行重点表达。最后，将自注意力机制引入 GAN 中，增强了不同模态图像像素与全局依赖关系的表达，并将像素依赖关系作为权重对输入特征进行处理，以增强关键特征的表达，提高特征的细节丰富度，进而提升融合图像的视觉效果。实验结果表明，本方法能有效解决红外与可见光图像融合中模态信息相互干扰导致的图像跨模态信息缺失和视觉表达不足导致的图像对比度低问题。

2 红外与可见光图像的融合方法

2.1 总体融合网络框架

本方法中的网络总框架如图 1 所示，该网络模型主要分为用于生成融合图像的生成网络和对融合图像进行判别的鉴别网络。生成网络和鉴别网络经过联合对抗训练，最终得到的 GAN 即为融合模型。该模型具有端到端的网络结构，其输入为红外（$\boldsymbol{I}_{\mathrm{R}}$）与可见光图像（$\boldsymbol{I}_{\mathrm{V}}$）对，输出为生成的融合图像（$\boldsymbol{I}_{\mathrm{F}}$）。双通路指网络有两个分支，生成网络采用双通路网络结构分别将红外与可见光图像输入两个信息流分支中，两个分支的权重不共享，可实现跨模态信息的特征提取。此外，生成网络还通过引入自注意力机制计算了特征内部的像素依赖关系，并以此为权重加强了对融合任务贡献大的特征，从而提升

生成的融合图像质量。鉴别器由全卷积层组成，其作用是对真实可见光图像和生成的融合图像进行判别区分，进而驱动生成网络生成与真实可见光图像细节纹理信息一致的融合图像。

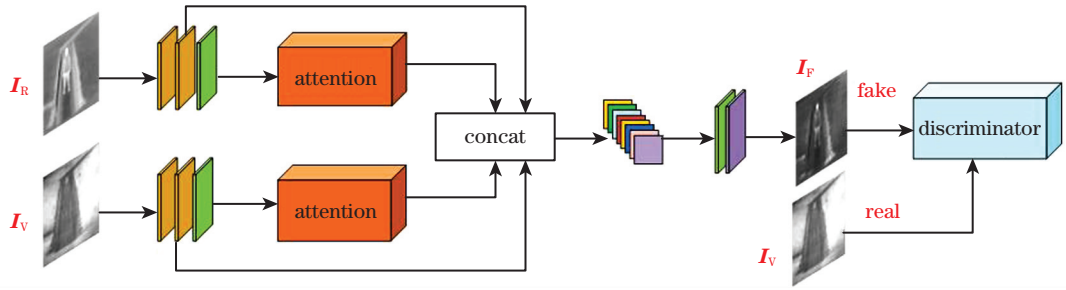

图 1　本方法的网络架构

Fig. 1　Network architecture of our method

2.2　基于双通路的生成器结构

生成器由特征提取模块、像素注意力模块、特征融合模块、输出模块组成，其网络结构如图 2 所示。其输入为红外与可见光图像对，经过 4 个模块输出得到融合图像。由于红外与可见光图像的特征存在较大差异，采用统一网络同时提取多模态图像特征时，多模态信息相互干扰，跨模态信息得不到很好的表达，从而造成信息损失。因此，在本网络中将特征提取模块分为两个通路，如图 2 中的虚线框所示。将红外与可见光图像分别作为两个通路的输入，特征提取模块的上下通路均由 3 个卷积(Conv)层、批归一化(BN)层和带泄露线性整流函数(LReLU)层组成，3 个卷积核的尺寸分别为 5×5，5×5，3×3，卷积核的个数 n 分别为 64、32、16，输出为 2 个原始图像的特征图。像素注意力模块利用自注意力机制获取原始图像特征像素与全局图像的依赖关系，在特征中映射出原始图像的显著像素并赋予其更多的权重，进而输出更能表达原始图像的特征增强图。特征融合模块可实现多模态输入图像的互补特征融合，将用通道连接后的浅层特征和增强后的特征作为特征融合模块的输入，以实现中间特征的有效利用，减少信息损失。特征融合模块由一个 3×3 卷积层、BN 层和 LReLU 层组成。输出模块的作用是输出生成的融合图像，由 1×1 卷积层、BN 层和 Tanh 激活层组成。

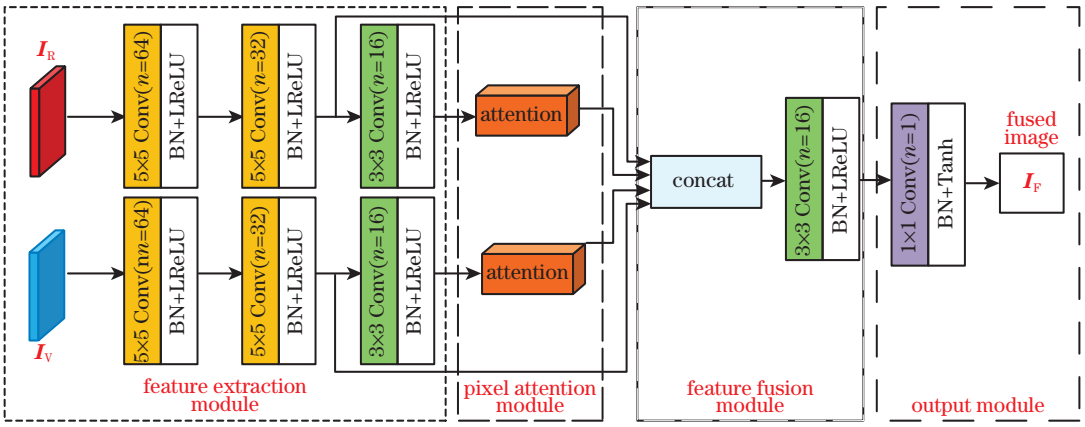

图 2　基于双通路的 GAN 结构

Fig. 2　GAN structure based on dual-channel

在整个生成器结构中，特征提取模块与特征融合模块都在每层卷积层后使用了 BN 层和 LReLU 层，使模型的训练更稳定，同时还能解决梯度消失问题。为了生成与输入图像尺寸相同的融合图像，对所有卷积层都进行了填充操作，卷积步长均设置为 1。

注意力机制模拟了人类视觉感知的情形，核心目标是从众多信息中选取对当前任务目标更关键的信息，并抑制其他无用信息。自注意力机制是注意力机制的特殊形式，其减少了对外部信息的依赖，更擅长捕捉数据或特征的内部相关性。本方法基于自注意力机制的像素注意力模块，利用自注意力机制捕捉多模态图像特征像素与全局信息的依赖关系，并根据特征依赖程度对特征内不同像素位置的信息分配权重，从而增强特征的表达能力。

本方法中的像素注意模块如图 3 所示，其输入

特征图为 $\boldsymbol{x}$，H、W、C 分别为输入特征图的高、宽以及通道数。将特征图 $\boldsymbol{x}$ 在两条通路上分别经过卷积核尺寸为 1×1、卷积核数目分别为 $C/8$ 和 C 的卷积层、BN 层后得到的输出分别为 $f(\boldsymbol{x})$、$g(\boldsymbol{x})$。其中，$f(\boldsymbol{x})$通过卷积操作实现降维的目的，减少了后续计算的复杂度。为了得到与输入特征图尺寸相同的自适应注意力图，将 $g(\boldsymbol{x})$设置为与输入特征图相同的尺寸，并用 BN 层达到稳定模型的作用。将 $f(\boldsymbol{x})$与其转置相乘，计算特征图中任意两个像素点之间的相关性，以捕捉特征图之间的长距离依赖关系，得到输入特征的自注意力图 $\boldsymbol{I}_{\mathrm{Am}}$。将表示像素与全局依赖关系的自注意力图作为权重，通过矩阵相乘分配给输入特征 $f(\boldsymbol{x})$，实现自适应特征细化，得到自适应注意力图 $\boldsymbol{I}_{\mathrm{AAm}}$。赋予自适应注意力图一定的权重，再与输入特征图在通道上进行相加操作，得到特征增强图 $\boldsymbol{I}_{\mathrm{Fem}}$。值得注意的是，该权重不是一个固定值，而是通过生成器损失反向传播训练的自适应权重，使自注意力机制能更好地应用于红外与可见光图像的融合任务中。$\boldsymbol{I}_{\mathrm{Fem}}$ 可表示为

$$\boldsymbol{I}_{\mathrm{Fem}}=\omega\boldsymbol{I}_{\mathrm{AAm}}+\boldsymbol{I}_{\mathrm{Fm}},\tag{1}$$

式中，$\boldsymbol{I}_{\mathrm{AAm}}=\boldsymbol{I}_{\mathrm{Am}}g(\boldsymbol{x})$，$\boldsymbol{I}_{\mathrm{Am}}=f(\boldsymbol{x})f(\boldsymbol{x})^{\mathrm{T}}$，$\omega$ 为训练得到的 $\boldsymbol{I}_{\mathrm{AAm}}$ 自适应权重。

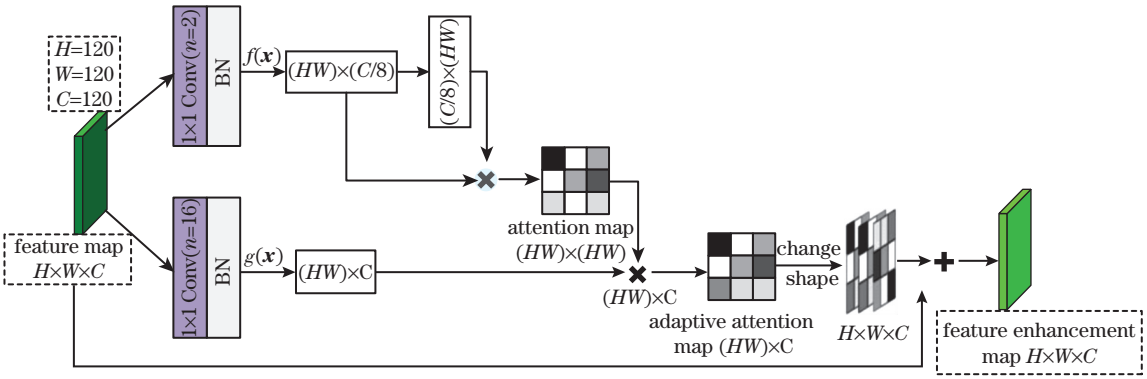

图 3　像素注意力模块的结构

Fig. 3　Structure of the pixel attention module

2.3　鉴别器的结构

由于缺乏真实的融合图像标签，生成器生成的图像在生成器损失函数的约束下，尽可能地保留红外图像的亮度信息和可见光图像的边缘结构信息。为了使融合图像的细节更丰富，提升融合效果，在网络中引入鉴别器，通过对抗训练在生成的融合图像中进一步融入原始图像的纹理信息。人眼对不同空间频率的敏感度不同，在一定的空间频率范围内，空间频率越大，人眼越敏感，获取的信息就越多。可见光图像的纹理信息丰富，空间频率较高，人眼的视觉敏感度较高。而同一局部区域的红外图像分辨率和空间频率较低，人眼的视觉敏感度也较低。这部分热信息在融合过程中，会降低融合图像的对比度并减弱人眼对局部空间频率信息的敏感程度，从而影响生成融合图像的质量。为了在融合图像中保留更多的可见光纹理信息，在鉴别器中将真实可见光图像作为正样本、生成图像作为负样本进行判别，通过生成器与鉴别器的对抗训练，使生成的融合图像中保留更多可见光图像的细节纹理信息，提高局部区域的空间频率。

本方法的鉴别器网络结构如图 4 所示，该网络是一个 5 层卷积层，输入的可见光图像和生成图像的尺寸均为 120 pixel×120 pixel，二者分别表示真样本和假样本。5 层卷积层中，前 4 层卷积层的尺寸均为 3×3，卷积核数量分别为 32、64、128、128，步长 s 为 2，最后一层为 LReLU 层。为了加强鉴别器的鉴别能力，将第 3、第 4 层卷积后得到的特征图进行通道连接，再输送至 LReLU 层进行逻辑约束及判别分类，输出为 $D(\boldsymbol{I}_{\mathrm{V}})$或 $D(\boldsymbol{I}_{\mathrm{F}})$，分别表示鉴别器网络对于输入可见光图像或生成图像的判别结果。

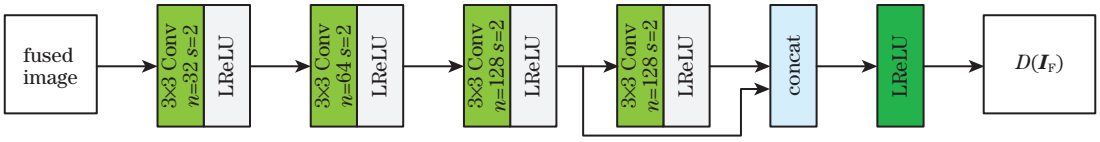

图 4　鉴别器网络的结构

Fig. 4　Structure of the discriminator network

2.4 损失函数

本方法通过文献[12]中设计的损失函数对整体网络进行优化，该损失函数由生成器(G)的内容损失 $L_{\text{content}}(G)$、生成器(G)和鉴别器(D)之间的对抗损失 $L_{\text{adv}}(G,D)$两部分组成。内容损失 $L_{\text{content}}(G)$可表示为

$$L_{\text{content}}(G)=L_{\text{IR}}(G)+L_{\text{VI}}(G)。\tag{2}$$

可以发现，内容损失 $L_{\text{content}}(G)$由红外强度损失 $L_{\text{IR}}(G)$和可见光结构损失 $L_{\text{VI}}(G)$两部分组成。用融合图像与红外图像像素强度分布的相似性约束保证融合图像的热辐射信息，用融合图像与可见光图像边缘结构的相似性约束保证融合图像的整体场景结构信息。红外强度损失 $L_{\text{IR}}(G)$和可见光结构损失 $L_{\text{VI}}(G)$约束融合图像的过程可表示为

$$L_{\text{IR}}(G)=\frac{1}{HW}\left\|\boldsymbol{I}_{\text{F}}-\boldsymbol{I}_{\text{R}}\right\|_2^2,\tag{3}$$

$$L_{\text{VI}}(G)=\frac{1}{HW}\left\|\nabla^2\boldsymbol{I}_{\text{F}}-\nabla^2\boldsymbol{I}_{\text{V}}\right\|_2^2,\tag{4}$$

式中，$\|\ \|_2^2$ 为 L_2 范数，$\nabla^2\boldsymbol{I}_{\text{F}}$、$\nabla^2\boldsymbol{I}_{\text{V}}$ 分别为生成的融合图像与可见光图像经拉普拉斯算子卷积运算后得到的二阶微分值，∇^2为拉普拉斯算子。红外图像的热辐射信息由像素强度表征，(3)式通过最小化融合图像和红外图像欧氏距离的均方根值在像素分布上对融合图像进行约束，使融合图像的像素分布向红外图像逼近，从而捕获红外图像的热辐射信息。(4)式中的拉普拉斯算子为二阶微分算子，能反映图像像素差异的变化情况，可用于表征图像的边缘等结构信息。通过最小化$\nabla^2\boldsymbol{I}_{\text{F}}$ 和$\nabla^2\boldsymbol{I}_{\text{V}}$ 欧氏距离的均方根值，可在图像结构上对融合图像进行约束，使融合图像的结构向可见光图像逼近，从而捕获可见光图像的结构信息。

对抗损失 $L_{\text{adv}}(G,D)$约束了生成器和鉴别器间的联合训练方向，分别将可见光图像与生成图像作为真假样本输入鉴别器，并进行生成器与鉴别器间的博弈，迫使生成图像获取更多的可见光信息，生成更好的融合图像。生成器需尽可能生成欺骗鉴别器的虚假可见光图像，而鉴别器需尽量将生成图像与真实可见光图像区分开来。为了最小化生成器的损失，最大化鉴别器的损失，将 GAN 的对抗损失表示为

$$\min_G\max_D L_{\text{adv}}(G,D)=E_{\boldsymbol{x}\sim P_{\text{data}}(\boldsymbol{x})}\{\lg[D(\boldsymbol{x})]\}+E_{\boldsymbol{z}\sim P_{\text{data}}(\boldsymbol{z})}\{\lg[1-D(\boldsymbol{z})]\},\tag{5}$$

式中，E 为期望，$\boldsymbol{z}$ 为生成器生成的融合图像，$D(\boldsymbol{x})$、$D(\boldsymbol{z})$分别为可见光图像、生成器生成的融合图像经过鉴别器得到的逻辑输出，$P_{\text{data}}(\boldsymbol{x})$和 $P_{\text{data}}(\boldsymbol{z})$分别为可见光图像的真实分布和融合图像的生成分布。

2.5 网络训练过程

本方法进行对抗训练的流程：首先，将红外和可见光图像进行通道连接后输入基于双通路的生成器网络中，生成融合图像；然后，将融合图像和可见光图像输入判别网络进行判别分类，并计算对抗损失 $L_{\text{adv}}(G,D)$以更新鉴别器 D 的参数；最后，通过计算红外图像的像素强度信息、可见光图像的结构信息设计内容损失 $L_{\text{content}}(G)$和对抗损失 $L_{\text{adv}}(G,D)$，进行生成器 G 的参数更新。迭代次数内循环 N 次，进而实现 G 和 D 的对抗训练，得到理想的模型参数。

3 分析与讨论

3.1 实验参数设置

从 TNO 公开数据集[20]中选取 35 对不同场景下的红外与可见光图像作为实验训练数据，将 20 对图像作为测试数据。通过滑窗方式对训练图像对进行裁剪，以扩充数据集。裁剪步长为 32，裁剪图像块的尺寸为 120 pixel×120 pixel。训练参数：批处理大小为 32，初始学习率为 0.0001，用 Adam Optimizer 进行网络模型参数的优化。实验平台：台式工作站的 GPU 为 GTX 1080，内存为 16 GB。

为了验证本方法的有效性，将本方法与多种传统融合方法和先进的深度学习方法的融合结果进行对比，包括 CVT 方法[9]、MSVD 方法[10]、Wavelet 方法[8]、基于卷积稀疏编码的 CSR 方法[3]、基于深度神经网络的 DLF 方法[11]、FusionGAN 方法[12]、DenseFuse 方法[13]、基于潜在低秩表示的 LatLRR 方法[21]、基于图像梯度与强度比例维持的 PMGI[14]方法，结果如图 5 所示。其中，IR 为红外图像，VI 为可见光图像，插图为融合图像中某个细节区域的放大图。可以发现，MSVD、Wavelet、CSR、LatLRR 方法生成的融合图像红外目标不显著，与周围环境的对比度较低；CVT 方法生成的融合图像局部信息存在失真现象，如图 5(c3)中人的面部、图 5(d3)中字母 AUTO 背后的纹理信息；FusionGAN、DenseFuse 方法生成的融合图像在红外热信息和可见光细节信息重叠区域存在纹理细节和红外热辐射信息相互干扰的情况，如图 5(c8)、图 5(c9)中人面部未保留可见光图像中存在的眼睛、鼻子等细节信息，图 5(a8)、图 5(a9)的热目标信息边缘不清晰，被可见光图像中的灯光信息干扰。CSR、DLF、PMGI

和本方法生成的融合图像红外目标突出，具有丰富的可见光图像细节纹理信息。总体来说，CSR、DLF方法背景区域生成的细节纹理对比度低、梯度小，主观视觉效果一般。PMGI与本方法在局部细节区域都可以保证较好的对比度，且本方法生成的融合图像背景细节更丰富自然，在非红外目标区域保留了更多的可见光纹理信息，局部区域的空间频率较高，更符合人类的视觉感知。如在图5(b11)、图5(b12)中，本方法生成的融合图像整体对比度略低于PMGI方法，但在细节纹理方面的效果更好。从插图中的放大区域可以看出，在热目标背后的山脉中，本方法融合了更多的可见光信息。同时，在图像的最底部保留了可见光图像信息，而这部分信息在PMGI方法生成的图像中完全消失了。在图5(d12)和图5(e12)中，本方法生成的融合图像树木纹理更丰富自然，更倾向于真实的可见光场景。

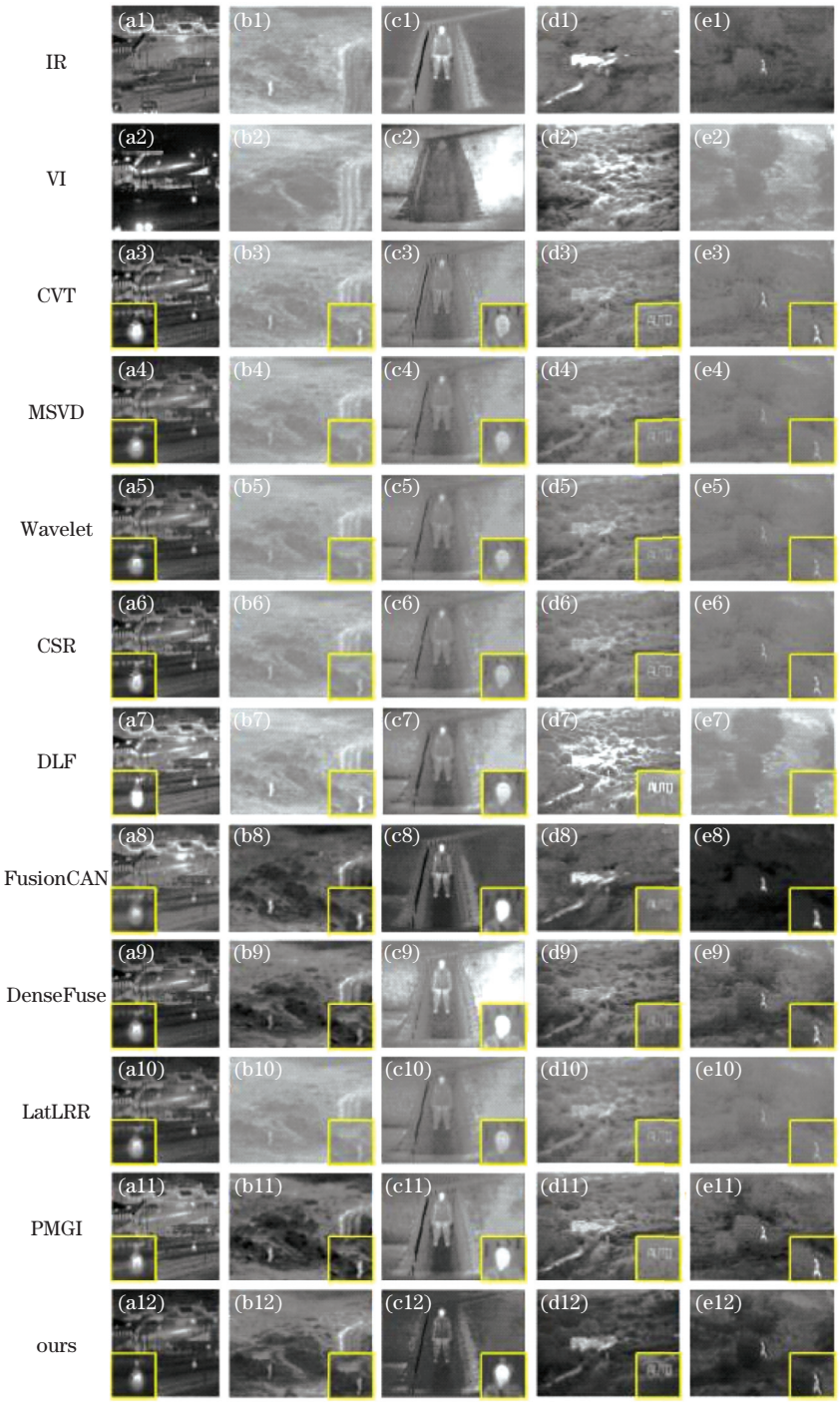

图5　不同方法的融合结果。(a)～(e)图像1～图像5

Fig. 5　Fusion results of different methods. (a)-(e) Image 1-image 5

图 6 为图 5(d)的局部细节放大图，可以发现，本方法在保留红外热信息的同时融合了大量可见光信息，使融合图像的热目标显著且细节纹理与可见光图像相似，图像整体具有人眼视觉更敏感的高空间频率，主观视觉效果较好。

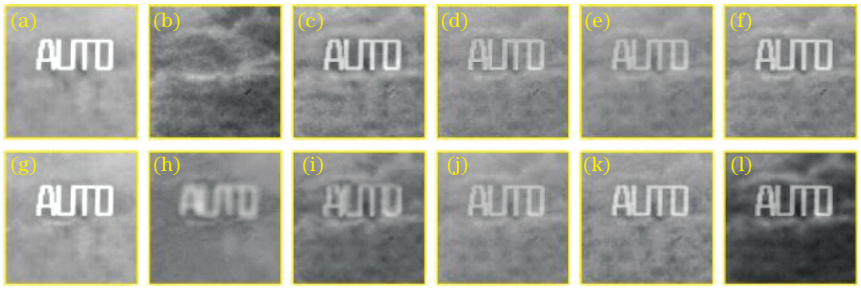

图 6　图 5(d)的局部放大图。(a) IR；(b) VI；(c) CVT；(d) MSVD；(e) Wavelet；(f) CSR；(g) DLF；(h) FusionGAN；(i) DenseFuse；(g) LatLRR；(k) PMGI；(l)本方法

Fig. 6　Partial enlarged view of Fig. 5(d). (a) IR; (b) VI; (c) CVT; (d) MSVD; (e) Wavelet; (f) CSR; (g) DLF; (h) FusionGAN; (i) DenseFuse; (g) LatLRR; (k) PMGI; (l) our method

3.2　融合结果的客观评价

为了进一步评价融合图像的质量，选取了熵(EN)、标准差(SD)、相关系数(CC)、结构相似度(SSIM)、基于人类视觉感知启发的质量度量(Q_CV)[22]、基于人类视觉系统模型的图像融合感知质量评估(Q_CB)[23]六个指标对融合图像的质量进行定量评估。其中，EN 是评价图像信息含量的指标，EN 越大，表明图像包含的信息越丰富；SD 能反映图像的分布和对比度，SD 越大，表明图像的质量越好；CC 可用于测量融合图像与原始图像的平均相关程度；SSIM 从相关度损失、亮度损失、对比度损失三个方面度量原始图像和融合图像之间的结构相似程度，SSIM 越大，表明图像的结构信息越相似；Q_CV、Q_CB 都是针对融合领域基于人类视觉系统的评价指标，Q_CV 通过计算融合图像和原始图像加权差分图像的均方误差，得到融合图像的质量测度值；Q_CB 是反映融合图像与原始图像感知对比度保留程度的评价指标。所有指标的数值均越大，表示对应的性能越好，不同融合方法的评价指标如表 1 所示。

表 1　不同融合方法的评价指标

Table 1　Evaluation indicators of different fusion methods

Evaluation	EN	SD	CC	SSIM	Q_CV	Q_CB
CVT	6.583	27.029	0.608	0.712	370.72	0.508
MSVD	6.408	23.744	0.630	0.743	401.314	0.491
Wavelet	6.321	23.194	**0.636**	0.773	356.350	0.487
CSR	6.415	25.030	0.626	0.743	390.768	0.491
DLF	6.535	35.279	0.635	0.716	317.850	0.461
FusionGAN	6.336	25.030	0.517	0.647	**627.313**	0.421
DenseFuse	**7.048**	**39.306**	0.594	0.670	350.909	0.489
LatLRR	6.343	23.510	0.635	0.772	365.984	0.483
PMGI	6.964	34.012	0.628	0.706	419.241	0.522
Ours	6.836	32.638	0.615	**0.779**	626.444	**0.525**

从表 1 可以发现，相比 CVT、MSVD、Wavelet、CSR、LatLRR 这 5 种传统方法，本方法在 EN、SD、SSIM、Q_CV、Q_CB 指标上均是最优的，原因是本方法基于深度神经网络进行特征提取，得到的融合图像信息丰富，具有较高的对比度和 SSIM，人眼视觉效果较好。相比 DLF、FusionGAN、DenseFuse、PMGI 这 4 种深度学习方法，本方法具有最优的 SSIM、Q_CB 和次优的 Q_CV 性能。这表明本方法生成的融合图像综合考虑了图像的亮度、对比度、相关度，保留了更多的结构信息，与原始图像在结构上更相似；同时，本方法生成的融合图像具有良好的视觉对比敏感度，在各局部区域上有效保留了可见光图像的纹理信息。此外，本方法的 Q_CV 指标仅次于 FusionGAN 方法，但明显优于其他融合方法。本方法的 EN 指标位于第 3 位，CC、SD 指标位于第 4 位。值得注意的是，CC、SSIM、Q_CV、Q_CB 指标都是以原始图像作为先验进行计算的，比单纯的统计指标 EN、SD 更强调与原始图像的关联性，较高的 SSIM、Q_CV、Q_CB 指标也进一步验证了本方法的有效性。相比主观视觉效果较好的 PMGI 方法，本方法的生成网络采用了基于自注意力机制的特征提取模块，通过加强不同模态图像内部的特征依赖关系，增强了特征的表达能力，得到的图像模态信息和结构信息更完整。从人眼对空间频率敏感程度的角度考虑，本方法通过引入鉴别器丰富了融合图像的纹理细节。相比 PMGI 方法，本方法的 EN、SD、CC 指标分别降低了 1.838%、4.040%、2.070%，SSIM、Q_CV、Q_CB 指标分别提高了 10.340%、49.423%、0.575%。这表明本方法可在保留原始图像最大有效信息的同时保证图像结构的

完整性，融合图像的细节纹理丰富、目标与背景对比度较高，更符合人眼视觉的感知特性。

3.3 消融实验的分析

为了进一步验证本方法中自注意力机制模块和双通路的有效性，进行了消融实验，包括无自注意力机制模块且特征提取使用单通路（no_att&single_ch）、无自注意力机制模块且特征提取使用双通路（no_att&double_ch）、有自注意力机制模块且特征提取使用单通路（att&single_ch）、有自注意力机制模块且特征提取使用双通路（att&double_ch）的网络模型。将各网络分别进行训练，并随机选取 3 组数据在相同迭代次数下进行融合测试，结果如图 7 所示。可以发现，no_att&single_ch 融合图像的对比度较低，图像整体与红外图像相似，可见光图像的细节纹理保留较少。相比 no_att&single_ch 融合图像，no_att&double_ch 融合图像保留了更多可见光图像的纹理细节，更符合真实的可见光场景。相比 no_att&single_ch 融合图像，att&single_ch 融合图像的红外热目标明显，且融入了可见光图像的部分纹理，但背景没有明显的梯度变化，可见光信息保留不充分，整体效果较差。att&double_ch 融合图像综合了 att&single_ch 与 no_att&double_ch 的优势，红外热目标突出，且背景细节纹理更丰富，整体风格与可见光图像更相似，更符合人类视觉感知。这表明相比单通路网络，双通路网络能更好地提取、融合红外与可见光图像的跨模态特征，有效保留各模态图像的高频信息，融合图像的整体效果好、信息量更丰富。而自注意力机制可以学习到像素与全局特征之间的依赖关系，关注多模态图像的局部显著区域并赋予其更多的权重，使融合图像中红外热目标与背景的对比较高，同时在局部细节上仍能保证必要的梯度变化。二者相互配合，可得到含有丰富原始图像信息并符合人类视觉感知的融合图像。

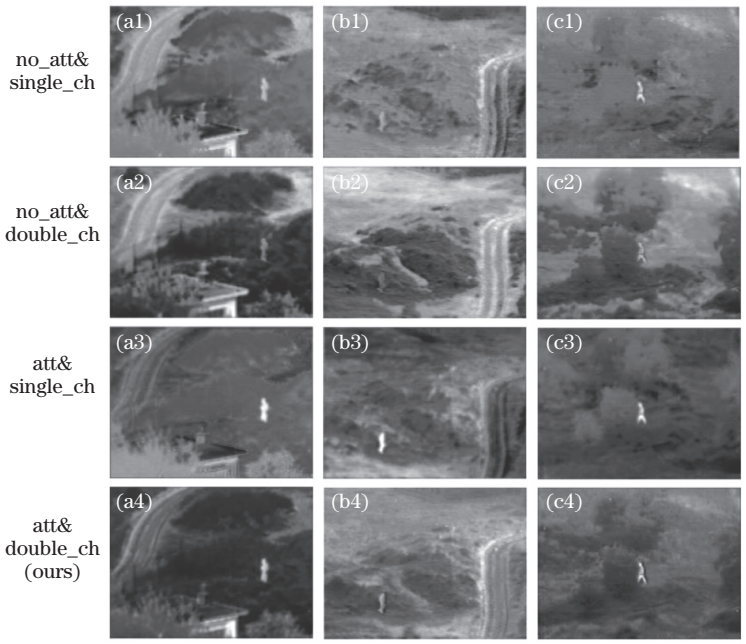

图 7　消融实验的结果。(a)～(c)图像 1～图像 3

Fig. 7　Results of ablation experiments. (a)-(c) Image 1-image 3

不同模块的定量评价指标如表 2 所示，可以发现，双通道网络比单通道网络生成的融合图像 EN、SD、CC、Q_CB 指标更高，这表明双通路提取的多模态特征能有效保留不同模态的差异信息，在融合图像中表现为高频信息丰富、对比度高，与原始图像信息更相似、视觉效果更好。有自注意力机制的网络比无自注意力机制网络生成的融合图像 CC、Q_CB、Q_CV 指标更高，这表明含注意力机制的网络可通过加强多模态特征的依赖关系加强特征表达能

表 2　不同方法的图像融合结果

Table 2　Image fusion results of different methods

Evaluation	EN	SD	CC	SSIM	Q_CV	Q_CB
no_att&single_ch	6.784	28.231	0.601	0.834	377.794	0.470
no_att&double_ch	**7.300**	**45.745**	0.646	0.799	405.544	0.496
att&single_ch	6.476	25.454	0.610	0.654	**1052.110**	0.479
att&double_ch	6.792	28.480	**0.654**	**0.864**	688.366	**0.513**

力，进而选择更有效的信息，使融合图像与原始图像的相关性更高，更符合人类视觉感知。在所有指标中，att&double_ch 模型生成的融合图像 CC、SSIM、Q_CB 指标最高，EN、SD、Q_CV 指标均位居第 2 位，综合表现最好。

4 结 论

针对现有红外与可见光图像融合方法中不同模态数据的特性差异表达不足、单模态图像内长距离依赖关系弱导致的融合图像背景细节纹理丢失、对比度低等问题，提出了一种基于双通路 GAN 的红外与可见光图像融合方法。为了使融合图像保留多模态图像的差异性信息，用双通路分别获取红外与可见光图像的特征，并在多源特征上利用自注意力机制计算特征的像素依赖关系，进一步得到增强特征。将生成器与鉴别器进行对抗训练，使最终的融合图像保留较好的对比度和更多的细节纹理信息。实验结果表明，相比其他方法，本方法得到的融合图像更符合人类视觉感知要求，既突出了红外热目标又尽可能地保留了可见光图像的细节纹理，同时具有较好的客观评价指标。

参 考 文 献

[1] Feng Y F, Yin H, Lu H Q, et al. Infrared and visible light image fusion method based on improved fully convolutional neural network [J]. Computer Engineering, 2020, 46(8): 243-249, 257.
冯玉芳, 殷宏, 卢厚清, 等. 基于改进全卷积神经网络的红外与可见光图像融合方法[J]. 计算机工程, 2020, 46(8): 243-249, 257.

[2] Ma J Y, Ma Y, Li C, et al. Infrared and visible image fusion methods and applications: a survey[J]. Information Fusion, 2019, 45: 153-178.

[3] Liu Y, Chen X, Ward R K, et al. Image fusion with convolutional sparse representation[J]. IEEE Signal Processing Letters, 2016, 23(12): 1882-1886.

[4] Li H, Zhang L M, Jiang M R, et al. An infrared and visible image fusion algorithm based on ResNet152 [J]. Laser & Optoelectronics Progress, 2020, 57(8): 081013.
李恒, 张黎明, 蒋美容, 等. 一种基于 ResNet152 的红外与可见光图像融合算法[J]. 激光与光电子学进展, 2020, 57(8): 081013.

[5] Ding L Y, Duan J, Song Y, et al. Image fusion based on residual learning and visual saliency mapping [J]. Laser & Optoelectronics Progress, 2020, 57(16): 161008.
丁罗依, 段锦, 宋宇, 等. 基于残差学习和视觉显著性映射的图像融合[J]. 激光与光电子学进展, 2020, 57(16): 161008.

[6] Li S T, Kang X D, Fang L Y, et al. Pixel-level image fusion: a survey of the state of the art[J]. Information Fusion, 2017, 33: 100-112.

[7] Zhao C, Huang Y D. Infrared and visible image fusion via rolling guidance filtering and hybrid multi-scale decomposition [J]. Laser & Optoelectronics Progress, 2019, 56(14): 141007.
赵程, 黄永东. 基于滚动导向滤波和混合多尺度分解的红外与可见光图像融合方法[J]. 激光与光电子学进展, 2019, 56(14): 141007.

[8] Qu G H, Zhang D L, Yan P F, et al. Medical image fusion by wavelet transform modulus maxima [J]. Optics Express, 2001, 9(4): 184-190.

[9] Choi M, Kim R Y, Nam M R, et al. Fusion of multispectral and panchromatic satellite images using the curvelet transform [J]. IEEE Geoscience and Remote Sensing Letters, 2005, 2(2): 136-140.

[10] Naidu V P S. Image fusion technique using multi-resolution singular value decomposition[J]. Defence Science Journal, 2011, 61(5): 479.

[11] Li H, Wu X J, Kittler J. Infrared and visible image fusion using a deep learning framework [C]// 2018 24th International Conference on Pattern Recognition (ICPR), August 20-24, 2018, Beijing, China. New York: IEEE Press, 2018: 2705-2710.

[12] Ma J Y, Yu W, Liang P W, et al. FusionGAN: a generative adversarial network for infrared and visible image fusion[J]. Information Fusion, 2019, 48: 11-26.

[13] Li H, Wu X J. DenseFuse: a fusion approach to infrared and visible images[J]. IEEE Transactions on Image Processing, 2019, 28(5): 2614-2623.

[14] Zhang H, Xu H, Xiao Y, et al. Rethinking the image fusion: a fast unified image fusion network based on proportional maintenance of gradient and intensity[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12797-12804.

[15] Tang C Y, Pu S L, Ye P Z, et al. Fusion of low-illuminance visible and near-infrared images based on convolutional neural networks [J]. Acta Optica Sinica, 2020, 40(16): 1610001.
唐超影, 浦世亮, 叶鹏钊, 等. 基于卷积神经网络的低照度可见光与近红外图像融合[J]. 光学学报, 2020, 40(16): 1610001.

[16] Zhao L. Research on multimodal data fusion methods [D]. Dalian: Dalian University of Technology, 2018: 23-24.
赵亮. 多模态数据融合算法研究[D]. 大连: 大连理

工大学, 2018: 23-24.

[17] He Q, Li Y, Song W, et al. Multimodal remote sensing image classification with small sample size based on high-level feature fusion [J]. Laser & Optoelectronics Progress, 2019, 56(11): 111001.
贺琪, 李瑶, 宋巍, 等. 小样本的多模态遥感影像高层特征融合分类[J]. 激光与光电子学进展, 2019, 56(11): 111001.

[18] Zhu D T. Traffic signs recognition based on deep learning [D]. Hefei: Anhui Jianzhu University, 2018: 30-58.
朱东涛. 基于深度学习的交通标志识别[D]. 合肥: 安徽建筑大学, 2018: 30-58.

[19] Kamnitsas K, Ledig C, Newcombe V F J, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation [J]. Medical Image Analysis, 2017, 36: 61-78.

[20] Toet A. TNO image fusion dataset [DB/OL]. (2014-04-26) [2020-08-10]. https://figshare.com/articles/TN_Image_Fusion_Dataset/1008029.

[21] Li H, Wu X J. Infrared and visible image fusion using latent low-rank representation [EB/OL]. (2019-08-09) [2020-08-10]. https://arxiv.org/abs/1804.08992.

[22] Chen Y, Blum R S. A new automated quality assessment algorithm for image fusion[J]. Image and Vision Computing, 2009, 27(10): 1421-1432.

[23] Chen H, Varshney P K. A human perception inspired quality metric for image fusion based on regional information[J]. Information Fusion, 2007, 8(2): 193-207.