改进生成对抗网络实现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]

(1. 沈阳建筑大学 机械工程学院, 辽宁 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, 辽宁 沈阳 110169)

摘 要: 红外与可见光图像融合技术能够同时提供红外图像的热辐射信息和可见光图像的纹理细节信息,在智能监控、目标探测和跟踪等领域具有广泛的应用。两种图像基于不同的成像原理,如何融合各自图像的优点并保证图像不失真是融合技术的关键,传统融合算法只是叠加图像信息而忽略了图像的语义信息。针对该问题,提出了一种改进的生成对抗网络,生成器设计了局部细节特征和全局语义特征两路分支捕获源图像的细节和语义信息;在判别器中引入谱归一化模块,解决传统生成对抗网络不易训练的问题,加速网络收敛;引入了感知损失,保持融合图像与源图像的结构相似性,进一步提升了融合精度。实验结果表明,提出的方法在主观评价与客观指标上均优于其他代表性方法,对比基于全变分模型方法,平均梯度和空间频率分别提升了 55.84% 和 49.95%。

关键词: 图像融合; 生成对抗网络; 语义信息; 谱归一化
中图分类号: TP391 **文献标志码:** A **DOI:** 10.3788/IRLA20210291

Infrared and visible image fusion using improved generative adversarial networks

Min Li[1], Cao Sijian[1], Zhao Huaici[2*], Liu Pengfei[2]

(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)

Abstract: The infrared and visible image fusion technology can provide both the thermal radiation information of infrared images and the texture detail information of visible images. It has a wide range of applications in the fields of intelligent monitoring, target detection and tracking. The two type of images are based on different imaging principles. How to integrate the advantages of each type of image and ensure that the image will not distorted is the key to the fusion technology. Traditional fusion methods only superimpose images information and ignore the semantic information of images. To solve this problem, an improved generative adversarial network was proposed. The generator was designed with two branches of part detail feature and global semantic feature to capture the detail and semantic information of source images; the spectral normalization module was introduced into the discriminator, which would solve the problem that traditional generation adversarial networks were not easy to train and accelerates the network convergence; the perceptual loss was introduced to maintain the structural similarity between the fused image and source images, and further improve the fusion accuracy. The experimental results show that the proposed method is superior to other representative methods in subjective

收稿日期: 2021-05-02; **修订日期:** 2021-07-14
基金项目: 国家重点研发计划 (2018YFB1105300); 装备预研重点基金 (JZX7Y2019025049301)
作者简介: 闵莉, 女, 副教授, 硕士生导师, 博士, 主要研究方向为模式识别与智能系统、图像处理与机器视觉。
通讯作者: 赵怀慈, 男, 研究员, 博士生导师, 博士, 主要研究方向为图像处理, 复杂系统建模与仿真技术, 指挥、控制、通信与信息处理技术。

evaluation and objective indicators. Compared with the method based on the total variation model, the average gradient and spatial frequency are increased by 55.84% and 49.95%, respectively.

Key words: image fusion; generative adversarial network; semantic information; spectral normalization

0 引 言

红外与可见光图像融合作为图像融合技术的重要分支,在军事侦察和民用监控等领域有着广泛应用[1]。红外成像探测器能够通过目标与背景的亮温差捕获目标,摆脱了可见光传感器对光源的依赖,可以在夜晚识别目标,具有能克服恶劣天气的优点,但通常图像分辨率低;而可见光成像传感器捕捉目标的反射信息,其图像适合人类的视觉感知系统,具有分辨率高、细节特征丰富等优点,但容易受到光照与天气因素影响。因此,这两种图像具有天然的互补性[2],融合后的图像可以同时提供高亮目标信息与高分辨率场景纹理细节信息。

在传统的红外与可见光图像融合方法中,通常将其他方法如显著性检测[3]加入多尺度变换框架,通过建立混合模型结合各方法优点,虽然提升了图像融合性能,但又需要手动设计融合规则,这让传统方法变得越来越复杂。

卷积神经网络 (Convolutional Neural Networks, CNN) 通过卷积操作分割图像并自动提取不同层次特征,近些年通过在空间利用、深度、多路径、宽度、特征图利用、通道提升和引入注意力机制等方面的改进,使自身学习能力得到显著提升,在红外与可见光图像融合领域也获得广泛应用[4]。如 An 等人[5]提出基于 CNN 的图像融合方法,使融合图像特征更加清晰;Pan 等人[6]使用密集连接的卷积神经网络 (Densely Connected Convolutional Networks, Densenet) 构建融合算法,充分利用了各卷积层提取的特征。但训练 CNN 需要大量标记数据,而红外与可见光图像融合任务无法定义融合标准,缺少 Ground Truth 指导融合框架训练,导致 CNN 融合性能较差。

生成对抗网络 (Generative Adversarial Networks, GAN)[7]在图像生成领域具有独到的优势,在无监督情况下可以任意逼近真实数据的分布。利用 GAN 的这种特性,Ma 等人[8]提出 FusionGAN 方法,建立生成器与判别器之间的对抗,使融合图像保留更丰富的信息,端到端的网络结构不再需要手动设计融合规则;之后,Ma 等人[9]做出改进,建立双判别器条件生成对抗网络模型 (A Dual-Discriminator Conditional Generative Adversarial Network, DDcGAN),同时保留两种源图像的信息,但是双判别器也让网络更复杂,导致难以平衡生成器与判别器,使融合图像出现伪影。

基于 GAN 的图像融合方法正致力于设计更复杂的生成器结构获取红外图像的热辐射强度和可见光图像的纹理细节,并对源图像采取单一特征提取方式,使融合图像通常在局部区域表现突出,但因为忽略了源图像包含的丰富语义信息,导致融合图像边缘模糊,并且存在 GAN 网络的通病,训练不稳定问题。

针对以上问题,文中提出了一种改进的生成对抗网络实现红外与可见光图像融合。通过在融合图像与源图像之间建立对抗博弈充分训练生成器,提升图像融合效果。首先,在生成器中建立局部细节特征分支 (Part Detail Feature Branch, PDFB) 和全局语义特征分支 (Global Semantic Feature Branch, GSFB),同时提取输入图像的细节和语义信息,使融合图像具有更清晰的纹理和边缘;其次,在判别器中引入谱归一化模块 (Spectral Normalization, SN),增强网络训练过程中的稳定性,使网络更易收敛;最后,损失函数中增加感知损失,提高融合图像与源图像的语义相似性。

1 相关工作

1.1 生成对抗网络

GAN 最初由 Goodfellow 等人提出,框架由两个对立模型:生成器 (Generator, G) 与判别器 (Discriminator, D) 构成。生成器将先验分布P_z的噪声作为输入,试图将生成样本分布 (P_G) 逼近真实样本分布 (P_{data}) 来欺骗判别器,判别器负责确定输入样本来自样本分布还是数据分布,训练过程中,生成器输出的样本分布不断逼近真实样本分布。

近年来,对于 GAN 的改进主要集中在目标函

数、训练技巧和网络结构三方面。

GAN 具有在无监督情况下学习真实数据分布的能力，非常适合无法定义图像融合标准的红外与可见光图像融合任务。而根据图像特点构建损失函数约束融合图像与源图像在像素强度和梯度等方面的差异，能够避免因缺少 Ground Truth 造成对源图像典型特征的缺失，端到端的网络结构也无需手动设计融合规则。

1.2 生成对抗网络的缺陷及改进方式

GAN 内部的对抗博弈可以表示为：

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

从公式 (1) 中可以看出，P_G不能显示地表示，且在训练过程中，如果判别器训练得太好或者不够好，生成器将学习不到有效的梯度，不能保持二者的同步更新，因此导致 GAN 训练困难。针对这一问题，解决方法是使判别器网络满足利普希茨连续条件 (Lipschitz continuity)：

$$|f(x_1)-f(x_2)| \leqslant K|x_1-x_2| \quad (2)$$

对于f，最小的常数K称为f的利普希茨常数。将判别器的梯度限制在一定范围内，使判别器小范围地逐步进行更新，可以保持生成器与判别器的同步训练。Arjovsky 等人[10]提出 WGAN(Wasserstein GAN) 进行判别器权重裁剪，限制判别器网络中所有参数不超过某个范围以满足 Lipschitz 的连续性，但破坏了参数之间的数值比例关系，进而降低了判别器的判别能力；Mi[11] 等人提出在判别器中建立谱归一化层，通过约束每一层参数矩阵的谱范数，使判别器网络满足 Lipschitz 条件，在提高网络训练稳定性的同时也提高了计算效率，并且在这一过程中对每一层网络的权重矩阵的归一化处理也不会破坏矩阵结构。

2 文中方法

2.1 网络整体框架

文中通过建立一个双判别器的生成对抗网络来解决相同分辨率的红外图像与可见光图像融合问题，整体框架结构如图 1 所示。

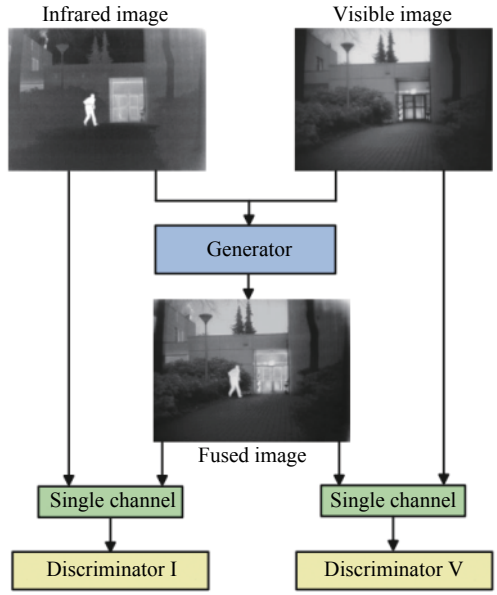


图 1 网络结构整体框架

Fig.1 Overall framework of network structure

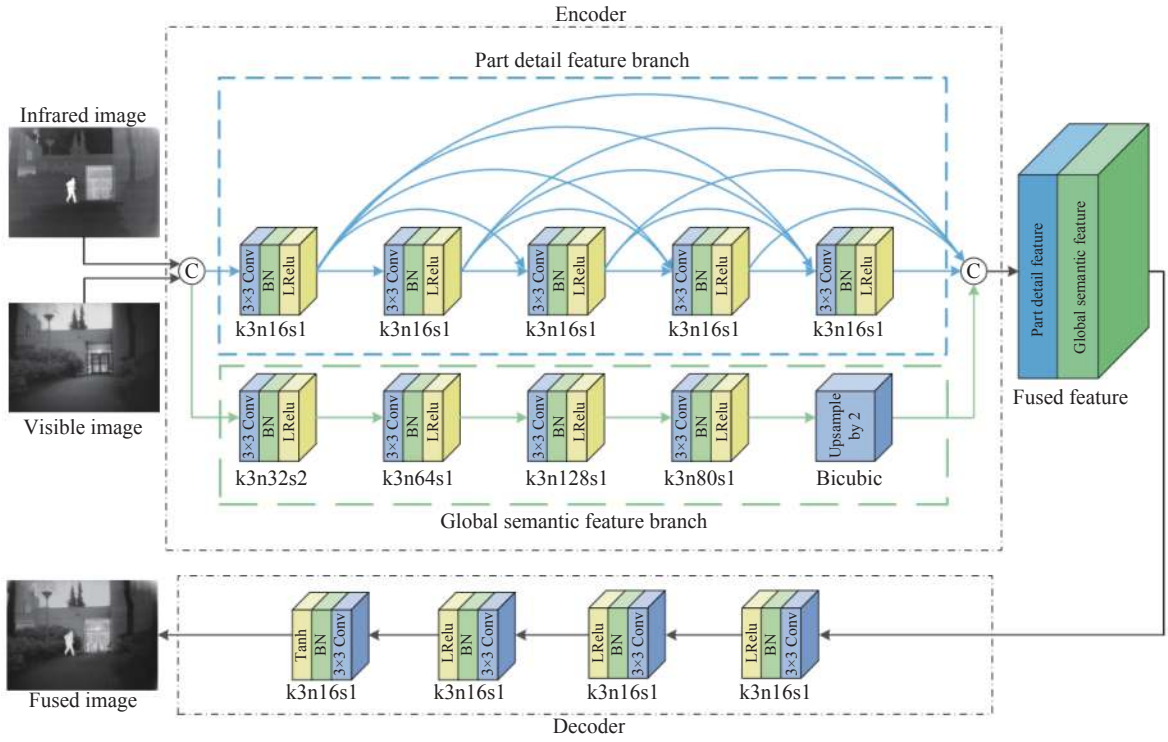
Generator 是输出融合图像的生成器，网络开始训练后，输入通道相连后的红外与可见光图像与融合图像并不组成数据对，而是以单通道形式输入到判别器中。双判别器 Discriminator I 与 Discriminator V 分别用于区分融合图像与红外图像、融合图像与可见光图像。经过生成器与判别器的不断对抗与迭代更新，得到训练完成的生成器，输入测试图像，输出图像融合结果。

Single Channel 代表每个判别器的输入是包含源图像与融合图像的单通道，因为如果输入是同时包含融合图像与对应的源图像作为条件信息的双通道，判别任务被简化为判别输入图像是否相同，而这对于判别器网络来说过于简单，将无法建立生成器与判别器之间的有效对抗关系，网络模型也失去意义。

2.2 生成器网络结构

红外与可见光图像融合任务属于无监督学习问题，而编码器-解码器网络结构在无监督学习情况下具有良好的特征重构特性，因此，生成器采用编码器-解码器结构，如图 2 所示。

通道相连后的红外与可见光图像需经过编码器的局部细节特征分支和全局语义特征分支进行特征提取，在特征融合后输入解码器对特征进行重构，输出最终的融合图像。其中局部细节特征分支采用 Densenet 结构，对于特征和梯度的传递更加有效，在充分利用图像的局部细节特征信息的同时又缓解了



Ⓒ: Concatenate Conv: Convolutional layer BN: Batch normalization Bicubic: Bicubic interpolation
k: kernel n: channel s: stride

图 2 生成器网络结构

Fig.2 Generator network structure

梯度消失。局部细节特征分支的基本单元是卷积层、批量归一化层和 Leaky Relu 激活函数。其中每个卷积层均使用 3×3 卷积核得到 16 个特征映射，通道数依次为 16、32、48、64、80。

输入图像除了包含红外图像的热辐射信息和可见光图像的细节纹理信息外，还包含丰富的语义信息。局部细节特征分支的网络结构是仅包含 1 个 5 层卷积的密集块 (Dense Block, DB)，相比于在语义分割中使用的网络模型，如 FC-Densnet103[12] 包含 4、5、7、10、12、15 等多层卷积共 11 个 DB，细节分支的网络结构过于简单，对输入图像包含的语义信息提取不充分。但如果继续提高细节分支的网络深度，尽管特征映射会包含更多的语义信息，也会让模型变得更加复杂和难以训练。

感受野反映了卷积神经网络每一层输出的特征图上的特征点在原始图像上映射的区域大小，降采样可以扩大感受野，能够提供更丰富的语义信息[13]。因此，建立全局语义特征分支，使用降采样层提取输入图像的语义信息。它由 5 个卷积层构成，通道相连的图像首先经过 1 次卷积得到 32 个通道的特征，然后再经历多次卷积，将通道数依次扩大到 64 和 128，再降到 80，最后一层采用双三次插值进行上采样恢复原始大小，使全局语义特征分支和局部细节特征分支最终的通道数相同，以便于两组特征的融合。

两支路输出的细节特征与语义特征经过融合后，特征图的通道数达到 160，再输入解码器中进行特征重构，最终通道数降至 1，输出融合图像。

2.3 判别器网络结构

两种源图像的成像原理不同，红外图像的热辐射信息主要以像素强度表达，而可见光图像的细节纹理信息主要以梯度表达。此外，两种源图像还各自包含语义信息，因此需分别设计判别器，引导生成器对源图像进行特征提取。判别器的网络结构如图 3 所示。

双判别器内部采用相同结构，是一个 5 层的卷积神经网络，在第 1 层~第 4 层的卷积层中使用 3×3 卷积核，判别器本质是从输入图像中提取特征映射并进行分类，工作方式与池化层相同，所以将步长设置为 2。最后的全连接层将前 4 层卷积后的特征进行整

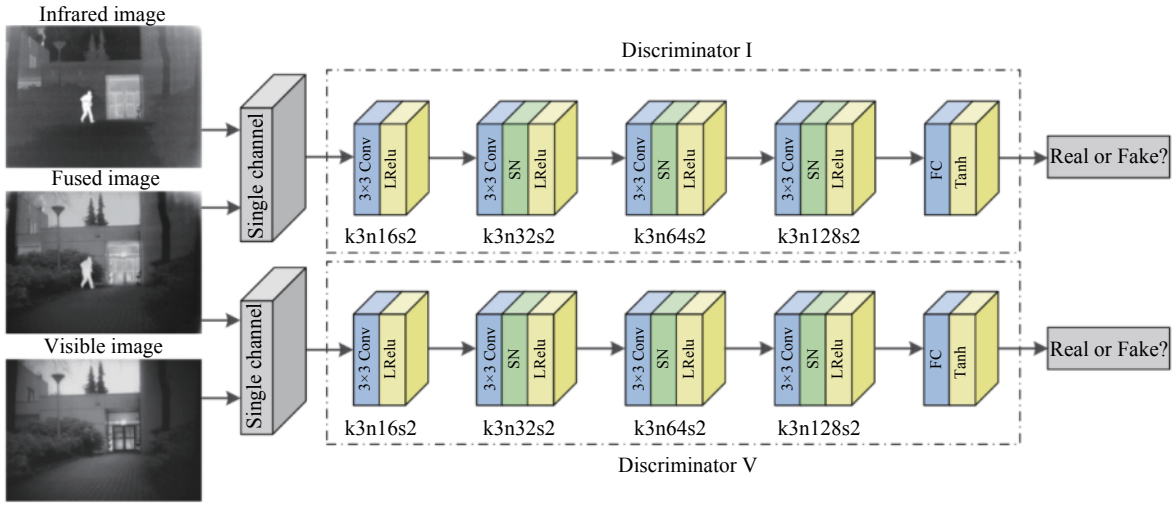


图 3 判别器网络结构

Fig.3 Discriminator network structure

合，并由 Tanh 激活函数生成标量，估计输入图像来自于源图像而非融合图像的概率。

为解决 GAN 训练不稳定的问题，在判别器网络中引入谱归一化，通过限制每一层网络权重矩阵的谱范数使判别器满足 Lipschitz 约束$\sigma(W)=1$：

$$\sigma(W_{SN})=\frac{W}{\sigma(W)} \tag{3}$$

权重矩阵的谱范数$\sigma(W)$等于矩阵的最大奇异值，通常采用幂迭代法近似求解谱范数，避免直接通过矩阵奇异值分解计算增加计算量。经过谱归一化的权重矩阵的最大奇异值均为 1，保证判别器网络满足 1−Lipschitz 连续。与和谱归一化配合使用的 Leaky ReLU 激活函数满足 Lipschitz 约束要求，不会破坏判别器网络已经满足的 1−Lipschitz 连续性。

2.4 损失函数设计

生成器与双判别器之间的损失构成总对抗损失L_{adv}，公式如下：

$$L_{adv-i,f}=E\left[-\log D_{\mathrm{I}}(i)\right]+E\left[-\log\left(1-D_{\mathrm{I}}(G(i,v))\right)\right] \tag{4}$$

$$L_{adv-v,f}=E\left[-\log D_{\mathrm{V}}(v)\right]+E\left[-\log\left(1-D_{\mathrm{V}}(G(i,v))\right)\right] \tag{5}$$

式中：G代表生成器；D_{I}和D_{V}代表双判别器；E为期望；i和v为红外与可见光图像。

内容损失L_{con}通过 Frobenius 范数和全变分模型分别从像素强度和梯度两方面约束融合图像与源图像的相似度；将红外图像、可见光图像和融合图像组成 RGB 三通道输入预训练的 VGG19[14]，由第 2、4、8、12 和 16 卷积层输出的特征图计算感知损失L_{per}，提升融合图像的视觉效果。

总损失函数表达式为：

$$\begin{aligned}L_{total}&=L_{adv}+\lambda_1 L_{con}+\lambda_2 L_{per}=L_{adv}+\\&\lambda_1\left\{E\left[\|G(i,v)-i\|_F^2+\xi\|G(i,v)-v\|_{TV}\right]\right\}+\\&\lambda_2\left\{E\left[\|\phi(G(i,v))-\phi(x)\|_2^2\right]\right\}\end{aligned} \tag{6}$$

式中：λ_1和λ_2分别为内容损失和感知损失的权重系数；F代表 Frobenius 范数；TV代表 Total Variation 范数；ξ为平衡像素强度损失和梯度损失的系数；$\phi(x)$代表提取的特征图。

3 实验设置与结果评价

3.1 数据集与实验过程

文中方法及其他深度学习类型对比方法使用的硬件平台配置如下：CPU 为 AMD Ryzen 53600，显卡为 GeForce RTX 2060 SUPER 8G，内存为 32 GB，采用 tensorflow 框架，在 Ubuntu18.04 系统上使用 python 3.6.5 实现文中模型。部分对比方法和客观评价指标在 Windows 10 系统上使用 Matlab R2018b 运行和计算。

在训练阶段采用 TNO 图像融合数据集，该数据集包含多种军事场景的多光谱夜间图像，波段涵盖可见光、近红外和长波红外。从中选取 36 对红外与可见光图像作为训练集，但图像数量不足以训练好网络

模型, 所以将它们裁剪为 84×84 分辨率的共 27264 对红外与可见光图像作为最终的训练集。

批数量设置为 24, 优化器选用 RMSProp, 训练阶段包含 920 个 epochs, 采用的初始学习率为 2×10^{-3}, 并在每个 epoch 后衰减到原始值的 0.85。其他参数设置如下: $\lambda_1=0.1$, $\lambda_2=1\times10^{-3}$, $\xi=1.2$。

在测试阶段, 首先从 TNO 数据集中选取 20 对红外与可见光图像作为测试集, 输入经过训练的生成器得到最终的融合图像; 其次在 RoadScene 数据集进行第 2 组实验, 该数据集涵盖典型的道路场景, 从中选取 20 对红外与可见光图像作为第二组测试集。

3.2 对比实验

首先, 使用 TNO 数据集进行对比实验, 选取 5 种典型的图像融合方法, 实验结果如图 4 所示。图像从上至下依次对应红外图像、可见光图像、全变分模型 (Different Resolutions via Total Variation Model, DRTV) 方法[15]、CNN 方法[16]、FusionGAN 方法、深度图像分解 (Deep Image Decomposition for Infrared and Visible Image Fusion, DIDFuse) 方法[17]、DDcGAN 方法及文中方法。红框代表对图像局部做截取放大处理, 可以更直观地观察图像细节。

主观方面, 从图 4 中可以看出, DRTV 因为全变

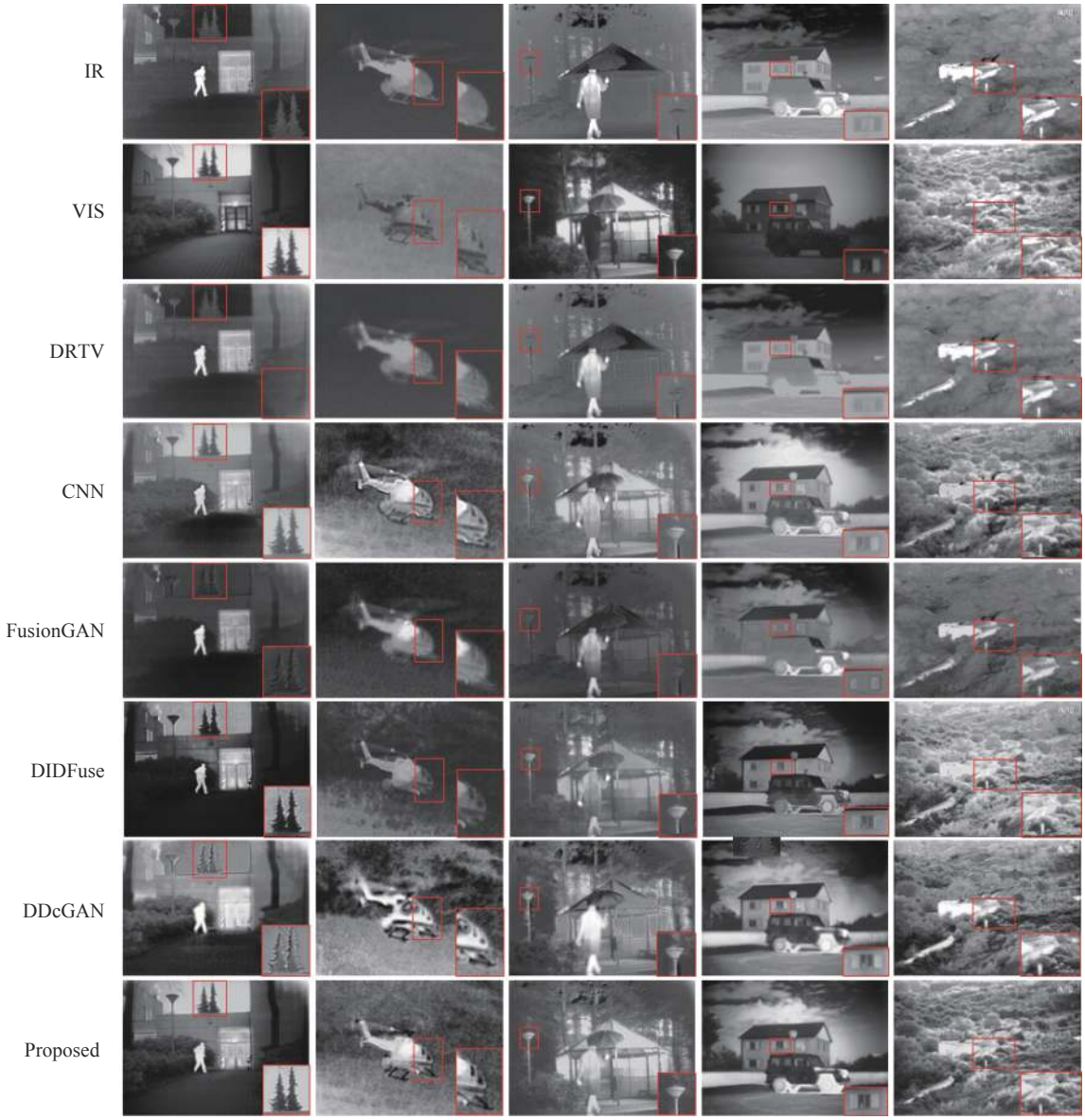


图 4 TNO 数据集对比实验结果

Fig.4 Comparative experimental results of the TNO dataset

分模型的特性，导致融合图像在平滑区域出现阶梯效应；CNN 未经过图像融合任务的训练，提取的特征不能有效融合，图像整体性较差；FusionGAN 使用单判别器，造成融合图像更趋近于红外图像的像素分布，背景区域纹理细节信息保留不足；DIDFuse 构建了基于深度学习的双尺度图像分解网络，但使用的编码器网络较浅，无法正确提取源图像典型特征；DDcGAN 的网络结构对源图像高层特征空间的语义信息提取不足，导致边缘信息的丢失，双判别器结构又加剧了网络训练的不稳定性，导致融合图像出现伪影。而文中方法通过在提取语义信息和稳定网络训练等方面做出的改进，获得的融合图像清晰度高，目标边缘清晰，背景区域纹理细节丰富。

第二组实验在 RoadScene 数据集上进行，为了更加充分和直观地对比不同算法在同一序列图像中的实验结果，从中选取同时包含道路、车辆和行人的图像，如图 5 所示。

图 5 RoadScene 数据集对比实验结果。(a) 红外图像；(b) 可见光图像；(c) DRTV；(d) CNN；(e) FusionGAN；(f) DIDFuse；(g) DDcGAN；(h) 文中方法

Fig.5 Comparative experimental results of the RoadScene dataset. (a) Infrared image; (b) Visible image; (c) DRTV; (d) CNN; (e) FusionGAN; (f) DIDFuse; (g) DDcGAN; (h) Proposed method

图 5(a)、(b) 分别为红外与可见光图像；图 5(c) 同时丢失了行人与车辆的热辐射信息；图 5(d)、(f) 并未体现出背景区域树木的纹理；图 5(e)、(g) 的目标车辆和行人边缘模糊；图 5(h) 中同时保留了目标行人和车辆热辐射信息、道路和树木的纹理细节信息，易于人眼识别。

主观评价方法依赖人类视觉敏感度，容易受到自身观感和外界环境的影响，需要引入客观指标定量衡量融合图像在纹理和边缘等方面的特性，以主客观结合的方式对图像融合算法的性能给出最终评价。

在客观方法中，选取 4 种用于评估红外与可见光图像融合算法性能的典型指标，包括：平均梯度 (Average Gradient, AG) 和空间频率 (Spatial Frequency, SF) 评价图像清晰度和纹理细节丰富程度；边缘信息保留 (Gradient-based fusion performance, $Q^{AB/F}$) 评价源图像传递到融合图像的边缘信息量；人类视觉敏感度 (Chen-Blum, Q_{CB}) 评价融合图像对源图像局部细节特征的保留程度。

客观评价结果如表 1 所示，文中方法在两组实验中的表现均优于其他对比方法，相比于 DDcGAN 方法，在 TNO 数据集中，4 项客观指标分别提升了 6.00%、10.81%、12.64% 和 10.53%；在 RoadScene 数据集中，4 项客观指标分别提升了 12.60%、12.25%、13.48% 和 7.16%。表明融合图像在边缘和纹理方面提升明显，客观上验证了算法改进的有效性。

表 1 两组对比实验客观评价结果

Tab.1 Objective evaluation results of two comparison experiment

Dataset	Methods	AG	SF	$Q^{AB/F}$	Q_{CB}
TNO	DRTV	3.761	9.639	0.319	0.411
	CNN	4.700	11.489	0.332	0.463
	FusionGAN	4.014	10.006	0.313	0.425
	DIDFuse	4.644	11.771	0.395	0.472
	DDcGAN	5.529	13.044	0.356	0.456
	Proposed method	5.861	14.454	0.401	0.504
Road scene	DRTV	3.221	8.696	0.368	0.384
	CNN	4.484	10.536	0.398	0.384
	FusionGAN	3.290	8.426	0.278	0.387
	DIDFuse	5.253	14.149	0.469	0.452
	DDcGAN	5.200	13.580	0.423	0.461
	Proposed method	5.855	15.243	0.480	0.494

综上，主观与客观评价结果相吻合，证明改进算法建立的局部与全局双分支编码器结构使源图像的细节和语义信息得到最大化保留，保证融合图像在目标和背景区域均具有清晰的纹理和边缘；引入谱归一，稳定网络训练，消除视觉伪影，进一步提升了图像融合结果。

3.3 消融实验

使用 TNO 数据集进行消融实验，分析网络结构改进的作用。实验包含 3 个部分：①DDcGAN + SN：在判别器中引入谱归一化；②DDcGAN + GSFB：在生成器中建立全局语义特征分支；③DDcGAN + GSFB + SN：同时对网络做出以上两种改进。谱归一化能够提高网络训练稳定性，可通过损失函数曲线验证效果，如图 6 所示。

曲线中 G-D2-Loss 代表生成器与用于区分可见光图像与源图像的判别器之间的对抗损失，在引入谱

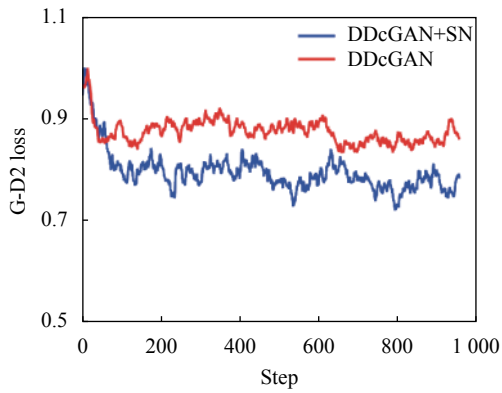


图 6 损失函数曲线

Fig.6 Loss function curve

归一化模块后，损失函数收敛更快，稳定后数值更小，代表网络训练更加稳定，图像融合精度更高。

消融实验评价结果如表 2 所示。

可以看出，全局语义分支结构通过降采样扩大感受野，能够挖掘出源图像高层特征空间包含的语义信息，为融合图像提供更清晰的边缘和纹理。与谱归一化相结合，最终使图像的融合效果产生了明显提升，证明改进有效。

表 2 消融实验评价结果

Tab.2 Evaluation results of ablation experiment

Methods	AG	SF	$Q^{AB/F}$	Q_{CB}
DDcGAN	5.529	13.044	0.356	0.456
DDcGAN+SN	5.670	13.062	0.368	0.469
DDcGAN+GSFB	5.746	13.841	0.388	0.484
DDcGAN+GSFB+SN	5.861	14.454	0.401	0.504

4 结 论

文中提出了一种基于语义信息和谱归一化改进的生成对抗网络，以端到端形式进行红外与可见光图像融合。首先，在生成器的编码器内部建立局部细节特征分支和全局语义特征分支，提取输入图像的细节和语义信息；其次，在判别器中使用谱归一化，提升网络训练的稳定性，并引入感知损失，提升融合图像的视觉效果；最后，通过对比实验和消融实验证明改进方法在主观融合图像边缘、纹理细节上及客观融合图像质量上都达到更优的效果。但文中方法也存在不足，处理测试图像的平均时间为 1.274 s，原因是方法基于深度学习，对硬件性能要求较高。因此，将在下一阶段重点研究轻量化网络方法解决上述问题，以便应用于便携式多光谱相机等设备。

参考文献：

[1] Shen Ying, Huang Chunhong, Huang Feng, et al. Infrared and visible image fusion: review of key technologies [J]. ***Infrared and Laser Engineering***, 2021, 50(9): 20200467. (in Chinese)

[2] Shen Yali. RGBT dual-model Siamese tracking network with feature fusion [J]. ***Infrared and Laser Engineering***, 2021, 50(3): 20200459. (in Chinese)

[3] Chen J, Wu K, Cheng Z, et al. A saliency-based multiscale approach for infrared and visible image fusion [J]. ***Signal Processing***, 2021, 182(4): 107936.

[4] Huan Kewei, Li Xiangyang, Cao Yutong, et al. Infrared and visible image fusion with convolutional neural network and NSST [J]. ***Infrared and Laser Engineering***, 2022, 51(3): 20210139. (in Chinese)

[5] An W B, Wang H M. Infrared and visible image fusion with supervised convolutional neural network [J]. ***Optik-International Journal for Light and Electron Optics***, 2020, 219(17): 165120.

[6] Pan Y, Pi D, Khan I A, et al. DenseNetFuse: A study of deep unsupervised DenseNet to infrared and visual image fusion [J]. ***Journal of Ambient Intelligence and Humanized Computing***, 2021(3): 02820.

[7] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks [J]. ***Advances in Neural Information Processing Systems***, 2014, 3: 2672-2680.

[8] Ma J, Wei Y, Liang P, et al. FusionGAN: A generative adversarial network for infrared and visible image fusion [J]. ***Information Fusion***, 2019, 48: 11-26.

[9] Ma J, Xu H, Jiang J, et al. DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion [J]. ***IEEE Transactions on Image Processing***, 2020, 29: 4980-4995.

[10] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN [J]. ***arXiv***, 2017: 1701. 07875v1.

[11] Miyato T, Kataoka T, Koyama M, et al. Spectral normalization for generative adversarial networks [C]//International Conference on Learning Representations, 2018.

[12] Jégou S, Drozdzal M, Vazquez D, et al. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2016.

[13] X Li, You A, Zhu Z, et al. Semantic flow for fast and accurate scene parsing [J]. ***arXiv***, 2020: 2002.10120.

[14] Karen S, Andrew Z. Very deep convolutional networks for large-scale image recognition[J]. ***arXiv***, 2014: 1409.1556.

[15] Du Qinglei, Xu Han, Ma Yongoing, et al. Fusing infrared and visible images of different resolutions via total variation model [J]. ***Sensors***, 2018, 18(11): 3827.

[16] Li H, Wu X J, Kittler J. Infrared and visible image fusion using a deep learning framework[J]. ***arXiv***, 2018: 1804.06992.

[17] Zhao Z, Xu S, Zhang C, et al. DIDFuse: Deep image decomposition for infrared and visible image fusion [C]//Twenty-Ninth International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence, 2020.