

基于生成对抗网络的多模态图像融合

杨晓莉¹, 蔺素珍^{1*}, 禄晓飞², 王丽芳¹, 李大威¹, 王斌¹

¹中北大学大数据学院, 山西 太原 030051;

²酒泉卫星发射中心, 甘肃 酒泉 735000

摘要 针对多模态图像融合中多尺度几何工具和融合规则设计困难的问题, 提出一种基于生成对抗网络(GANs)的图像融合方法, 实现了多模态图像端到端的自适应融合。将多模态源图像同步输入基于残差的卷积神经网络(生成网络), 通过网络的自适应学习生成融合图像; 将融合图像和标签图像分别送入判别网络, 通过判别器的特征表示和分类识别逐渐优化生成器, 在生成器和判别器的动态平衡中得到最终融合图像。与具有代表性的融合方法相比, 实验结果表明, 本文方法的融合结果更干净, 没有伪影, 提供了更好的视觉质量。

关键词 图像处理; 图像融合; 多模态图像; 深度学习; 生成对抗网络

中图分类号 TP391

文献标识码 A

doi: 10.3788/LOP56.161004

Multimodal Image Fusion Based on Generative Adversarial Networks

Yang Xiaoli¹, Lin Suzhen^{1*}, Lu Xiaofei², Wang Lifang¹, Li Dawei¹, Wang Bin¹

¹School of Big Data, North University of China, Taiyuan, Shanxi 030051, China;

²Jiuquan Satellite Launch Center, Jiuquan, Gansu 735000, China

Abstract This study proposes a new network based on generative adversarial networks to achieve an end-to-end image adaptive fusion, thus solving the difficulties in designing multiscale geometric tools and fusion rules in multimodal image fusion. First, the multimodal source image is synchronously input into the generative network, whose structure is created based on a residual-based convolutional neural network proposed herein. The network can generate the fused image through adaptive learning. Second, the fused and label images are sent to the discriminant network. The generator is gradually optimized through the feature representation and classification identification of the discriminator. The final fused image is obtained in the dynamic balance of the generator and discriminator. In comparison with the existing representative fusion methods, the proposed algorithm obtains more cleaner fusion results and has no artifacts, thereby providing a better visual quality.

Key words image processing; image fusion; multimodal image; deep learning; generative adversarial networks

OCIS codes 100.4996; 100.2960

1 引言

多波段/多模态成像系统在军事、医疗、工业检测等领域已有广泛应用, 图像融合是这些系统实现高精度智能探测的关键技术之一, 其目的是将同一场景的多幅图像或序列图像合成一幅信息更完整、更全面的图像^[1]。

目前的图像融合方法大致可分为两大类: 基于

模型驱动的传统融合法^[2-9], 如基于非下采样剪切波变换(NSST)融合法^[7]、基于稀疏表示融合法^[8]、导向滤波融合法^[9]等; 基于数据驱动的神经网络融合法^[10-14], 如基于卷积神经网络的图像融合法^[11]、基于编码和解码网络的图像融合法^[12]等。基于模型驱动的传统融合法首先对源图像进行分解, 再选用合适的融合规则对分解后的子图像进行融合, 最后将子图像重构获得最终融合图像。此类方法依赖先

收稿日期: 2019-01-09; 修回日期: 2019-01-18; 录用日期: 2019-03-22

基金项目: 山西省重点研发计划(201603D321128)、山西省应用基础研究计划(201701D121062)、中北大学第十五届研究生科技科技立项(2018514)

* E-mail: lsz@nuc.edu.cn

验知识选取具有针对性的分解方法和融合规则,并需人工设定模型参数,具有很大的局限性。神经网络具有强大的特征提取和数据表示能力^[10],因此,新近有研究探索了基于数据驱动的神经网络融合法,相比传统融合法,其更具优越性。文献[11]中将卷积神经网络(CNNs)用于图像融合,使用包含源图像不同模糊程度的图像块来训练网络并获得决策图,最后通过决策图和源图像组合获得融合图像。此方法首次将CNNs用于图像融合,实现了活动水平测量和融合规则的统一,但仅适用于多聚焦图像融合,且因仅使用CNNs最后层计算的结果,丢失了中间层获得的许多有用信息,当网络加深时,信息丢失更加严重。文献[13]中提出一种更深的网络结构和多层深度特征图像融合方法,首先利用导向滤波对源图像进行高低频分解,通过加权平均融合规则获取低频融合子图像,然后利用VGG-19网络提取高频细节的多层特征,通过取大规则获得高频融合子图像,最后重构融合图像。该方法仍需使用传统方法进行图像分解,虽然利用了VGG-19网络的中间层特征,但多个中间层特征组合策略过于简单,仍然丢失了许多有用的信息。同时,以上基于深度学习的融合法均只针对双波段/模态图像融合,不能满足多波段/模态探测的需要。

针对以上问题,本文提出了一种新的神经网络架构,并用于多波段/模态图像融合,该网络由生成网络和判别网络组成。生成网络由残差块和卷积层构成,用于提取多波段源图像特征。由于残差块的结构特性,深层特征已包含多层信息,因此本文只使用最后一层生成融合图像。判别网络通过与生成网络的对抗学习提升融合图像质量。主要创新点有:1)图像特征提取和融合一并通过网络学习实现,活动水平的测量和融合规则的统一均未经人为设定,减少了对先验知识的依赖,减小了融合策略选择的不确定性;2)使用生成对抗网络框架(GANs)同步融合三波段图像,获得更清晰、更真实的数据,真正实现了端到端融合。

2 相关工作

2.1 深度残差网络(ResNet)

在2016年的计算机视觉和模式识别国际会议(CVPR)中,He等^[15]提出了一种新颖的网络架构来解决网络退化问题。以捷径方式连接和残差表示,网络比以前更容易优化;通过增加深度,网络精度亦得到提高。残差块体系结构如图1所示,图中

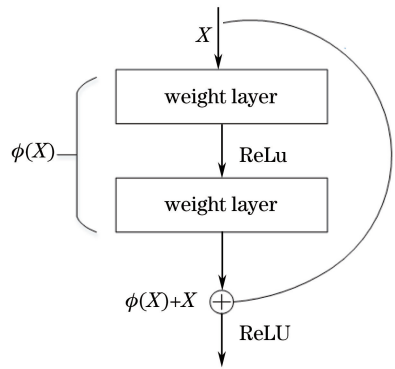


图1 残差块结构图

Fig. 1 Structure of residual block

X 表示网络的输入, $\phi(X)$ 表示包含两个权重层的网络操作,ReLU表示修正线性单元^[16],残差块的输出由 $\phi(X)+X$ 计算得到。通过这种结构,可以利用多层信息。此外,在图像重建任务^[17-18]中,使用残差块可以取得更好的性能。考虑到残差网络良好的特征提取能力能够提升图像的融合效果,本文将引入多模态图像融合。

2.2 生成对抗网络(GANs)

2014年,Goodfellow等^[19]首次提出GANs概念,其基于对抗思想,提供了一种有效方式来估计目标分布,并生成新的样本。GANs包括生成器和判别器两个模型:生成器用 $G(z)$ 表示输入变量到数据空间的映射, z 为随机噪声;判别器用 $D(x)$ 表示来自标签而非输入数据的概率, x 为真实数据。生成器和判别器在动态博弈中达到平衡,遵循

$$\min_G \max_D V(D, G) = E_{x \sim p_{D_{\text{ata}}}(x)} \{\lg[D(x)]\} + E_{z \sim p_{z(z)}} \{\lg\{1 - D[G(z)]\}\} \quad (1)$$

实现生成功能,式中: $V(D, G)$ 为值函数; $E_{x \sim p_{D_{\text{ata}}}(x)}$ 、 $E_{z \sim p_{z(z)}}$ 分别为真实数据和生成数据的期望; $p_{z(z)}$ 为输入噪声变量先验概率; $p_{D_{\text{ata}}}(x)$ 为真实数据分布概率。就原始的GANs来说,输入的噪声变量一般为服从高斯分布的随机数,虽然高斯分布能够在复杂的情况下拟合任何一种数据分布,但在实际训练过程中,往往会由于图像数据流形成的维度太高而无法准确拟合,从而导致模式崩塌,且由于生成器和判别器不能很好地同步训练,易造成GANs训练不稳定。各种GANs变体被相继提出,如深度卷积GANs(DCGANs)^[20]、最小二乘GANs(least squares GANs)^[21]等。DCGANs将CNN引入GANs,弥补了用于监督学习的CNN与用于无监督学习的GANs之间的差距,使传统GANs网络的训练更稳定。文献[22]中首次使用对抗学习的方

式训练深度卷积网络,实现了超分辨率复原,将原先网络输入的随机噪声扩展到单幅图像。受此启发,如果能同时将多幅图像输入,并且约束其生成,则可设计网络来实现融合,而多波段/多模态图像融合可以看作将一幅图像以多通道形式输入网络,输出一幅理想图像的过程;所以,从逻辑上讲 GAN 可以用于图像融合。

3 基于 GANs 的图像融合方法

3.1 方法框架

本文方法框架如图 2 所示。训练网络时,将多波段/多模态图像以一幅图像多通道的形式输入生成网络,将生成的单幅融合图像和标签图像输入判别模型,经过网络训练,完成二分类任务,最终输出高质量的融合图像。由于图像融合没有标准输出,故从模型驱动法中选取性能优良的多种算法融合多模态图像,再经多指标评价优选高质量的融合图像作为最终标签,以解决有监督训练中缺乏标签数据的问题。

3.2 GANs 生成模型的构建

本文使用的基于残差块的卷积神经网络如图 3

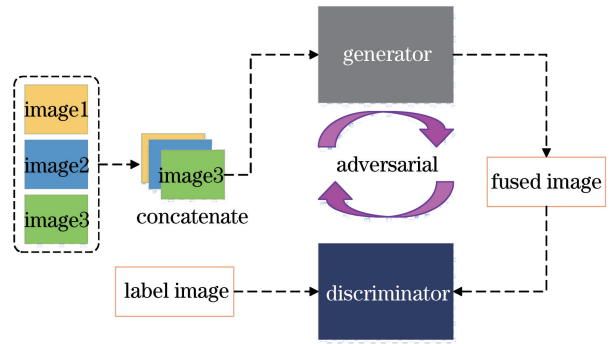


图 2 方法框架图

Fig. 2 Framework of method

所示。将多波段图像在通道维度上通过 concatenate 函数连接输入神经网络,充分利用各层的特征 map,保持图像信息的传递。这里以 3 通道为例,输入 $128 \text{ pixel} \times 128 \text{ pixel} \times 3$ 的图像,进行一次卷积后,输入 7 个残差单元块,再进行两层卷积操作,提取高维度特征。具体网络参数如表 1 所示。为保持图像大小不变,设步长为 1,尽可能减少源图像信息的丢失。每个卷积层之后使用 ReLU^[16],以提升网络的非线性程度。

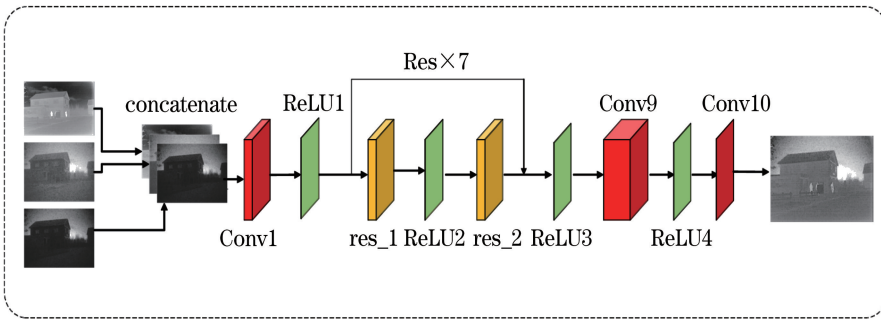


图 3 生成模型的网络结构

Fig. 3 Network structure of generative model

表 1 生成器参数

Table 1 Parameters of generator

Layer	Filter size /step	Output size
Conv1	$3 \times 3 / 1$	$128 \times 128 \times 64$
Res(7 units)	$3 \times 3 / 1$	$128 \times 128 \times 64$
	$3 \times 3 / 1$	$128 \times 128 \times 64$
Conv9	$3 \times 3 / 1$	$128 \times 128 \times 256$
Conv10	$3 \times 3 / 1$	$128 \times 128 \times 1$

3.3 GANs 判别模型的构建

本文使用的判别网络结构如图 4 所示。判别器使用跨步长卷积分别提取生成图像和真实图像的深度特征。在残差单元块中,由于数据维度的改变,先使用 1×1 卷积核进行数据降维,再通过 3×3 卷积核提取特征。介于网络结构较深,除 Conv1 外,每

层卷积之后都使用 BatchNorm(BN)^[23]对数据归一化,BN后,使用 LeakyReLU(LReLU)^[24]激活函数提升网络的非线性程度,最后以全连接层(Fc)输出。具体网络参数如表 2 所示。

3.4 GANs 损失函数的构建

使用损失函数来衡量网络输出和数据标签之间的差距。其中,最简单的是使用均方误差损失,虽然有一定的效果,但其不能智能地学习目标细节,故生成的图片稍放大后即变模糊。为此,引入梯度损失优化图像边缘,再加入 GAN 对抗损失。为明确梯度损失对图像的影响,这里设置权重来提高融合图像质量。

1) 均方误差(MSE)损失

MSE损失^[25]通过网络训练过程中输出的预测

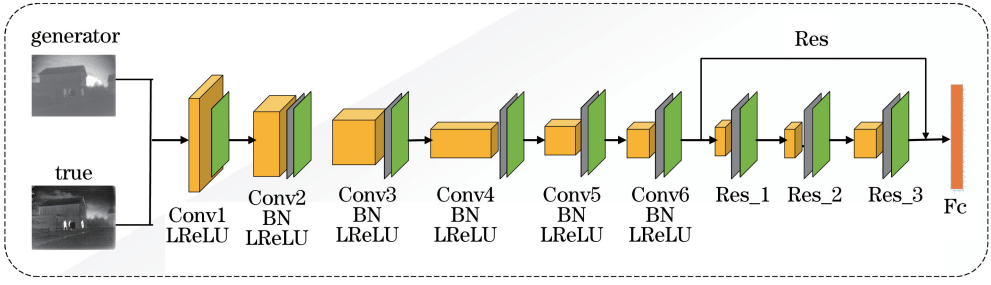


图 4 判别模型的网络结构

Fig. 4 Network structure of discriminative model

表 2 判别器参数

Table 2 Parameters of discriminator

Layer	Filter size/step	Output size
Conv1	3×3 /1	128×128×64
Conv2	3×3 /2	64×64×128
Conv3	3×3 /2	32×32×256
Conv4	3×3 /2	16×16×512
Conv5	3×3 /1	16×16×256
Conv6	1×1 /1	16×16×128
Res	1×1 /1	16×16×64
	3×3 /1	16×16×64
	3×3 /1	16×16×128
Fc	—	1

值与真实数据之间的差值促进多尺度图像内容的一致性,计算公式为

$$L_{\text{mse}} = \frac{1}{W \times H} \sum_{m=1}^W \sum_{n=1}^H |I_{m,n}^{\text{pred}} - I_{m,n}^{\text{true}}|, \quad (2)$$

式中: L_{mse} 为 MSE 损失; W 、 H 分别为图像的宽和高; $I_{m,n}^{\text{pred}}$ 为生成图像中某点 (m,n) 的像素值; $I_{m,n}^{\text{true}}$ 为真实图像中某点 (m,n) 的像素值。

2) 梯度损失

梯度是描述图像边缘信息的主要指标之一。由于训练集中图像尺度较小且边缘不清晰,故引入梯度损失来优化边缘,具体实现方法如下。

$$\mathbf{G}_h = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * \mathbf{A}, \quad (3)$$

$$\mathbf{G}_v = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * \mathbf{A}, \quad (4)$$

$$G = |\mathbf{G}_h| + |\mathbf{G}_v|, \quad (5)$$

$$L_{\text{grad}} = \frac{1}{W \times H} \sum_{m=1}^W \sum_{n=1}^H |\mathbf{G}_{m,n}^{\text{pred}} - \mathbf{G}_{m,n}^{\text{true}}|, \quad (6)$$

式中: L_{grad} 为梯度损失; \mathbf{A} 为原图像;矩阵表示的是 sobel 算子; \mathbf{G}_m 、 \mathbf{G}_n 分别为横、纵边缘检测的灰度

值; $\mathbf{G}_{m,n}^{\text{pred}}$ 为生成图像点 (m,n) 的梯度; $\mathbf{G}_{m,n}^{\text{true}}$ 为真实图像点 (m,n) 的梯度; G 为图像梯度。

3) 对抗损失

将对抗损失^[19]加入到感知损失中,鼓励生成器适应真实数据的分布,计算公式为

$$L_{\text{gen}} = \sum_1^N -\lg\{D[G(I^{\text{pred}})]\}, \quad (7)$$

式中: L_{gen} 为对抗损失; $D[G(I^{\text{pred}})]$ 为重构图像 $G(I^{\text{pred}})$ 是真实图像的估计概率; N 为图像像素个数。

4) 联合损失函数

生成器最终的损失函数是以上 3 项损失之和,即

$$L_{\text{gan}} = L_{\text{pixel}} + L_{\text{gen}} + \lambda L_{\text{grad}}. \quad (8)$$

对结果施加 L2 正则化^[26],即在原来损失函数的基础上加上权重参数的平方和,用以约束要优化的参数,防止模型过拟合。

4 分析与讨论

4.1 数据集制作

本文 GANs 网络的数据集有 14000 张图像,多波段图像来自公开 TNO 图像融合数据集^[27],包含可见光(390~700 nm)、近红外(700~1000 nm)和红外长波(8~12 μm)。医学图像来自哈佛医学院的 Neuroimaging Primer。标签图像制作方法如下:将配准好的多模态图像用 128×128 的滑动窗口以步长 14 划分成子块,使用基于拉普拉斯金字塔(LP)、离散小波变换(DWT)、非下采样轮廓波变换(NSCT)、NSST 传统融合法进行融合,融合规则统一使用“高频梯度取大,低频加权平均”^[28]。这时每组多模态子图像均会有多个融合图像,使用包含基于统计特性的评价指标——标准差(SD)、平均梯度(AG)、对比度(Con)、相关系数(CC),基于信息量的评价指标——信息熵(IE)、互信息(MI),以及表示人类视觉感知的视觉信息保真度(VIFF)等多种质

量评价指标对融合图像进行评价^[29],选取融合效果最好的图像作为最终标签。

从 14000 张标签图像中随机选取第 50 张进行可视化展示,如图 5 所示。通过上述 4 种传统融合方法对三波段源图像进行融合,然后使用上述 7 个质量评价指标对这 4 幅图像进行评价,结果如表 3 所示。可以看出,LP 方法融合的图像有 5 项指标值

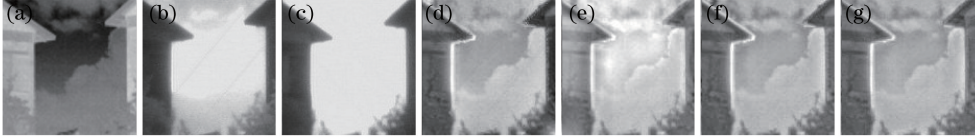


图 5 标签图像预选图。(a)红外长波;(b)红外短波;(c)可见光;(d) LP;(e) DWT;(f) NSCT;(g) NSST
Fig. 5 Pre-selection maps of label images. (a) Longwave infrared; (b) shortwave infrared; (c) visible light; (d) LP; (e) DWT; (f) NSCT; (g) NSST

表 3 标签图像选取依据表
Table 3 Label image selection table

Fusion method	SD	AG	Con	CC	IE	MI	VIFF
LP	49.019	4.211	40.812	0.424	7.223	5.608	0.466
DWT	44.772	3.826	36.650	0.342	7.175	5.415	0.442
NSCT	41.123	4.061	34.816	0.437	6.953	5.222	0.469
NSST	40.904	4.149	34.902	0.441	6.932	5.201	0.467

4.2 训练参数设置

1) 学习率选取

学习率是衡量网络梯度下降快慢的因子,合适的学习率能使网络快速收敛,减少训练时间。但学习率过大时会出现振荡现象,使网络无法收敛;过小时,网络迭代更新速度慢,训练速度降低。不同学习率对损失函数的影响如图 6 和图 7 所示。训练选取 100 次迭代,图 6 中隔点取值,由图可知,当学习率为 0.0002 时,损失值下降很缓慢,达到相同效果耗费时间较长;学习率为 0.02 时,损失值下降很快,但出现振荡现象,影响网络收敛;当学习率为 0.002 时效果最好,损失值既能快速收敛又比较稳定。在同一训练过程中选取全部点得到图 7 所示结果。当学习率为 0.02 时,损失值无法收敛;而学习率为 0.002 时比 0.0002 更快收敛。由此可知,若学习率再增大,损失曲线只会比学习率为 0.02 时更振荡,故将其作为学习率上限;若学习率低于 0.0002,则收敛会更缓慢,故将其作为学习率下限。

2) λ 权重选取

借鉴文献[22]及文献[25]实验成果,分别设置梯度损失的权重 λ 为 0、0.01、0.1、1 进行测试。从大量实验数据中选取具有代表性的 3 组图像进行说明,实验结果如图 8~10 所示。

优于其他 3 种方法,因此选取最优的 LP 融合图像作为该子图像的训练标签。这样,用于网络训练的训练集为 14000×3 张多模态切块图像,它们用 concatenate 连接,以一幅图像的三通道形式输入网络,标签为 14000 张融合图像。网络模型训练好后,只需将待融合的多模态图像输入网络便可获得理想的融合图像。

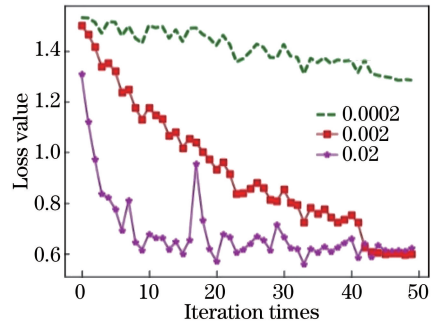


图 6 学习率对生成器损失的影响
Fig. 6 Effect of learning rate on generator loss

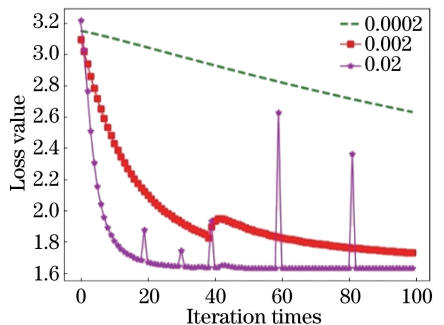


图 7 学习率对判别器损失的影响
Fig. 7 Effect of learning rate on discriminator loss

图 8 为 λ 选择不同值时的融合结果,为了便于观察,将要对比的地方放大,可以看出,当 λ 为 0 和

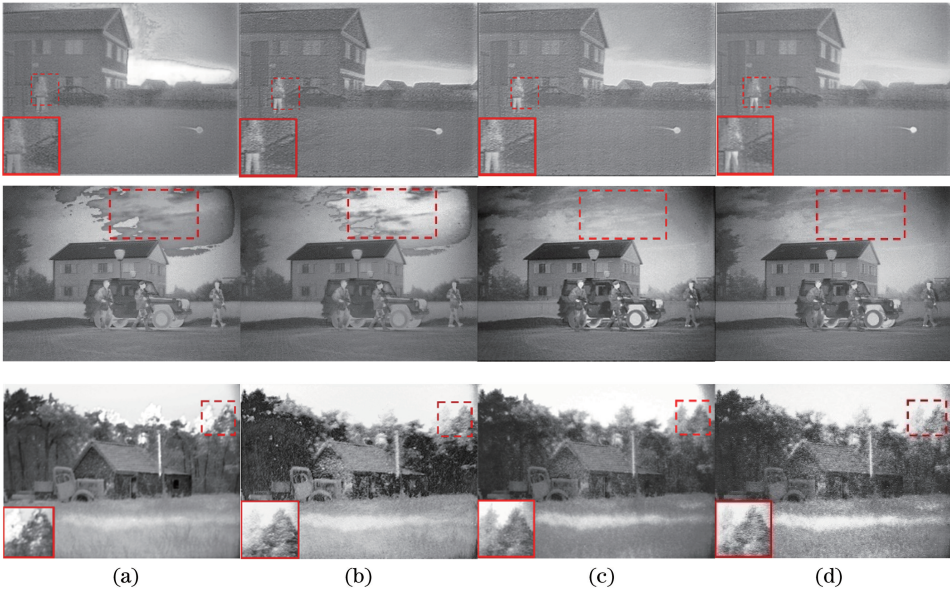


图 8 λ 取值对图像质量的影响。(a) $\lambda=0$;(b) $\lambda=0.01$;(c) $\lambda=0.1$;(d) $\lambda=1$

Fig. 8 Effect of different λ on image quality. (a) $\lambda=0$; (b) $\lambda=0.01$; (c) $\lambda=0.1$; (d) $\lambda=1$

0.01 时,第 2 组融合结果出现了失真,随着 λ 增加,3 组融合结果的主观对比度整体变高,边缘变清晰,但当 $\lambda=1$ 时,效果又开始下降。图 9 展示了 λ 取不同值时生成器的损失变化,相比之下, λ 为 0.1 时损失值最小最稳定。图 10 为 λ 取值对 3 组图像融合结果客观指标的影响,雷达图面积越大,表示效果越好,由图可知, $\lambda=0.1$ 时雷达图面积最大。综合主观和客观评价,选择 $\lambda=0.1$ 。

4.3 融合结果分析

本实验的硬件平台为 NVIDIA 英伟达 TeslaM40 服务器,Matlab 版本为 R2016a,Python 版本为 3.6.3。图 11 给出 8 组测试图像系网络

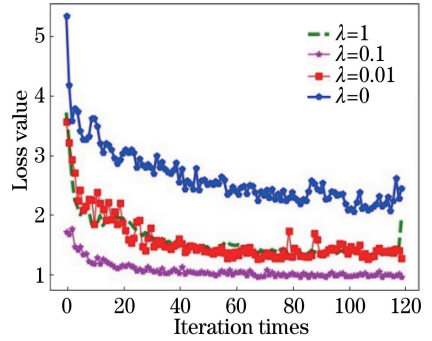


图 9 λ 取值对生成器损失的影响

Fig. 9 Effect of different λ on generator loss

训练时随机留存的融合结果。

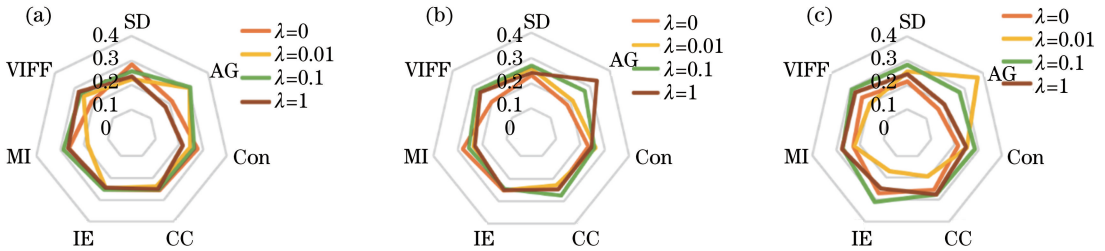


图 10 λ 取值对融合图像客观评价指标的影响。(a)第 1 组融合图像;(b)第 2 组融合图像;(c)第 3 组融合图像

Fig. 10 Effect of λ on objective evaluation index of fused image. (a) The first set of fused images; (b) the second set of fused images; (c) the third set of fused images

由于多尺度变换融合方法有很多,各有优劣,这里选择文献中认为效果超过单一多尺度变换方法的混合式融合方法及其他深度学习方法进行比较^[4,11-14],包括基于双树复小波变换和稀疏结合的 DTCWT_SR 方法、基于 NSST 与 NSCT 结合的

NSST_NSCT,以及基于 CNNs 的 CNN 方法、基于卷积稀疏表示的卷积稀疏表示(CSR)神经网络方法。

观察实验结果可知,DTCWT_SR 和 CNN 方法得到的图像目标更突出,但有明显失真现象,且边缘较模糊;CSR 方法得到的图像整体偏暗,对比度较

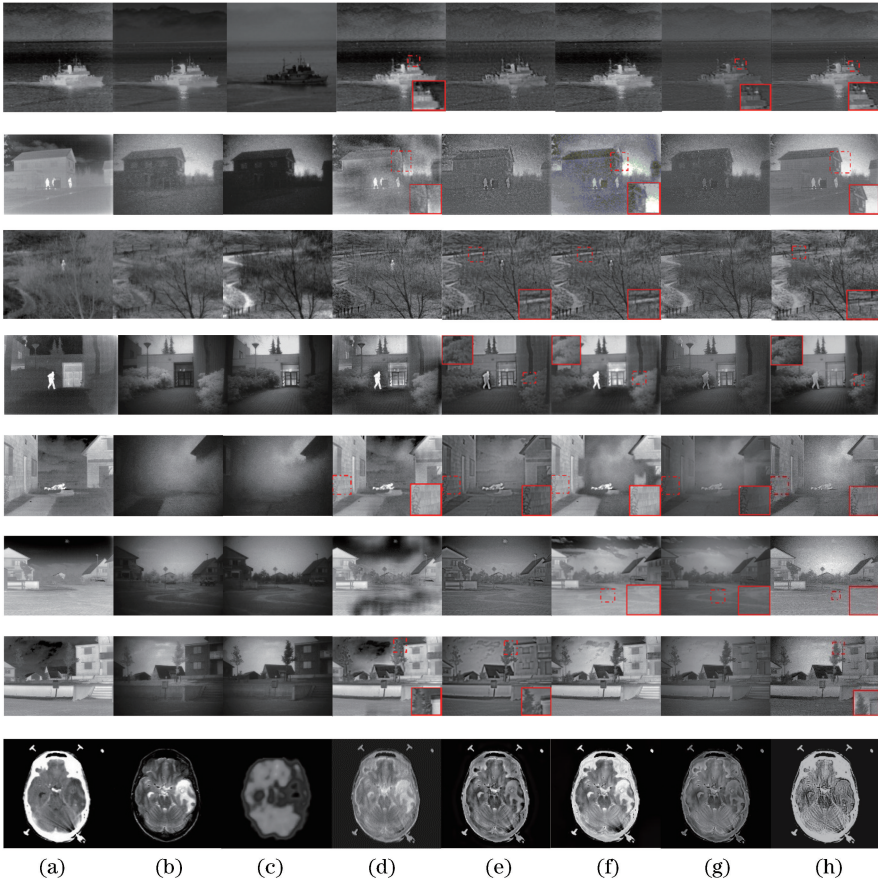


图 11 图像融合结果。(a)红外长波;(b)红外短波;(c)可见光;(d) DTCWT_SR;
(e) NSST_NSCT;(f) CNN;(g) CSR;(h)本文方法

Fig. 11 Image fusion results. (a) Longwave infrared; (b) shortwave infrared; (c) visible light; (d) DTCWT_SR;
(e) NSST_NSCT; (f) CNN; (g) CSR; (h) proposed method

低,不利于人眼观察;NSST_NSCT 方法得到的图像虽然细节较丰富,但图像中热目标出现伪影;本文方法得到的图像整体更为干净,边缘清晰且没有伪影。为使对比结果更便于观察,将有代表性的部分进行了放大,如图中红框所示。相比之下,本文方法得到的图像中船身轮廓、房屋和树桩边缘,以及草丛

纹理等细节信息都较为丰富,且图像清晰。同时,本文方法得到的医学图像中包含的细节纹理特征也更丰富。

由于人的视觉感知存在差异,故采用客观评价指标对融合结果进行评价(表 4)。这些指标的值越大,表示图像融合效果越好。

表 4 融合结果评价指标比较

Table 4 Comparison of evaluation index of fusion results

Image	Fusion method	SD	AG	Con	CC	IE	MI	VIFF
No. 1	DTCWT_SR	35.471	3.150	23.916	0.409	6.968	4.869	0.505
	NSCT_NSST	19.621	3.163	11.911	0.406	6.147	2.853	0.522
	CNN	34.928	2.885	23.254	0.408	6.961	4.610	0.363
	CSR	12.081	1.953	7.570	0.418	5.534	2.663	0.359
	Proposed method	38.011	8.129	27.793	0.431	6.977	2.552	0.301
No. 2	DTCWT_SR	26.274	7.245	25.927	0.143	5.984	1.492	0.367
	NSCT_NSST	23.268	7.270	14.815	0.306	6.026	1.542	0.376
	CNN	29.696	4.985	22.754	0.016	2.373	1.312	0.211
	CSR	25.271	5.543	15.989	0.322	6.032	2.540	0.322
	Proposed method	25.527	4.374	13.731	0.381	6.057	2.567	0.461

Image	Fusion method	SD	AG	Con	CC	IE	MI	VIFF
No. 3	DTCWT_SR	21.854	5.187	15.038	0.427	6.475	1.256	0.415
	NSCT_NSST	22.692	5.370	15.7166	0.471	6.806	1.542	0.419
	CNN	38.590	4.961	25.609	0.441	6.910	2.799	0.392
	CSR	23.054	3.868	15.726	0.499	6.450	2.391	0.382
	Proposed method	41.089	4.310	30.929	0.593	6.938	3.093	0.420
No. 4	DTCWT_SR	47.492	3.545	23.176	0.422	6.315	2.773	0.298
	NSCT_NSST	35.373	3.605	15.136	0.445	6.831	2.531	0.317
	CNN	54.597	3.163	29.778	0.436	6.283	2.862	0.351
	CSR	38.587	2.217	17.656	0.452	6.796	3.331	0.375
	Proposed method	32.400	8.452	22.767	0.457	6.938	2.055	0.384
No. 5	DTCWT_SR	40.006	5.058	32.554	0.025	7.308	2.976	0.407
	NSCT_NSST	24.886	5.147	18.059	0.489	6.615	1.812	0.428
	CNN	40.559	3.951	34.331	0.271	7.316	2.133	0.346
	CSR	26.239	2.526	20.707	0.515	6.642	3.007	0.277
	Proposed method	40.677	5.974	33.890	0.569	7.321	2.106	0.473
No. 6	DTCWT_SR	54.458	3.038	44.020	0.042	7.744	2.986	0.524
	NSCT_NSST	25.990	3.047	18.728	0.455	6.686	2.254	0.552
	CNN	45.549	2.538	35.269	0.193	7.333	1.839	0.208
	CSR	28.803	1.523	21.125	0.429	6.691	2.858	0.160
	Proposed method	46.425	5.643	36.462	0.476	7.453	2.993	0.648
No. 7	DTCWT_SR	54.848	3.555	36.984	0.044	7.420	3.485	0.484
	NSCT_NSST	27.527	3.550	21.490	0.460	6.813	1.807	0.504
	CNN	45.880	2.945	37.129	0.334	7.104	2.627	0.440
	CSR	27.853	1.834	21.168	0.472	6.701	2.968	0.326
	Proposed method	47.024	4.259	37.523	0.500	7.487	2.461	0.606
No. 8	DTCWT_SR	55.890	2.940	47.062	0.771	4.357	3.612	0.169
	NSCT_NSST	52.610	4.872	47.647	0.868	4.832	3.906	0.380
	CNN	87.236	6.208	73.907	0.721	5.688	3.933	0.294
	CSR	54.208	4.753	47.214	0.794	4.245	3.383	0.285
	Proposed method	92.152	15.279	76.870	0.943	4.333	4.947	0.505

可以看出,在 7 项评价指标中,本文方法在第 3 组和第 8 组图像上有 6 项指标优于其他方法;在其余图像上也有 4~5 项指标优于其他方法。总体来说,本文方法在梯度、相关系数、信息熵、互信息、人眼视觉特性这些客观指标上较其他方法均有较大改观,体现出更好的综合性能。

5 结 论

提出了一种基于生成对抗网络的多波段/多模态图像融合方法,实现了端到端的自适应图像融合。研究表明:利用深度残差块替代传统 GANs 模型的卷积操作,可以提升图像信息的传递能力;利用传统的模型驱动法可以缓解深度学习中缺乏标签数据的问题。在 GANs 网络训练过程中,除了传统的对抗损失外,还引入了基于像素的 MSE 损失和图像梯度损失,使网络训练更稳定。与目前效果突出

的几种方法相比,本文方法获得的图像整体更干净,边缘强度更大,且包含更多的细节信息。需要指出的是,由于标签是选用融合效果较好的图像进行分块得到的,因此标签图像的质量限制了融合结果的质量,下一步将重点改进此问题。

参 考 文 献

- [1] Ma J Y, Ma Y, Li C. Infrared and visible image fusion methods and applications: a survey [J]. Information Fusion, 2019, 45: 153-178.
- [2] Ranchin T, Wald L. The wavelet transform for the analysis of remotely sensed images[J]. International Journal of Remote Sensing, 1993, 14(3): 615-619.
- [3] Kingsbury N. A dual-tree complex wavelet transform with improved orthogonality and symmetry properties [C] // Proceedings 2000 International Conference on Image Processing (Cat. No. 00CH37101),

- September 10-13, 2000, Vancouver, BC, Canada. New York: IEEE, 2000: 375-378.
- [4] Liu Y, Liu S P, Wang Z F. A general framework for image fusion based on multi-scale transform and sparse representation[J]. *Information Fusion*, 2015, 24: 147-164.
- [5] Yi W, Zeng Y, Yuan Z. Fusion of GF-3 SAR and optical images based on the nonsubsampling contourlet transform[J]. *Acta Optica Sinica*, 2018, 38(11): 1110002.
易维, 曾湧, 原征. 基于 NSCT 变换的高分三号 SAR 与光学图像融合[J]. *光学学报*, 2018, 38(11): 1110002.
- [6] Hu J W, Li S T. The multiscale directional bilateral filter and its application to multisensor image fusion [J]. *Information Fusion*, 2012, 13(3): 196-206.
- [7] Ding W S, Bi D Y, He L Y, *et al.* Fusion of infrared and visible images based on shearlet transform and neighborhood structure features [J]. *Acta Optica Sinica*, 2017, 37(10):1010002.
丁文杉, 毕笃彦, 何林远, 等. 基于剪切波变换和邻域结构特征的红外与可见光图像融合 [J]. *光学学报*, 2017, 37(10): 1010002.
- [8] Zhang Q, Liu Y, Blum R S, *et al.* Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: a review[J]. *Information Fusion*, 2018, 40: 57-75.
- [9] Zhu D R, Xu L, Wang F B, *et al.* Multi-focus image fusion algorithm based on fast finite shearlet transform and guided filter [J]. *Laser & Optoelectronics Progress*, 2018, 55(1): 011001.
朱达荣, 许露, 汪方斌, 等. 基于快速有限剪切波变换与引导滤波的多聚焦图像融合算法[J]. *激光与光电子学进展*, 2018, 55(1): 011001.
- [10] Liu Y, Chen X, Wang Z F, *et al.* Deep learning for pixel-level image fusion: recent advances and future prospects[J]. *Information Fusion*, 2018, 42: 158-173.
- [11] Liu Y, Chen X, Peng H, *et al.* Multi-focus image fusion with a deep convolutional neural network[J]. *Information Fusion*, 2017, 36: 191-207.
- [12] Lin S Z, Han Z. Images fusion based on deep stack convolutional neural network[J]. *Chinese Journal of Computers*, 2017, 40(11): 2506-2518.
蔺素珍, 韩泽. 基于深度堆叠卷积神经网络的图像融合[J]. *计算机学报*, 2017, 40(11): 2506-2518.
- [13] 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, 2018: 2705-2710.
- [14] 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.
- [15] 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 (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.
- [16] Jiang X H, Pang Y W, Li X L, *et al.* Deep neural networks with elastic rectified linear units for object recognition[J]. *Neurocomputing*, 2018, 275: 1132-1139.
- [17] Cai J R, Gu S H, Zhang L. Learning a deep single image contrast enhancer from multi-exposure images [J]. *IEEE Transactions on Image Processing*, 2018, 27(4): 2049-2062.
- [18] Zhang H, Dana K. Multi-style generative network for real-time transfer [M] // Leal-Taixé L, Roth S. *Computer Vision-ECCV 2018 Workshops. Lecture notes in computer science*. Cham: Springer, 2019, 11132: 349-365.
- [19] Goodfellow I J, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets[C] // *Proceeding NIPS'14 Proceedings of the 27th International Conference on Neural Information Processing Systems*, December 08-13, 2014, Montreal, Canada. Cambridge: MIT Press, 2014, 2: 2672-2680.
- [20] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J/OL]. (2016-01-07) [2018-12-25]. <https://arxiv.org/abs/1511.06434>.
- [21] Mao X D, Li Q, Xie H R, *et al.* Least squares generative adversarial networks [C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2813-2821.
- [22] Ledig C, Theis L, Huszár F, *et al.* Photo-realistic single image super-resolution using a generative adversarial network[C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 105-114.
- [23] Li Y, Wang N, Shi J, *et al.* Adaptive Batch Normalization for practical domain adaptation [J]. *Pattern Recognition*, 2016, 80: 109-117.

- [24] Wang S H, Phillips P, Sui Y X, *et al.* Classification of alzheimer' s disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling [J]. *Journal of Medical Systems*, 2018, 42(5): 85.
- [25] Huang R, Zhang S, Li T Y, *et al.* Beyond face rotation: global and local perception GAN for photorealistic and identity preserving frontal view synthesis [C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2458-2467.
- [26] Shi J G, Liu X, Zong Y, *et al.* Hallucinating face image by regularization models in high-resolution feature space [J]. *IEEE Transactions on Image Processing*, 2018, 27(6): 2980-2995.
- [27] Toet A. The TNO multiband image data collection [J]. *Data in Brief*, 2017, 15: 249-251.
- [28] Huang F S. Comparison of multiscale transform fusion methods for multiband image [D]. Taiyuan: The North University of China, 2018.
黄福升. 多波段图像多尺度变换融合方法比较 [D]. 太原: 中北大学, 2018.
- [29] Jagalingam P, Hegde A V. A review of quality metrics for fused image [J]. *Aquatic Procedia*, 2015, 4: 133-142.