

基于监督学习的全卷积神经网络多聚焦图像融合算法

李恒^{1*}, 张黎明^{2,3**}, 蒋美容², 李玉龙¹

¹兰州交通大学电子与信息工程学院, 甘肃 兰州 730070;

²兰州交通大学测绘与地理信息学院, 甘肃 兰州 730070;

³地理国情监测技术应用国家地方联合工程研究中心, 甘肃 兰州 730070

摘要 为进一步提高多聚焦图像的融合质量, 提出一种基于监督学习的全卷积神经网络多聚焦图像融合算法。该算法旨在运用神经网络学习源图像不同聚焦区域的互补关系, 即选择源图像中不同的聚焦位置合成一张全局清晰图像。该算法构造聚焦图像作为训练数据, 网络采用稠密连接和 1×1 卷积以提高网络的理解能力和效率。实验结果表明, 本文算法在主观视觉评估和客观评价两方面均优于其他对比算法, 图像的融合质量得到进一步提升。

关键词 图像处理; 监督学习; 全卷积; 多聚焦图像; 图像融合

中图分类号 TP391

文献标志码 A

doi: 10.3788/LOP57.081015

Multi-Focus Image Fusion Algorithm Based on Supervised Learning for Fully Convolutional Neural Networks

Li Heng^{1*}, Zhang Liming^{2,3**}, Jiang Meirong², Li Yulong¹

¹School of Electronics and Information Engineering, Lanzhou Jiaotong University, Lanzhou, Gansu 730070, China;

²Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou, Gansu 730070, China;

³National-Local Joint Engineering Research Center of Technologies and Applications for National Geographic State Monitoring, Lanzhou, Gansu 730070, China

Abstract To improve the quality of multi-focus image fusion, a fully convolutional neural network multi-focus image fusion algorithm based on supervised learning is proposed. The proposed algorithm aims to use neural networks to learn the complementary relationship between different focus areas of the source image, that is, to select different focus positions of the source image to synthesize a global clear image. In this algorithm, the focus images are constructed as training data, and the dense connection and 1×1 convolution are used in the network to improve the understanding ability and efficiency of the network. The experimental results show that the proposed algorithm is superior to other contrast algorithms in both subjective visual evaluation and objective evaluation, and the quality of image fusion is significantly improved.

Key words image processing; supervised learning; fully convolution; multi-focus image; image fusion

OCIS codes 100.2980; 100.2000

1 引言

由于光学原件景深限制, 很难捕获到图像中所有的清晰目标, 只有距离相机特定位置的目标才能被聚焦, 在聚焦平面之前或之后的目标将失去聚焦

变得模糊^[1]。为解决这一问题, 研究者提出了多聚焦图像融合算法, 其目的是将不同景深图像融合成一张全聚焦图像, 从而获取更全面、更可靠的场景描述^[2]。目前, 多聚焦图像融合技术在目标识别、机器视觉、数字摄影和显微成像等领域应用广泛^[3-4]。

收稿日期: 2019-07-22; 修回日期: 2019-09-05; 录用日期: 2019-09-16

基金项目: 国家自然科学基金(41761080)、兰州市人才创新创业科技计划(2016-RC-59)、甘肃高等学校产业支撑引导项目(2019C-04)

* E-mail: 453984016@qq.com; ** E-mail: zhanglm8@gmail.com

近年来有多种多聚焦图像融合算法被提出,根据其融合策略不同,这些融合算法可以分成两类:基于变换域的融合方法和基于空间域的融合方法。基于变换域的方法有拉普拉斯金字塔^[5]、离散小波分解^[6]、非上采样剪切波变换(NSCT)^[7-8]等,此类方法都是将源图像分解成多尺度域,再融合分解后的多尺度域,最后使用融合后的多尺度域来重建图像。基于空间域的融合方法主要利用像素级的梯度信息或图像块来融合图像,然而这一过程会引入伪影块,近年来基于导向滤波(GF)^[9-10]、基于图像提取(IM)^[11]、基于密集尺度不变变换(DSIFT)^[12]和基于同质相似^[13]等像素级融合方法在提取和表现图像细节方面均表现较好。但这些方法的活动测度和融合规则难以进行人工设计,并且有很多因素不能完全考虑在内。

由于深度学习具有很强的特征提取和数据表示能力,在图像处理和计算机视觉任务中表现优异。基于深度学习的多聚焦图像融合领域由于缺少标记数据,大多使用的神经网络模型都是基于非监督学习或将神经网络作为图像融合算法的一部分。Mustafa等^[14]提出基于非监督学习的多尺度卷积神经网络多聚焦图像融合方法,该方法使用3种不同尺度的卷积核来提取源图像的低频和低频信息,取得了较好的融合效果,但该方法并未充分利用网络中间层信息。Liu等^[1]提出了一种基于卷积神经网络的图像融合算法,将神经网络融合多聚焦图像的方法视为分类问题,使用分块后图像及其模糊处理来训练网络,使网络具有分类能力,网络输出为分类得分表,得分表经过平均重叠块、二值化和两个一致性检验策略得到决定映射,最后将决定映射作为源图权重来重建图像,但这样并未做到端到端的映射。

为解决以上问题,本文提出一种基于监督学习的全卷积神经网络多聚焦图像融合算法,该算法将多聚焦图像融合视为像素点预测问题,通过监督学习来训练全卷积网络,使网络学习到源图像中不同聚焦区域的互补关系,从而合成一张全局清晰的图像。

2 相关工作

2.1 稠密连接

在传统的深度学习网络中,多数中间层信息并未充分得到利用,且随着网络深度的增加,网络会出现梯度消失问题。为解决这一问题,He等^[15]提出了深度残差网络,为了进一步充分利用网络中间层提取到的信息,Huang等^[16]提出了稠密神经网络,

该网络采用的稠密连接可以从任意层连向其他层。稠密连接有3个优点:1)充分利用网络中间卷积层提取到的信息;2)加速网络中的信息流和梯度流从而使网络更易于训练;3)具有归一化效果可以防止网络过拟合。因此,本文采用这种结构可以提取更多有用信息提升网络学习能力。

2.2 1×1 卷积核

为了提高神经网络的理解能力,Lin等^[17]提出网中网模型。在该模型中加入卷积核为 1×1 的卷积层,提升了网络的深度。在本文的网络模型中多处采用 1×1 的卷积结构,主要目的有两个:1)使用该结构对神经网络进行降维,可大幅减少神经网络的参数,使网络更易于训练;2)在保持神经网络特征映射尺度不变的情况下,可以大幅提高网络的非线性特性,使网络的深度增加,理解能力增强。

3 本文方法

本文主要考虑已配准的两幅多聚焦图像融合,而多幅多聚焦图像的融合可以依次两两融合实现。图1为本文方法框架图。在训练阶段,本文神经网络在训练数据集 S 上学习到一个多聚焦图像互补关系映射 Φ 。训练完成后,网络输入源图像 I_1, I_2 ,输出融合图像 F 。这是一个基于监督学习的端到端的融合过程。

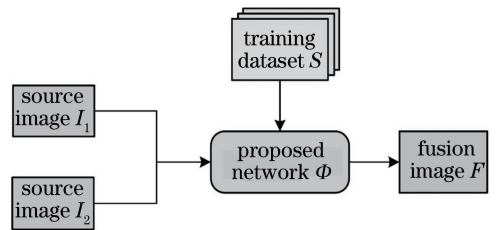


图1 本文方法的框架图

Fig. 1 Framework of proposed method

3.1 网络结构

本文网络结构如图2所示,神经网络采用 3×3 和 1×1 两种卷积核,步长均为1,未使用全连接层,因此神经网络输入端图像可以为任意大小。为尽量保留源图像的信息,整个网络均未使用池化层,除最后一层外,其他每个卷积层之后均使用ReLU作为激活函数,最后一层使用Tanh函数激活。该网络由三部分组成:特征提取、特征融合和图像重建。特征提取部分由两个网络分支组成,每个网络分支包含3个卷积层,每个卷积层用于提取源图像的不同特征。特征融合部分是将特征提取部分中的两个网络分支的输出在通道方向上

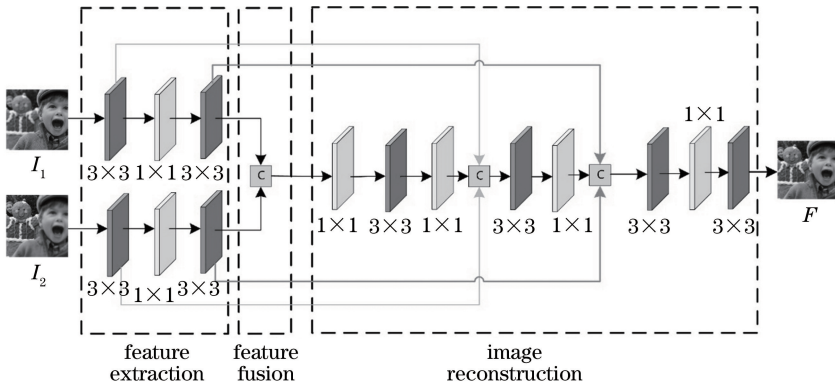


图 2 网络结构图

Fig. 2 Network structure diagram

进行连接而得到融合特征,之后融合特征作为图像重建部分网络的输入。图像重建部分包括 8 个卷积层,使用稠密连接分别将特征提取部分的第 1、3 层在通道方向上连接到图像重建部分的第 3、5 输出层上,这样可以充分利用不同特征层的信息,从而得到更好的融合效果。

3.2 训练数据集

训练基于监督学习的神经网络需要大量带标签的数据,然而多聚焦图像数据集中带标签数据集缺乏,标记一个巨大的多聚焦图像数据集不仅成本高而且难度大。本文基于公共数据集 VOC2007^[18] 构造了一个带有标签的多聚焦图像数据集 S , VOC2007 中有 5011 张训练图像,选取其中 2000 张,尺寸裁剪为 256×256 作为标签图像。神经网络

的输入为两幅不同聚焦区域的图像,本文对每张标签图像做不同区域的高斯模糊处理,高斯模糊半径取为 3。采用两组不同的模糊方式,分别为平行模糊和交叉模糊。交叉模糊相对于平行模糊,能更好地模拟真实的多聚焦图像。如图 3 所示, P, \bar{P} 为互补的平行模糊组, C, \bar{C} 为互补的交叉模糊组,分别在两组上进行模糊密度为 2、4、8 和 16 的模糊处理。实验结果如表 1 所示,平行模糊和交叉模糊均在模糊密度为 8 时达到最优结果,此时,除结构相似性 (SSIM) 外,交叉模糊在峰值信噪比 (PSNR)、相关系数 (CC)、通用质量指标 (UQI) 三种指标方面均优于平行模糊,总体取得了较优的结果,故本文选取交叉模糊方式,模糊密度为 8 来构造训练数据集。

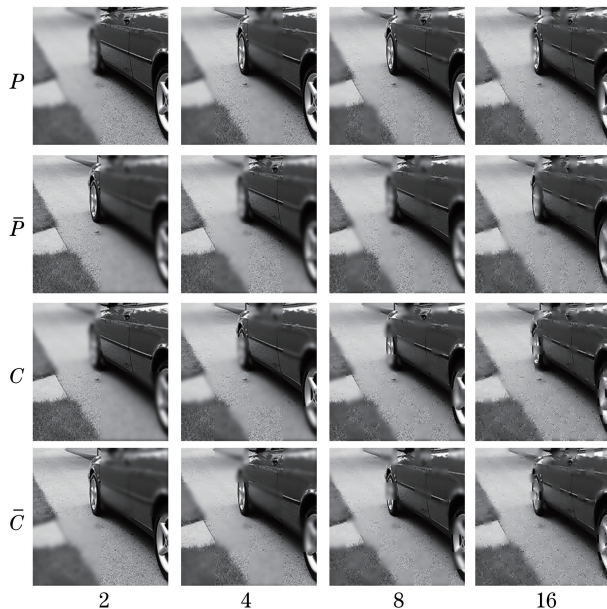


图 3 构造的数据

Fig. 3 Constructed data

表 1 不同模糊设置下的 27 幅测试图像融合指标的平均值

Table 1 Average values of quality metrics for 27 test images under different fuzzy settings

| Index | SSIM | PSNR | CC | UQI |
|-------|---------------|----------------|---------------|---------------|
| P2 | 0.8663 | 54.0889 | 0.9753 | 0.7928 |
| P4 | 0.8842 | 57.1438 | 0.9793 | 0.8194 |
| P8 | 0.8842 | 57.6086 | 0.9793 | 0.8195 |
| P16 | 0.8842 | 57.6086 | 0.9793 | 0.8195 |
| C2 | 0.8631 | 53.5615 | 0.9743 | 0.7863 |
| C4 | 0.8813 | 56.8458 | 0.9788 | 0.8136 |
| C8 | 0.8837 | 57.6728 | 0.9795 | 0.8200 |
| C16 | 0.8837 | 57.6728 | 0.9795 | 0.8200 |

3.3 网络训练

为使重建图像更加准确,通过最小化损失函数 L 来训练网络,该损失函数联合了网络输出与标签图像的 SSIM 和其像素点的均方误差,表示为

$$L = L_{SSIM} + \lambda L_{MSE}, \quad (1)$$

$$L_{SSIM} = 1 - SSIM(O, T), \quad (2)$$

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^N [O_i(x, y), T_i(x, y)]^2, \quad (3)$$

式中: λ 为 L_{SSIM} 和 L_{MSE} 的权重,本文 $\lambda = 10$; O 为神经网络的输出; T 为标签图像; $O_i(x, y)$ 为神经网络输出的第 i 位置的像素值; $T_i(x, y)$ 为标签图像第 i 位置的像素值; N 为像素点的总数。

在网络的训练过程中,本文使用 Adam 作为优化函数,学习率为 0.0005,批处理大小为 2,训练 100 轮。图 4 为使用交叉模糊,模糊密度为 8,分别在第 1、25、50、75、100 轮的训练实验结果, T 为标签图像, O 为神经网络的输出结果。图 5 为训练 100 轮损失函数 L 的变化曲线图,横坐标为神经网络训练的轮次,纵坐标为损失函数 L 的值。从图 4 中可以看出,训练第 1 轮神经网络输出 O 较模糊,说明神经网络权重还未收敛,损失函数值较大,而第 25 轮神经网络输出 O 与其标记图像基本一致,肉眼很难区分,说明神经网络权重基本收敛,损失函数值较小,这和图 5 所示结果相吻合。

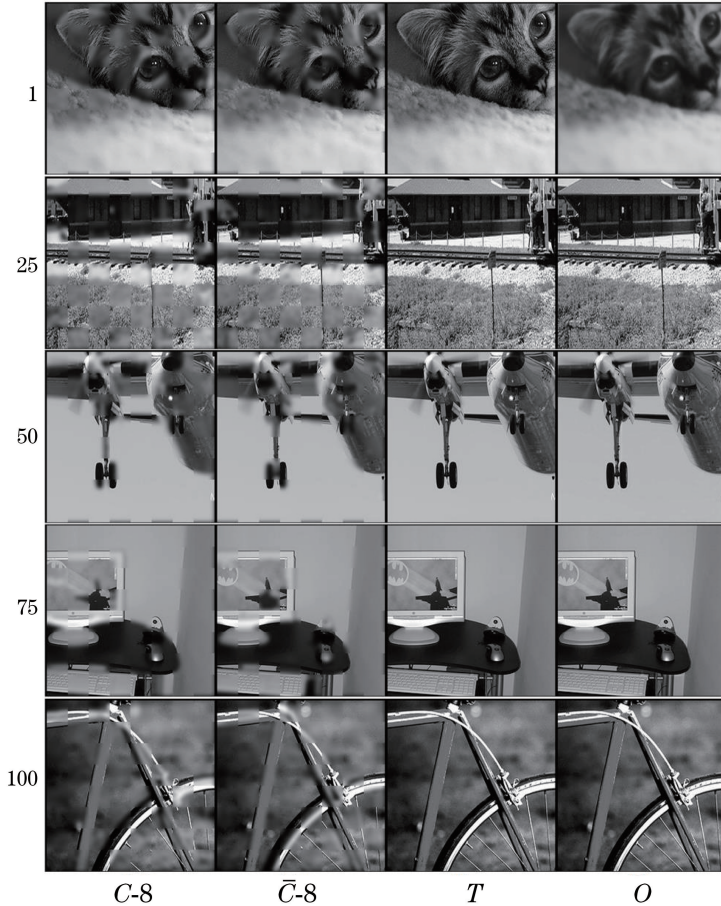


图 4 训练第 1、25、50、75、100 轮的实验结果

Fig. 4 Training results of the 1st, 25th, 50th, 75th, and 100th epochs

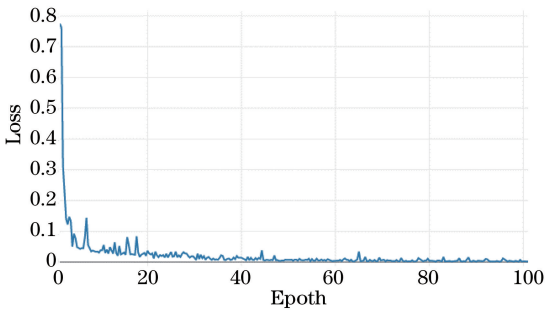


图5 训练100轮神经网络损失函数的变化曲线

Fig. 5 Neural network loss function curve of training 100 epochs

4 实验结果和分析

4.1 实验设定

为验证本文方法的有效性,分别选取文献[7]方法(NSCT)、文献[19]方法(DTCWT)、文献[1]方法(CNN)、文献[20]方法(NSST)作为对比方法。选



图6 实验中用到的部分多聚焦测试图像

Fig. 6 Partial of multi-focus test images used in experiment

2) PSNR 表示源图像和融合图像的峰值信噪比,公式为

$$L_{\text{PSNR}} = 20 \log_{10} \left\{ \frac{L}{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N [X(i,j) - Y(i,j)]^2} \right\}, \quad (5)$$

式中: L 为图像的灰度级,PSNR 值越大,说明图像在融合过程中失真越少,融合后的图像效果越好。

3) CC 表示源图像和融合图像光谱特征的相似性,公式为

$$L_{\text{CC}} = \frac{\sum_{i=1}^M \sum_{j=1}^N (X_{i,j} - \bar{X})(Y_{i,j} - \bar{Y})}{\sqrt{\sum_{i=1}^M \sum_{j=1}^N (X_{i,j} - \bar{X})^2 \sum_{i=1}^M \sum_{j=1}^N (Y_{i,j} - \bar{Y})^2}}, \quad (6)$$

取了 27 幅多聚焦图像进行实验,其中 19 幅来自公开数据集“Lytro”^[21],其余 8 幅被广泛应用于多聚焦图像的融合研究中,图 6 为实验中用到的部分多聚焦测试图像。

客观评价在图像融合中起到重要的作用,融合方法效果主要由多个指标定量体现,本文选取 SSIM^[22]、PSNR、CC^[23]、UQI^[24] 4 种客观评价指标来定量评估算法的性能。

1) SSIM 表示源图像和融合图像的结构相似性,公式为

$$L_{\text{SSIM}} = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)}, \quad (4)$$

式中: μ_X 、 μ_Y 分别表示源图像和融合图像的平均值; σ_{XY} 表示源图像和融合图像的协方差; σ_X 、 σ_Y 分别表示源图像和融合图像的标准差。SSIM 的取值范围为 $[-1, 1]$,其值越接近 1,说明融合图像和源图像越相似。

式中: X 、 Y 分别表示源图像和融合图像。CC 值越接近 1,说明融合图像与源图像越相似。

4) UQI 表示源图像与融合图像相似信息的数量,公式为

$$L_{\text{UQI}} = \frac{4\sigma_{XY}(\mu_X + \mu_Y)}{(\mu_X^2 + \mu_Y^2)(\sigma_X^2 + \sigma_Y^2)}, \quad (7)$$

式中: μ_X 、 μ_Y 分别表示源图像和融合图像的均值; σ_X 、 σ_Y 分别表示源图像和融合图像的标准差; σ_{XY} 表示源图像和融合图像的协方差。UQI 的取值范围为 $[-1, 1]$,其值越接近 1,说明融合图像与源图像越相似。

4.2 主观评估

首先从视觉感知的角度比较不同融合算法的表现,从 27 幅实验结果中选取 2 组图像进行说明,为便于观察对两幅图片中部分区域进行放大处理。

图 7 为第一组图像不同方法的融合结果。由图可见,图 7(c)、(d)中木质花纹区域模糊严重,部分区域、纹理和边缘细节信息丢失;图 7(e)、(f)中木质花纹区域比较清晰;图 7(e)中融合边界产生不同程度的伪影,融合效果不佳;图 7(f)在熊偶衣服边缘发白严重,观感不佳;图 7(g)中聚焦物体突出,主观效果较好,木质花纹区域清晰无伪影产生,说明本文算法较大程度保留了源图像的纹理细节信息,使融

合图像的质量得到提升。

图 8 为第二组图像不同方法的融合结果。由图可见,图 8(c)中数字 5 模糊严重,融合效果较差;图 8(d)中数字 5 模糊且整体偏暗,视觉效果较差;图 8(e)中数字 5 较清晰,但数字 6 模糊严重;图 8(f)中数字 5 模糊且字母数字周围伪影严重,部分细节和边缘信息丢失;图 8(g)中数字 5、6 均较清晰,聚焦物体周围无伪影产生,较好地保留了源图像的边缘及纹理信息。

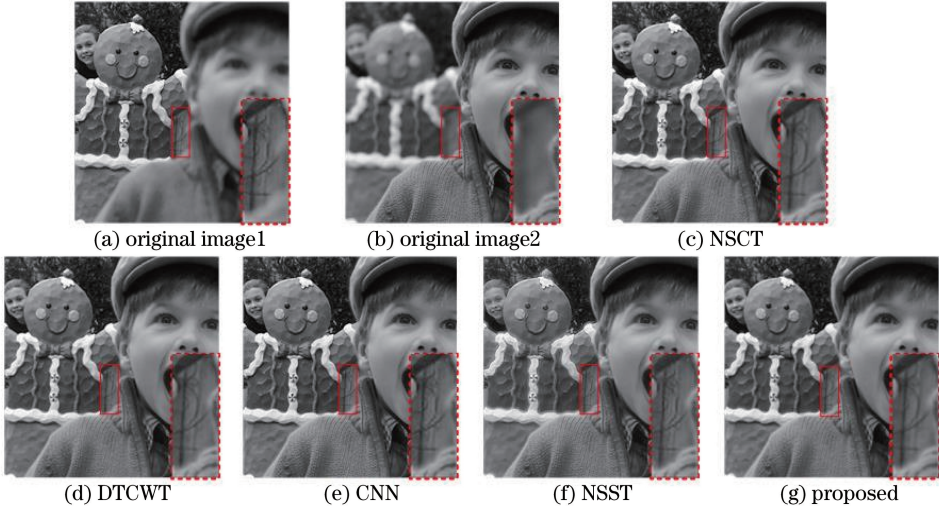


图 7 第一组图像不同方法的融合结果

Fig. 7 Experimental results of first group images with different algorithms

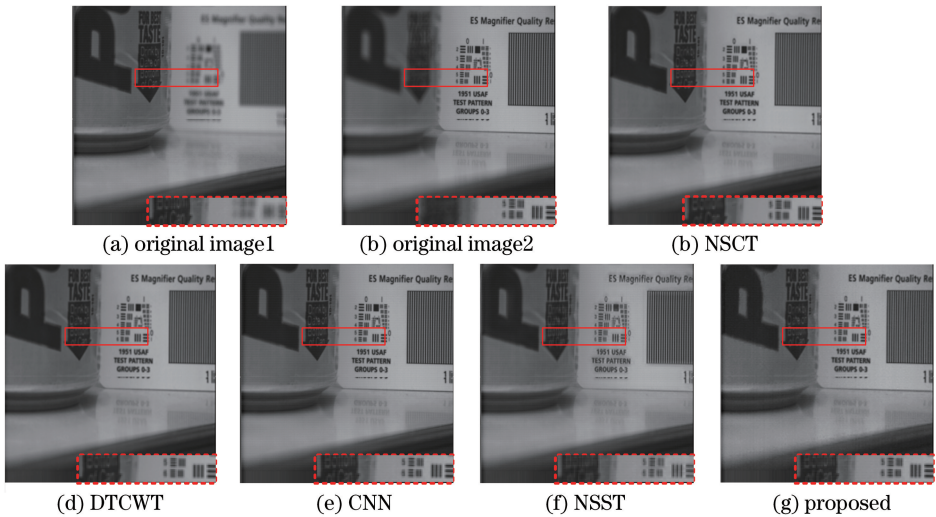


图 8 第二组图像不同方法的融合结果

Fig. 8 Experimental results of second group images with different algorithms

4.3 客观评估

为了更加全面客观地评价不同方法的融合效果,实验中采用 SSIM、PSNR、CC、UQI 进行定量分析,表 2 为 27 幅融合图像在 4 种指标下的平均值。可知,本文方法在 SSIM、CC、UQI 方面均优于其他对比方法,NSST 方法中仅 PSNR 高于本文方法,

其他 3 种指标均不如本文方法。因此,在客观上说明本文方法的融合效果要优于其他对比方法。

结合主观评估和客观评估可知,本文方法很好地保留了源图像中的细节和纹理信息,提升了图像的清晰度,并且聚焦物体边缘处自然、无伪影产生,

表2 27幅测试图像融合指标的平均值

Table 2 Average values of quality metrics for 27 test images

| Method | SSIM | PSNR | CC | UQI |
|----------|---------------|----------------|---------------|---------------|
| NSCT | 0.8758 | 57.1025 | 0.9778 | 0.8160 |
| DTCWT | 0.8724 | 56.9115 | 0.9769 | 0.8069 |
| CNN | 0.8742 | 56.9447 | 0.9765 | 0.8155 |
| NSST | 0.8806 | 57.9811 | 0.9793 | 0.8175 |
| Proposed | 0.8837 | 57.6728 | 0.9795 | 0.8200 |

总体上取得了较好的融合效果。

5 结 论

本文提出一种基于监督学习的全卷积神经网络多聚焦图像融合算法,该网络可以学习到源图像不同聚焦区域的互补关系,从而融合成一张全局清晰的图像。实验结果表明,该算法有效保留了源图像的区域、纹理和边缘细节信息,清晰度得到进一步提升,且图像融合边界无伪影产生,取得了较好的融合效果。

参 考 文 献

- [1] 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.
- [2] Chen Q J, Li Y, Chai Y Z. A multi-focus image fusion algorithm based on depth learning[J]. Laser & Optoelectronics Progress, 2018, 55(7): 071015. 陈清江, 李毅, 柴昱洲. 一种基于深度学习的多聚焦图像融合算法[J]. 激光与光电子学进展, 2018, 55(7): 071015.
- [3] Zhan K, Li Q Q, Teng J C, et al. Multifocus image fusion using phase congruency [J]. Journal of Electronic Imaging, 2015, 24(3): 033014.
- [4] Liu Y, Wang Z F. Simultaneous image fusion and denoising with adaptive sparse representation [J]. IET Image Processing, 2015, 9(5): 347-357.
- [5] Burt P, Adelson E. The Laplacian pyramid as a compact image code [J]. IEEE Transactions on Communications, 1983, 31(4): 532-540.
- [6] Li H, Manjunath B S, Mitra S K. Multisensor image fusion using the wavelet transform [J]. Graphical Models and Image Processing, 1995, 57(3): 235-245.
- [7] Zhang Q, Guo B L. Multifocus image fusion using the nonsubsampling contourlet transform[J]. Signal Processing, 2009, 89(7): 1334-1346.
- [8] Li J, Yang Y C, Dang J W, et al. Multi-focus image

fusion based on NSCT and guided filtering[J]. Laser & Optoelectronics Progress, 2018, 55(7): 071007.

李娇, 杨艳春, 党建武, 等. 基于 NSCT 与引导滤波的多聚焦图像融合[J]. 激光与光电子学进展, 2018, 55(7): 071007.

- [9] Li S T, Kang X D, Hu J W. Image fusion with guided filtering [J]. IEEE Transactions on Image Processing, 2013, 22(7): 2864-2875.
- [10] Wang K, Li Z W, Zhu C D, et al. Local stereo matching algorithm based on secondary guided filtering [J]. Laser & Optoelectronics Progress, 2019, 56(8): 116-122. 王凯, 李志伟, 朱成德, 等. 基于二次引导滤波的局部立体匹配算法[J]. 激光与光电子学进展, 2019, 56(8): 116-122.
- [11] Li S T, Kang X D, Hu J W, et al. Image matting for fusion of multi-focus images in dynamic scenes[J]. Information Fusion, 2013, 14(2): 147-162.
- [12] Liu Y, Liu S P, Wang Z F. Multi-focus image fusion with dense SIFT[J]. Information Fusion, 2015, 23: 139-155.
- [13] Li H F, Chai Y, Yin H P, et al. Multifocus image fusion and denoising scheme based on homogeneity similarity [J]. Optics Communications, 2012, 285(2): 91-100.
- [14] Mustafa H T, Yang J, Zareapoor M. Multi-scale convolutional neural network for multi-focus image fusion[J]. Image and Vision Computing, 2019, 85: 26-35.
- [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] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 2261-2269.
- [17] Lin M, Chen Q, Yan S. Network in network [J/OL]. (2014-03-04) [2019-07-21]. <https://arxiv.xilesou.top/abs/1312.4400>.
- [18] The PASCAL visual object classes challenge 2007 [EB/OL]. [2019-07-21]. <http://host.robots.ox.ac.uk/pascal/VOC/voc2007/>.
- [19] 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.

- [20] Yin M, Liu X N, Liu Y, et al. Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampling shearlet transform domain [J]. IEEE Transactions on Instrumentation and Measurement, 2019, 68(1): 49-64.
- [21] Nejati M, Samavi S, Shirani S. Multi-focus image fusion using dictionary-based sparse representation [J]. Information Fusion, 2015, 25: 72-84.
- [22] 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.
- [23] Kumar M, Dass S. A total variation-based algorithm for pixel-level image fusion [J]. IEEE Transactions on Image Processing, 2009, 18(9): 2137-2143.
- [24] Wang Z, Bovik A C. A universal image quality index [J]. IEEE Signal Processing Letters, 2002, 9(3): 81-84.