

基于三支对抗学习和补偿注意力的红外和可见光图像融合

邸敬¹, 任莉¹, 刘冀钊², 郭文庆¹, 廉敬¹

(1. 兰州交通大学 电子与信息工程学院, 甘肃 兰州 730070; 2. 兰州大学 信息科学与工程学院, 甘肃 兰州 730000)

摘要: 针对现有深度学习图像融合方法依赖卷积提取特征, 并未考虑源图像全局特征, 融合结果容易产生纹理模糊、对比度低等问题, 本文提出一种基于三支对抗学习和补偿注意力的红外和可见光图像融合方法。首先, 生成器网络采用密集块和补偿注意力机制构建局部-全局三支提取特征信息。然后, 利用通道特征和空间特征变化构建补偿注意力机制提取全局信息, 更进一步提取红外目标和可见光细节表征。其次, 设计聚焦双对抗鉴别器, 以确定融合结果和源图像之间的相似分布。最后, 选用公开数据集 TNO 和 RoadScene 进行实验并与其他 9 种具有代表性的图像融合方法进行对比, 本文提出的方法不仅获得纹理细节更清晰、对比度更好的融合结果, 而且客观度量指标优于其他先进方法。
关键词: 红外可见光图像融合; 局部-全局三支; 局部特征提取; 补偿注意力机制; 对抗学习; 聚焦双对抗鉴别器;

中图分类号: TP391

文献标识码: A

文章编号: 1001-8891(2024)05-0510-12

Infrared and Visible Image Fusion Based on Three-branch Adversarial Learning and Compensation Attention Mechanism

DI Jing¹, REN Li¹, LIU Jizhao², GUO Wenqing¹, LIAN Jing¹

(1. School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China;

2. School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China)

Abstract: The existing deep learning image fusion methods rely on convolution to extract features and do not consider the global features of the source image. Moreover, the fusion results are prone to texture blurring, low contrast, etc. Therefore, this study proposes an infrared and visible image fusion method with adversarial learning and compensated attention. First, the generator network uses dense blocks and the compensated attention mechanism to construct three local-global branches to extract feature information. The compensated attention mechanism is then constructed using channel features and spatial feature variations to extract global information, infrared targets, and visible light detail representations. Subsequently, a focusing dual-adversarial discriminator is designed to determine the similarity distribution between the fusion result and source image. Finally, the public dataset TNO and RoadScene are selected for the experiments and compared with nine representative image fusion methods. The method proposed in this study not only obtains fusion results with clearer texture details and better contrast, but also outperforms other advanced methods in terms of the objective metrics.

Key words: infrared-visible image fusion, local-global three-branch, local feature extraction, compensated attention mechanism, adversarial learning, focused dual adversarial discriminator

0 引言

图像融合是图像处理领域中最具挑战且长期存在的任务。由于成像技术或传感器的限制, 单个传感

器获取的图像很难表现出全面或详细的场景。例如, 红外图像可以在各种照明条件下检测目标。然而, 它通常以低空间分辨率提供有限的场景细节。相比之下, 可见光图像是通过捕捉可见光来生成的, 它揭示了

收稿日期: 2023-09-07; 修订日期: 2023-12-11.

作者简介: 邸敬 (1979-), 女, 硕士生导师, 副教授, 主要研究方向为图像检测识别、信号处理技术和宽带无线通信方面的研究。

通信作者: 任莉 (1997-), 女, 硕士研究生, 主要研究方向为深度学习与图像融合。E-mail: 1427594911@qq.com。

基金项目: 国家自然科学基金 (62061023); 甘肃省杰出青年基金 (21JR7RA345); 甘肃省科技计划资助项目 (22JR5RA360)。

丰富的纹理细节, 并且与人类视觉系统一致。然而, 可见光图像受天气的影响很大。因此, 红外与可见光图像融合可以克服硬件设备的限制, 并已广泛应用于遥感、目标检测、军事侦察等领域^[1-3]。

红外和可见光图像融合方法大致分为两类: 传统方法和深度学习方法^[4-7]。传统方法通过一定的数学变换将输入图像变形到特定的变换域中, 然后手动设置融合规则, 如多尺度变换^[8]、稀疏表示^[9-10]、基于显著性检测^[11]等。但传统方法没有考虑图像的内在差异性, 采用统一的数学模型来无差别地提取图像特征, 导致在某些情况下融合效果不佳, 如一种基于显著性多尺度分解的红外和可见光图像融合方法^[12]和一种基于卷积稀疏编码 (convolutional sparse coding, CSC) 和耦合特征学习 (coupled feature learning, CFL) 的 CCFL 红外和可见光图像融合方法^[13]。近年来, 随着深度学习在计算机视觉任务中的快速发展, 已经提出了大量基于深度学习的图像融合方法。Liu 等^[14]首次将卷积神经网络应用于图像融合, 是将传统方法与深度学习方法相结合的有效尝试, 但需要进行有监督的训练。Prabhakar 等^[15]提出了一种无监督深度学习融合框架, 虽然该框架可以通过端到端的方式进行训练, 但是仅利用最后一层的特征进行融合, 导致没有用到所有层信息, 融合效果较差。Ma 等^[16]构建提出一种生成式对抗网络 FusionGAN, 将红外和可见光图像融合转化为保持红外信息与可见光信息的对抗博弈, 在一定程度上提高了融合性能, 但融合图像丢失了可见光图像中的细节信息。为了保留可见光图像中更多的细节信息, Zhang 等^[17]提出了一种新型的生成对抗网络 GAN-FM, 生成器是一种全尺寸跳跃连接生成器, 用来提取浅层特征, 鉴别器则是使用两个马尔可夫判别器, 通过与生成器进行对抗博弈, 可以充分保留红外显著目标和可见光纹理信息。此外, 还设计了一种新的强度掩膜生成对抗网络 IMGAN^[18]和一种无监督连续学习生成对抗网络 UIFGAN^[19]以补充多模态图像信息。然而却未能将提取的特征进行高效的整合, 融合图像容易产生纹理模糊和图像整体对比度低的融合结果。

针对以上问题, 本文提出一种三支对抗学习和补偿注意力的红外和可见光图像融合方法。生成器采用密集块和交互注意力机制构建了一个三支路径, 即红外局部路径、可见光局部路径和红外可见光全局路径。红外路径和可见光路径通过密集块充分提取图像中的像素信息和梯度信息, 减少中间层信息丢失。红外与可见光全局路径通过级联建立长期依赖和远程传输为红外局部路径和可见光局部路径提供额外

的像素信息和强度信息, 从而可以在融合图像中保留更多的红外目标和可见光细节纹理。此外, 不仅设计双鉴别器对融合信息进行平衡, 而且在此基础上引入补偿注意力模块, 使鉴别器能够精准地鉴别图像, 通过对生成器优化, 融合出纹理清晰和对比度更好的融合图像。

1 图像融合方法

1.1 网络整体框架

红外图像具有很强的抗干扰能力, 可见光图像具有较好的纹理细节和空间分辨率。为了提高红外与可见光图像融合质量, 提出了一种无监督端到端红外和可见光图像融合模型, 网络总体框架如图 1 所示。

本文总体网络框架由生成器和鉴别器两部分组成。生成器网络考虑源图像来自不同的传感器具有不同的性质, 设计了三支网络提取红外和可见光图像信息。鉴别器网络的目的是通过不断地区分融合图像和源图像迫使最终融合图像包含更丰富的场景信息, 设计了聚焦双对抗鉴别器提高鉴别能力。在生成器网络中, 采用密集块构建两条局部特征提取分支, 避免卷积操作过程中中间层信息丢失; 再利用通道特征和空间特征变化构建补偿注意力机制, 其为第三分支, 对全局信息进行建模后将补偿注意力机制生成的显著目标馈送入局部特征提取分支, 从而更进一步提取红外目标和可见光细节表征。在鉴别器网络中, 分别构建红外鉴别器和可见光鉴别器对融合图像进行判别, 且引入补偿注意力机制使鉴别器更精准地聚焦典型区域信息。通过双鉴别器和生成器对抗性博弈获得纹理清晰、对比度高的融合图像。

1.2 生成器网络

1.2.1 局部-全局三支生成器

局部-全局三支生成器网络主要由红外与可见光源图像输入、特征提取、融合 3 部分组成。在特征提取阶段, 考虑到红外和可见光图像来源于不同的传感器具有不同的差异。本文从红外和可见光图像的本质出发解决现有图像融合存在对比度低、纹理细节丢失等融合问题。本文采用三条路径进行源图像特征提取, 其中两条特征提取路径采用密集块设计两条局部路径提取红外目标信息和可见光纹理信息。两条路径具有相同的结构、共享参数, 并分别接受红外和可见光作为输入。密集连接不会丢失中间层大量信息, 可以充分提取红外和可见光图像的局部特征信息, 但无法获得全局关联信息, 因此, 本文设计一种补偿注意力机制作为第三条特征提取路径捕获聚焦红外和可见光全局特征信息, 并且在每次进行局部特征提取

时，将补偿注意力机制生成的显著红外目标和背景纹理信息馈送入两条局部特征提取中，经过 3 次的馈送输入，确保有用信息得以充分提取。在融合阶段，则是将不同来源的特征按照通道维度进行 **concat** 拼接，再使用卷积对不同来源的特征进行融合，使其形成对比度高、纹理细节清晰的融合图像。

本文设计的局部-全局三支生成器网络具体框架如图 2 所示。首先，将红外与可见光源图像分别通过 4 个密集连接块部分，其中每一块通过核为 3×3 的卷积后采用 **Leaky ReLU** 为激活函数进行特征提取，前面的各个模块再作为后一模块的输入，依次往后，

密集块将最大程度地提取红外图像的背景和目标信息以及可见光的高分辨率和高清晰度的纹理信息。其次，由空间注意力和通道注意力构成的补偿注意力机制捕捉两幅源图像共同的显著信息，这为图像融合提供了额外的像素和梯度信息。然后，将提取的信息馈送入局部特征的每一阶段，使其特征提取更加全面。最后，对 3 条路径提取的红外目标信息和可见光纹理细节信息进行融合，并采用 **concat** 和 3×3 的卷积策略实现这一目的，沿着通道维度连接其特征图，提高融合质量。

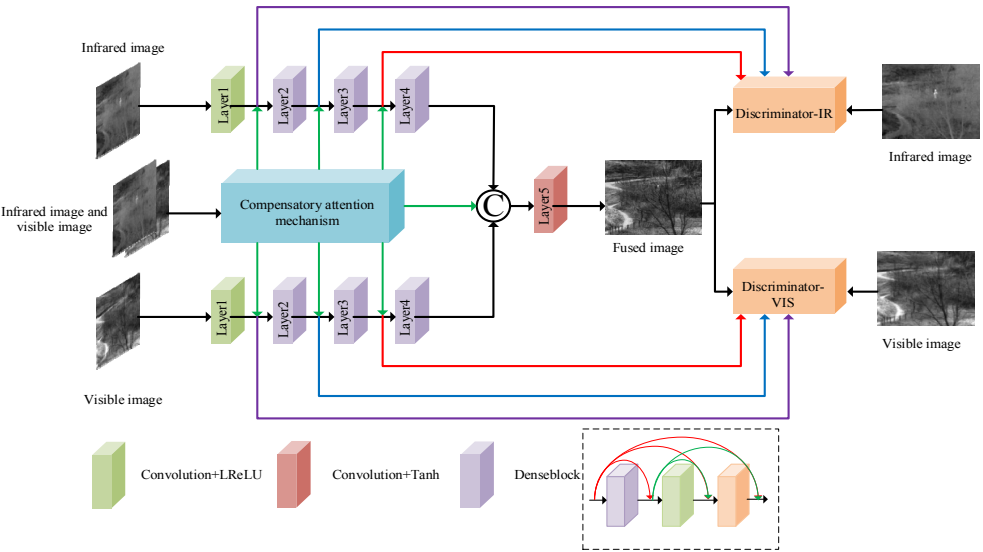


图 1 网络总体框架图

Fig.1 General framework diagram of the network

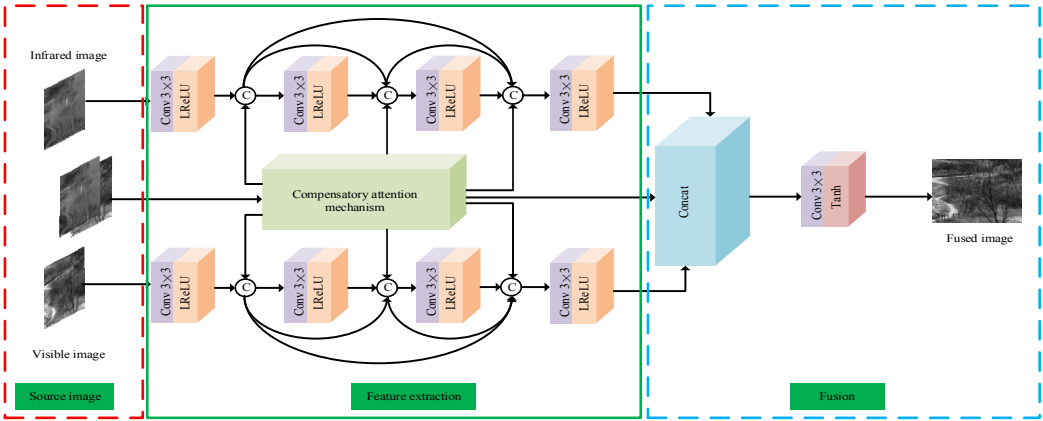


图 2 局部-全局三支生成器网络模型

Fig.2 Local-global three-branch generator network model

1.2.2 补偿注意力机制

受现有混合注意力机制的启发，本文重新设计并构建了补偿注意力模块，以传达红外显著目标和可见光背景纹理信息。补偿注意力机制模块的框架如图 3 所示。对于中间特征 $\phi_s \in R^{H\times W\times C}$ ，先对其进行全局平均池化和最大池化，之后利用 **concat** 操作将输出的空间

特征图级联在一起并馈送到空洞卷积层和 **Sigmoid** 激活层中，获得它们各自的初始空间加权系数 ρ_1 、 ρ_2 。随后，将相应的空间注意力图作为通道注意力模型的输入，并分别采用全局平均池化和最大池化将特征映射聚合到通道中，再通过两个核大小为 3×3 卷积层和一个 **LReLU** 激活层，输出特征向量进行连接，并

通过洞卷积层和 sigmoid 激活层。补偿注意力模块将红外或可见光图像的特征馈送到通道和空间注意力模型中以产生它们各自的注意力图,这些注意力图用于补偿特征重建的信息。

1.3 聚焦双对抗鉴别器

鉴别器网络的主要作用在于区分生成的融合图像与源图像之间的差异。通过不断地与生成器进行对抗博弈,获得更丰富的场景信息,更进一步地提高融合质量。本文设计的聚焦双对抗鉴别器如图 4 所示。红外鉴别器和可见光鉴别器具有相同的网络架构,都由一个公共卷积层和 4 个卷积组成。所有卷积层都选择 3×3 大小的卷积核和 Leaky Relu 激活函数。在训练过程中,首先将初始融合图像通过公共卷积层提取

渐层信息特征,然后再将补偿注意力机制引入剩下的卷积层中进行深层特征提取,使鉴别器能够更精准地从显著信息区域而非整幅图像来鉴别图像。聚集双对抗鉴别器网络以原可见光与红外图像作为标签,输出为每类图像的判别得分,希望对其红外图像和可见光图像输出的判别值更大,同时希望对初步融合图像的判别值更小。红外鉴别器迫使融合图像逐渐保留越来越多的红外像素信息,而可见光鉴别器迫使融合图像逐渐包含越来越多的可见细节信息。当生成器和双鉴别器的对抗博弈达到平衡时,表明生成器已经欺骗了双鉴别器,得到了期望的融合结果,可以同时保持更相似的红外像素强度和更精细的可见纹理细节。

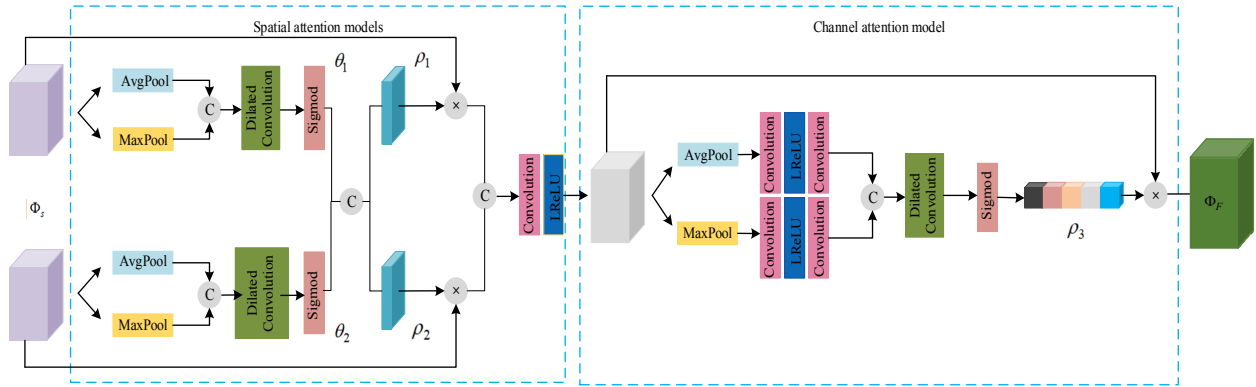


图 3 补偿注意力机制

Fig.3 Compensatory attention mechanism

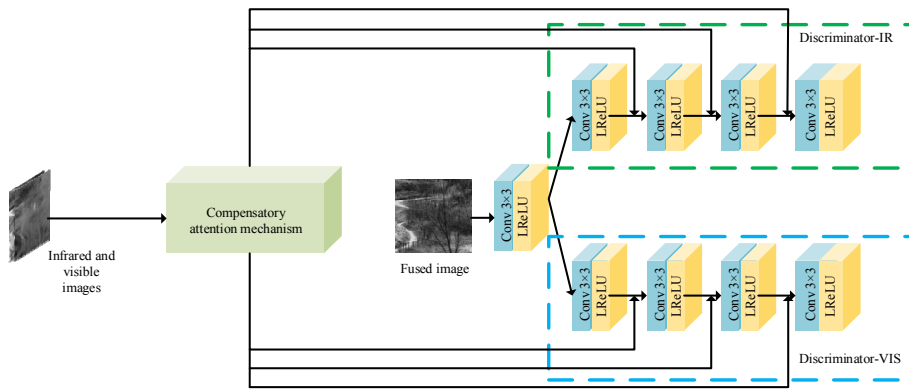


图 4 聚焦双对抗鉴别器网络模型

Fig.4 Focused dual adversarial discriminator network model

1.4 损失函数

本文总体框架主要由局部-全局三支生成器和聚焦双对抗鉴别器组成,生成器是通过特征提取融合成单幅图像,鉴别器则是通过红外鉴别器和可见光鉴别器获得更多的场景信息和目标信息,相应的损失函数也由生成器损失 L_G 和鉴别器损失 L_D 组成,它们的定义如下:

$$L = L_G + L_D \quad (1)$$

生成器损失 L_G 由内容损失 L_{Gcon} 和对抗损失 L_{Gadv} 组成。由于 GAN 的不稳定性,内容损失的引入为生成器增加了一系列约束以实现融合目标,而对抗性损失则允许融合图像满足更严格的要求。 L_G 定义如下:

$$L_G = L_{Gcon} + L_{Gadv} \quad (2)$$

式中: L_{Gcon} 内容包括像素损失 L_{pix} 、梯度损失 L_{grad} 、

一致性损失 L_{com} ，可以表示为：

$$L_{\text{Gcon}} = \alpha_1 L_{\text{pix}} + \alpha_2 L_{\text{grad}} + \alpha_3 L_{\text{com}} \quad (3)$$

式中： α_1 、 α_2 、 α_3 用于控制平衡。

梯度损失 L_{grad} 迫使融合图像包含丰富的纹理细节，本文设计在梯度损失中引入了自适应决策块，自适应决策块先利用高斯低通滤波，尽可能减少噪声对决策块的影响，通过基于梯度丰富度来评估相应像素的重要性，以便生成像素级权重块引导融合图像中的纹理自适应的向源图像中的纹理靠近。梯度损失函数定义为：

$$L_{\text{grad}} = \frac{1}{HW} \sum_i \sum_j S_{1i,j} \cdot (\nabla I_{\text{fused},i,j} - \nabla I_{1,i,j})^2 + \quad (4)$$

$$S_{2i,j} \cdot (\nabla I_{\text{fused},i,j} - \nabla I_{2,i,j})^2$$

$$S_{1i,j} = \text{sign}(|\nabla(L(I_{1i,j}))|) - \min(|\nabla(L(I_{1i,j}))|, |\nabla(L(I_{2i,j}))|) \quad (5)$$

$$S_{2i,j} = 1 - S_{1i,j} \quad (6)$$

式中： i 和 j 表示像素在第 i 行和第 j 列决策图； H 和 W 代表图像的高度和宽度； I_1 和 I_2 为源图像； I_{fused} 为融合图像； $|\cdot|$ 表示绝对值函数； $\nabla(\cdot)$ 为拉普拉斯算子； $L(\cdot)$ 表示高斯低通滤波器函数； $\min(\cdot)$ 表示最小函数；符号 (\cdot) 为符号函数； S_{ij} 表示基于源图像的梯度水平生成的决策块。

像素损失会影响融合图像的对比度，像素损失通过改进 Frobenius 范数，改进的 Frobenius 范数不同于欧式距离，它是矩阵中每一个位置的元素的平方和取平均数再开方，其作用是衡量融合图像像素和源图像像素的距离，有效调整融合图像和两个源图像之间的权重，迫使网络选择更有效的信息。公式如下：

$$L_{\text{pix}} = \sqrt{\frac{1}{HW} \sum_i \sum_j \|I_{\text{fused},i,j} - I_{1,i,j}\|_F^2 + \beta \|I_{\text{fused},i,j} - I_{2,i,j}\|_F^2} \quad (7)$$

式中： I_1 和 I_2 是原始图像； I_{fused} 是融合图像； $\|\cdot\|_F$ 表示 Frobenius 范数。

在此基础上，本文采用一致性损失函数，以避免融合图像的色差，并确保其整体的自然性。具体地定义如下：

$$L_{\text{com}} = \frac{1}{HW} \sum_i \sum_j (I_{\text{fused},i,j} - I_{1i,j})^2 + (I_{\text{fused},i,j} - I_{2i,j})^2 \quad (8)$$

生成器的对抗损失 L_{Gadv} 用于迫使融合图像实现更高的质量，其公式如下：

$$L_{\text{Gadv}} = \frac{1}{N} \sum_{n=1}^N (D(I_f^n) - a)^2 \quad (9)$$

式中： N 表示融合图像的数目，并且本文采用 a 作为生成器期望鉴别器对融合图像进行判断的概率标签。

鉴别器损失起着判别融合图像与源图像之间的真假性，通过与生成器的不断博弈，使生成的融合结果更趋向于源图像的真实数据分布，最终形成场景信息丰富、纹理背景细节清晰的融合图像。红外鉴别器和可见光鉴别器损失函数定义如下：

$$L_{D_r} = \frac{1}{N} \sum_{n=1}^N [D_r(I_{r,f}^n) + \lambda(1 - \|\nabla D_r(I_r^n)\|_2)^2] \quad (10)$$

$$L_{D_v} = \frac{1}{N} \sum_{n=1}^N [D_v(I_{v,f}^n) + \lambda(1 - \|\nabla D_v(I_v^n)\|_2)^2] \quad (11)$$

式中： λ 是正则化参数； $\|g\|_2$ 表示 L2 范数。第一项表示融合结果与红外或可见光图像之间的 wasserstein 距离，而第二项是梯度惩罚，这限制了鉴别器的学习能力。

2 实验结果及分析

2.1 实验设置和环境

本文融合方法选用公开数据集 TNO^[20] 和 RoadScene^[21] 进行图像融合任务。在训练时，将原始的 TNO 数据集通过裁剪分解的方法对数据进行扩张，得到 32484 张大小为 120×120 的图像。在测试阶段选择源图像本身大小的红外和可见光图像进行测试。

本实验在 Windows10 操作系统下的 Intel Core i5-1035G1 CPU 下进行。主要采用 tensorflow、imageio 库在 Pycharm 编译器中训练和测试网络性能。实验设置 epoch=15、batch_size=32、learning_rate=1e-4。在优化算法方面，选择了适应性强、收敛性好的自适应优化算法 Adam。 α_1 、 α_2 、 α_3 比例系数经过反复实验后设置为 1，80，1。正则化参数 λ 设置为 10。此外，在等式(7)中， β 为 0.5，使网络从红外图像中获取主要的强度信息，以保持显著的对比度。

2.2 TNO 数据集实验分析

为了验证本文算法的有效性，本文从 TNO 测试数据集中选取了 Nato camp、airplane_in_trees、helicopter、Kaptein_1654、Marne_04、Movie_18、sandpath 7 种场景作为红外与可见光图像融合任务的源图像，并将其与最流行的 9 种融合方法进行比较，包括 FusionGAN^[16]、BF (bayesian fusion)^[22]、SDNet (squeeze-and-decomposition network)^[21]、SMVIF (saliency-based multiscale approach for infrared and visible image fusion)^[12]、RTVD (relative total variation decomposition)^[23]、U2Fusion (unified unsupervised image fusion network)^[24]、SwinFusion^[25]、PIAFusion (progressive infrared and visible image fusion)^[26]、FLFusion (fast and lightweight infrared and visible image

fusion network) [27]。不同融合方法不同场景的融合结果如图 5 所示。通过对比可以看出, 传统融合方法 BF 和 SMVIF 方法整体亮度偏低且红外目标不突出, 背景纹理细节缺失严重, 导致整体融合效果差。相比之下 U2Fusion 和 FLFusion 方法提高了整体亮度, 但红外目标依然不突出。FusionGAN 方法整体亮度较好、



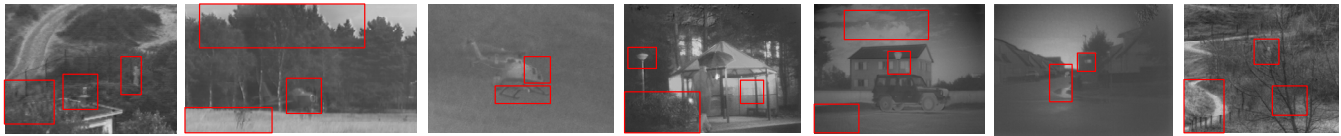
(a)Infrared images



(b)Visible images



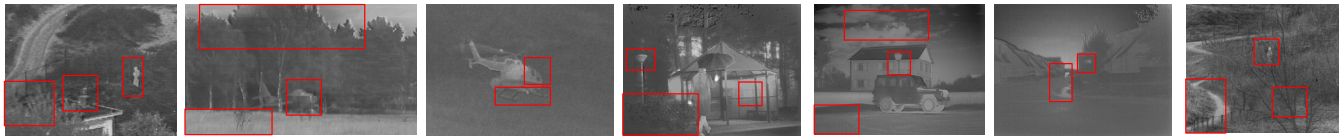
(c)FusionGAN



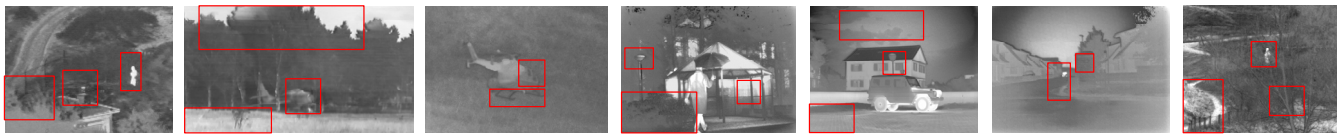
(d)BF



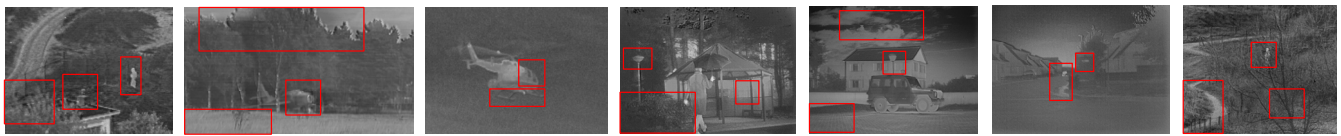
(e)SDNet



(f)SMVIF



(g)RTVD



(h)U2Fusion

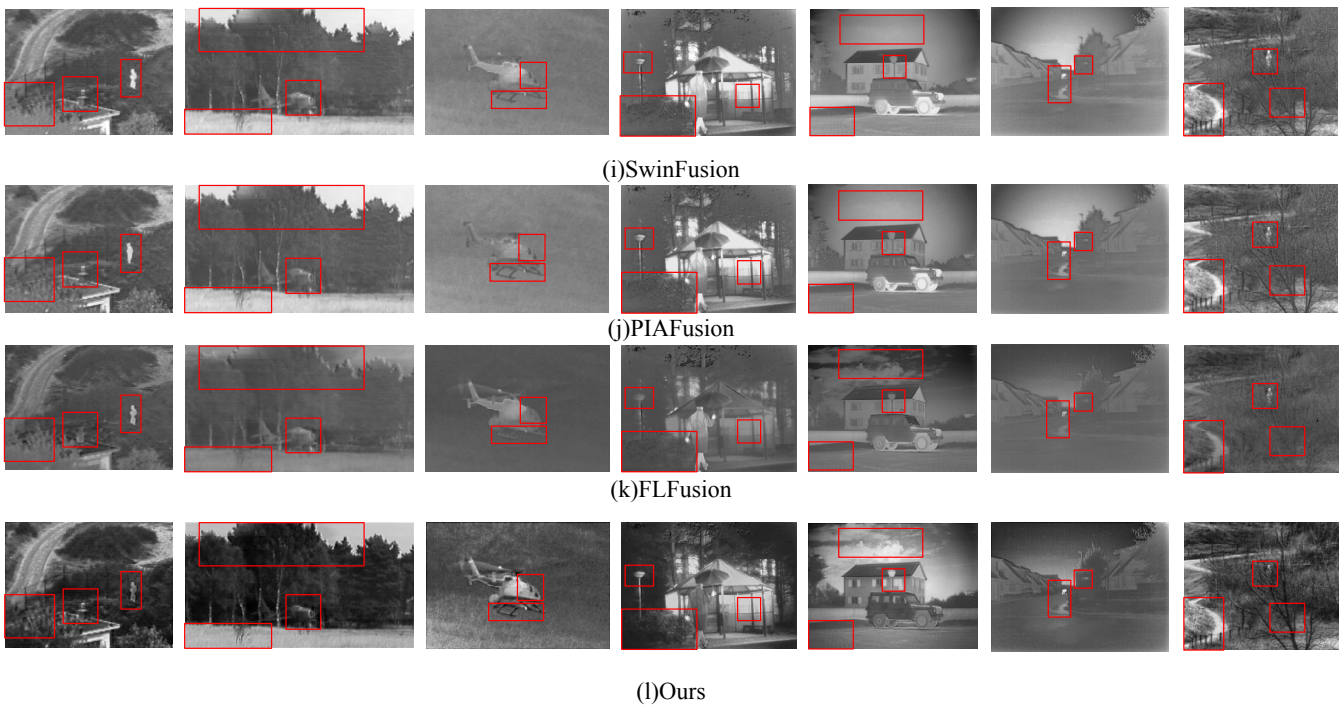


图 5 七组场景的融合结果

Fig.5 Fusion results for seven groups of scenes

红外目标突出，但整体融合图像模糊，纹理细节信息丢失严重。RTVD、SwinFusion 和 PIAFusion 方法虽然在一定程度上将红外和可见光源图像信息都融入进来，但噪声现象严重，部分融合场景出现了失真，如第五列融合图像的天空部分，这 3 种融合方法都出现了不同程度上的失真。SDNet 算法很大程度上解决了上述问题，整体对比度高、目标突出，但一些重要信息仍无法准确体现，如第四组图像中，左下角的数目纹理细节信息缺失。本文融合方法整体对比度高、目标人物突出、背景纹理信息清晰，场景信息丰富，视觉效果最佳。

为了使融合结果更具有说服力，本文选取信息熵（entropy, EN）、平均梯度（average gradient, AG）、标准差（standard deviation, SD）、空间频率（spatial frequency, SF）、基于视觉信息保真度（visual information fidelity for fusion, VIFF）、相关系数（Correlation coefficient, CC）、互信息（mutual information, MI）和基于多尺度结构相似性（multi-scale structural similarity, MSSSIM）8 种评价指标进行客观评价。从融合的 7 幅场景中选择一幅最具代表性的融合结果进行评价，指标如表 1 所示，同时也将 7 幅场景的客观评价指标进行平均如表 2 所示。

表 1 代表性图像的客观评价结果（第四列图像）

Table 1 Objective evaluation results of representative images (Fourth column images)

Methods	EN	AG	SD	SF	VIFF	CC	MI	MSSSIM
FusionGAN	6.0183	1.6860	4.7715	4.6687	0.1133	0.0905	12.0366	0.6645
BF	6.6117	2.9497	5.9747	7.9090	0.3070	0.9662	13.2234	0.8852
SDNet	6.2120	4.3359	4.9257	10.8688	0.2373	0.3034	12.4239	0.8042
SMVIF	6.3927	3.8549	5.2895	10.3887	0.3384	0.8488	12.7855	0.8822
RTVD	6.7321	3.5092	6.2998	9.5507	0.2946	0.6309	13.4642	0.7830
U2Fusion	6.5800	3.7273	5.3552	8.6307	0.3528	0.8178	13.1599	0.8911
SwinFusion	6.7706	4.4139	6.5156	11.0276	0.4470	0.7718	13.5411	0.8891
PIAFusion	6.6548	4.0317	6.5047	10.0641	0.4158	0.8453	13.3097	0.9094
FLFusion	6.1153	2.5540	4.7810	6.5569	0.1681	0.6325	12.2308	0.7440
Ours	7.0854	4.2056	7.0897	11.1093	0.4646	0.9682	14.1709	0.9298

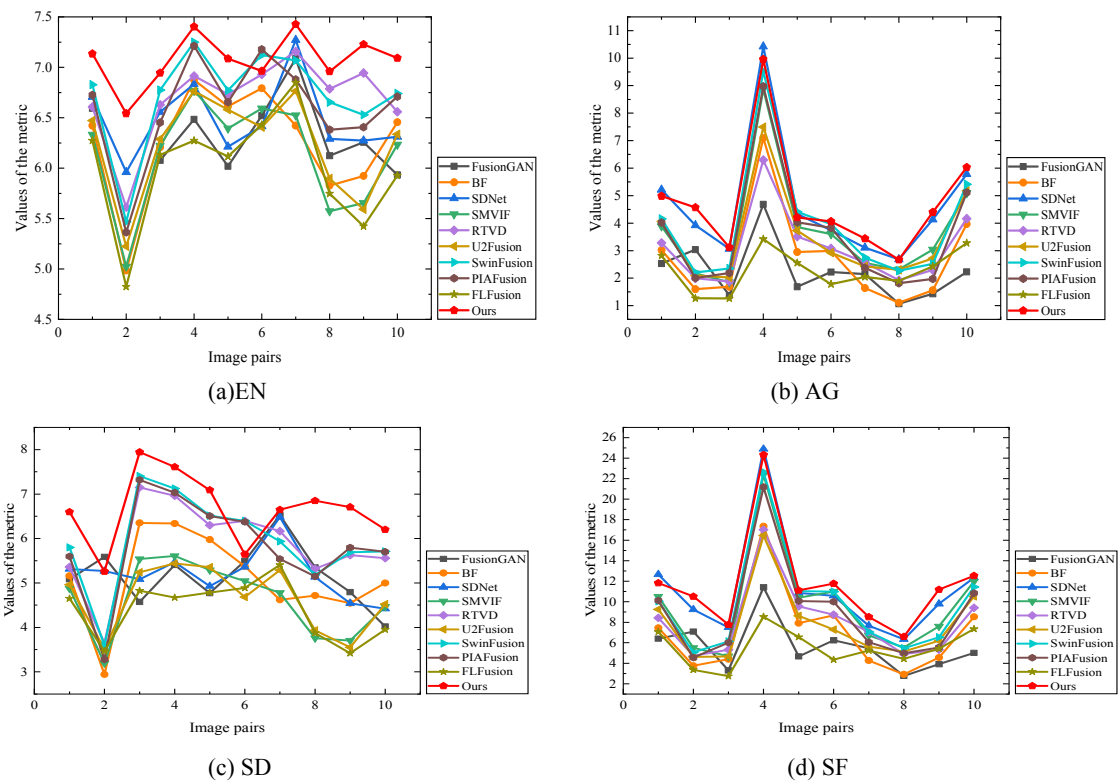
表 2 7 组图像的客观评价指标平均值
Table 2 Average of objective evaluation indexes of 7 groups of images

Methods	EN	AG	SD	SF	VIFF	CC	MI	MSSSIM
FusionGAN	6.3272	2.0612	5.0497	5.1198	0.2267	0.1786	12.6718	0.7276
BF	6.1464	2.3461	4.9419	5.8386	0.2103	0.8886	12.2927	0.8789
SDNet	6.4704	4.2226	5.1469	10.0102	0.3489	0.3938	12.9409	0.8584
SMVIF	6.0544	3.2066	4.5489	8.1944	0.2917	0.6290	12.1088	0.9014
RTVD	6.6056	2.8126	5.6856	7.1157	0.2328	0.5575	13.2111	0.7889
U2Fusion	6.1796	3.1756	4.6321	7.0709	0.3423	0.5475	12.3592	0.9186
SwinFusion	6.5968	3.4010	5.8062	8.1872	0.3592	0.6646	13.1935	0.8958
PIAFusion	6.4558	3.1007	5.6763	7.5976	0.2652	0.8107	12.9116	0.8712
FLFusion	5.9349	2.2358	4.3565	5.3895	0.1653	0.4108	11.8698	0.7768
Ours	7.0646	4.3918	6.6350	10.4934	0.5808	0.8746	14.1293	0.9173
AMIR	11.996%	48.804%	30.256%	46.366%	114.001%	54.906%	11.980%	8.377%

从表 2 中可以看出, 本文方法的 EN 值高, 说明融合后图像包含的信息量丰富; AG 值高, 说明融合后图像的细节表征和纹理特征明显; SF 值高, 说明融合后图像清晰度高; SD 值高, 说明图像灰度值离散度高, 图像携带的细节信息越丰富; VIFF 值高, 说明图像与人眼视觉特性一致; CC 值高说明融合图像与源图像之间的相似性越高, 融合图像获得了源图像中的大量信息; MI 值高, 说明融合后图像获取了大量源图像信息。MSSSIM 值略低于 SDNet 是因为 SDNet 融合结果中纹理细节信息存在模糊不清的地方, 而 MSSSIM 对于测定模糊降质图像的敏感度低。根据平均提升率 (average mean increase rate, AMIR) 数值也

可以看出本文算法的 8 种评价指标相比于其他 9 种流行算法平均分别提升了 11.996%、48.804%、30.256%、46.366%、114.001%、54.906%、11.980%、8.377%。此外, 为了进一步说明本文方法的优越性, 挑选了 10 幅红外与可见光图像融合结果进行客观评价对比, 如图 6 所示。通过折线图更直观了解数据变化趋势, 证明本文所提方法在纹理细节信息和红外目标特征提取等方面都优于其他融合方法。

运行效率也是评价性能的重要指标之一。本文统计了不同方法在生成 21 张红外和可见光融合图像时的平均运行时间, 如表 3 所示。



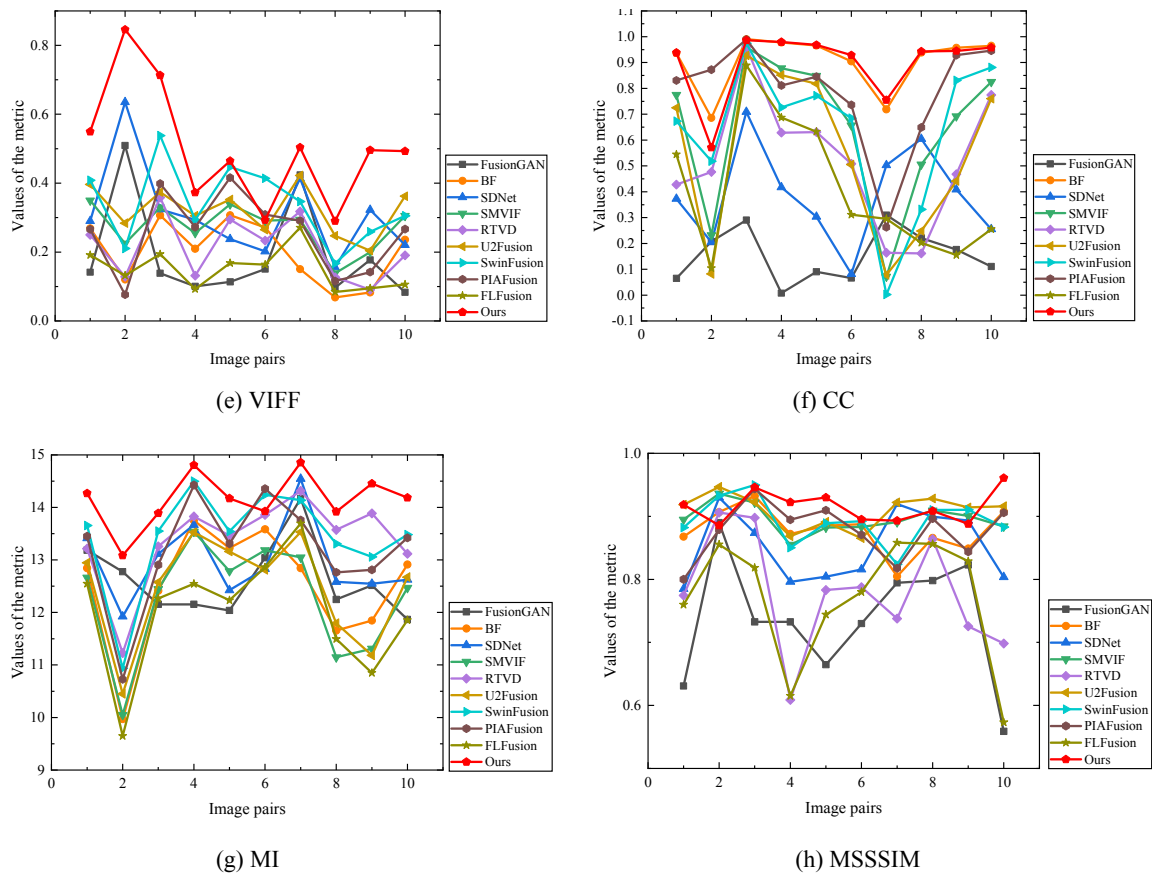


图6 10组源图像的8种客观评价指标对比

Fig.6 Comparison of eight objective evaluation metrics for 10 sets of source images

表3 不同融合算法的运行时间

Table 3 Running time of different fusion methods

Methods	FusionGAN	BF	SDNet	SMVIF	RTVD	U2Fusion	SwinFusion	PIAFusion	FLFusion	Ours
Time	1.251	1.029	0.992	0.803	3.561	2.499	48.063	1.136	1.072	1.058

从表3可以看出,本文运行时间较传统融合方法SMVIF略高一点,但通过一系列定性、定量评价以及时间性能的综合分析可以看出本文方法的性能最佳。

2.3 RoadScene数据集实验分析

为了更进一步验证本文所提方法的广泛性,挑选了RoadScene数据集集中的32组红外和可见光图像进行测试,将其与CNN^[14]、MLGCF(multi-level Gaussian curvature filtering image decomposition)^[28]、SwinFusion^[26]、CDL(multimodal image fusion via coupled feature learning)^[29]、Fractal^[30]、PCNN(pulse coupled neural network)^[31]进行对比,从中选取一组典型场景(FLIR_00006)场景作为主客观评价对比,如图7和表4所示。从中可以看出,相比于其他对比算法,本文方法可以有效融合红外信息和可见光信息,融合结果图纹理细节清晰,目标信息显著,客观评价指标更进一步证明本文方法融合图像不仅符合人眼视觉效

果,而且有六个客观评价指标优于其他对比方法。

2.4 消融实验

为了验证本文所提各模块各框架方法的有效性,设计了一系列消融实验。第一组实验将密集块换为普通卷积,网络其他结构保持不变,记为Denseblock。第二组实验将补偿注意力机制换为CBAM(convolutional block attention module)注意力机制,其他保持不变,记为CBAM。第三组实验将三支变分支为二分支且去掉加入双鉴别器的补偿注意力机制,记为Attention-generator。第四组实验将引入双鉴别器的补偿注意力机制去掉,其他网络结构保持一致,记为Attention-discriminator。第五组实验则是将全部的模块嵌入网络模型中,记为All。

从图8和表5可以看出,本文提出的融合框架在主观视觉角度和客观度量角度都优于其他4种替换或者是降级的模型框架,这充分证明本文提出的融合框

架具有高效的融合性能, 设计的三支路径、密集块、注意力机制框架的有效性。
补偿注意力机制、双鉴别器以及双鉴别器引入补偿注

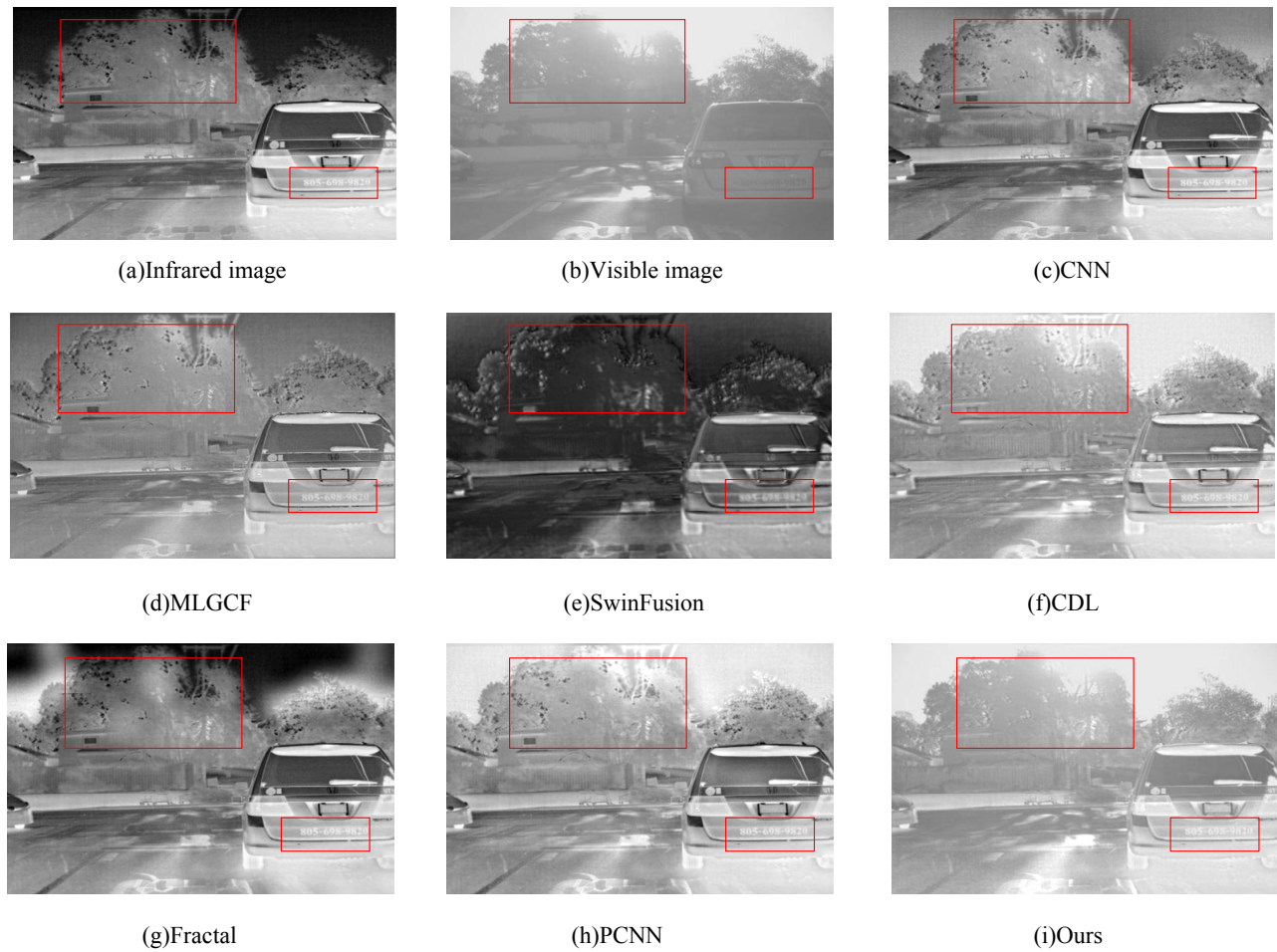


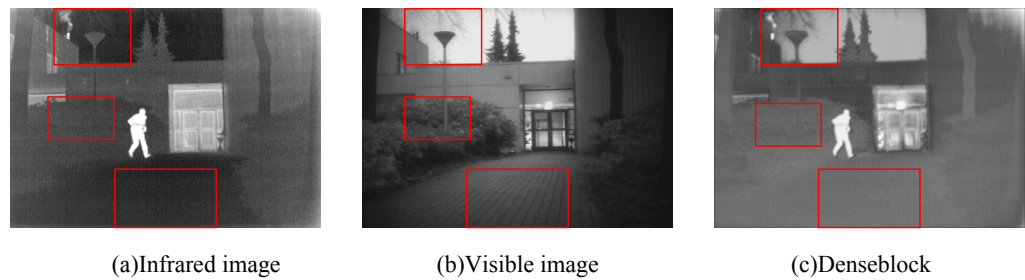
图 7 FLIR_00006 场景的主观评价对比结果

Fig.7 Comparative results of subjective evaluation of FLIR_00006 scenes

表 4 FLIR_00006 场景的客观评价对比结果

Table 4 Comparative results of the objective evaluation of the FLIR_00006 scenario

Methods	EN	AG	SD	SF	VIFF	MI	MSSSIM
CNN	6.9642	6.4636	6.5088	23.0067	0.5664	2.4576	0.9442
MLGCF	7.0043	6.5462	5.7386	23.0275	0.4041	2.7186	0.9227
SwinFusion	6.9047	5.2062	6.5814	18.5056	0.2202	2.1103	0.3850
CDL	6.9642	5.7096	6.1676	18.8730	0.2764	2.8324	0.8055
Fractal	7.7249	6.5918	6.4663	23.0894	0.6032	3.2194	0.8915
PCNN	7.2290	6.3944	6.5351	22.7765	0.4827	2.6052	0.9070
Ours	7.6876	6.8140	6.8379	23.7251	0.6675	3.2313	0.9513



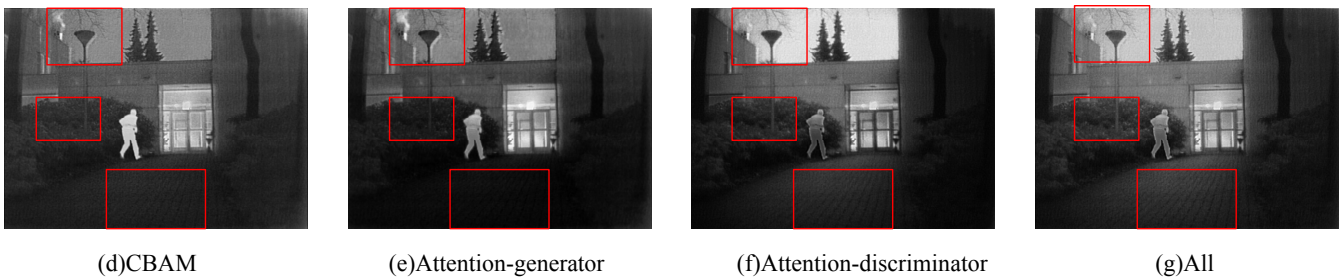


图 8 消融实验结果图

Fig.8 Results of ablation experiments

表 5 消融实验客观评价结果

Table 5 Objective evaluation results of ablation experiments

Models	EN	AG	SF	MI	SCD
Denseblock	6.3184	1.9436	6.2444	12.6369	0.8091
CBAM	6.8433	3.0356	7.9358	13.6865	1.8115
Attention-generator	6.8484	2.4442	7.0260	13.6971	1.8593
Attention-discriminator	6.9290	2.8020	7.5908	13.8580	1.6908
All	7.0877	3.1903	8.3740	14.1755	1.9091

3 结论

为了提高人眼或机器视觉特性的融合效果，本文利用密集块和补偿注意力机制构建了一个新的端到端对抗学习的红外与可见光图像融合方法。在生成器模型中，构造三支路路径，充分提取红外和可见光源图像中的像素信息和梯度信息，并且构建补偿注意力机制，构建长期依赖关系，更深层次地提取红外显著目标和可见光纹理细节。此外，设计双鉴别器以确定融合结果和源图像之间的相似分布。在此基础上引入补偿注意力模块，使得鉴别器能够更精准地从聚焦区域而非整幅图像来鉴别图像，提高鉴别能力。实验表明，本文算法的客观评价指标 EN、AG、SF、SD、VIFF、CC、MI、MSSSIM 对比其他 9 种代表性算法平均提高了 11.996%，48.804%，30.256%，46.366%，114.001%，65.906%，11.980%，8.377%，有效融合了红外和可见光图像，达到单一图像无法提供全面信息的优势。相比于现有算法，本算法在细节保留、纹理对比度提高方面具有一定的优势。

参考文献：

[1] YANG B, JIANG Z, PAN D, et al. Detail-aware near infrared and visible fusion with multi-order hyper-Laplacian priors[J]. *Information Fusion*, 2023, **99**: 101851.

[2] ZHANG X, DAI X, ZHANG X, et al. Joint principal component analysis and total variation for infrared and visible image fusion[J]. *Infrared Physics & Technology*, 2023, **128**: 104523.

[3] TANG L, XIANG X, ZHANG H, et al. DIVFusion: Darkness-free infrared and visible image fusion[J]. *Information Fusion*, 2023, **91**: 477-493.

[4] WANG Z, SHAO W, CHEN Y, et al. Infrared and visible image fusion via interactive compensatory attention adversarial learning[J]. *IEEE Transactions on Multimedia*, 2022, **25**: 1-12.

[5] 唐丽丽, 刘刚, 肖刚. 基于双路级联对抗机制的红外与可见光图像融合方法[J]. *光子学报*, 2021, **50**(9): 0910004.

TANG Lili, LIU Gang, XIAO Gang. Infrared and visible image fusion method based on dual-path cascade adversarial mechanism[J]. *Acta Photonica Sinica*, 2021, **50**(9): 0910004.

[6] 王志社, 邵文禹, 杨风暴, 等. 红外与可见光图像交互注意力生成对抗融合方法[J]. *光子学报*, 2022, **51**(4): 0410002.

WANG Zhishe, SHAO Wenyu, YANG Fengbao, et al. Infrared and visible image fusion method via interactive attention-based generative adversarial network [J]. *Acta Photonica Sinica*, 2022, **51**(4): 0410002.

[7] RAO D, XU T, WU X. Tgfuse: An infrared and visible image fusion approach based on transformer and generative adversarial network[J]. *IEEE Transactions on Image Processing*, 2023, **5**: 1-11.

[8] XIE Q, MA L, GUO Z, et al. Infrared and visible image fusion based on NSST and phase consistency adaptive DUAL channel PCNN[J]. *Infrared Physics & Technology*, 2023, **131**: 104659.

[9] 江泽涛, 蒋琦, 黄永松, 等. 基于潜在低秩表示与复合滤波的红外与弱可见光增强图像融合方法[J]. *光子学报*, 2020, **49**(4): 0410001.

JIANG Zetao, JIANG Qi, HUANG Yongsong, et al. Infrared and low-light-level visible light enhancement image fusion method based on latent low-rank representation and composite filtering [J]. *Acta Photonica Sinica*,

- 2020, **49**(4): 0410001.
- [10] WANG C, WU Y, YU Y, et al. Joint patch clustering-based adaptive dictionary and sparse representation for multi-modality image fusion [J]. *Machine Vision and Applications*, 2022, **33**(5): 69.
- [11] PANG Y, ZHAO X, ZHANG L, et al. Multi-scale interactive network for salient object detection[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 9413-9422.
- [12] CHEN J, WU K, CHENG Z, et al. A saliency-based multiscale approach for infrared and visible image fusion[J]. *Signal Processing*, 2021, **182**: 107936.
- [13] VESHKI F G, VOROBYOV S A. Coupled feature learning via structured convolutional sparse coding for multimodal image fusion[C]//*ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022: 2500-2504.
- [14] LIU Y, CHEN X, CHENG J, et al. Infrared and visible image fusion with convolutional neural networks[J]. *International Journal of Wavelets, Multiresolution and Information Processing*, 2018, **16**(3): 1850018.
- [15] Prabhakar K R, Srikanth V S, Babu R V. DeepFuse: a deep unsupervised approach for exposure fusion with extreme exposure image pairs[C]//*Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017: 4714-4722.
- [16] MA J, YU W, LIANG P, et al. FusionGAN: A generative adversarial network for infrared and visible image fusion[J]. *Information Fusion*, 2019, **48**: 11-26.
- [17] ZHANG H, YUAN J, TIAN X, Ma J. GAN-FM: Infrared and visible image fusion using GAN with full-scale skip connection and dual Markovian discriminators[J]. *IEEE Transactions on Computational Imaging*, 2021, **7**: 1134-1147.
- [18] SUN X, HU S, MA X, et al. IMGAN: Infrared and visible image fusion using a novel intensity masking generative adversarial network[J]. *Infrared Physics & Technology*, 2022, **125**: 104221.
- [19] LE Z, HUANG J, XU H, et al. UIFGAN: An unsupervised continual-learning generative adversarial network for unified image fusion[J]. *Information Fusion*, 2022, **88**: 305-318.
- [20] Toet Alexander. The TNO Multiband image data collection[J]. *Data in Brief*, 2017, **15**: 249-251.
- [21] ZHANG H, MA J. SDNet: a versatile squeeze-and-decomposition network for real-time image fusion[J]. *International Journal of Computer Vision*, 2021, **129**: 2761-2785.
- [22] ZHAO Z, XU S, ZHANG C, et al. Bayesian fusion for infrared and visible images[J]. *Signal Processing*, 2020, **177**: 107734.
- [23] CHEN J, LI X, WU K. Infrared and visible image fusion based on relative total variation decomposition[J]. *Infrared Physics & Technology*, 2022, **123**: 104112.
- [24] XU H, MA J, JIANG J, et al. U2Fusion: a unified unsupervised image fusion network [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, **44**(1): 502-518.
- [25] MA J, TANG L, FAN F, et al. SwinFusion: cross-domain long-range learning for general image fusion via swin transformer [J]. *IEEE/CAA Journal of Automatica Sinica*, 2022, **9**(7): 1200-1217.
- [26] TANG L, YUAN J, ZHANG H, et al. PIAFusion: a progressive infrared and visible image fusion network based on illumination aware [J]. *Information Fusion*, 2022, **83**: 79-92.
- [27] XUE W, WANG A, ZHAO L. FLFuse-Net: A fast and lightweight infrared and visible image fusion network via feature flow and edge compensation for salient information [J]. *Infrared Physics & Technology*, 2022, **127**: 104383.
- [28] TAN W, ZHOU H, SONG J, et al. Infrared and visible image perceptive fusion through multi-level Gaussian curvature filtering image decomposition [J]. *Applied optics*, 2019, **58**(12): 3064-3073.
- [29] VESHKI FG, OUZIR N, VOROBYOV SA, et al. Multimodal image fusion via coupled feature learning[J]. *Signal Processing*, 2022, **200**: 108637.
- [30] Panigrahy C, Seal A, Mahato N K. Fractal dimension based parameter adaptive dual channel PCNN for multi-focus image fusion[J]. *Optics and Lasers in Engineering*, 2020, **133**: 106141.
- [31] Panigrahy C, Seal A, Mahato N K. Parameter adaptive unit-linking dual-channel PCNN based infrared and visible image fusion[J]. *Neurocomputing*, 2022, **514**: 21-38.