

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 (Generative Adversarial Networks, GANs) 的图像融合方法, 实现了多模态图像端到端自适应融合。首先将多模态源图像同步输入生成网络, 网络结构采用本文提出的一种基于残差的卷积神经网络, 通过网络自适应学习生成融合图像; 再将该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。通过与目前具有代表性的融合方法相比, 实验结果表明, 所提出方法的融合结果更干净且没有伪影, 提供了更好的视觉质量。

关键词 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

Multi-Modal Image Fusion Based On Generative Adversarial Networks

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

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

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

Abstract In order to solve the excessive reliance on prior knowledge problems of multi-scale geometric tool selection and the designing fusion rules in multi-model image fusion, a new network based on Generative Adversarial Networks (GANs) is proposed, which can achieve end-to-end image fusion. Firstly, the multi-modal source image is synchronously input into the generative network which structure is created by a residual-based convolutional neural network proposed in this paper. The network can generate the fused image through adaptive learning. Secondly, the fused image and the label image are respectively sent to the discriminant. Through the feature representation and classification identification of the discriminator network, gradually optimizes the generator. And then the final fused image is obtained in the dynamic balance of the generator and the discriminator. Compared with the existing representative fusion methods, the experimental results demonstrate that the fusion results of the proposed algorithm are cleaner and have no artifacts, providing better visual quality.

Key words image fusion; multi-modal image; deep learning; generative adversarial networks

OCIS codes 100.4996; 100.2960;

收稿日期: 年-月-日; 修回日期: 年-月-日; 录用日期: 年-月-日

基金项目: 山西省重点研发计划资助项目 (201603D321128); 山西省应用基础研究项目 (201701D121062);

中北大学第十五届科技立项资助项目 (2018514)

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

1 引 言

多波段/多模态成像系统在军事、医疗、工业检测等众多领域已有广泛应用，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（Non-Subsampled Shear Transform, 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。此类方法需依赖先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。由于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，其更具有优越性。文献[11]将卷积神经网络（Convolutional neural networks, CNNs）用于图像融合，其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，但此方法仅适用于多聚焦图像融合且仅使用 CNNs 最后层计算的结果，中间层获得的许多有用信息丢失，当网络加深时，信息丢失往往更严重；文献[13]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且虽然利用了 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用于多波段/模态图像融合，该网络由生成网络（generator）和判别网络（discriminator）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，实现融合图像融合质量的提升。我们的主要创新点有：1）图像特征提取和融合一并通过网络学习实现，未经人为设定活动水平测量和融合规则，减少了对先验知识的依赖和对融合策略选择的不确定性；2）

使用生成对抗网络框架（Generative Adversarial Networks, 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 CVPR 2016 中，He 等人^[15] 提出了一种新颖的网络架构来解决网络退化问题。通过捷径方式连接和残差表示，网络比以前更容易优化并通过增加深度来提高精度。残差块体系结构如图 1 所示。

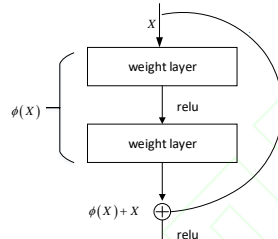


图1 残差块结构图

Fig.1 The architecture 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。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，通过使用残差块，取得了更好的性能。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 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^[19]。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，判别器用 $D(x)$ 表示来自标签而非输入数据的概率，生成器和判别器在动态博弈中达到平衡，遵循函数 $V(D, G)$ ，如公式（1）所示，实现生成功能。

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log(D(x))] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

其中， $p_{z(z)}$ 为输入噪声变量先验概率， $p_{data}(x)$ 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 G 和判别器 D 不能很好地同步训练，因此易造成

GANs 训练不稳定。之后各种 GANs 变体被相继提出，如深度卷积 GANs (Deep convolutional GANs, DCGANs) [20]、最小二乘 GANs (Least squares GANs) [21]等。DCGANs 将 CNN 成功引入 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]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做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

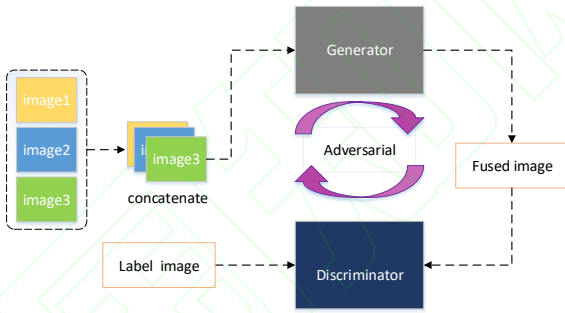


图2 方法框架图

Fig.2 Framework of the method

本文方法框架如图 2 所示，我们在训练网络时，将多波段/多模态图像以一幅图像多通道形式输入生成网络，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，最终输出高质量融合图像。由于图像融合没有标准输出，本文提出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 所示。将多波段图像在通道维度上通过 concatenate 函数联接输入神经网络，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 $128 \times 128 \times 3$ 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，减少对源图像信息的丢失。每个卷积层之后使用修正线性单元 (Rectified Linear Unit,

ReLU)^[16]，提升了网络的非线性程度。

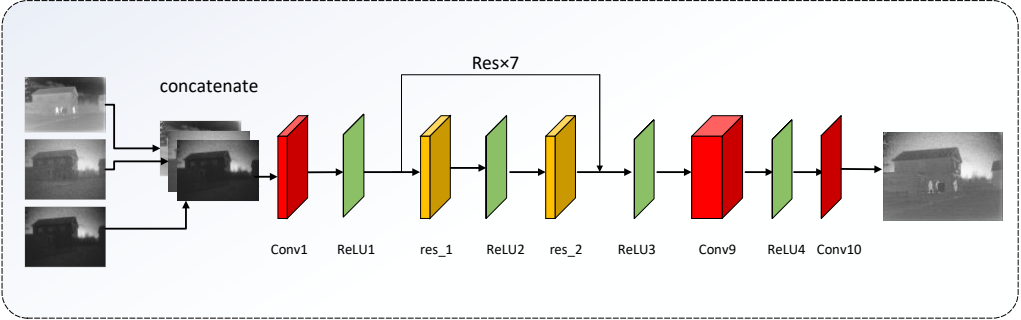


图3 生成模型网络结构

Fig.3 Generative model network architecture

表1 生成器参数

Table1 Generator parameter

Layer	Filter Size/step	Output Szie
Conv1	3 × 3 /1	128×128×64
Res(7 units)	3 × 3 /1	128×128×64
	3 × 3 /1	128×128×64
conv9	3 × 3 /1	128×128×256
conv10	3 × 3 /1	128×128×1

表2 判别器参数

Table 2 Discriminator parameter

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

3.3 GANs 判别模型的构建

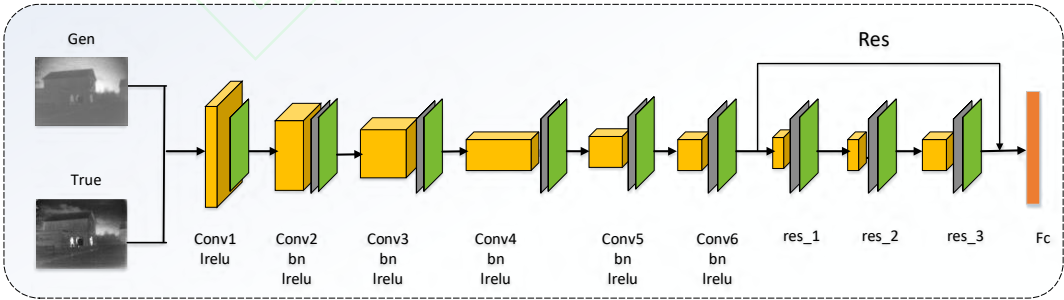


图4 判别模型网络结构

Fig.4 Discriminative model network architecture

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，分别对生成图像和真实图像进行深度特征提取。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卷积核特征提取。鉴于网络结构较深，除Conv1外，每层卷积之后都使用BatchNorm（BN）^[23]给数据归一化，BN后，使用LeakyReLU（LReLU）^[24]激活函数提升网络非线性程度。具体网络参数如上表2所示。

3.4 GANs 损失函数的构建

损失函数用来衡量网络输出和数据标签之间的差距。最简单的是均方误差，有一定效果，但是不能“智能”地学习目标细节，使得生成图片稍放大后变模糊。为此本文引入梯度损失，优化图像边缘，再加入 GAN 对抗损失。为证明梯度损失对图像影响，这里设置权重提高融合图像质量。

（1）均方误差（Mean Square Error, MSE）损失

MSE 损失^[25]通过网络训练过程中输出的预测值与真实数据之间的差值促进多尺度图像内容的一致性，计算公式如（2）所示，其中， W 、 H 分别表示图像的宽和高， $I_{x,y}^{pred}$ 是生成图像点 (x,y) 的像素值， $I_{x,y}^{true}$ 表示真实图像点 (x,y) 像素值。

$$L_{pixel} = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H |I_{x,y}^{pred} - I_{x,y}^{true}| \quad (2)$$

（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，故本文引入梯度损失来优化边缘，具体实现方法如公式（3）所示，其中， A 表示原图像，矩阵表示的是 sobel 算子， G_x, G_y 分别代表横纵边缘检测的灰度值， $G_{x,y}^{pred}$ 表示生成图像点 (x,y) 的梯度， $G_{x,y}^{true}$ 表示真实图像点 (x,y) 的梯度。

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * A, \quad G_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * A, \quad G = |G_y| + |G_x|$$

$$L_{grad} = \frac{1}{W \times H} \sum_{x=1}^W \sum_{y=1}^H |G_{x,y}^{pred} - G_{x,y}^{true}| \quad (3)$$

（3）对抗损失

将对抗损失^[21]加入到感知损失中，鼓励生成器适应真实数据的分布，计算如公式（4）所示，其中， $D(G(I^{pred}))$ 表示重构图像 $(G(I^{pred}))$ 是真实图像的估计概率， N 表示图像像素个数。

$$L_{gen} = \sum_i^N -\log D(G(I^{pred})) \quad (4)$$

（4）联合损失函数

生成器最终的损失函数是以上三项损失之和，计算如公式（5）所示，最后对结果施加 L2 正则化^[26]，用以约束要优化的参数，从而防止模型过拟合。

$$L_{gan}=L_{pixel}+L_{gen}+\lambda L_{grad} \quad (5)$$

4 分析与讨论

4.1 数据集制作

本文 GANs 网络的数据集有 14,000 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（390-700nm）、近红外（700-1000 nm）和红外长波（8-12μm）。医学图像来自 Havard Medical School 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（Laplacian Pyramid,LP）、DWT、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会有多个融合图像，我们使用包含基于统计特性的评价指标：标准差 (Standard Deviation,SD)、平均梯度（Average Gradient,AG）、对比度（Contrast, Con）、相关系数(Correlation coefficient, cc)；基于信息量的评价指标：信息熵（Information Entropy,IE）和互信息(Mutual Information,MI);表示人类视觉感知的视觉信息保真度（Visual Information Fidelity for Fusion, VIFF ）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。

我们从 14, 000 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，通过上述四种传统融合方法对三波段源图像进行融合，然后使用上述所提 7 个质量评价指标对这四幅图像进行评价，结果如表 3 所示，观察表格可以看出，LP 方法融合的图像有 5 个指标值优于其它三种方法，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 14,000×3 张多模态切块图像，它们用 concatenate 联接，以一幅图像的三通道形式输入网络，标签为 14,000 张融合图像。网络模型训练好后，我们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。

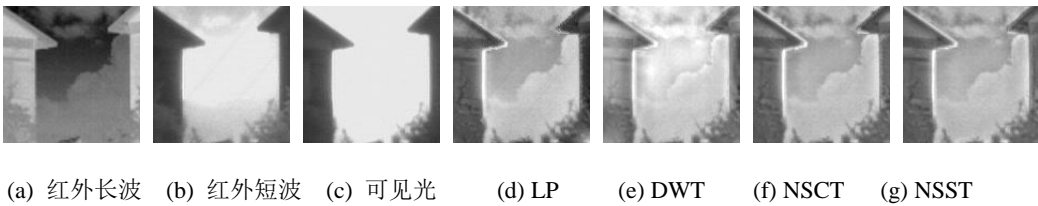


图5 标签图像预选图

Fig.5 Label image pre-selection map

表 3 标签图像选取依据表

Table 3 Label image selection table

	Fusion methods	SD	AG	Con	CC	IE	MI	VIFF
image	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。当学习率过大时会出现震荡现象，使网络无法收敛；过小时，网络迭代更新速度慢，降低训练速度。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 隔点取值，由 6 可见学习率为 0.0002 时，图中绿色线的下降很缓慢，达到相同效果耗费时间较长；蓝色虚线代表的 0.02 学习率下降很快，但出现震荡现象，影响网络收敛；而红色星线学习率 0.002，效果最好，既能很快收敛又比较稳定。在同一训练过程，若选取全部点得到图 7 结果，图 7 的蓝色虚线无法收敛；而红色线表示的 0.002 学习率比绿色线表示的 0.0002 更快收敛。由此可知学习率若再增大损失曲线只会比 0.02 更震荡，固将其作为学习率上限，而低于 0.0002 时收敛则会更缓慢，因此将其作为学习率下限。

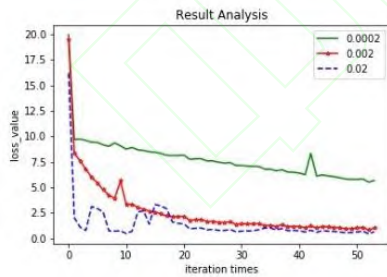


图 6 学习率对生成器损失的影响

Fig.6 The effect of learning rate on generator loss

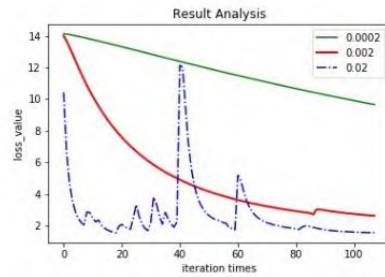


图 7 学习率对判别器损失的影响

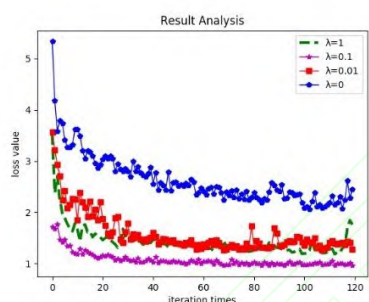
Fig.7 The effect of learning rate on discriminator loss

(2) λ 权重选取

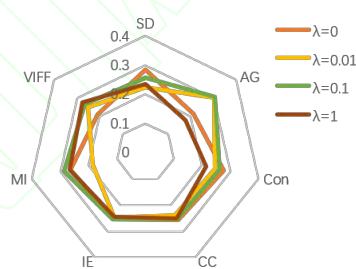
借鉴文献[22]及文献[25]实验成果，我们将梯度损失的权重 λ 分别从 0、0.01、0.1、1 进行测试，现从大量实验数据中选取有代表性的三组图像进行说明，实验结果如图 8 所示。



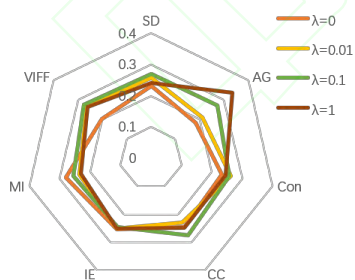
(a) $\lambda = 0$ (b) $\lambda = 0.01$ (c) $\lambda = 0.1$ (d) $\lambda = 1$



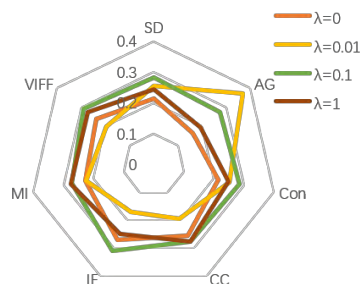
(e) λ 大小对生成器损失的影响
The effect of size λ on generator loss



(f) λ 大小对第一组融合图像客观指标影响
The influence of λ value on the objective index of the first set of fused images



(g) λ 大小对第二组融合图像客观指标影响
The influence of λ value on the objective index of the second set of fused images



(h) λ 大小对第三组融合图像客观指标影响
The influence of λ value on the objective index of the third set of fused images

图 8 λ 取值对 GANs 的影响

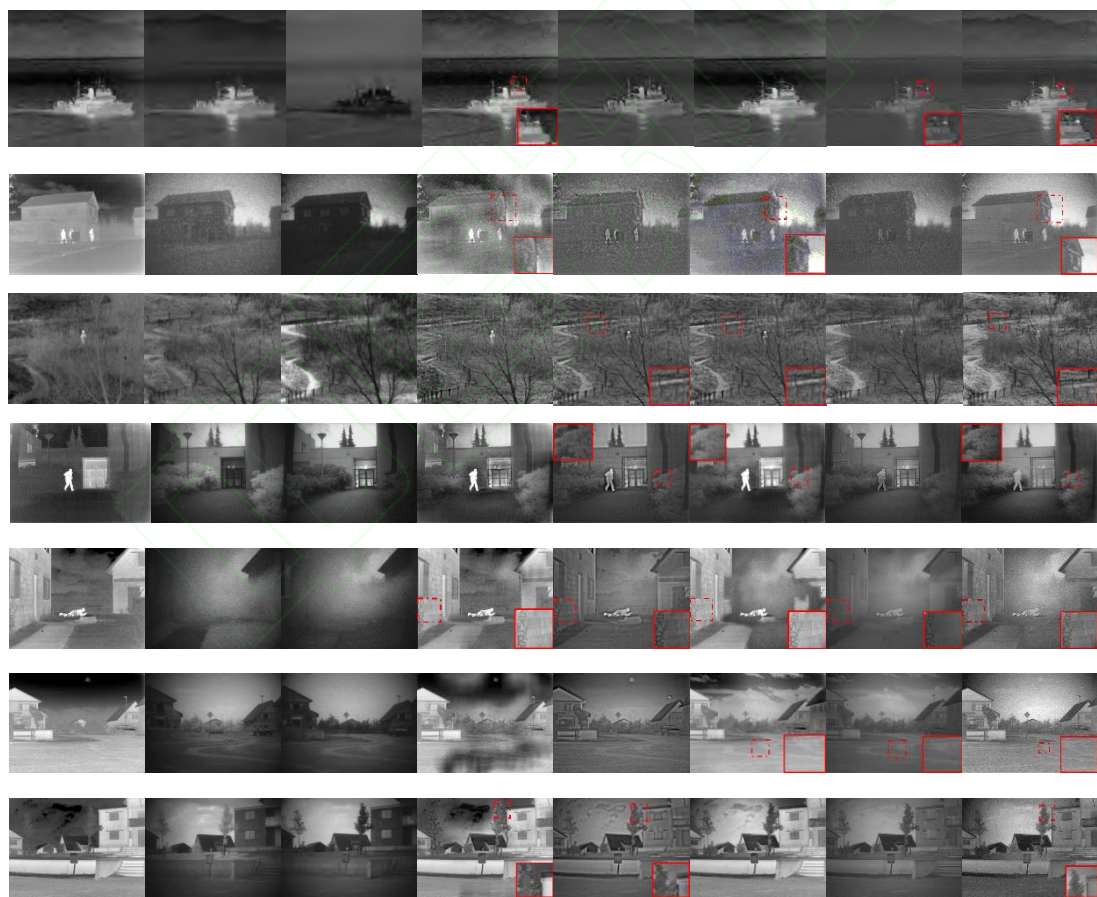
Fig.8 The effect of different λ on GANs

其中, 图 8(a-d) 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,

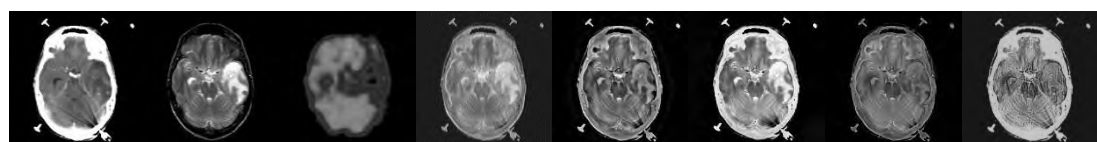
由图可以看出, 当 λ 为 0 和 0.01 时, 第二组融合结果出现了失真, 随着 λ 的增加, 三组融合结果的主观对比度整体变高, 边缘变清晰, 但当达到 1 时, 效果又开始下降; 图 8 (e) 为 λ 不同值时生成器损失变化, 由图可知, 相比之下, λ 为 0.1 时损失值最小最稳定; 图 8 (f-h) 为不同 λ 取值对三组图像融合结果的客观指标影响, 雷达图面积越大, 效果越好, 由图可知 λ 值为 0.1 时面积最大。因此综合主观和客观评价选择 $\lambda=0.1$ 。

4.3 融合结果分析

实验部分的 8 组测试图像由网络训练时随机留存。由于多尺度变换融合方法很多, 各有优劣, 这里与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以及其它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、基于 NSST 与 NSCT 结合的 NSST_NSCT 方法以及基于 CNNs、卷积稀疏表示的 CSR 神经网络方法, 实验结果如图 9 所示。



(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)Our method



(a)CT (b) MRI_T2 (c)SPECT_Tc (d) DTCWT_SR(e)NSST_NSCT (f)CNN (g)CSR (h) Our method

图 9 图像融合结果

Fig 9 image fusion result

观察实验结果可知，DTCWT_SR 和 CNN 方法目标更突出，但图像有明显失真现象且边缘较模糊；基于 CSR 的方法图像整体偏暗，对比度较低，不利用人眼视觉观察；基于 NSST_NSCT 融合法虽然细节较为丰富，但图像中热目标出现伪影；相比之下本文方法获得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：相比之下本文方法的船身轮廓、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；对于医学图像，本文方法包含的细节纹理特征更为丰富。

由于人的视觉感知存在差异，现通过前文提到的几组客观评价指标对融合结果进行评价，以上这些指标值越大，表示图像融合效果越好，具体指标值如表 4 所示。

表 4 融合结果评价指标比较

Table 4 Comparison of evaluation index of fusion results

	Fusion methods	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
	Our 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
	Our method	25.527	4.374	13.731	0.381	6.057	2.567	0.461
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
	Our method	41.089	4.310	30.929	0.593	6.938	3.093	0.420
	DTCWT_SR	47.492	3.545	23.176	0.422	6.315	2.773	0.298

No.4	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
	Our 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
	Our 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
	Our 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
	Our 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
	Our method	92.152	15.279	76.870	0.943	4.333	4.947	0.505

综合以上实验结果可看出，在 7 个评价指标中，文本方法第三组和第八组图像在 6 个指标上均优于其它方法；其余组图像均在 4-5 个指标上超过其它方法。总的来说，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均有较大改观，体现出更好的综合性能。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。

5 结论

本文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,引入基于像素的 MSE 损失和图像梯度损失,会使网络训练更稳定。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,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] Thierry R, Lucien W. 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]//International Conference on Image Processing, 2000: 375-378.
- [4] Liu Y, Liu S, Wang Z. 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 Nonsubsampled 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, Li S. The multiscale directional bilateral filter and its application to multisensory 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] Qiang Z, Yi L, Rick S.B, *et al.* Sparse representation based muti-sensor image fusion for multi-focus and multi-modality images:a review[J]. Information Fusion, 2018, 40: 47-75.
- [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(10): 011001.
朱达荣,许露,汪方斌,等. 基于快速有限剪切波变换域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(10): 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]//International Conference on Pattern Recognition, 2018.
- [14] Liu Y, Chen X, Ward R K, *et al.* Image Fusion With Convolutional Sparse Representation. IEEE Signal Processing Letters[J], 2016, 23(12): 1882-1886.
- [15] He K, Zhang X, Ren S, *et al.* Deep Residual Learning for Image Recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770778.
- [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.
- [17] Cai J, Gu S, Zhang L. Learning a deep single image contrast enhancer from multi-exposure images[J], IEEE Trans on Image Process, 2018, 27(4): 20492062.
- [18] Zhang H, Dana K. Multi-style Generative Network for Real-time Transfer. arXiv1703.06953, 2017.
- [19] Goodfellow I J, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets[C]//International Conference on Neural Information Processing Systems: MIT Press, 2014: 2672-2680.
- [20] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv:1511.06434v1, 2015.
- [21] Mao X D, Li Q, Xie H, *et al.* Least Squares Generative Adversarial Networks[C]//IEEE Conference on Computer Vision, 2017: 2813-2821.
- [22] Ledig C, Theis L, Huszár F, *et al.* Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network[C]//IEEE Conference on Computer Vision and Pattern Recognition: IEEE Computer Society, 2017: 105-114.
- [23] Santurkar S, Tsipras D, Ilyas A, *et al.* How Does Batch Normalization Help Optimization? : No, It Is Not About Internal Covariate Shift[J], 2018.
- [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]//IEEE International Conference on Computer Vision, 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 Trans on Image Processing, 2018, pp(99): 1-1.
- [27] Toet A. The TNO multiband image data collection[J]. Data in Bridf, 2017, 15: 249-251.
- [28] Huang F S. Comparision of multiscale transform fusion methods for multiband images[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[C]//International conference on water resources coastal and ocean engineering, 2015: 133-142.

网络首发:

标题: 基于生成对抗网络的多模态图像融合

作者: 杨晓莉, 蔺素珍, 禄晓飞, 王丽芳, 李大威, 王斌

收稿日期: 2019-01-09

录用日期: 2019-03-22

DOI: 10.3788/lop56.161004

引用格式:

杨晓莉, 蔺素珍, 禄晓飞, 王丽芳, 李大威, 王斌. 基于生成对抗网络的多模态图像融合[J]. 激光与光电子学进展, 2019, 56(16): 161004.

网络首发文章内容与正式出版的有细微差别, 请以正式出版文件为准!

您感兴趣的其他相关论文:

基于深度学习航拍图像检测的梯度聚类算法

解博 朱斌 张宏伟 马旗 张扬

国防科技大学脉冲功率激光技术国家重点实验室, 安徽 合肥 230037

激光与光电子学进展, 2019, 56(6): 061007

基于深度学习和人眼视觉系统的遥感图像质量评价

刘迪 李迎春

航天工程大学电子与光学工程系, 北京 101416

激光与光电子学进展, 2019, 56(6): 061101

基于改进的全卷积神经网络的资源三号遥感影像云检测

裴亮 刘阳 谭海 高琳

辽宁工程技术大学测绘与地理科学学院, 辽宁 阜新 123000

激光与光电子学进展, 2019, 56(5): 052801

基于卷积神经网络的遥感图像目标检测

欧攀 张正 路奎 刘泽阳

北京航空航天大学仪器科学与光电工程学院, 北京 100191

激光与光电子学进展, 2019, 56(5): 051002

深度卷积网络压缩算法在焊缝识别中的应用

刘美菊 运勃

沈阳建筑大学信息与控制工程学院, 辽宁 沈阳 110168

激光与光电子学进展, 2019, 56(5): 051004