红外与可见光图像融合的 U-GAN 模型

陈卓¹, 方明^{1,2}, 柴旭¹, 付飞蚬¹, 苑丽红¹

(1. 长春理工大学 计算机科学技术学院, 吉林 长春 130022; 2. 长春理工大学 人工智能学院, 吉林 长春 130022)

摘 要: 红外与可见光图像进行融合是解决单一传感器成像不足的有效手段, 目的是得到适合人眼并有利于下一步应用和处理的融合图像。为解决大部分方法特征提取不全面, 细节纹理丢失及公共数据集样本较少不利于训练等问题, 提出一种用于图像融合的端到端网络结构。将 U-net 特有的卷积结构用于图像融合, 最大程度地提取并保留源图像的重要特征信息。再通过生成对抗网络得到最后的融合结果, 将 U-net 提取的特征输入生成器与包含红外图像的鉴别器进行对抗, 得到训练模型。实验结果表明, 所提算法能够得到轮廓清晰、纹理突出、目标明显的融合图像, SD、SF、SSIM、AG 等指标明显得到提升。

关 键 词: 图像融合; U-net 特征提取; 生成对抗网络; 红外图像
中图分类号: TG391   **文献标志码:** A   **文章编号:** 1000-2758(2020)04-0904-09

红外探测器对温度敏感、能够在可见光不足的条件下对物体成像, 但一般图像对比度较低, 成像质量不高, 视觉分辨能力差; 可见光探测器对物体的分辨率较好, 但需要更多的可见光光线才能获得理想的成像效果。将两者进行融合是低照度环境下获得高质量图像的有效手段。近年来, 图像融合在目标识别、目标检测、图像增强、监控和遥感等方面均有广泛应用^[1-2]。

随着融合需求不断增加, 许多思想被相继提出, 并取得较好的融合结果。主要分为两大类, 第一类为传统方法, 通过对图像像素分布的情况进行统计与分析, 主要有多尺度变换、小波变换、显著性分析等^[3]。林子慧等^[4]提出一种基于显著性的图像融合方法, 将 FT 算法进行改进, 提取红外图像的显著性, 通过自适应直方图均衡化, 增强可见光图像的对比度, 再将 NSCT 分解后的低频与高频部分进行融合, 取得较好的视觉效果。刘琰煜等^[5]提出一种低秩表示和字典学习的融合算法, 通过获取源图像低秩和稀疏成分, 更好地表示细节部分, 用最大范数规则提取目标, 重建后得到最终融合图像, 融合效果较为显著。第二类为基于深度学习的图像融合, 主要

是利用神经网络的特有结构对图像进行特征提取, 依据其强大的计算能力和处理速度, 可以有效避免传统方法中多尺度分解思想导致数据量增多、只根据图像的单一特征进行融合导致图像信息提取不全面的问题^[6]。冯玉芳等^[7]将经过 NSST 分解得到的高频子带输入 FCN 模型, 高低频分别进行融合, 通过逆变换得到融合图像, 虽取得较好的实验结果, 但特征提取过程有些复杂。Prabhakar 等^[8]在 ICCV 上提出基于 CNN 的曝光融合问题。对输入的图像进行编码, 获得特征序列图, 再通过加权策略融合, 最终由解码器重建。该网络在融合中表现良好, 但深层网络的特征充分利用上略显不足。Ma 等^[9]提出一种基于生成对抗网络的图像融合框架 Fusion-GAN, 将红外图像与可见光图像进行连接后输入生成器中, 为了使融合图像中包含更多可见光图像中的细节信息, 与包含可见光图像的鉴别器进行对抗, 通过博弈论的思想, 得到最终的融合图像。Fusion-GAN 得到的融合结果可以突出目标并保留细节信息。但是对小目标细节的提取和图像整体清晰度有待提高。

基于上述问题, 提出一种用于图像融合的端到

端新型网络框架 U-GAN,本文所选数据集多为夜间场景,由于场景较暗,红外图像只能对目标进行突出,可见光图像虽然包含大量的背景信息,但不能得到较好的体现,不利于人眼观察。所以将常用于图像分割的 U-net^[10] 作为对源图像特征提取的一部分,不仅可以提取红外图像中目标的轮廓信息,又可以通过其内部多层卷积结构及其特性,多尺度提取并最大程度保留图像各层的融合特征,也能对可见光图像中包含的背景及细节信息进行提取,有利于对图像中小目标进行识别,再根据 GAN 网络中生成器与鉴别器的对抗思想,得到成像自然的融合图像。

这样,在不丢失目标信息的前提下,使融合图像增加了大量的纹理、背景信息,对夜间侦察和目标识别领域具有较好的应用前景。

1 网络结构

1.1 U-GAN 网络模型

将已配准的红外与可见光图像分别进行输入,为使融合过程中增加更多的细节及纹理信息,通过 U-net 内部特殊“U”型卷积结构进行特征的提取,如图 1 所示。

图 1 U-GAN 网络架构

将下采样过程(max pool 2×2)中得到的高像素特征在上采样过程(up conv 2×2)中与新的特征图进行结合(copy and crop),最大程度保留与之对应下采样中的一些重要特征信息。通过上述操作,达到对 2 种光谱图像信息充分利用的目的。由于生成对抗网络可以较好地保留源图像的红外信息及梯度信息、成像自然的优点。将提取到的红外与可见光图像特征,维度转换后进行拼接(concat),作为对抗网络生成器的输入,产生初始融合图像,再将红外图像放入鉴别器中。通过生成器与鉴别器的对抗,得到训练模型,用于得到最终融合图像。本文网络命名为 U-GAN,可分为特征提取和生成器模型训练两部分,下文将进行详细介绍。

1.2 U-net 红外可见光图像特征提取模型

U-net 网络是一种轻量级全卷积神经网络。其网络左边与右边呈对称结构,如图 2 所示。左侧为收缩路径,右侧为扩展路径。收缩路径由卷积层和

池化层组成,扩展路径由卷积层和反卷积层组成。分别输入红外与可见光图像(输入设置为 240×240×1),经收缩路径中 4 个卷积和池化组合模块对红外与可见光图像的特征进行充分提取,每个模块包含 2 次 3×3 的卷积操作,步长设置为 1,一个 2×2 的最大池化层,其步长设置为 2,用于减少网络参数,并在每个卷积层使用 Relu 激活函数对源图像进行下采样操作,同时,通道数也会逐层增加一倍(64, 128, 256, 512)。再经 2 次 3×3 卷积操作后,特征图大小为 15×15,通道数为 1 024。由于下采样会降低图像分辨率,再利用扩展路径中的 4 个卷积层与反卷积层组合模块对其进行上采样操作,每个模块包含 1 个 2×2 卷积层及 2 个 3×3 反卷积层,并在上采样过程中,与其对应收缩路径上相同尺寸的特征图进行拼接。以此避免池化过程中细节信息的丢失。最后,通过 1×1 卷积核和 sigmoid 激活函数,得到输出特征。上采样卷积操作后,通道数也逐层减少

50% (512, 256, 128, 64), 得到的输出参数与输入相同。通过 U-net 红外可见光图像特征提取网络, 对输入源图像特征进行充分利用及融合, 作为后面处

理的输入, 目的是在生成器内部卷积网络处理之前, 就已经得到融合特征向量, 对取得较好的融合结果有重要作用。

图 2 U-net 特征提取模型

1.3 生成器训练模型

由于生成对抗网络博弈论的思想, 使生成器中的像素分布逐渐趋向于鉴别器输入的图像, 产生的融合结果更接近真实图像。所以将用于图像融合的 FusionGAN 基础结构用于最终图像生成。由于 U-net 网络可对任意大小图像进行卷积操作, 适用于样本较少、内容复杂的图像集, 符合本文选用小样本数据集。FusionGAN 架构如图 3a) 所示, 通过对图像进行分块处理, 达到增加数据集的目的。将 U-net 模型提取的红外与可见光图像的融合特征向量, 作为生成器的输入, 如图 3b) 所示。由于原始可见光图像包含的信息更丰富, 通过 U-net 特征提取模型的多层卷积操作后, 生成器产生的初始融合图像中可见光图像信息比重较大, 红外图像特征较少。基于此, 将红外图像作为鉴别器的输入。使对抗训练过程中, 初始融合图像逐渐包含更多红外信息以起到突出目标和提高融合图像亮度的作用。生成器具有 5 个卷积层, 4 个批量化标准层, 4 个 leaky relu 激活层及 tanh 激活层。鉴别器具有 4 个卷积层, 1 个线性层, 4 个批量归一化层及 4 个 leaky ReLU 激活

图 3 生成器训练模型

层。经训练后得到包含 U-net 特征提取的生成器模型,输入红外与可见光图像,即可得到最终的融合图像。

2 实验结果及分析

为了说明算法的有效性和优势,采用军事场景公开数据集 TNO image fusion dataset^[11]和针对不同环境下拍摄的可见光和红外视频数据集 INO^[9]进行对比实验,TNO image fusion dataset 为 39 对红外与可见光图像,INO database 为视频序列,选用其中包含红外与可见光视频的 INO_ClosePerson、INO_MainEntrance、TNO_TreeAndRunner、INO_VisitorPark、INO_ParkingSnow,对其进行帧拆分,共 298 对红外与可见光图像进行实验。本研究实验环境为: Intel(R) xeon(R) CPU E31245@ 3.30GHZ,16G 运行内存,NVIDIA Geforce GTX1060,ubuntu16.04,64 位操作系统,Python 3.7 版本。初始迭代次数为 10,学习率设为 1×10^{-4} 。

2.1 FusionGAN 与 U-GAN 对比结果

一般对图像融合算法的评价可以分为主观评价和客观评价。主观评价根据人的主观意识进行评判,属于定性评估。客观评价根据对融合的指标值大小,评判算法的优劣。属于定量评估。由于 U-GAN与 FusionGAN 方法都是基于生成对抗网络的融合框架,本文增加了 U-net 特征提取模型,用于提高源图像利用率,旨在增加图像中的梯度信息,使融合图像中的物体目标更利于观察,背景纹理更加突出。所以得到融合结果后,在针对细节、纹理的相关的指标上进行测量,分别为平均梯度(AG)、图像清晰度(FD)、边缘强度(EI)及相对标准差(RSD)。

以此对 2 种数据集验证是否在目标突出、细节增强方面有所提升。

TNO image fusion dataset 中,选择 30 张图像进行测试,从中随机选择 10 组图像,将本文方法与 FusionGAN 结果进行对比。实验结果如图 4 所示,从左至右分别为红外图像、可见光图像、FusionGAN 融合结果及本文实验结果。

通过图 4,主观上观察发现,本文方法在没有损失原始图像精度的基础上,对图像中一些较小目标细节的提取有所改进。对图中建筑物细节特征及表面的纹理能更清楚地体现,并且对于复杂场景,如空中的云及茂密的树叶,也能避免背景复杂导致的融合图片模糊的问题。多个云朵的轮廓,较小且繁密的树叶纹理特征也可以提取到(参考图 4 中第二行和第七行方框内)。并且由于 U-net 网络是常用于图像分割的基本模型,所以能对图像中目标的边缘轮廓进行较好描绘,以此达到突出目标的目的。

TNO image fusion dataset 在指标值平均梯度(AG)、图像清晰度(FD)、边缘强度(EI)、及相对标准差(RSD)对比情况如表 1 所示,所有指标值均为值越大,融合效果越好。

表 1 TNO Image Fusion dataset 指标值对比

融合方法	融合指标			
	AG	FD	EI	RSD
FusionGAN	2.381 0	3.151 7	23.842 2	0.211 0
U-GAN(本文方法)	4.873 8	6.945 5	49.949 7	0.345 1

表 1 中可以发现,本文方法在上述指标均有明显的提升,尤其是关于梯度和边缘的提取,并且也增加了图像清晰度,符合主观的判断。

图 4 TNO image fusion dataset 实验结果对比

INO database 测试样本数量为 200 张图像,由于该数据集视频样本是一段长时间针对同一场景的视频录像,图中目标相对静止,同一个视频拆分的图像相

似,所以本文在选用的 5 个视频样本中,每个视频样本只选用一张包含较多目标信息的图像与 FusionGAN 在该数据集复现结果进行对比。实验结

果如图 5 所示。

图 5 INO database 实验结果对比

从图 5 中可以发现,针对 INO 数据集中不同场景的数据集样本。本文方法的融合结果较为清晰,对比度更高,对行人、小汽车边缘轮廓的提取有明显的提升,对背景信息中树木和建筑物的体现较好。

INO database 在 AG、FD、EI、RSD 的指标值对比如表 2 所示。

表 2 INO database 指标值对比

融合方法	融合指标			
	AG	FD	EI	RSD
FusionGAN	4.084 6	4.375 6	31.235 7	0.289 7
U-GAN(本文方法)	4.528 5	4.959 7	33.053 7	0.372 2

通过表 2 验证,本文算法在 INO database 的融合效果上也有所提升,说明本文算法在边缘、纹理的提取上是有效可行的。

对 2 种算法在不同数据集的测试时间进行测量。由于对训练样本的处理方式不同,对比训练时间不具有客观性,所以,在本文实验环境下对 2 种模型的测试时间进行对比,对比结果如表 3 所示。每个值为当前数据集测试时间的平均值。

表 3 2 种数据集测试时间对比

融合方法	数据集	
	TNO	INO
FusionGAN	6.75×10^{-1}	1.00×10^{-1}
U-GAN(本文方法)	7.02×10^{-1}	1.29×10^{-1}

FusionGAN 模型进行测试时,只需通过生成器模型得到融合结果,U-GAN 需要经过 U-net 特征提取模型及生成器模型,所以测试时间有所增加。

2.2 本文方法与多种实验对比结果

FusionGAN 方法较为完善,已与多种方法进行对比,并取得了较好的融合结果,在此基础上再进行比较。如图 6 所示,从上到下依次为红外图像、可见光图像、ASR^[12]、CVT^[13]、DTCWT^[14]、FPDE^[15]、GFF^[16]、LPP^[17]、TSIFVS^[18]、GTF^[19]、FusionGAN 及本文方法融合结果。

从图 6 中可以发现,U-GAN 的融合结果在 2 种公共数据集上都取得了较好的实验结果,相比于其他方法对细节特征的提取更加丰富,对目标包含的特征也能较好地进行提取与保留,前景和背景的分更明显。如云彩中的建筑物,在没有丢失云彩特征的同时,对建筑物也能很好的体现。对树干和树叶的特征提取融合后没有发生模糊的情况,而其余方法大部分针对密集连接的小目标容易发生连成一片的弊端,不利于对物体的识别。路灯和车轮等部分在多数融合方法中可见度不高。而本文方法对图像中包含的场景信息都可以清晰地体现,对比度较高,符合人眼视觉特性。

本文又选用对融合图像整体质量评价的融合指标信息熵(EN)、标准差(SD)、结构相似性(SSIM)、及空间频率(SF),对图 6 融合结果进行定量评估。

表 4 U-GAN 与多种融合方法指标值对比结果

融合方法	融合指标			
	EN	SD	SSIM	SF
GFF	6.668 2	31.197 1	0.258 0	4.057 8
LPP	6.406 8	23.888 0	0.295 2	6.089 4
CVT	6.344 2	22.115 1	0.300 3	4.759 3
TSIFV	6.346 8	21.566 3	0.300 0	3.834 2
DTCWT	6.258 8	20.799 1	0.298 7	4.490 4
FPDE	6.153 1	19.501 8	0.294 4	3.588 9
ASR	6.142 9	19.564 2	0.293 9	3.768 0
GTF	7.019 3	49.619 4	0.240 7	3.682 0
FusionGAN	7.303 3	51.954 2	0.335 2	6.460 0
U-GAN(本文方法)	7.109 2	60.454 1	0.421 8	11.564 0

对比结果如表 4 所示,所有指标均为值越大效果越好(所有数值均为迭代 10 次的平均值)。其中指标标准差(SD)、结构相似性(SSIM)与空间频率(SF),分别比 FusionGAN 提高了 16.4%、25.8%,

79%,对于指标信息熵(EN),由于本文方法提取较多细节信息,图像整体偏暗,红外图像信息包含较

少,图像的复杂度减少,导致信息熵指标值提高不明显。

图 6 多种融合方法对比实验

(从上到下依次为红外图像、可见光图像、ASR、CVT、DTCWT、FPDE、GFF、LPP、TSIFVS、GTF、FusionGAN 及本文方法融合结果)

3 结 论

本研究中,将提出的特征提取模型与生成对抗网络进行结合,得到 U-GAN 端到端网络框架,在突出目标信息的同时,又保留了目标中一些重要的纹理细节,对可见光图像包含的背景信息也能进行较

好体现。本文还对红外与可见光视频进行拆分,丰富了数据集,在 2 种数据集上进行了对比实验。主观上看,本文融合结果边缘轮廓明显,细节纹理清晰。满足预期效果。客观上看,多数指标有明显优势,尤其是关于边缘及梯度的指标,具有大幅度提升。

参考文献:

- [1] 王峰程,咏梅,李辉. 基于 SHT 域 TAM-SCM 与聚焦区域检测的图像融合算法[J]. 西北工业大学学报, 2019, 37(1): 114-121
WANG Fengcheng, CHENG Yongmei, LI Hui. Image Fusion Algorithm of Focal Region Detection and TAM-SCM Based on SHT Domain[J]. Journal of Northwestern Polytechnical University, 2019, 37(1): 114-121 (in Chinese)
- [2] MA J Y, MA Y, LI C. Infrared and Visible Image Fusion Methods and Applications: a Survey[J]. Information Fusion, 2019, 45: 153-178
- [3] 刘战文,冯燕,李旭,等. 一种基于 NSST 和字典学习的红外和可见光图像[J].西北工业大学学报, 2017, 35(3): 408-413
LIU Zhanwen, FENG Yan, LI Xu, et al. A Fusion Algorithm for Infrared and Visible Images Based on Dictionary Learning and NSST[J]. Journal of Northwestern Polytechnical University, 2017, 35(3): 408-413 (in Chinese)
- [4] 林子慧,魏宇星,张建林,等. 基于显著性图的红外与可见光图像融合[J]. 红外技术, 2019, 41(7): 640-645
LIN Zihui, WEI Yuxing, ZHANG Jianlin, et al. Image Fusion of Infrared and Visible Images Based on Saliency Map[J]. Infrared Technology, 2019, 41(7): 640-645 (in Chinese)
- [5] 刘琰煜,周冬明,聂仁灿,等. 低秩表示和字典学习的红外与可见光图像融合算法[J]. 云南大学学报, 2019, 41(4): 689-698
LIU Yanyu, ZHOU Dongming, NIE Rencan, et al. Infrared and Visible Image Fusion Scheme Using Low Rank Representation and Dictionary Learning[J]. Journal of Yunnan University, 2019, 41(4): 689-698 (in Chinese)
- [6] 黄福升,蔺素珍. 基于拉普拉斯金字塔变换方法的多波段图像融合规则比较[J]. 红外技术, 2019, 41(1): 64-71
HUANG Fusheng, LIN Suzhen. Comparison of Multi-Band Image Fusion Rules Based on Laplacian Pyramid Transform Method [J]. Infrared Technology, 2019, 41(1): 64-71 (in Chinese)
- [7] 冯玉芳,殷宏,卢厚清,等. 基于改进卷积神经网络的红外与可见光图像融合[J/OL].(2019-08-18)[2019-09-10]. http://www.ecice06.com/CN/10.196781/j_issn_1000-3428.0055034
FENG Yufang, YAN Hong, LU Houqing, et al. Infrared and Visible Image Fusion Based on Improved Convolutional Neural Network[J/OL].(2019-08-18)[2019-09-10]. http://www.ecice06.com/CN/10.196781/j_issn_1000-3428.0055034 (in Chinese)
- [8] PRABHAKAR K R, SRIKAR V S, BABU R V. DeepFuse: a Deep Unsupervised Approach for Exposure Fusion with Extreme Exposure Image Pairs[C]//2017 International Conference on Computer Vision, Venice, Italy, 2017: 4724-4732
- [9] MA J Y, YU W, LIANG P W. FusionGAN: a Generative Adversarial Network for Infrared and Visible Image Fusion[J]. Information Fusion, 2019, 48: 11-26
- [10] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional Networks for Biomedical Image Segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich German, 2015, 234-241
- [11] TOET Alexander. TNO Image Fusion Dataset [EB/OL]. (2014-04-26) [2019-05-01]. https://figshare.com/articles/TNO_Image_Fusion_Dataset/1008029
- [12] LIU Y, WANG Z F. Simultaneous Image Fusion and Denoising with Adaptive Sparse Representation[J]. Image Processing Let, 2015, 9(5): 347-357
- [13] NENCINI F, GARZELLI A, BARONTI S, et, al. Remote Sensing Image Fusion Using the Curvelet Transform[J]. Information Fusion, 2007, 8(2): 143-156
- [14] LEWIS J J, O'CALLAGHAN R J, NIKOLOV S G, et, al. Pixel-and Region-Based Image Fusion with Complex Wavelets[J].

Information Fusion, 2007, 8(2): 119-130

- [15] BAVIRISETTI D P, XIAO G, LIU G. Multi-Sensor Image Fusion Based on Fourth Order Partial Differential Equations[C]// Proceedings of the International Conference on Information Fusion, 2017: 1-9
- [16] LI S, KANG X D, HU J W. Image Fusion with Guided Filtering[J]. IEEE Trans on Image Processing, 2013, 22(7): 2864-2875
- [17] KONG W, ZHANG L, LEI Y. Novel Fusion Method for Visible Light and Infrared Images Based on NSST-SF-PCNN[J]. Infrared Physics & Technology, 2014, 65: 103-112
- [18] BAVIRISETTI D P, DHULI R. Two-Scale Image Fusion of Visible and Infrared Images Using Saliency Detection[J]. Infrared Physics & Technology, 2016, 76: 52-64
- [19] MA J, CHEN C, LI C. Infrared and Visible Image Fusion via Gradient Transfer and Total Variation Minimization[J]. Inform Fusion, 2016, 31: 100-109

U-GAN Model for Infrared and Visible Images Fusion

CHEN Zhuo¹, FANG Ming^{1,2}, CHAI Xu¹, FU Feiran¹, YUAN Lihong¹

(1.School of Computer Science and Technology, Changchun University of Science and Technology, Changchun 130022, China;)
(2.School of Artificial Intelligence, Changchun University of Science and Technology, Changchun 130022, China)

Abstract: Infrared and visible image fusion is an effective method to solve the lack of single sensor imaging. The purpose is that the fusion images are suitable for human eyes and conducive to the next application and processing. In order to solve the problems of incomplete feature extraction, loss of details, and less samples of common data sets, it is not conducive to training, an end-to-end network architecture for image fusion is proposed. U-net is introduced into image fusion, and the final fusion result is obtained by using the generative adversarial network. Through its special convolution structure, the important feature information is extracted to the maximum extent, and the sample does not need to be cut to avoid the problem of reducing the fusion accuracy, but also to improve the training speed. Then the U-net extracted feature is confronted with the discriminator containing infrared image, and the generator model is obtained. The experimental results show that the present algorithm can obtain the fusion image with clear outline, prominent texture and obvious target. SD, SF, SSIM, AG and other indicators are obviously improved.

Keywords: image fusion; U-net feature extraction; generative adversarial network; infrared image