

控制与决策

Control and Decision

改进残差密集生成对抗网络的红外与可见光图像融合

闵莉, 曹思健, 赵怀慈, 刘鹏飞, 郇炳昌

引用本文:

闵莉, 曹思健, 赵怀慈, 刘鹏飞, 郇炳昌. 改进残差密集生成对抗网络的红外与可见光图像融合[J]. *控制与决策*, 2023, 38(3): 721–728.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2021.1580>

您可能感兴趣的其他文章

Articles you may be interested in

[基于内生长机制和卷积稀疏表示的红外与可见光图像融合](#)

Infrared and visible light image fusion based on internal generative mechanism and convolution sparse representation
控制与决策. 2022, 37(1): 167–174 <https://doi.org/10.13195/j.kzyjc.2020.1080>

[基于多粒度特征融合的边缘一致性图像补全](#)

Edge consistent image completion based on multi-granularity feature fusion
控制与决策. 2022, 37(12): 3240–3250 <https://doi.org/10.13195/j.kzyjc.2021.0665>

[基于FRC框架的红外与可见光图像融合方法](#)

Infrared and visible image fusion based on FRC algorithm
控制与决策. 2021, 36(11): 2690–2698 <https://doi.org/10.13195/j.kzyjc.2020.0669>

[基于混合模型驱动的红外与可见光图像融合](#)

Infrared and visible image fusion based on hybrid model driving
控制与决策. 2021, 36(9): 2143–2151 <https://doi.org/10.13195/j.kzyjc.2019.1749>

[基于自注意力生成对抗网络的图像超分辨率重建](#)

Image super-resolution reconstruction based on self-attention GAN
控制与决策. 2021, 36(6): 1324–1332 <https://doi.org/10.13195/j.kzyjc.2019.1290>

改进残差密集生成对抗网络的红外与可见光图像融合

闵莉¹, 曹思健¹, 赵怀慈^{2†}, 刘鹏飞², 邵炳昌³

1. 沈阳建筑大学 机械工程学院, 沈阳 110168;
2. 中国科学院 沈阳自动化研究所光电信息处理重点实验室, 沈阳 110169;
3. 卓越新时代认证有限公司, 沈阳 110013)

摘要: 基于深度学习的红外与可见光图像融合算法通常无法感知源图像显著性区域, 导致融合结果没有突出红外与可见光图像各自的典型特征, 无法达到理想的融合效果. 针对上述问题, 设计一种适用于红外与可见光图像融合任务的改进残差密集生成对抗网络结构. 首先, 使用改进残差密集模块作为基础网络分别构建生成器与判别器, 并引入基于注意力机制的挤压激励网络来捕获通道维度下的显著特征, 充分保留红外图像的热辐射信息和可见光图像的纹理细节信息; 其次, 使用相对平均判别器, 分别衡量融合图像与红外图像、可见光图像之间的相对差异, 并根据差异指导生成器保留缺少的源图像信息; 最后, 在 TNO 等多个图像融合数据集上进行实验, 结果表明所提方法能够生成目标清晰、细节丰富的融合图像, 相比基于残差网络的融合方法, 边缘强度和平均梯度分别提升了 64.56% 和 64.94%.

关键词: 图像融合; 残差密集块; 生成对抗网络; 注意力机制; 显著性区域; 相对平均判别器

中图分类号: TP391

文献标志码: A

DOI: 10.13195/j.kzyjc.2021.1580

开放科学(资源服务)标识码(OSID):



引用格式: 闵莉, 曹思健, 赵怀慈, 等. 改进残差密集生成对抗网络的红外与可见光图像融合[J]. 控制与决策, 2023, 38(3): 721-728.

Infrared and visible image fusion with improved residual dense generative adversarial network

MIN Li¹, CAO Si-jian¹, ZHAO Huai-ci^{2†}, LIU Peng-fei², TAI Bing-chang³

1. School of Mechanical Engineering, Shenyang Jianzhu University, Shenyang 110168, China;
2. Key Laboratory of Optical-Electronics Information Processing, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110169, China;
3. Excellence Xin Shi Dai Certification Co., Ltd, Shenyang 110013, China)

Abstract: The current infrared and visible image fusion methods based on deep learning usually cannot perceive the saliency region of source images. As a result, the result of image fusion fails to highlight the respective typical features of infrared and visible images, and the ideal fusion effect cannot be reached. In response to these issues, an improved residual dense generative adversarial network structure suitable for infrared and visible image fusion tasks is designed. First of all, the improved residual dense block is used as the basic network to construct the generator and the discriminator respectively; the squeeze and excitation network based on attention mechanism is introduced to capture the salient features in the channel dimension, which will preserve the thermal radiation information of infrared images and the textures information of visible images adequately. Then, the relativistic average discriminator is used to measure the relative average difference between the fusion image and two source images respectively, so as to instruct the generator to preserve the missing information of source images based on the difference. Finally, the experimental results on multiple image fusion datasets such as the TNO, prove that the proposed method can generate fused images with clear targets and rich details. Compared with the fusion method based on the residual network, the edge intensity and average gradient are increased by 64.56% and 64.94%, respectively.

Keywords: image fusion; residual dense block; generative adversarial network; attention mechanism; salient region; relativistic average discriminator

0 引言

红外图像能够突出热红外目标显著特征, 可以抵御极端环境的影响, 但图像分辨率低且缺少纹理细

节; 可见光图像符合人眼视觉特性, 分辨率高且纹理细节丰富, 但在恶劣天气条件下会丢失场景信息. 融合两种图像的优势可以全面地反映同一场景信息, 已

收稿日期: 2021-09-10; 录用日期: 2021-11-26.

基金项目: 装备预研重点基金项目(41401040105).

责任编辑: 程龙.

†通讯作者. E-mail: hczhao@sia.cn.

应用于军事侦察、遥感探测和目标识别等领域^[1]。

目前,红外与可见光图像融合方法可以分为传统方法和深度学习方法。传统方法通常在空间域或变换域对图像进行处理^[2],如多尺度变换^[3]和稀疏表示,但这些传统方法通常需要手动设计复杂的图像分解和融合规则,融合算法整体效率不高。基于深度学习的图像融合方法,利用神经网络强大的特征学习能力对图像进行特征提取和融合,具有更高的效率和融合精度。如Zhang等^[4]构建了一种基于卷积神经网络的通用图像融合框架(image fusion framework based on convolutional neural networks, IFCNN),并与变换域算法相结合,使网络模型具有更强的泛化能力;Li等^[5]提出一种融合结构性约束的卷积神经网络,有效消除了融合图像的视觉伪影。但以上基于CNN的方法仍未摆脱手动设计复杂融合规则的限制。

为了实现从源图像到融合图像的端到端映射,Ma等^[6]将生成对抗网络(generative adversarial networks, GAN)^[7]引入到红外与可见光图像融合任务,生成器负责融合红外图像的像素强度和附加的可见光图像的梯度,判别器负责区分融合图像与可见光图像,使融合图像在对抗博弈中保留足够的梯度信息;在此基础上,Fu等^[8]在生成器中利用跳跃连接融合可见光图像的梯度信息,进一步增强了融合图像的纹理细节。但这两种算法无法有效感知源图像显著性区域,前者缺失了背景区域的纹理细节,后者红外目标对比度较弱,生成的融合图像均无法同时突出两种源图像的典型特征。

本文针对上述问题,提出一种基于注意力机制和残差密集生成对抗网络的红外与可见光图像融合算法。首先,将基于特征通道的注意力机制集成到残差密集块中,引导生成器和判别器聚焦于源图像的显著性区域,同时捕获红外图像的前景目标信息和可见光图像的背景信息;其次,构建双路相对平均判别器结构,分别计算融合图像与红外图像在像素强度方面、融合图像与可见光图像在梯度方面的偏移量,有针对性地指导生成器融合两种源图像信息,获得更优的图像融合精度和视觉效果。

1 基本原理

1.1 生成对抗网络

GAN是继CNN之后另一种流行的深度学习模型,主要思想是在生成器(generator, G)和判别器(discriminator, D)之间建立一个极小极大的博弈。在基于GAN的红外与可见光图像融合任务中,生成器负责最小化融合图像(P_G)与源图像(P_{data})之间的数

据分布差异,判别器负责最大限度地区分两种图像。 G 与 D 的对抗关系由下式给出:

$$\min_G \max_D V_{GAN}(G, D) = E_{x \sim P_{data}} [\log D(x)] + E_{x \sim P_G} [\log(1 - D(G(x)))], \quad (1)$$

其中 E 代表期望。

1.2 残差密集连接网络

Radford等^[9]提出将卷积神经网络作为基础结构,搭建生成对抗网络的生成器和判别器,利用CNN强大的特征提取能力有效提升了GAN的效果。但传统CNN随着网络深度增加会出现网络退化问题,中间层提取的信息未得到充分利用。

针对上述问题,残差网络(residual network, ResNet)^[10]通过建立残差连接能够有效增强信息特征流和融合局部特征的能力;在密集连接网络(densely connected convolutional network, DenseNet)^[11]中,每层都与通道维度中的所有先前层进行连接,作为下一层的输入,能够进一步加强学习局部特征的能力。Zhang等^[12]提出残差密集连接网络(residual densely network, RDN)同时利用残差学习和密集连接,通过联合增强信息流、学习残差特征和改进局部特征融合提高特征学习能力。为了充分提取源图像重要特征信息,本文提出一种融合注意力机制改进的残差密集块(residual dense attention block, RDAB),结构如图1所示。

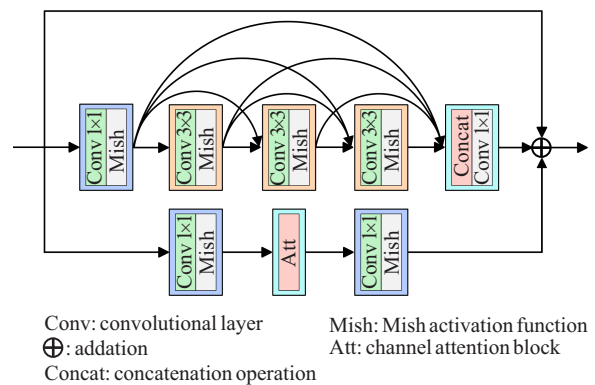


图1 残差密集注意力块

每个RDAB包含3个分支,保证能够充分利用从每个卷积层提取的有效特征。分支的第1层卷积使用 1×1 卷积核生成低维特征映射,降低模块内部的卷积计算成本;然后经过3层卷积核为 3×3 的密集连接层,充分提取局部特征;最后一层由 1×1 卷积核对通道相连后的特征图作降维处理,实现跨通道信息交互。为了充分利用残差密集模块在提取图像特征方面的优势,避免不加区分地集成特征造成的信息冗余,引入

基于通道的注意力模块(channel attention block, Att)感知源图像显著性区域,过滤特征信息.

在模块内使用 Mish 激活函数^[13]代替常见的 ReLU 激活函数,提供更高的准确率和稳定性,公式为

$$\text{Mish}(x) = x \cdot \tanh(\ln(1 + e^x)), \quad (2)$$

其中 tanh 代表 Tanh 激活函数.

1.3 注意力机制

注意力机制符合人类的视觉系统,能够根据像素

的重要程度赋予它们不同的权重,并以此为依据指导网络模型保留高权重的显著性特征,抑制低权重的非典型特征.

引入挤压激励网络 (squeeze-and-excitation networks, SENet)^[14]构建基于通道的注意力模块,通过挤压 (squeeze, F_{sq})、激励 (excitation, F_{ex}) 和特征重标定 (scale, F_{scale}), 显式地建模通道间的相互依赖关系,筛选出针对特征通道的注意力,指导网络提取源图像显著性特征. 模块的网络结构如图2所示.

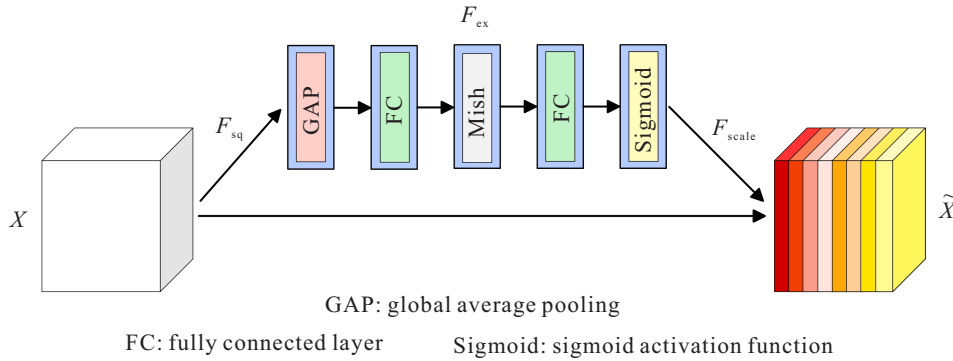


图2 通道注意力块

首先,通过全局平均池化在空间维度压缩特征,获取全局感受野;然后,使用包含 Mish 激活函数的两级非线性全连接层 (fully connected layer, FC) 为每个特征通道生成权重,并由 Sigmoid 激活函数对权重做归一化处理;最后,将权重逐通道加权到先前的特征 X 上,输出结果 \tilde{X} ,完成对重要特征的增强和对非重要特征的削弱.

模块产生的注意力可视化效果如图3所示,左侧图为红外图像,右侧图为可见光图像.对于融合任务关注的显著性区域,如红外图像的目标和可见光图像的典型纹理细节,赋予更高的权重,以高亮颜色标记,网络模型也将重点提取和融合这两部分特征信息;而红外图像的背景和可见光图像的目标属于非典型特征,被赋予低权重并受到抑制.

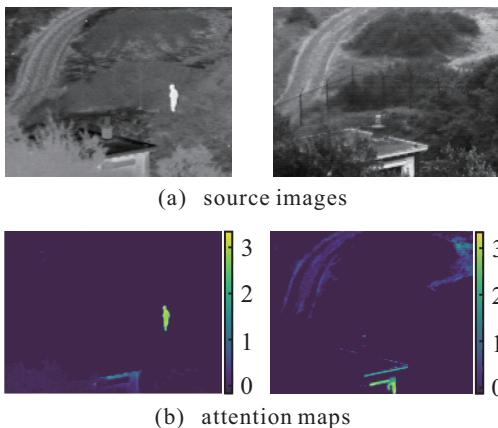


图3 注意力可视化效果

2 本文方法

2.1 融合框架概述

算法框架如图4所示,包含一个生成器和两个判别器.红外图像与可见光图像通过生成器 generator 完成融合任务.判别器 discriminator_ir 负责区分红外图像与融合图像,与生成器进行对抗,约束融合图像保留红外图像的热辐射信息;同理,判别器 discriminator_vis 负责约束融合图像保留可见光图像的纹理细节信息.

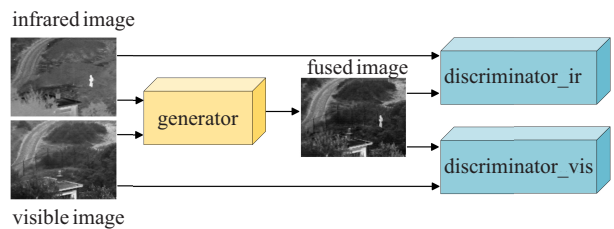


图4 算法框架

采用伪代码形式表示网络模型的具体训练过程,如算法1所示.保证生成器输出高质量融合图像的前提是:判别器具有准确区分融合图像与源图像的能力.因此,采用先训练双判别器 k 次、再训练生成器 1 次的分批循环训练形式,持续迭代到最大训练次数 M .训练过程中采用优化器 RMSProp 更新模型参数.

算法1 网络模型的训练过程.

input: 红外图像 I_i , 可见光图像 I_v ;

output: 融合图像 I_f .

step 1: for M epochs do
 step 2: for k steps do
 step 3: 从红外图像分布中采样 n 个样本 $\{I_i^1, I_i^2, \dots, I_i^n\}$;
 step 4: 从融合图像分布中采样 n 个样本 $\{I_f^1, I_f^2, \dots, I_f^n\}$;
 step 5: 训练判别器 discriminator_ir , 并更新模型参数;
 step 6: 从可见光图像分布中采样 n 个样本 $\{I_v^1, I_v^2, \dots, I_v^n\}$;
 step 7: 训练判别器 discriminator_vis , 并更新模型参数;

step 8: end for
 step 9: 训练生成器 generator , 并更新模型参数;
 step 10: end for

2.2 生成器

生成器采用编码器-解码器结构实现特征提取、融合和重构, 网络结构如图5所示. 考虑到红外与可见光图像的不同特性, 在特征提取阶段建立双分支结构. 两路分支使用架构相同、参数独立的网络结构, 能够有效降低模型复杂度; 结合设计的混合损失函数, 保证从源图像中充分提取出显著目标特征和纹理细节特征.

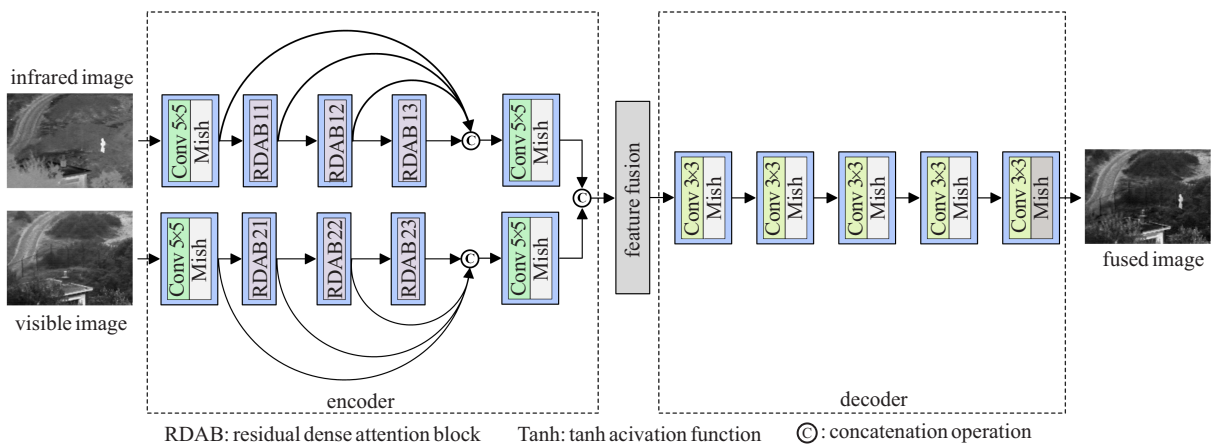


图5 生成器网络结构

特征提取网络由两个普通卷积层和3个RDAB组成. 第1层卷积使用 5×5 卷积核提取输入图像的浅层特征; 为了避免丢失源图像重要特征和进一步提高网络框架的图像融合效果, 使用连续的3层RDAB结构提取图像的深层显著特征; 浅层与深层的特征映射通道数均为64, 以便完成最后一层的全局特征融合.

将双分支的全局特征通道相连后, 通过5层卷积的解码器从融合特征中重构出融合图像. 在网络结构中统一删除了批量归一化层 (batch normalization, BN), 因为BN层通常会引入视觉伪影并限制网络的泛化能力, 而且删除BN层也能有效降低计算复杂性和显存占用.

2.3 判别器

在基于GAN的图像融合任务中, 判别器一般只是用于估计输入数据为真实样本的概率, 即定义源图像为真、融合图像为假, 未能有效指出融合图像缺少的具体源图像信息. 因此, 引入相对平均判别器^[15], 通过建立判别器负责最大化源图像与融合图像的平均差距、生成器负责最小化平均差距的对抗博弈, 预

测源图像比融合图像相对更真实的概率, 从而有针对性地指导融合图像保留源图像信息.

在判别器 discriminator_ir 中, 通过将红外图像的 batch 均值与融合图像的 batch 均值作差, 再与 Sigmoid 函数相乘映射到 0-1 之间, 得出二者的平均差距 $D_{\text{Ra1}}(I_i, I_f)$ 为

$$D_{\text{Ra1}}(I_i, I_f) = \sigma\{C(I_i) - E[C(I_f)]\}. \quad (3)$$

其中: I_i 代表红外图像, I_f 代表融合图像, σ 代表 Sigmoid 激活函数, C 代表非线性变换.

类似地, 在 discriminator_vis 中, 可见光图像与融合图像的平均差距 $D_{\text{Ra2}}(I_v, I_f)$ 可表示为

$$D_{\text{Ra2}}(I_v, I_f) = \sigma\{C(I_v) - E[C(I_f)]\}, \quad (4)$$

其中 I_v 代表可见光图像.

相对平均判别器的优势在于能够指出融合图像与源图像之间的相对差异, 生成器依据差异, 不断地提取和融合缺失的源图像信息.

使用基于RDAB的网络结构搭建双判别器, 双判别器的网络结构相同但参数独立, 达到区分两种源图像与融合图像的目的. 网络结构如图6所示. 相比于

简单的卷积层,使用更复杂的网络结构能够提升判别器的能力,使判别器更多地关注融合图像与源图像之间的细节和结构差异,并在对抗博弈中提高生成器输出的样本质量^[16].

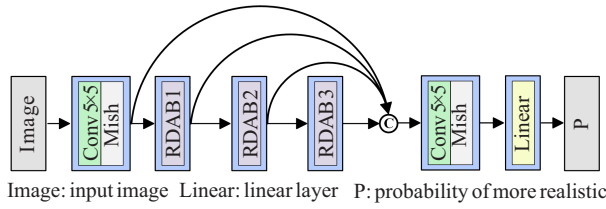


图6 判别器网络结构

2.4 损失函数

生成器损失函数由3部分构成,包括对抗损失 L_{adv} 、结构相似性损失 L_{ssim} 和感知损失 L_{per} ,表示为

$$L_G = L_{adv} + \lambda_1 L_{ssim} + \lambda_2 L_{per}, \quad (5)$$

其中 λ_1 和 λ_2 代表损失函数的权重系数。

对抗损失是GAN的核心,代表通过对抗博弈使融合图像特征分布不断逼近源图像特征分布,公式为

$$L_{adv_i}^{Ra} = -E_{I_i}[\log(1 - D_{Ra1}(I_i, I_f))] - E_{I_f}[\log(D_{Ra1}(I_f, I_i))], \quad (6)$$

$$L_{adv_v}^{Ra} = -E_{I_v}[\log(1 - D_{Ra2}(I_v, I_f))] - E_{I_f}[\log(D_{Ra2}(I_f, I_v))], \quad (7)$$

其中 E_{I_i} 、 E_{I_v} 和 E_{I_f} 分别代表对batch中红外图像、可见光图像和融合图像取均值。

通过结构相似性损失从亮度、对比度和结构方面约束融合图像与源图像的相似性,有

$$L_{ssim} = E[(1 - S_{I_f, I_i}) + (1 - S_{I_f, I_v})], \quad (8)$$

其中 $S_{x,y}$ 代表融合图像与源图像之间的SSIM值。

GAN本身训练并不稳定,在缺少像素级约束的情况下,融合图像会产生视觉伪影。因此,增加使用预训练的VGG16构建的感知损失,提升融合图像与源图像在像素级和语义级的相似性^[17],有

$$L_{per} = E[\|\phi(f) - \phi(i)\|_2^2 + \|\phi(f) - \phi(v)\|_2^2], \quad (9)$$

其中 $\phi(x)$ 代表从源图像和融合图像提取的特征图。

3 实验与结果分析

3.1 实验设置

选取TNO、RoadScene和INO三个典型的红外与可见光图像融合数据集,在GeForce RTX 2080Ti 11 GB显卡上使用Tensorflow深度学习框架进行实验。上述数据集中存在部分相同背景或噪声较多的图像,会影响算法性能的准确评估。因此,从3个数据集中各筛选出40对红外与可见光图像,并统一裁剪

为 $84 \text{ pixel} \times 84 \text{ pixel}$ 的图像块组成训练集,再各选取20对图像作为测试集。epoch为1 600, batch size为12,学习率为0.002。参数 $\lambda_1 = 0.2$, $\lambda_2 = 0.001$ 。

3.2 对比实验

对比实验中选取了6种典型的图像融合方法,分别是FusionGAN方法、多分类约束(GAN with multiclassification constraints, GANMcC)方法^[18]、双分支网络(dual-branch network, DBN)方法^[19]、IFCNN方法、梯度和强度比例保持(proportional maintenance of gradient and intensity, PMGI)方法^[20]以及残差融合网络(residual fusion network, RFN-Nest)方法^[21]。从3个测试集中选取的部分图像融合结果如图7所示,部分结果可能存在偶然性。

充分对比图像中截取标记的突出红外目标和典型可见光纹理细节,并结合各算法原理进行深入分析,可以得出以下结论:FusionGAN方法使用单一的对抗架构,未能充分提取可见光图像信息,导致融合图像背景区域缺失关键纹理细节;GANMcC方法使用多分类约束在融合图像中平衡同时估计两种源图像的特征分布,但缺少对图像中非显著特征的抑制,未突出融合图像的边缘和纹理信息,背景区域的树木边缘模糊;DBN方法构建的双分支自动编码器结构缺少对非典型特征的抑制,使融合图像信息冗余,人物目标的对比度几乎与背景相同,未能突出红外目标;IFCNN方法使用的卷积层较浅,在提取特征过程中不可避免地丢失了源图像重要信息,输出的融合图像存在红外图像的噪声干扰,导致清晰度较差;PMGI方法基于均方误差损失构建了像素强度和梯度约束,缺少了感知损失,使融合图像趋于平滑,视觉效果较差;RFN-Nest方法使用的自动编码器更注重纹理细节,导致融合图像红外目标的对比度较低。本文算法通过注意力机制有效引导残差密集生成对抗网络,充分提取和融合了源图像典型特征,生成的图像具有更显著的目标对比度和背景纹理细节,在图像融合精度和人眼舒适性两方面均优于其他对比算法。

除了主观方面的视觉效果外,本文选取边缘强度(edge intensity, EI)、平均梯度(average gradient, AG)、空间频率(spatial frequency, SF)、互信息量(mutual information, MI)、标准差(standard deviation, SD)和视觉保真度(visual information fidelity, VIF)作为客观指标。定量衡量各算法在保持边缘、纹理细节、源图像重要特征信息、融合质量和视觉效果方面的性能^[22]。上述客观指标均为正向指标,评价结果如表1所示。

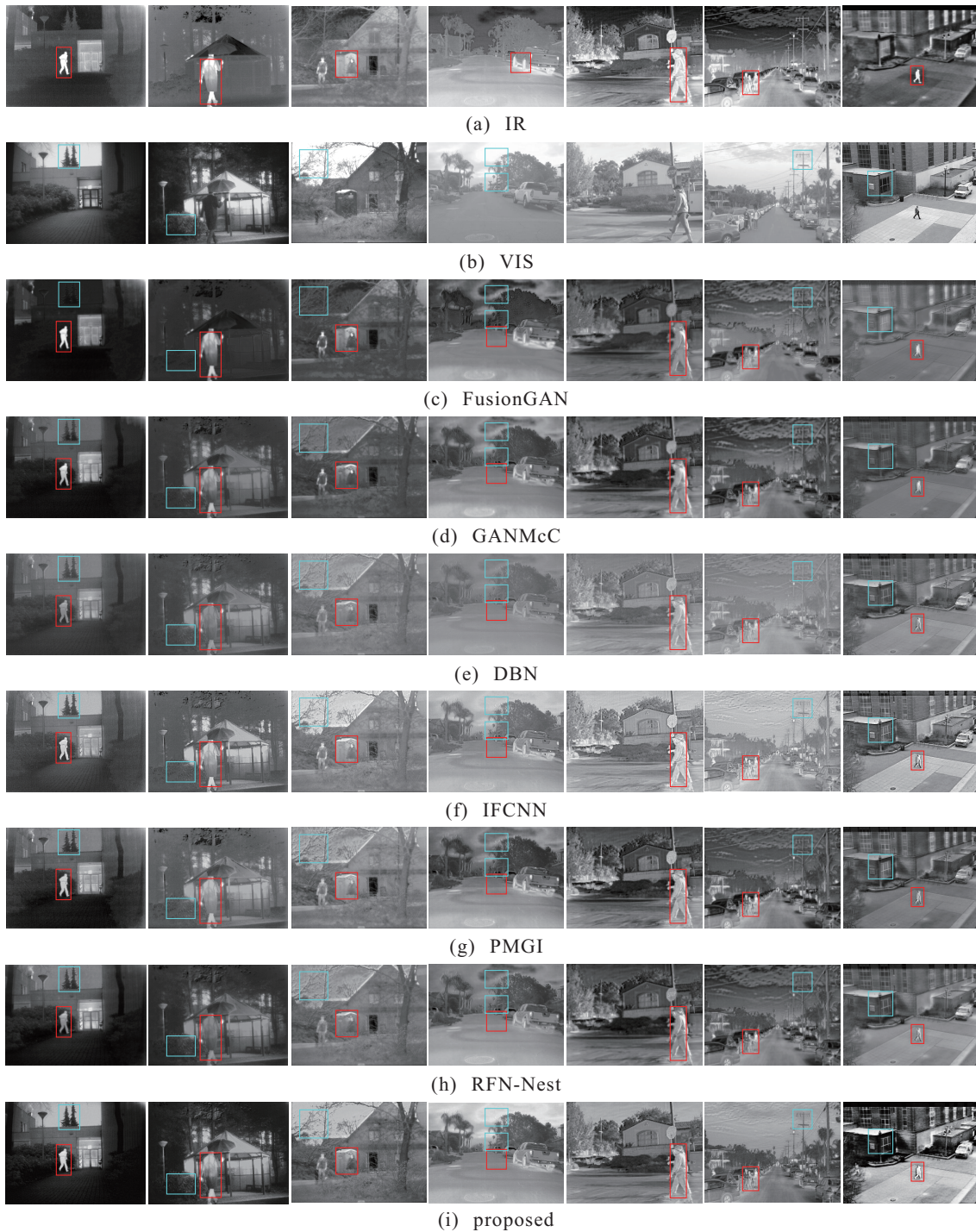


图7 对比实验结果

本文方法在不同数据集的6项指标中均取得最优结果,与主观评价结果相吻合.其中,在TNO数据集中,相比于RFN-Nest方法,指标EI、AG和SF提升明显,分别提升了64.56%、64.94%和83.71%,其他3项指标也具有2.78%、24.07%和27.78%的提升幅度.客观指标的评价结果表明了所提出的改进融合方法能够显著提升融合图像的边缘纹理细节和清晰度,并保留源图像的显著特征,且融合图像符合人类的视觉系统.

此外,在测试集上对比了各算法的平均运行时间,结果如表2所示.与其他算法相比,本文方法需要的图像融合时间更长.造成这种结果的原因主要在于算法网络结构复杂程度的不同.前5种算法的网络模型均使用基础的卷积模块,RFN-Nest算法的基础结构由残差模块构成,而本文算法为了充分提取源图像的深层次特征,使用了更复杂的改进残差密集模块,增加了计算时间;模块内包含实现密集连接的通道相连操作,对显存的负担较大,会降低计算效率.

表1 对比实验客观评价结果

数据集	指标	方法						
		FusionGAN	GANMcC	DBN	IFCNN	PMGI	RFN-Nest	proposed
TNO	EI	25.78	27.81	27.17	49.79	40.33	31.86	52.43
	AG	2.57	2.73	2.68	5.09	3.94	3.28	5.41
	SF	6.36	6.43	6.24	12.52	9.24	7.49	13.76
	MI	13.01	13.45	12.77	13.52	14.11	14.03	14.42
	SD	28.96	32.85	24.52	33.89	37.32	37.56	46.60
	VIF	0.52	0.61	0.30	0.62	0.89	0.72	0.92
RoadScene	EI	37.49	40.42	37.37	54.35	52.72	42.30	56.00
	AG	3.53	3.76	3.59	5.27	4.94	3.89	5.53
	SF	8.67	8.60	8.83	12.10	11.52	8.75	13.52
	MI	13.98	14.28	13.02	13.91	14.52	14.34	14.86
	SD	36.05	39.96	25.49	34.16	42.80	40.88	47.42
	VIF	0.36	0.50	0.26	0.52	0.58	0.52	0.62
INO	EI	35.87	37.01	39.34	56.62	49.62	43.45	58.81
	AG	3.40	3.46	3.73	5.44	4.66	3.98	5.76
	SF	9.98	9.49	10.85	14.24	12.81	10.47	15.24
	MI	13.26	13.85	13.37	14.00	14.06	14.32	14.57
	SD	30.49	38.07	32.43	41.71	39.47	44.28	52.60
	VIF	0.41	0.57	0.41	0.69	0.67	0.66	0.87

表2 各算法的平均运行时间

	FusionGAN	GANMcC	DBN	IFCNN	PMGI	RFN-Nest	proposed
mean/s	0.55	0.90	0.04	0.02	0.72	0.26	1.16

3.3 消融实验

为了进一步验证算法改进的有效性,从残差密集块和注意力机制两方面进行消融实验.实验共设置4组,包括:1)生成器和判别器使用普通卷积且无注意力机制($G_{CNN} + D_{CNN}$);2)生成器和判别器分别使用残差密集块和普通卷积且均无注意力机制($G_{RDB} + D_{CNN}$);3)生成器和判别器同时使用无注意力机制的残差密集块($G_{RDB} + D_{RDB}$);4)生成器和判别器同时使用融合注意力机制的残差密集块($G_{RDAB} + D_{RDAB}$).实验结果如图8所示.

第1)组实验因传统CNN的网络退化问题,融合图像仅能保留红外图像的部分热辐射信息;第2)组实验在生成器中引入残差密集块后,有效避免了中间层特征的丢失,在背景区域增强了纹理细节,但红外目标边缘模糊,原因是判别器网络结构简单,判别能力较弱,导致生成器未得到充分训练;第3)组实验在判别器中引入残差密集块后,融合图像目标边缘清晰,表明复杂结构的判别器具有更强的判别能力,能从细节层面提升生成器输出的融合图像质量,但融合图像整体清晰度仍有待提升;第4)组实验引入了注意力机制,同时关注红外图像的目标区域和可见光图像的背景区域,有效引导网络融合源图像中显著的跨模态特征,使融合图像整体近似于具有高对比度目标的可见光图像,综合表现最好,表明所作改进有效.

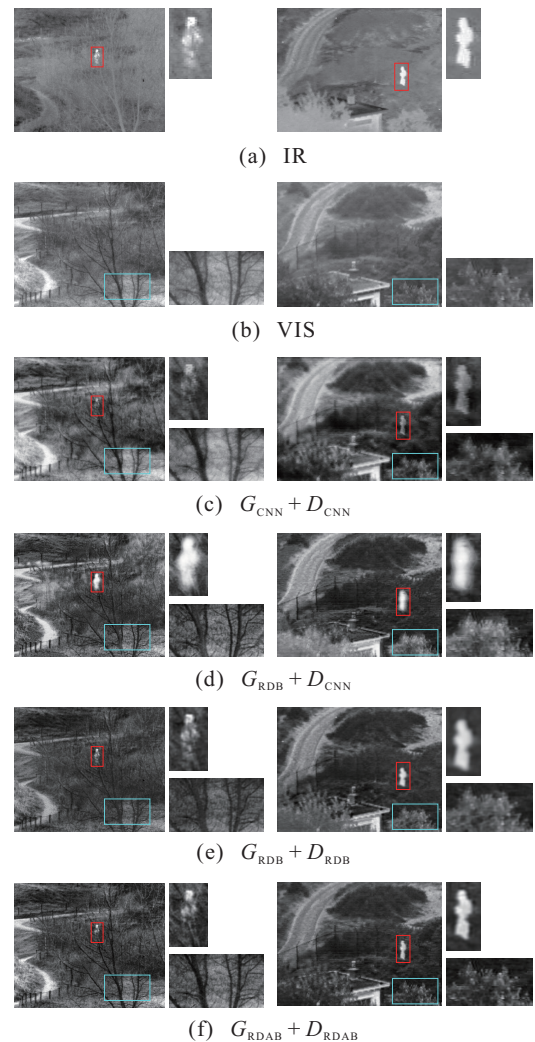


图8 消融实验结果

4 结论

本文提出了一种改进的端到端的红外与可见光图像融合方法,简化了算法设计流程.基于残差密集注意力模块构建网络基础结构,使生成器和判别器聚焦于红外图像的前景目标和可见光图像的局部细节特征;使用相对平均判别器来判断融合图像的相对真实性,强化判别器能力的同时又提升了生成器的图像融合性能;通过在多个公开数据集上进行的对比和消融实验,表明了改进后的方法具有更强的图像融合性能.所提算法适用于离线双模图像融合系统等一系列对图像融合精度要求较高、对处理时间不敏感的大型设备应用场景,但在数据的实时处理能力方面存在不足.未来将通过引入基于自注意力机制的transformer改进GAN网络基础结构,以期能够在增强性能的同时降低模型复杂度,支持并行化的优势又可以有效降低显存负担、提升效率.

参考文献(References)

- [1] Dai L Y, Liu G, Xiao G, et al. Infrared and visible image fusion based on FRC algorithm[J]. Control and Decision, 2021, 36(11): 2690-2698.
- [2] Feng X. Infrared and visible light image fusion based on internal generative mechanism and convolution sparse representation[J]. Control and Decision, 2022, 37(1): 167-174.
- [3] Shen Y, Chen X P, Liu C, et al. Infrared and visible image fusion based on hybrid model driving[J]. Control and Decision, 2021, 36(9): 2143-2151.
- [4] Zhang Y, Liu Y, Sun P, et al. IFCNN: A general image fusion framework based on convolutional neural network[J]. Information Fusion, 2020, 54: 99-118.
- [5] Li L, Xia Z Q, Han H J, et al. Infrared and visible image fusion using a shallow CNN and structural similarity constraint[J]. IET Image Processing, 2020, 14(14): 3562-3571.
- [6] 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.
- [7] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [8] Fu Y, Wu X J, Durrani T. Image fusion based on generative adversarial network consistent with perception[J]. Information Fusion, 2021, 72: 110-125.
- [9] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J/OL]. 2015, arXiv: 1511.06434
- [10] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 770-778.
- [11] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 2261-2269.
- [12] Zhang Y L, Tian Y P, Kong Y, et al. Residual dense network for image super-resolution[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 2472-2481.
- [13] Misra D. Mish: A self regularized non-monotonic neural activation function[J/OL]. 2019, arXiv: 1908.08681.
- [14] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 7132-7141.
- [15] Jolicoeur-Martineau A. The relativistic discriminator: A key element missing from standard GAN[J/OL]. 2018, arXiv: 1807.00734.
- [16] Schönfeld E, Schiele B, Khoreva A. A U-net based discriminator for generative adversarial networks[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020: 8204-8213.
- [17] Xu H, Ma J Y, Le Z L, et al. FusionDN: A unified densely connected network for image fusion[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12484-12491.
- [18] Ma J Y, Zhang H, Shao Z F, et al. GANMcC: A generative adversarial network with multiclassification constraints for infrared and visible image fusion[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-14.
- [19] Fu Y, Wu X J. A dual-branch network for infrared and visible image fusion[C]. The 25th International Conference on Pattern Recognition (ICPR). Milan, 2021: 10675-10680.
- [20] 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.
- [21] Li H, Wu X J, Kittler J. RFN-Nest: An end-to-end residual fusion network for infrared and visible images[J]. Information Fusion, 2021, 73: 72-86.
- [22] Zhang X C, Ye P, Xiao G. VIFB: A visible and infrared image fusion benchmark[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, 2020: 468-478.

作者简介

闵莉(1974—),女,副教授,博士,从事图像处理与机器视觉等研究, E-mail: mlsia@126.com;

曹思健(1995—),男,硕士生,从事图像融合与图像配准的研究, E-mail: cao_sijian@126.com;

赵怀慈(1974—),男,研究员,博士生导师,从事复杂系统建模与仿真技术等研究, E-mail: hczhao@sia.cn;

刘鹏飞(1988—),男,助理研究员,博士,从事图像处理与图像仿真等研究, E-mail: liupengfei@sia.cn;

邵炳昌(1964—),男,高级工程师,博士,从事航空装备等研究, E-mail: tbc9981@163.com.