

doi: 10.3969/j.issn.1003-3106.2022.04.005

引用格式: 刘锃亮, 张宇, 吕恒毅. 基于生成对抗网络的可见光与红外图像融合[J]. 无线电工程, 2022, 52(4): 555-561. [LIU Zengliang, ZHANG Yu, LYU Hengyi. Visible and Infrared Image Fusion Based on Generative Adversarial Network[J]. Radio Engineering, 2022, 52(4): 555-561.]

基于生成对抗网络的可见光与红外图像融合

刘锃亮^{1,2}, 张宇¹, 吕恒毅¹

(1. 中国科学院长春光学精密机械与物理研究所, 吉林 长春 130033;

2. 中国科学院大学 光电学院, 北京 100039)

摘要: 图像融合是图像处理领域中非常重要的分支, 可见光图像与红外图像的融合在机器感知、目标检测与追踪、监控、遥感和图像去雾等方面扮演着十分重要的角色。针对目前一些融合算法时效性差、复杂程度高、泛化程度低和融合后图片信息丢失量大等问题, 在神经网络 FusionGAN 的基础上进行了改进。在其中引入了一种多尺度卷积 PSConv 和一种轻量化注意力模块 ECA-Net, 前者能够在更细粒度角度进行多尺度特征融合, 后者能自适应地选择一维卷积核大小, 从而实现性能上的提优。实验采用经典的红外与可见光数据集 TNO 和 NIO 数据集, 经实验表明, 改进后的算法在主观评价与客观评价下, 与原算法和其他算法相比有着明显提高。

关键词: 图像融合; 红外图像; 可见光图像; FusionGAN; PSConv; ECA-Net

中图分类号: TP183

文献标志码: A

开放科学(资源服务)标识码(OSID):



文章编号: 1003-3106(2022)04-0555-07

Visible and Infrared Image Fusion Based on Generative Adversarial Network

LIU Zengliang^{1,2}, ZHANG Yu¹, LYU Hengyi¹

(1. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China;

2. College of Optoelectronics, University of Chinese Academy of Sciences, Beijing 100039, China)

Abstract: Image fusion is a very important branch in the field of image processing. The fusion of visible image and infrared image plays a very important role in machine perception, target detection and tracking, monitoring, remote sensing, image defogging and so on. In order to solve the problems of some current fusion algorithms such as poor timeliness, high complexity, low degree of generalization, and large amount of information loss after fusion, the neural network FusionGAN is improved. A multi-scale convolution PSConv and a lightweight attention module ECA-Net are introduced. The former can perform multi-scale feature fusion at a more fine-grained angle and the latter can adaptatively select the size of one-dimensional convolution kernel to achieve performance optimization. Classical infrared and visible data sets TNO and NIO are used in experiment. Experimental results show that the improved algorithm is significantly improved as compared with the original algorithm and other algorithms under subjective and objective evaluation.

Keywords: image fusion; infrared image; visible image; FusionGAN; PSConv; ECA-Net

0 引言

图像融合属于图像增强技术, 目的是为了将不同的图像融合在一起生成信息丰富的图像, 从而方便进行进一步处理。为了满足人们生产生活中在各种复杂环境下的需要, 多源图像的融合引起了广泛的重视。红外图像可基于辐射差异将目标与背景区分开来, 同时可见光图像可提供符合人类视觉感知的具有高空间分辨率和清晰度的纹理细节^[1]。为

了取得令人满意的融合效果, 关键是有有效的图像信息提取和合适的融合原则。

图像融合算法目前已经发展多种不同的方案, 包括多尺度变换^[2]、稀疏表示^[3]、神经网络^[4]、子空间^[5]、混合模型^[6]和其他方法^[7]。现有方法通常在

收稿日期: 2021-12-23

基金项目: 国家青年科学基金(62005269)

Foundation Item: National Science Fund for Young Scholars of China (62005269)

融合过程中对不同的源图像使用相同的变换或表示,然而它不适用于红外和可见光图像,因为红外图像中的热辐射和可见光图像中的细节纹理是2种不同现象的表现。此外,大多数现有方法中的图像信息提取和融合规则都是手工设计的,并且变得越来越复杂,受实现难度和计算成本的限制^[8]。

受最近两年在 CVPR 会议上香港科技大学李铎、陈启峰团队^[9]提出的一种多尺度卷 PSConv 和天津大学王启龙团队^[10]改进的轻量级注意力模块 ECA-Net 的启发,在融合生成对抗网络 FusionGAN^[11]的基础上,在其残差网络中引入了 PSConv,以提升特征提取的细粒度与深度,再经过 ECA 网络增强对有用信息的收集,从而使最终融合后的图像具有更丰富的细节特征和纹理信息。

1 算法与改进

1.1 FusionGAN 原理

FusionGAN 的原理是将红外与可见光融合的过程公式化成一个对抗的过程,其结构包括生成器 G (Generate) 和判别器 D (Discriminator),训练过程与测试过程如图 1 和图 2 所示。训练时,首先将红外图像 I_r 与可见图像 I_v 叠加在一起传送给生成器 G,融合后的图像既包含了红外的热辐射信息,又保留可见光图像的梯度信息;再将生成融合后的图像 I_f 与可见图像 I_v 一同发送给判别器 D,让其区分二者;最后将辨别的结果形成一个反馈,输送回生成器 G,形成一个对抗的反馈网络。经过大量的训练之后,当判别器无法辨别 I_r 与 I_v 时,说明生成的融合图像已经达到了好的效果,训练完成。再进行测试时,只需要用到已经训练好的生成器 G 即可。

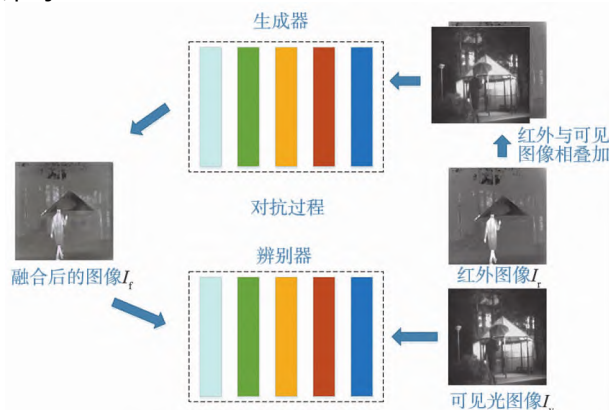


图 1 训练过程
Fig.1 Training process

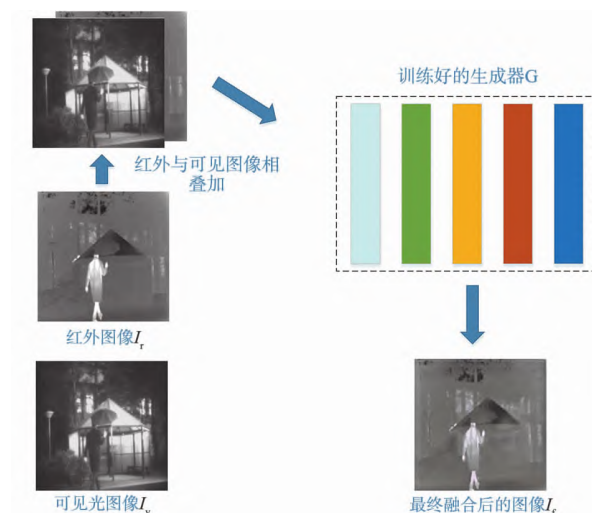


图 2 测试过程
Fig.2 Testing process

1.2 损失函数

FusionGAN 的损失函数主要包括生成器 G 的损失函数与判别器 D 的损失函数两部分。

1.2.1 生成器损失函数

生成器损失函数为:

$$\mathcal{L}_G = V_{\text{FusionGAN}}(G) + \lambda \mathcal{L}_{\text{content}}, \quad (1)$$

其主要由 2 个部分组成, $V_{\text{FusionGAN}}(G)$ 代表生成器与判别器之间的对抗损耗,即:

$$V_{\text{FusionGAN}}(G) = \frac{1}{N} \sum_{n=1}^N (D(I_f^n) - c)^2, \quad (2)$$

式中, I_f 为融合后的图像; N 为图像融合的数量, $n \in N$; c 为生成器希望判别器相信假数据的值。系数 λ 用于平衡 $V_{\text{FusionGAN}}(G)$ 与 $\mathcal{L}_{\text{content}}$, $\mathcal{L}_{\text{content}}$ 代表内容的损失:

$$\mathcal{L}_{\text{content}} = \frac{1}{HW} (\|I_f - I_r\|_F^2 + \xi \|\nabla I_f - \nabla I_v\|_F^2), \quad (3)$$

式中, $\|\cdot\|_F$ 为矩阵范数; H 和 W 为输入图像的高度和宽度; ∇ 为梯度算子; $\|I_f - I_r\|_F^2$ 为保留红外图像的热辐射信息; $\|\nabla I_f - \nabla I_v\|_F^2$ 为保留可见光图像的梯度信息; ξ 为控制 2 项之间权衡的参数。

1.2.2 判别器损失函数

判别器基于可见图像中提取的特征来区分融合图像和可见图像,使用最小二乘作为损失函数使训练过程更加稳定,损失函数收敛速度更加迅速:

$$\mathcal{L}_D = \frac{1}{N} \sum_{n=1}^N (D(I_v) - b)^2 + \frac{1}{N} \sum_{n=1}^N (D(I_f) - a)^2 \quad (4)$$

式中, a 和 b 为融合后图像和可见图像的标签; $D(I_v)$ 与 $D(I_f)$ 为可见图像和融合图像的分类结果。

1.3 算法的改进

虽然 FusionGAN 算法能很好地平衡红外与可

见图像中的有效信息,与其他方法相比图像融合的质量更高,但是在细节纹理和深度特征提取上还不够,因此新增加了一个即插即用卷积 PSConv 与一个超轻量级的注意力模块 ECA Module。改进后的算法网络结构如图 3 所示。输入的红外图像与可见图像分别经过各自的 3 层卷积和 1 层残差网络后,在第 2 层残差块中引入上述 2 个模块,加深了网络的深度,有利于深度特征的提取,最后再经过 3 次反卷积生成最后融合图像。虽然增加了 2 个模块,但是由于都是轻量级网络,使整个模型增加的运算量不大,接下来将分别介绍这 2 个模块。

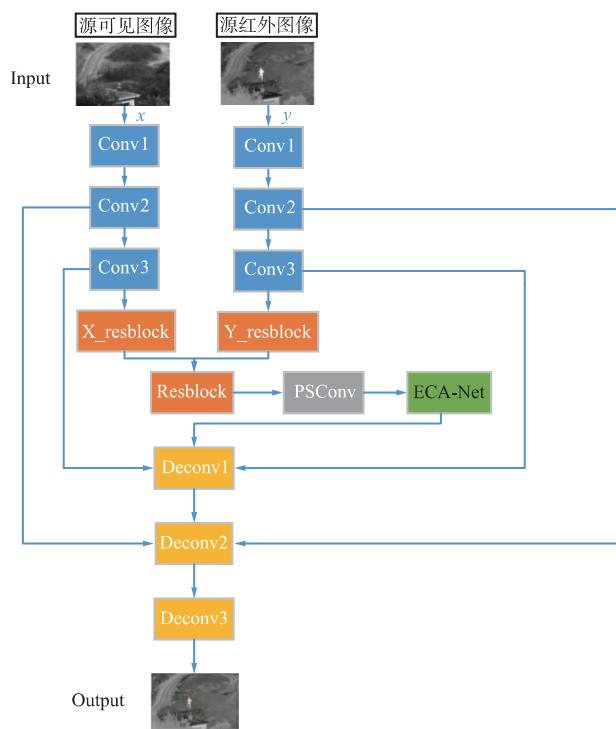


图 3 改进后的算法结构

Fig.3 Improved algorithm structure

1.3.1 PSConv

PSConv 是一种紧凑的多尺度卷积层,能够将输入的特征按尺度分成多块,相比已有的多尺度特征融合模型,它能从更细粒度角度进行特征探索,在具有同样计算复杂度的同时,具有更强的特征提取能力,示意图如图 4 所示。图 4 中, H 和 W 分别为高度和宽度,内核大小为 $K \times K$, $\mathcal{F} \in \mathcal{R}^{C_{in} \times H \times W}$ 表示输入的特征, $\mathcal{G} \in \mathcal{R}^{C_{out} \times C_{in} \times K \times K}$ 表示卷积核, $\mathcal{H} \in \mathcal{R}^{C_{out} \times H \times W}$ 表示特征的输出。普通卷积定义如下:

$$\mathcal{H}_{c \times \delta} = \sum_{k=1}^{C_{in}} \sum_{i=-\frac{K-1}{2}}^{\frac{K-1}{2}} \sum_{j=-\frac{K-1}{2}}^{\frac{K-1}{2}} \mathcal{G}_{c \times K \times i \times j} \mathcal{F}_{K \times i \times j} \quad (5)$$

扩张卷积则可描述为:

$$\mathcal{H}_{c \times \delta} = \sum_{k=1}^{C_{in}} \sum_{i=-\frac{K-1}{2}}^{\frac{K-1}{2}} \sum_{j=-\frac{K-1}{2}}^{\frac{K-1}{2}} \mathcal{G}_{c \times K \times i \times j} \mathcal{F}_{K \times i \times j} \quad (6)$$

而 PSConv 则可描述为:

$$\mathcal{H}_{c \times \delta} = \sum_{k=1}^{C_{in}} \sum_{i=-\frac{K-1}{2}}^{\frac{K-1}{2}} \sum_{j=-\frac{K-1}{2}}^{\frac{K-1}{2}} \mathcal{G}_{c \times K \times i \times j} \mathcal{F}_{K \times i \times j} \odot \quad (7)$$

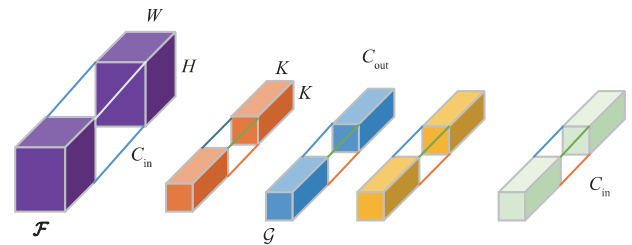


图 4 PSConv 示意

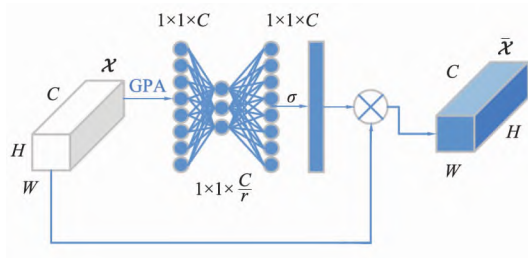
Fig.4 PSConv schematic diagram

从上述卷积计算公式可以看出,PSConv 将多尺度卷积归入同一个计算过程中,且不同尺度卷积计算按通道交替执行,是一种更细粒度的多尺度操作。

1.3.2 ECA-Net

注意力机制^[12]在自然语言处理(Natural Language Processing, NLP)领域大放异彩,近年来许多人将通道注意力机制用于深度卷积神经网络的性能提升,ECA-Net 是一种高效的通道注意力模块。现有的许多注意力模块为了获得更好的性能,往往将结构复杂化,而 ECA 模块只涉及少量的参数,却取得了明显的效果提升。其核心思想是让神经网络使用全局信息来增强有用的信息,同时抑制无用的信息。其架构来自于对 SE-Net (Squeeze-and-Excitation Networks)^[13]结构的改进,二者的架构对比如图 5 所示。图中 H 表示图像高度; W 表示图像宽度; C 表示通道数; $\mathcal{X}$ 表示张量; $\bar{\mathcal{X}}$ 表示经过处理后的张量,SE-Net 网络结构为一个全平均池化层(Global Average Pooling, GAP)然后再通过一个 1×1 全连接层(Fully Connect, FC)再进行降维处理,处理后的结果再通过一个 1×1 的全连接层。生成的数据 σ ,最后再通过一个一维的卷积进行特征权重的提取,使用 Sigmoid 函数作为激活函数,将特征信息映射到 0,1 之间,再将映射后的特征信息与原图像相乘。经过这样的处理,无用的信息将会被抑制到 0,而有用的信息将会被保留下来,从而提高图像中有效数据的提取。而 ECA-Net,在此基础上提出了

一种不降维的局部跨信道交互方法,该方法可以通过一维卷积有效地实现。在中间环节引入一种自适应选择一维卷积核大小的方法,以确定局部跨信道交互的覆盖率。ECA 通过执行大小为 K 的快速一维卷积来生成通道权值,其中 K 通过通道维 C 的函数相应地确定,二者的关系为:



$$K = \psi(C) = \left\lfloor \frac{\ln(C)}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}}, \quad (8)$$

式中, $\lfloor t \rfloor_{\text{odd}}$ 表示离 t 最近的奇数; γ 和 b 在本文中取 2 和 1; K 取 5。相比之下, ECA-Net 结构更加轻量化增加可以忽略不计的参数量的同时,带来性能明显的提升。

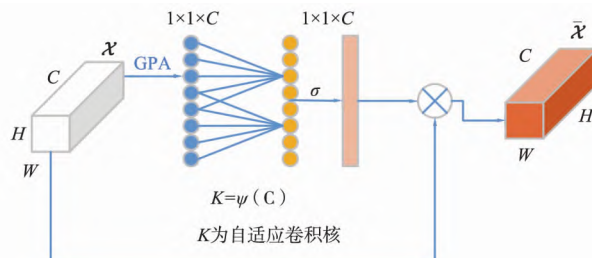


图 5 SE 与 ECA 结构对比

Fig.5 Structure comparison between SE and ECA

2 实验结果与评价

2.1 数据集

为了增强实验结果的准确性与可靠性,选用公开并且校准好的红外与可见光数据集 TNO 与 INO。其中 TNO 数据集包含军事场景和其他一些不同场景下的近红外和长波红外或热红外夜间图像与可见光图像,适用于复杂场景下的图像融合算法研究。INO 数据集来自加拿大国家光学研究所,包含了许许多在不同天气条件下拍摄的不同的城市道路街景。

2.2 结果

选取来自 TNO 和 INO 数据集中的 1 200 张红外图像与可见光图像作为训练集,为了验证算法的性能,再从 TNO 数据集中单独挑选出 20 对可见与红外图像作为测试集。训练集融合的结果如图 6 所示,模型训练好之后测试集融合的结果如图 7 所示。



图 6 训练集结果

Fig.6 Training set results

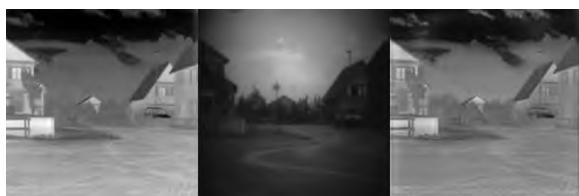


图 7 测试集结果

Fig.7 Testing set results

2.3 评价

本实验将几种目前比较常用的先进的图像融合方法与本文的方法进行比较,其中包括曲波变换 (CWT)^[14]、双树复小波变换 (DTCWT)^[15]、加权最小二乘优化法 (WLS)^[16] 和原融合生成对抗网络 FusionGAN 等图像融合方法,并且采用主观评价与客观评价相结合的方法,使实验结果更具有真实性和可靠性。

2.3.1 主观评价

选取了 5 对红外与可见图像的融合结果作为主观评价指标,上述不同的融合方法对来自 TNO 数据集中 5 对图像的融合结果如图 8 所示。为了体现改进的算法与 FusionGAN 的不同,在图 8 (f) 和 (g) 中加入了一些方框用于细节的比较,并且将其放大,放在原图的右下角。通过融合后的结果可以看出,上述方法均能对红外图像与可见图像进行成功融合,融合后的图像均能包含红外与可见图像的特征信息。虽然 CWT 和 DTCWT 方法融合的结果含有足够的细节特征,但红外的目标不够显著。而与 CWT 和 DTCWT 方法相比, WLS 方法具有更强的目标追踪性,但在背景部分损失了较多的红外信息。FusionGAN 方法在目标追踪和细节纹理特征保留之间取得了较好的实现,但是背景信息中的细节特征仍不够丰富。与上述方法相比,本文采用的方法既保留了红外目标的显著性,又在背景中包含了足够的细节纹理与边缘信息。

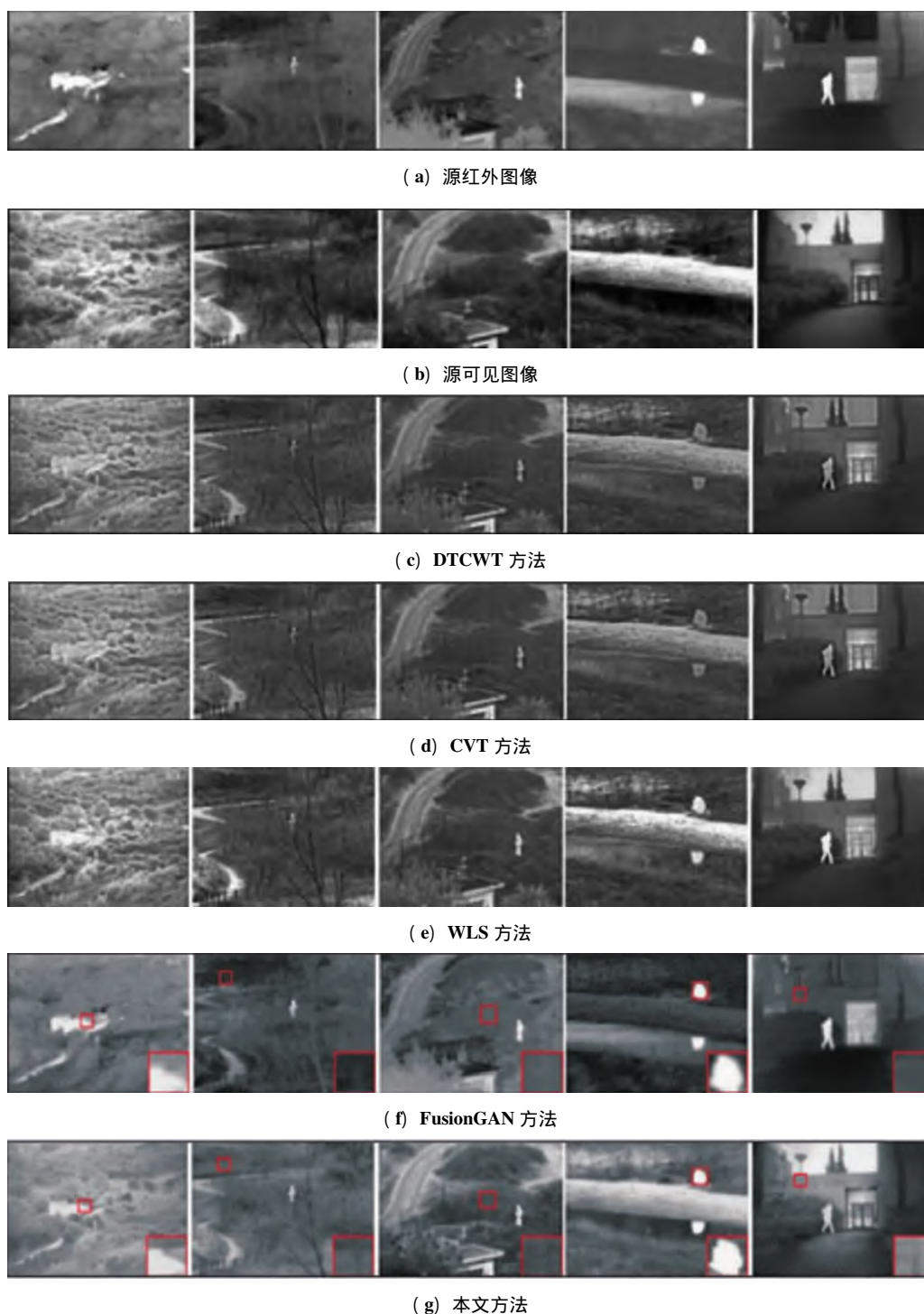


图 8 5 种方法融合后的对比

Fig.8 Comparison of five method safter fusion

2.3.2 客观评价

为了增强实验的准确性与客观性,采用多种评价指标进行定量分析。选取熵(EN)^[17]、平均梯度(AG)^[18]、空间频率(SF)^[19]和结构相似指数(SSIM)^[20]这4种常用评价指标。EN主要是度量图像包含信息量多少的一个客观量,熵值越大,表明

融合图像中的信息越丰富,而噪声也会对EN的结果造成影响,一般不单独使用。AG度量融合图像中包含的梯度信息,反映了细节和纹理,AG值越大,表明融合图像中所含梯度信息越多。SF可以有效衡量图像的梯度分布,SF越大,融合的图像具有更丰富的边缘信息和纹理特征。SSIM是用于模拟

图像畸变和失真的一种评价方法,SSIM 越大,说明融合算法的效果越好,失真与畸变越小。4 种指标下不同方法的平均性能如表 1 所示,20 对图像的数据结果用 Matlab 绘制成折线图,如图 9~图 12 所示。从表和图中可以看出,在上述 4 种评价指标下,5 种方法在客观的定量分析上有差距,本文所用的方法在 AG、SF、SSIM 这 3 种指标上取得了不错的效果。结果表明,本文方法的试验结果中包含了更多的细节纹理与梯度信息,得到的融合图像与源图像之间的畸变小于 CVT、DTCWT 和 WLS,FusionGAN 虽然能包含更多的信息熵,但在空间频率信息与结构相似性上远不如改进后的方法。

表 1 5 种算法在指标下的平均性能

Tab.1 Average performance of five algorithms under the index

方法	EN	AG	SF	SSIM
CVT	6.426 6	2.660 8	8.900 1	0.696 4
DTCWT	6.403 1	2.530 2	8.838 3	0.693 6
WLS	6.624 9	3.050 8	10.027 6	0.722 1
FusionGAN	6.640 6	3.935 5	8.946 5	0.697 8
本文	6.566 1	4.180 6	10.238 5	0.834 9

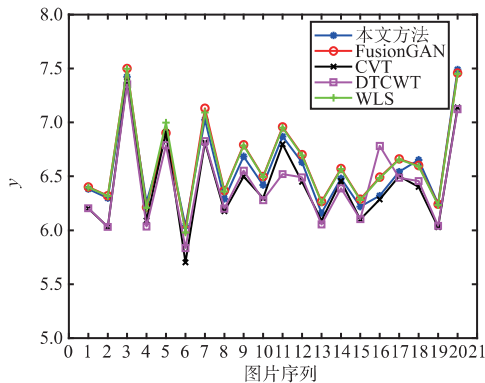


图 9 EN 指标下对 TNO 数据集中 20 对图像对的试验结果
Fig.9 Experimental results of 20 image pairs in TNO dataset under EN index

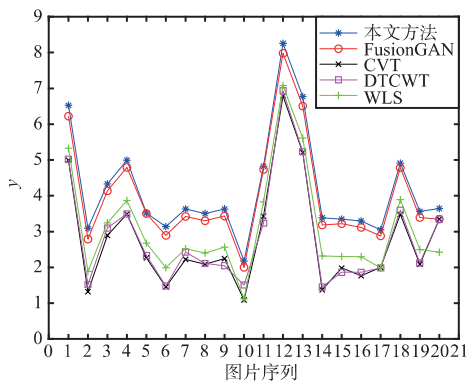


图 10 AG 指标下对 TNO 数据集中 20 对图像对的试验结果
Fig.10 Experimental results of 20 image pairs in TNO dataset under AG index

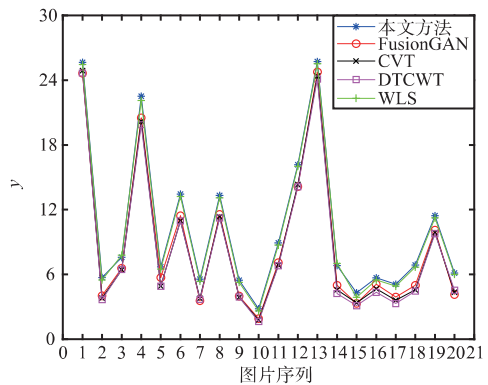


图 11 SF 指标下对 TNO 数据集中 20 对图像对的试验结果
Fig.11 Experimental results of 20 image pairs in TNO dataset under SF index

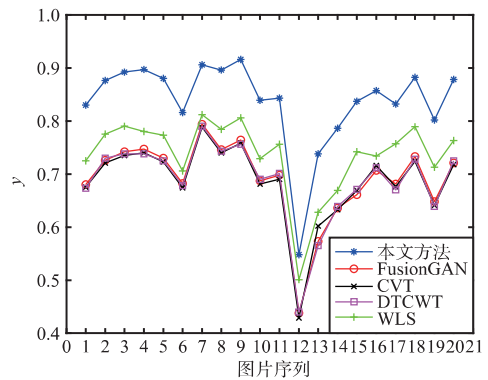


图 12 SSIM 指标下对 TNO 数据集中 20 对图像对的试验结果
Fig.12 Experimental results of 20 image pairs in TNO dataset under SSIM index

3 结束语

针对可见光与红外图像融合中所存在的问题,提出了一种改进的 FusionGAN 方法。分析了 FusionGAN 的原理和 2 个引入模块 PSConv 与 ECA-Net 的架构。PSConv 对图像细节处理更好,ECA-Net 能更好地提取图像中的有用信息,使用了数据集验证了新算法的可行性与效果,加入的模块使改进后的算法与原算法相比在 AG 上提升了 6.2%,在 SF 上提升了 14.4%,在 SSIM 上提升了 18.6%。本文引入注意力机制模块与插入式卷积,为图像融合算法的改进与提高提供了新的思路,更适用于复杂场景下的红外与可见图像融合,为今后继续研究深度神经网络用于图像融合打下了基础。

✦

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] LI S T, YANG B, HU J W. Performance Comparison of Different Multi-resolution Transforms for Image Fusion [J]. Information Fusion 2011, 12(2): 74–84.
- [3] WANG J, PENG J Y, FENG X Y, et al. Fusion Method for Infrared and Visible Images By Using Non-negative Sparse Representation [J]. Infrared Physics & Technology, 2014, 67: 477–489.
- [4] XIANG T Z, YAN L, GAO R R. A Fusion Algorithm for Infrared and Visible Images Based on Adaptive Dual-channel Unit-linking PCNN in NSCT Domain [J]. Infrared Physics and Technology, 2015, 69: 53–61.
- [5] BAVIRISETTI D P, XIAO G, LIU G. Multi-sensor Image Fusion Based on Fourth Order Partial Differential Equations [C] // 2017 20th International Conference on Information Fusion (Fusion). Xi'an: IEEE, 2017: 1–10.
- [6] ZHANG X Y, MA Y, FAN F, et al. Infrared and Visible Image Fusion via Saliency Analysis and Local Edge-preserving Multi-scale Decomposition [J]. Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 2017, 34(8): 1400–1410.
- [7] ZHAO J F, CUI G M, GONG X L, et al. Fusion of Visible and Infrared Images Using Global Entropy and Gradient Constrained Regularization [J]. Infrared Physics & Technology, 2017, 81: 201–209.
- [8] 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.
- [9] LI D, YAO A B, CHEN Q F. PSConv: Squeezing Feature Pyramid into One Compact Poly-Scale Convolutional Layer [C] // Computer Vision and Pattern Recognition. Hawaii: CVPR, 2020: 615–632.
- [10] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks [C] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 1–10.
- [11] 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.
- [12] SHAW P, USZKOREIT J, VASWANI A. Self-attention with Relative Position Representations [C] // 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers). New Orleans: NAACL, 2018: 464–468.
- [13] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-Excitation Networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011–2023.
- [14] NENCINI F, GARZELLI A, BARONTI S. Remote Sensing Image Fusion Using the Curvelet Transform [J]. Information Fusion, 2007, 8(2): 143–156.
- [15] 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.
- [16] MA J L, ZHOU Z Q, BO W, et al. Infrared and Visible Image Fusion Based on Visual Saliency Map and Weighted Least Square Optimization [J]. Infrared Physics and Technology, 2017, 82: 8–17.
- [17] ROBERTS W J, VAN AARDT J A, AHMED F B. Assessment of Image Fusion Procedures Using Entropy, Image Quality, and Multispectral Classification [J]. Journal of Applied Remote Sensing, 2008, 2(1): 1–28.
- [18] CUI G M, FENG H J, XU Z H, et al. Detail Preserved Fusion of Visible and Infrared Images Using Regional Saliency Extraction and Multi-scale Image Decomposition [J]. Optics Communications, 2015, 341: 199–209.
- [19] ESKICIOGLU A M, FISHER P S. Image Quality Measures and Their Performance [J]. IEEE Transactions on Communications, 1995, 43(12): 2959–2965.
- [20] WANG Z, BOVIK A C, SHEIKH H R, et al. Image Quality Assessment: from Error Visibility to Structural Similarity [J]. IEEE Transactions on Image Processing, 2004, 13(4): 600–612.

作者简介



刘程亮 男 (1998—), 就读于中国科学院长春光学精密机械与物理研究所机械电子专业, 硕士研究生。主要研究方向: 神经网络算法与图像处理。



张宇 男 (1981—), 博士, 研究员, 硕士生导师。主要研究方向: 航天遥感器成像电子学设计。



吕恒毅 男 (1984—), 博士, 副研究员, 博士生导师。主要研究方向: AI 智能先进成像、航天遥感器相机成像。