

基于超像素级卷积神经网络的多聚焦图像融合算法

聂茜茜 肖斌* 毕秀丽 李伟生

(重庆邮电大学计算智能重庆市重点实验室 重庆 400065)

摘要: 该文提出了基于超像素级卷积神经网络(sp-CNN)的多聚焦图像融合算法。该方法首先对源图像进行多尺度超像素分割, 将获取的超像素输入sp-CNN, 并对输出的初始分类映射图进行连通域操作得到初始决策图; 然后根据多幅初始决策图的异同获得不确定区域, 并利用空间频率对其再分类, 得到阶段决策图; 最后利用形态学对阶段决策图进行后处理, 并根据所得的最终决策图融合图像。该算法直接利用超像素分割块进行图像融合, 其相较以往利用重叠块的融合算法可达到降低时间复杂度的目的, 同时可获得较好的融合效果。

关键词: 多聚焦图像融合; 卷积神经网络; 超像素分割; 空间金字塔池化

中图分类号: TN911.73; TP751

文献标识码: A

文章编号: 1009-5896(2021)04-0965-09

DOI: 10.11999/JEIT191053

Multi-focus Image Fusion Algorithm Based on Super Pixel Level Convolutional Neural Network

NIE Xixi XIAO Bin BI Xiuli LI Weisheng

(Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: This paper proposes a multi-focus image fusion algorithm based on super pixel-level Convolutional Neural Network (sp-CNN). In this method, multi-scale super pixel segmentation is firstly applied to the source image to obtain the super pixels. Secondly, the sp-CNN is proposed to acquire the initial decision maps. Thirdly, according to the similarities and differences of the multiple initial decision maps, the uncertain region is reclassified by spatial frequency to obtain the phase decision map. At last, the final decision map is achieved to fuse the source images by post-processing the phase decision graph with morphology. Experimental results show that the proposed method achieves the goal of reducing time complexity and attains better fusion effect compared with the state-of-the-art fusion methods which utilize overlapping blocks.

Key words: Multi-focus image fusion; Convolutional Neural Network (CNN); Super pixel segmentation; Spatial pyramid pooling

1 引言

从光学理论来说, 当光学镜头聚焦于被摄物的一点时, 只能对焦点及其附近一定范围的区域形成较为清晰的图像, 为得到全景都清晰的图像, 多聚焦图像融合技术被提出。多聚焦图像融合是指在同一时间利用多种传感器对同一场景的不同目标拍摄多幅图像, 利用图像中信息的互补性和时空上的相关性合成一幅对目标信息描述更清晰可靠的融合图

像。该技术让融合后的图像更利于计算机处理与分析, 且更加适合人类的视觉感知, 因此具有非常重要的理论和现实意义。目前, 图像融合技术在遥感处理^[1]、医学成像^[2]和计算机视觉^[3]等领域中发挥着重要的作用。从研究方法来看, 常见方法大致可分为3类: 基于变换域、空间域和神经网络的多聚焦融合算法。

基于变换域的多聚焦图像融合思路是先对多源图像进行图像变换得到变换系数, 再将其进行组合后融合图像, 其中变换域的选择和高低频的融合规则至关重要。该算法具有较强的鲁棒性, 但时间复杂度较高, 目前大体可分为3大类: 金字塔分解^[4], 小波变换^[5]和多尺度几何分析^[6]。

基于空间域的多聚焦图像融合思路是直接源图像上进行融合, 根据融合规则大致可分为3类:

收稿日期: 2019-12-30; 改回日期: 2020-10-28; 网络出版: 2020-12-12

*通信作者: 肖斌 xiaobin@cqupt.edu.cn

基金项目: 国家重点研发计划(2016YFC1000307-3), 国家自然科学基金(61976031, 61806032)

Foundation Items: The National Key Research and Development Project of China (2016YFC1000307-3), The National Natural Science Foundation of China (61976031, 61806032)

基于像素、区域和图像块的多聚焦图像融合方法。基于像素的方法^[7]在预处理时需要严格配准的图像,这表现出对噪声和边缘的高度敏感性。基于区域的多聚焦图像融合在一定程度上能够降低算法对噪声的敏感度和配准误差对融合性能的影响,但该类方法直接以像素作为处理对象,因此算法的计算复杂度较高,且在融合图像边缘处易产生模糊现象。基于图像块的方法的最优分块大小设定是一个难点,很难完全恰好将清晰和模糊区域分隔出来,因此该算法容易产生块效应。目前几种典型算法有:基于引导滤波(GF)^[8]、Image Matting(IM)^[9]和基于尺度不变特征变换(DSIFT)^[10]。

基于神经网络的多聚焦图像融合思想是将源图像分割成重叠块输入网络模型中,并根据所得的聚焦测度判断结果融合图像。例如,基于脉冲耦合网络(PCNN)^[11]的融合方法通过模拟生物大脑皮层的脉冲现象,选择较清晰的像素点进行融合。文献^[12]首次提出将卷积神经网络(Convolutional Neural Networks, CNN)应用于多聚焦图像融合算法中,学习辨别多聚焦图像中聚焦和非聚焦区域的能力。文献^[13]提出p-CNN (pixel-wise Convolutional Neural Network)的模型通过卷积的方式将滑动图像块输入网络,拟解决时间复杂度过高的问题,但选择 32×32 图像块大小作为输入,以致影响图像聚焦判定的精确度。该类方法是基于学习的方式,能够解决基于空间域与变换域的融合方法需要手工制作聚焦测度的问题。

本文提出了基于超像素级卷积神经网络(super pixel based CNN, sp-CNN)的多聚焦图像融合算法,主要流程如下:首先,借助由粗到细的思想,对源图像进行多尺度的超像素分割^[14],并用sp-CNN作为超像素^[15]的聚焦评价方法;然后,利用空间频率(Spatial Frequency, SF)^[16]对不确定区域进行再分类得到阶段决策图;最后,依据对其作后处理所得到的最终决策图进行图像融合。

本文结构如下:第2节描述sp-CNN模型的设计结构以及训练数据的制作方法;第3节详细讲述本文提出的融合算法;第4节阐述实验结果和分析;第5节为本文的结论。

2 sp-CNN

分类网络在图像融合应用中,需将图像分割成连续的重叠矩形图像块作为网络模型的输入,并利用全连接层作为决策分类器。基于重叠图像块作为网络输入的方式,由于固定图像块大小限制了感受野的特征信息容量,从而导致分类精度的下降。其中大部分CNN以固定尺寸的图像数据作为输入,

是由于全连接层的权重矩阵为固定大小,需保证输入的特征映射图的大小与其匹配。去除全连接层的全卷积神经网络^[17]解决了该问题,但该网络是对图像进行像素级的分类,训练较为复杂,并且对图像的细节不够敏感,分类精度较分类网络低。因此,本文借助空间金字塔池化^[18]的思想,提取出固定长度的特征向量作为全连接层的输入。

在传统基于块的图像融合算法中,由于规则图像块可能同时包含清晰与模糊区域,致使图像边缘融合效果降低,因此本文提出利用不规则块替换规则图像块的思想,选取不规则的超像素^[15]作为图像块,以减少图像块大小选择对融合效果的影响。超像素能够保留图像分割的有效信息,并且基本不会破坏图像中物体的边缘信息。由于Achanta等人^[14]提出的简单线性迭代聚类算法(Simple Linear Iterative Clustering, SLIC)在边缘精度和运行速度上具有良好的表现,因此本文借助其思想,实现不规则块划分。为此本文采用超像素替换像素的方式,更加有利于局部特征的提取与结构信息的表达,并将其作为预处理步骤可以减少后处理复杂度。

基于上述问题,本文借助空间金字塔池化与超像素的思想,提出一个基于超像素级的卷积神经网络模型——sp-CNN。在本文研究中,多聚焦图像融合被看作分类问题,使用sp-CNN对超像素进行分类。sp-CNN为一个端到端的图像分类网络模型,首先将图像数据输入网络;然后隐性地从训练图像数据中进行有监督学习,其整个分类过程是由数据驱动,避免了手工提取特征;最后输出一个概率分布,表示每个类别的概率。

2.1 金字塔池化

空间金字塔池化^[18](Spatial Pyramid Pooling, SPP)可以将任意尺度图像的卷积特征转化为固定大小的特征向量,不仅避免了裁剪和变形操作,且有利于减少信息的丢失。借助金字塔池化的思想,并依据2.2节对数据集Cifar-10从全局与局部两方面进行处理的介绍,因此本文从全局($n=1$)和局部($n=2$)两个层次中对卷积特征采用平均池化去提取各层特征,最后拉伸平铺成一个固定长度的1维特征向量作为全连接层的输入。为此,本文对每一层的池化参数运算方法为

$$(h_K, w_K) = \left[\frac{h_{in}}{n}, \frac{w_{in}}{n} \right], (h_S, w_S) = \left[\frac{h_{in}}{n}, \frac{w_{in}}{n} \right] \quad (1)$$

$$P = \left[\frac{h_K \times n + 1 - h_{in}}{2}, \frac{w_K \times n + 1 - w_{in}}{2} \right] \quad (2)$$

其中,核大小为 (h_K, w_K) ,步幅为 (h_S, w_S) ,填充方式为 P ,池化层数为 n ,特征图的高、宽度分别为

h_{in}, w_{in} 。该方法大致具有3个优点: (1)输入图像可以为任意尺寸; (2)将特征映射图从不同角度提取特征, 再聚合成一个固定长度的特征表示, 因此具有较高的鲁棒性; (3)提高了分类平均精度, 同时大大减少了总训练时间。

2.2 训练数据集

一个分类精度较高的网络模型, 需要大量数据来支撑训练。由于现有图像数据库中无可直接利用的聚焦和非聚焦标签的数据库, 本文在大规模图像数据库Cifar-10¹⁾的基础上创建了一个适用于所提出的多聚焦图像融合算法的训练图像集。该训练集将清晰的原始图像标记为聚焦图像类, 而后在原始图像上添加高斯模糊, 并将其标记为非聚焦图像类。

为使得添加的高斯模糊更加贴近于多聚焦图像中真实的模糊现象, 本文对20组多聚焦图像数据集(Lytro)²⁾进行采样, 并利用估算的方式得到高斯模糊参数, 从而提升sp-CNN的聚焦分类效果。高斯模糊参数的计算方法如下: (1)源图像A与B为一组多聚焦图像对, 在源图像A的清晰区域中随机选取图像大小为 32×32 和 128×128 相应的清晰图像块A1-32和A1-128, 在模糊区域中选取模糊图像块A2-32和A2-128, 并以同样的方式在源图像B中选取清晰图像块B1-32和B1-128和模糊图像块B2-32和B2-128; (2)对清晰块添加规律递增的高斯模糊参数(模糊半径为 $1 \sim 8$, 标准差为 $1 \sim 32$), 得到高斯模糊图像块A1-32b, A1-128b, B1-32b和B1-128b; (3)通过比较源图像的模糊图像块与对应高斯模糊图像块之间(A2-32与A1-32b, A2-128与A1-128b, B2-32与B1-32b, B2-128与B1-128b)的梯度直方图的相似度, 根据最大相似度选取合适的高斯模糊参数。根据模糊参数计算试验得到标准差为3及滤波大小为 16×16 的高斯模糊参数, 图像块的选取方式如图1所示。

本文所提方法是基于超像素级的多聚焦图像融合算法, 为了提升sp-CNN的聚焦判定精度, 将

Cifar-10中每幅图像均分成4个 16×16 的图像块, 并将其添加到训练数据集中, 通过两种尺寸(局部 16×16 和全局 32×32)的训练数据集以适应超像素的不规则性。因此, 基于Cifar-10的图像训练数据集中清晰类图像数据包括50000幅 32×32 大小以及200000幅 16×16 大小的图像。采用上述高斯参数对原始图像进行模糊操作, 模糊类图像数据包含同样数量和尺寸的图像。对余下的10000幅图像作同样的方式处理, 将其作为验证数据集, 图像数据集的创建方式如图2所示。

2.3 sp-CNN网络结构

在文献[13]提出的算法中, 验证了深层网络模型较浅层模型, 其聚焦分类精度的提升并不明显, 并且本文考虑到网络参数学习效率等问题, 故将sp-CNN构造为一个浅层的卷积神经网络。sp-CNN首先利用卷积层提取图像高维特征信息, 同时探究聚焦与非聚焦区域之间的隐藏信息; 其次利用最大池化层减少信息维度, 进而扩大感受野的特征信息容量; 再次利用金字塔池化层以确保全连接层的输入数据为固定长度的特征向量; 最后通过全连接层对图像进行分类, 输出聚焦类与非聚焦类的概率分布。在sp-CNN网络中, 3层卷积层的卷积步长都为1, 并在前两层卷积层后连接一个步长为2的最大池化层, 对特征信息进行降维操作, 再将金字塔池化连接在第3层卷积层之后, 最后利用两层全连接层作为分类器。实验表明, sp-CNN卷积神经网络可达到99.95%的分类精度。sp-CNN网络结构模型如图3所示。

3 基于sp-CNN的多聚焦图像融合算法

基于sp-CNN的多聚焦图像融合算法的流程图如图4所示, 此算法主要分为5个步骤: (1)对两幅源图像进行加权平均(Weighted Average, WA), 得到预融合图像; (2)对预融合图像进行多尺度的超像素分割, 将得到的多尺度分割图像中的超像素输入sp-CNN网络, 以进一步获得不同尺度的初始

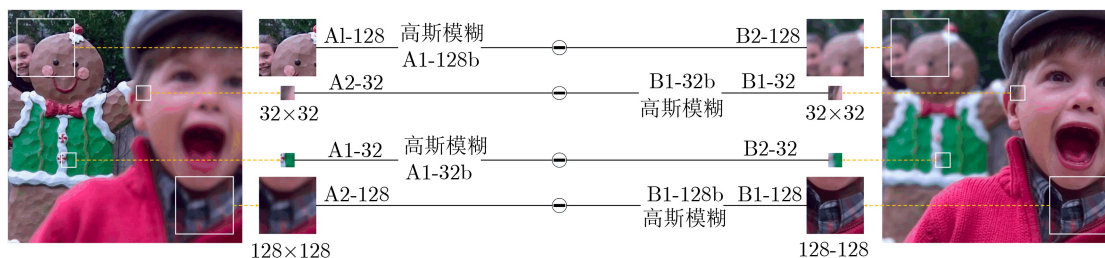


图1 图像块选取

¹⁾ Cifar-10: <<http://www.cs.toronto.edu/~kriz/cifar.html>>

²⁾ Lytro: <<https://github.com/xudif/Multi-focus-Image-Fusion-Dataset>>

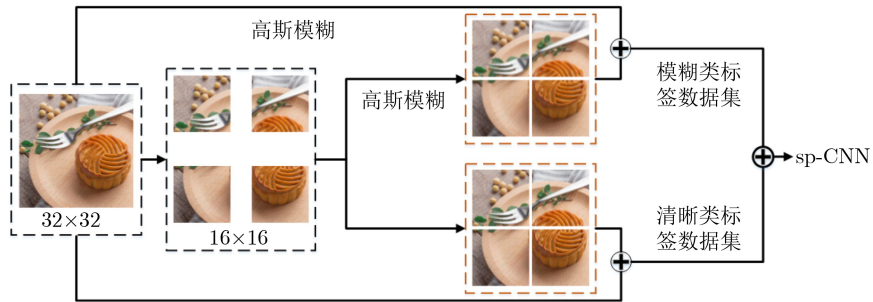


图2 图像数据集的创建

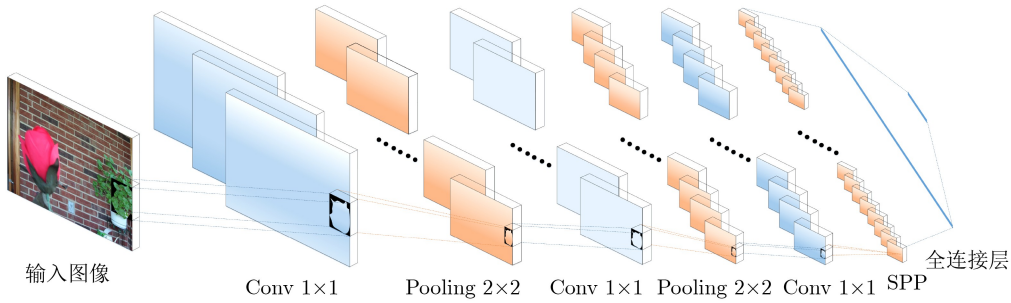


图3 sp-CNN网络结构

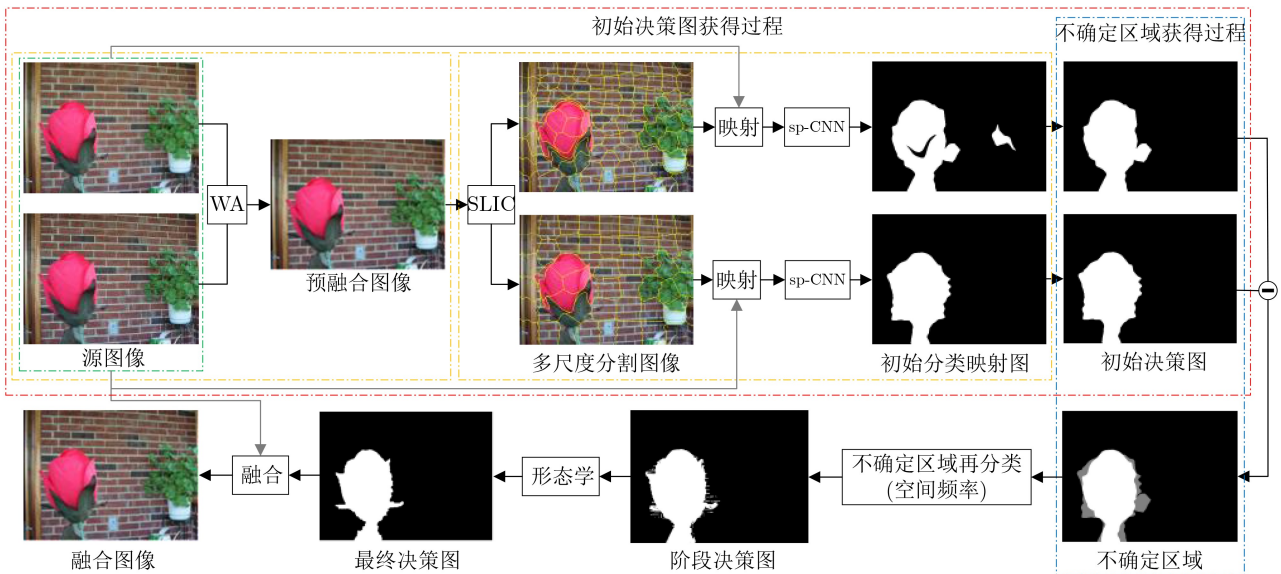


图4 融合算法流程图

分类映射图；(3)利用连通域操作筛除初始分类映射图中的误检部分，得到初始决策图；(4)根据多尺度初始决策图的异同以定位不确定区域，并利用空间频率^[16]操作对其进行再分类，得到阶段决策图；(5)对阶段决策图进行简单的形态学操作以得到最终决策图，并依据融合规则获得最终的多聚焦融合图像。为了简单起见，本文仅对两幅多聚焦图像做融合分析，多幅多聚焦图像可以依据两幅图像的融合扩展得到，融合效果分析见4.4节。图4中，绿色边框为两幅源图像；黄色边框分别为预融合图像与初始分类映射图的获得过程；蓝色边框为不确

定区域的获得过程；红色边框为初始决策图的获得过程。

3.1 初始决策图的获得

本文对两幅源图像进行加权平均得到预融合图像，利用SLIC对其进行分割，然后将得到的分割图映射到两幅多聚焦源图像以获得分割后的源图像。经过超像素分割之后，源图像被划分为不规则的超像素，而由于卷积神经网络是以规则图像块作为输入，因此本文采用补零的方式将不规则图像块补全为规则的矩形图像块，此填补操作不会引入外部信息，且不会增加额外计算消耗，补零方式如图5所示。

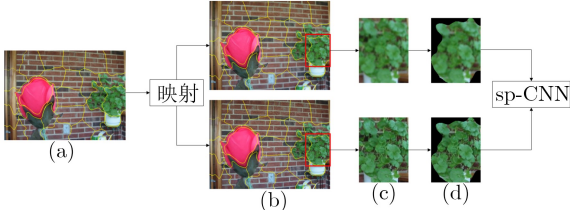


图5 超像素补零, (a)预融合图像的分割图; (b)两幅源图像的分割图; (c)局部区域; (d)不属于同一区域的像素补零

在将补零图像块输入sp-CNN网络后, 可得聚焦与非聚焦的概率分布。由于直接比较聚焦概率值难以得到准确的初始分类映射图, 因此本文采用对数函数的方法对图像进行聚焦评估。聚焦评价计算如式(3)

$$w = a \ln p + b \ln (1 - p) \quad (3)$$

其中, p 表示聚焦概率值, w 为超像素的聚焦评价, a 和 b ($a + b = 1, a = 0.6b = 0.4$)均为经验值, 权重系数 a 起加强作用。根据聚焦分类规则逐个比较源图像的超像素对的聚焦评价, 得到初始分类映射图。本文超像素的聚焦分类规则如式(4)所示

$$M = \begin{cases} 1, & w_1(i) > w_2(i) \\ 0, & w_1(i) \leq w_2(i) \end{cases} \quad (4)$$

其中, M 为初始分类映射图, w_i 表示第 i 幅的超像素聚焦评价。初始分类映射图可能包含一些误检区域, 在聚焦区域中包含非聚焦点, 或者在非聚焦区域包含聚焦点。本文采用简单连通域方法滤掉误检的超像素, 具体来说是将初始分类映射图小于给定区域阈值的聚焦区域移除, 而小于相同阈值的空洞被填充为聚焦区域, 其中区域阈值设置为一幅源图像大小的1%。初始分类映射图再经过连通域处理, 得到初始决策图 D_S 。

3.2 阶段决策图的获得

模糊数学是一种探究和处理模糊性现象的数学方法论, 用该方法来描述多聚焦图像在目标边缘存在聚焦与非聚焦之间的模糊界限, 即边界存在分类模糊概念。本文将不确定区域再分类的方式应用于多聚焦图像融合算法中, 以达到更好的融合效果。较小尺度的超像素分割精度较高, 但容易出现误检区域, 而较大尺度的超像素分割精度较低, 但聚焦分类精度高, 因此结合这两种尺度的优劣, 本文将

利用两种尺度进行超像素分割并补零成规则图像块。将上述图像块分别输入到sp-CNN中得到两种尺度下的初始决策图, 根据分类规则以获得初始决策图中的不确定区域(如图5中灰色区域所示)。分类规则如式(5)所示

$$D_{un} = \begin{cases} 1, & D_{S_1}(x, y) = 1, \quad D_{S_2}(x, y) = 1 \\ 0, & D_{S_1}(x, y) = 0, \quad D_{S_2}(x, y) = 0 \\ 0.5, & \text{其它} \end{cases} \quad (5)$$

其中, D_{un} 为不确定区域, $D_{S_i}(x, y)$ 为第 i 幅初始分类映射图中像素 (x, y) 的聚焦值。然后采用空间频率对不确定像素进行再次分类, 最终得到阶段决策图。

3.3 图像融合

最后, 对阶段决策图进行简单的形态学操作^[19], 再根据所得的最终决策图 D 融合源图像 A 和 B , 得到多聚焦融合图像 F 。融合规则如式(6)

$$F(x, y) = D(x, y)A(x, y) + (1 - D(x, y))B(x, y) \quad (6)$$

4 实验结果分析

在这一节中, 为了证明本文所提出的多聚焦图像融合算法的实效性, 将与当前6种具有代表性的多聚焦图像融合算法(IM^[9], GF^[8], DCT+C+V^[20], DSIFT^[10], PCNN^[11]和p-CNN^[13])进行比较, 上述方法的参数设置均为相应论文中的最优值。其中, 多聚焦图像融合的评价方法主要包括主观和客观两种评价标准。

4.1 对比方法的主观评价

图6展示了两幅“高尔夫”不同景深的多聚焦源图像, 以及6种对比方法和本文方法所得的融合图像。

在图7中, 将图像的某一局部进行放大, 并在图像中用红色矩形标记各种算法所产生的视觉偏差区域, 便于主观评价其融合效果。从图7(a)中可以发现DCT+C+V算法的融合图像能够清晰地看到白色斑块, 并有明显的颜色失真。IM算法的多聚焦融合图像如图7(d), 出现了不属于源图像所提供的图像信息, 这是由模糊引起的伪像。还可以观察DSIFT算法的多聚焦融合图像如图7(b)中有伪边缘现象, 这种边缘模糊是由于错误配准而产生的。同



图6 源图像和融合图像

样地，也可以在PCNN算法的多聚焦融合图像如图7(e)中观察到相同的现象。GF可以有效防止伪像出现在未校准区域中，如图7(c)所示，但该方法在非聚焦区域与聚焦区域之间的边界倾向于过度锐化。在图7(f)中p-CNN算法的多聚焦融合图像出现了错误分类的聚焦与非聚焦区域，使得图像中出现明显的模糊区域。从图7中，能够轻易地发现本文算法能够很好地处理这些问题，展现出较好的融合效果。

3对多聚焦融合图像的源图像如图8所示，利用本文方法以及现有的6种对比方法对其进行融合，得到的多聚焦融合图像如图9所示。经过人眼观察，可以从主观方面观测出本文方法在融合效果上的良好表现。

4.2 对比方法的客观评价

客观评价指标作为判断多聚焦图像融合结果好坏的衡量标准，为图像融合算法性能和图像融合质量的优劣性提供客观评价依据。在文献[21]中将已有的12种客观评价方法进行了归纳总结，根据各种方法的性能可大致分为4类。为保证实验的客观及

公正性，本文在4类评价方法中各选取一种作为融合结果的客观评价标准。分别是从信息论角度的 Q_{MI} 表示融合图像从源图像提取的信息量；从图像特征边缘角度的 Q_P 表示包含的显著信息量；从图像结构相似度角度的 Q_w 表示融合图像中保存源图像的结构信息量；以及从人类视觉感知角度的 Q_{af} 表示从源图像中获取的梯度信息量。

融合算法经过4种评价指标所得计算结果，其值越大表示融合效果越好。表1为3组多聚焦融合图像的客观评价指标值，其中客观评价表现最优的结果用加粗表示，展现出本文方法在客观评价上的优越性。

4.3 计算效率对比

本文通过与其它6种方法进行比较，为验证所提出方法在时间复杂度上面的优劣表现。实验所用计算机的硬件条件为：Intel(R) Core(TM) i5-4590CPU @ 3.30 GHz和16.0GB RAM, GPU: NVIDIA Quadro k2200。对比的6种方法所用实验平台均为MATLAB，而本文方法使用Python作为编程语言，将TensorFlow作为实现平台，因此在

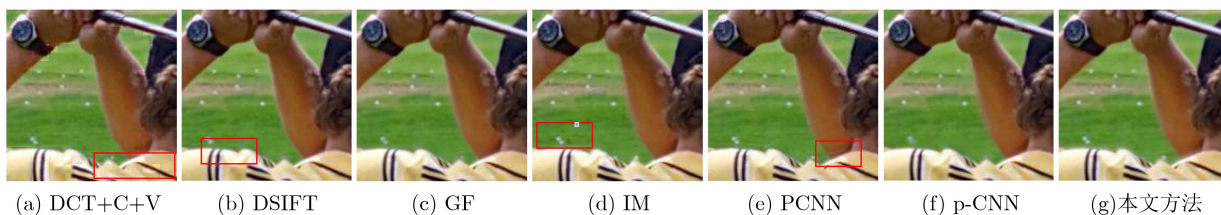


图7 各种方法融合图像的局部放大图

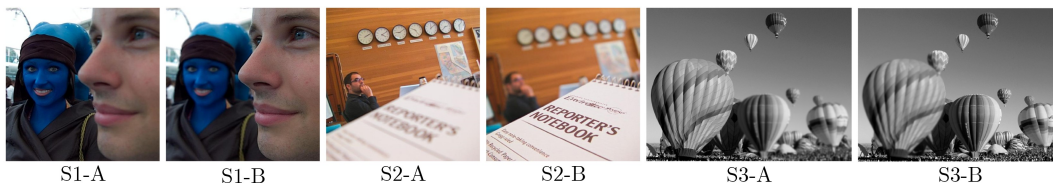


图8 3对多聚焦源图像

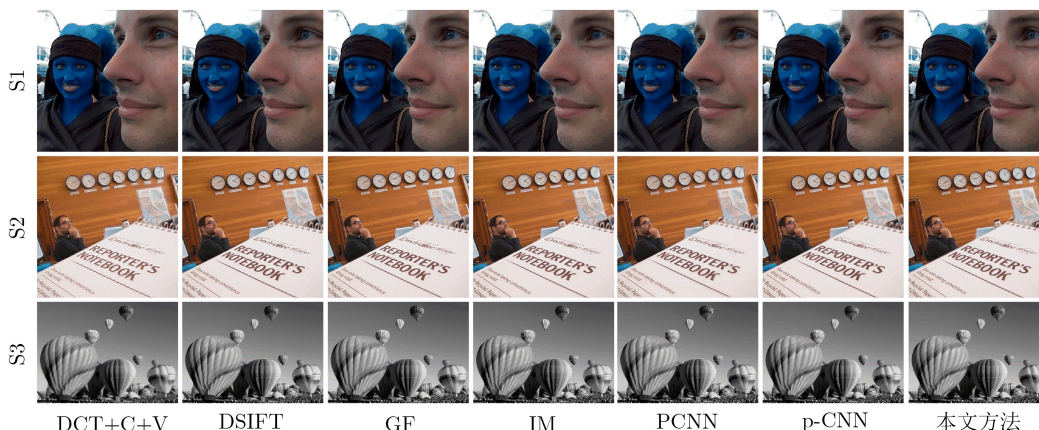


图9 各方法的融合图像

表1 融合图像的客观评价

算法	S1				S2				S3			
	Q_{MI}	Q_P	Q_w	Q_{af}	Q_{MI}	Q_P	Q_w	Q_{af}	Q_{MI}	Q_P	Q_w	Q_{af}
DCT+C+V	8.3869	0.5897	0.8130	0.6342	7.0161	0.6164	0.7704	0.6604	9.3605	0.7533	0.9469	0.7790
DSIFT	10.5221	0.7177	0.8416	0.7408	10.5992	0.7687	0.8226	0.7827	11.1847	0.8153	0.9476	0.8296
GF	10.0879	0.7138	0.8425	0.7377	9.6205	0.7577	0.8184	0.7755	11.1296	0.8153	0.9478	0.8295
IM	10.1618	0.7082	0.8392	0.7332	9.7418	0.7495	0.8112	0.7682	11.0551	0.8115	0.9451	0.8262
PCNN	9.7614	0.6153	0.8078	0.6519	9.8895	0.6927	0.7186	0.7179	11.1311	0.7805	0.9318	0.8033
p-CNN	10.4885	0.7171	0.8410	0.7403	10.6064	0.7655	0.8207	0.7797	11.1851	0.8154	0.9476	0.8296
本文方法	10.4750	0.7184	0.8426	0.7417	10.5767	0.7689	0.8245	0.7832	11.1783	0.8156	0.9477	0.8298

运算效率对比上可能存在偏差。平均运行时间的计算方法是选择两组相同尺寸(320×240, 480×360和640×480)的图像, 每组图像测试5次, 最终将10次数据取平均得到运行时间。在这7种对比方法中, 可以看出PCNN的运行时间最短, 而IM的运行效率最低。本文算法的平均运行时间是由每个块经过sp-CNN网络的时间构成, 其中每个块可以并行处理。每种多聚焦融合算法的平均运行时间如表2所示。

4.4 多聚焦图像序列融合

多聚焦图像序列的融合示例¹⁾, 含有4幅320×240的“书”多聚焦源图像(图10(a1~a4))和1幅参考图像(图10(b)), 每幅源图像都有1个模糊区域。在图10中, 图10(c)~图10(i)分别显示了对比方法的融合图(本文方法, DCT+C+V, DSIFT, GF, IM, PCNN和p-CNN)。

表2 对比方法的平均运行时间(s)

方法	320×240	480×360	640×480
DCT+C+V	0.82	0.93	1.12
DSIFT	1.93	3.89	6.48
GF	0.85	2.81	6.74
IM	10.59	24.25	38.18
PCNN	0.71	1.82	3.16
p-CNN	1.97	3.91	6.62
本文方法	1.15	2.53	4.14

IM, PCNN和p-CNN)。由于本实验中有参考图像, 因此本文使用4个全参考目标图像质量指标(峰值信噪比(PSNR)^[22]、结构相似度(SSIM)^[23]、均方根误差(RMSE)^[21]和梯度相似度(GS)^[24])。其中, RMSE值越小, 说明融合后的图像越接近参考图像, 而SSIM和GS值越接近1, 以及PSNR值越大, 则表示融合效果越好。从表3可以发现, 在GS指标上, 本文方法相较于最好的GF仅相差0.00001, 难以在视觉上识别其间的差异。此外, 在PSNR和SSIM上, 本文方法表现最优。

5 结论

本文提出了一种基于超像素级卷积神经网络的多聚焦图像融合算法。利用超像素分割方法划分源图像为语义相关的不规则块, 尽可能让图像块只包含聚焦或者非聚焦区域, 一定程度上弥补了基于单个像素点之间相关性小的不足, 进而有效减少融合像素的错误选取, 降低对边缘的敏感性, 同时也减少了块效应。本文直接利用超像素分割块进行融合, 致使融合图像中灰度值不会改变, 实现了信息的相互补充, 同时减少了重复数据检测, 降低了时间复杂度。由于使用自学习方式的sp-CNN作为聚焦评价方法, 实现了输入图像为任意大小的目的, 且避免了手工制作聚焦测度的影响, 因此显著提高

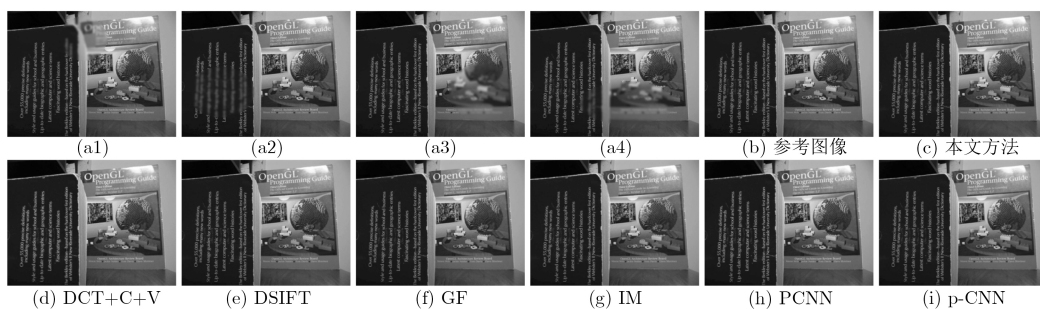


图10 对比方法的融合图像

¹⁾ 融合示例: <<https://github.com/sametaymaz/Multi-focus-Image-Fusion-Dataset>>

表3 融合图像的客观评价

	PSNR	SSIM	RMSE	GS
DCT+C+V	26.15017	0.88263	0.01457	0.98793
DSIFT	28.09234	0.90525	0.01552	0.98890
GF	28.10849	0.90569	0.01534	0.98894
IM	27.89434	0.90344	0.01527	0.98885
PCNN	27.66842	0.90257	0.01458	0.98872
p-CNN	28.09490	0.90537	0.01551	0.98890
本文方法	28.10925	0.90571	0.01532	0.98893

了算法的鲁棒性。在采用对数函数作为融合规则中，细节信息强度的度量不再仅仅依赖于某一点，而是根据待融合系数局部区域的统计特性来选取像素系数，因此整体融合效果较好。

参考文献

- [1] RAO Yizhou, HE Lin, and ZHU Jiawei. A residual convolutional neural network for pan-sharpening[C]. 2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP), Shanghai, China, 2017: 1–4. doi: [10.1109/RSIP.2017.7958807](https://doi.org/10.1109/RSIP.2017.7958807).
- [2] YIN Ming, LIU Xiaoning, LIU Yu, *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. doi: [10.1109/TIM.2018.2838778](https://doi.org/10.1109/TIM.2018.2838778).
- [3] 朱浩然, 刘云清, 张文颖. 基于灰度变换与两尺度分解的夜视图像融合[J]. *电子与信息学报*, 2019, 41(3): 640–648. doi: [10.11999/JEIT180407](https://doi.org/10.11999/JEIT180407).
ZHU Haoran, LIU Yunqing, and ZHANG Wenying. Night-vision image fusion based on intensity transformation and two-scale decomposition[J]. *Journal of Electronics & Information Technology*, 2019, 41(3): 640–648. doi: [10.11999/JEIT180407](https://doi.org/10.11999/JEIT180407).
- [4] PETROVIC V S and XYDEAS C S. Gradient-based multiresolution image fusion[J]. *IEEE Transactions on Image Processing*, 2004, 13(2): 228–237. doi: [10.1109/TIP.2004.823821](https://doi.org/10.1109/TIP.2004.823821).
- [5] 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. doi: [10.1016/j.inffus.2005.09.006](https://doi.org/10.1016/j.inffus.2005.09.006).
- [6] ZHANG Qiang and GUO Baolong. Multifocus image fusion using the nonsubsampling contourlet transform[J]. *Signal Processing*, 2009, 89(7): 1334–1346. doi: [10.1016/j.sigpro.2009.01.012](https://doi.org/10.1016/j.sigpro.2009.01.012).
- [7] LI Shutao, KANG Xudong, FANG Leyuan, *et al.* Pixel-level image fusion: A survey of the state of the art[J]. *Information Fusion*, 2017, 33: 100–112. doi: [10.1016/j.inffus.2016.05.004](https://doi.org/10.1016/j.inffus.2016.05.004).
- [8] LI Shutao, KANG Xudong, and HU Jianwen. Image fusion with guided filtering[J]. *IEEE Transactions on Image Processing*, 2013, 22(7): 2864–2875. doi: [10.1109/TIP.2013.2244222](https://doi.org/10.1109/TIP.2013.2244222).
- [9] LI Shutao, KANG Xudong, HU Jianwen, *et al.* Image matting for fusion of multi-focus images in dynamic scenes[J]. *Information Fusion*, 2013, 14(2): 147–162. doi: [10.1016/j.inffus.2011.07.001](https://doi.org/10.1016/j.inffus.2011.07.001).
- [10] LIU Yu, LIU Shuping, and WANG Zengfu. Multi-focus image fusion with dense SIFT[J]. *Information Fusion*, 2015, 23: 139–155. doi: [10.1016/j.inffus.2014.05.004](https://doi.org/10.1016/j.inffus.2014.05.004).
- [11] WANG Zhaobin, MA Yide, and GU J. Multi-focus image fusion using PCNN[J]. *Pattern Recognition*, 2010, 43(6): 2003–2016. doi: [10.1016/j.patcog.2010.01.011](https://doi.org/10.1016/j.patcog.2010.01.011).
- [12] LIU Yu, CHEN Xun, PENG Hu, *et al.* Multi-focus image fusion with a deep convolutional neural network[J]. *Information Fusion*, 2017, 36: 191–207. doi: [10.1016/j.inffus.2016.12.001](https://doi.org/10.1016/j.inffus.2016.12.001).
- [13] TANG Han, XIAO Bin, LI Weisheng, *et al.* Pixel convolutional neural network for multi-focus image fusion[J]. *Information Sciences*, 2018, 433/434: 125–141. doi: [10.1016/j.ins.2017.12.043](https://doi.org/10.1016/j.ins.2017.12.043).
- [14] ACHANTA R, SHAJI A, SMITH K, *et al.* SLIC superpixels compared to state-of-the-art superpixel methods[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(11): 2274–2282. doi: [10.1109/TPAMI.2012.120](https://doi.org/10.1109/TPAMI.2012.120).
- [15] REN Xiaofeng and MALIK J. Learning a classification model for segmentation[C]. The 9th IEEE International Conference on Computer Vision, Nice, France, 2003: 10–17. doi: [10.1109/ICCV.2003.1238308](https://doi.org/10.1109/ICCV.2003.1238308).
- [16] ESKICIOGLU A M and FISHER P S. Image quality measures and their performance[J]. *IEEE Transactions on Communications*, 1995, 43(12): 2959–2965. doi: [10.1109/26.477498](https://doi.org/10.1109/26.477498).
- [17] SHELHAMER E, LONG J, and DARRELL T. Fully convolutional networks for semantic segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 640–651. doi: [10.1109/TPAMI.2016.2572683](https://doi.org/10.1109/TPAMI.2016.2572683).
- [18] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904–1916. doi: [10.1109/TPAMI.2015.2389824](https://doi.org/10.1109/TPAMI.2015.2389824).
- [19] 肖斌, 唐翰, 徐韵秋, 等. 基于Hess矩阵的多聚焦图像融合方法[J]. *电子与信息学报*, 2018, 40(2): 255–263. doi: [10.11999/JEIT170497](https://doi.org/10.11999/JEIT170497).
XIAO Bin, TANG Han, XU Yunqiu, *et al.* Multi-focus image fusion based on Hess matrix[J]. *Journal of Electronics*

- & *Information Technology*, 2018, 40(2): 255–263. doi: [10.11999/JETT170497](https://doi.org/10.11999/JETT170497).
- [20] CAO Liu, JIN Longxu, TAO Hongjiang, *et al.* Multi-focus image fusion based on spatial frequency in discrete cosine transform domain[J]. *IEEE Signal Processing Letters*, 2015, 22(2): 220–224. doi: [10.1109/LSP.2014.2354534](https://doi.org/10.1109/LSP.2014.2354534).
- [21] LIU Zheng, BLASCH E, XUE Zhiyun, *et al.* Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(1): 94–109. doi: [10.1109/TPAMI.2011.109](https://doi.org/10.1109/TPAMI.2011.109).
- [22] WANG Zhou and LI Qiang. Information content weighting for perceptual image quality assessment[J]. *IEEE Transactions on Image Processing*, 2011, 20(5): 1185–1198. doi: [10.1109/TIP.2010.2092435](https://doi.org/10.1109/TIP.2010.2092435).
- [23] ZHAO Yuxin, JIA Renfeng, and SHI Peng. A novel combination method for conflicting evidence based on inconsistent measurements[J]. *Information Sciences*, 2016, 367/368: 125–142. doi: [10.1016/j.ins.2016.05.039](https://doi.org/10.1016/j.ins.2016.05.039).
- [24] LIU Anmin, LIN Weisi, and NARWARIA M. Image quality assessment based on gradient similarity[J]. *IEEE Transactions on Image Processing*, 2012, 21(4): 1500–1512. doi: [10.1109/TIP.2011.2175935](https://doi.org/10.1109/TIP.2011.2175935).
- 聂茜茜: 女, 1992年生, 博士, 研究方向为图像处理、深度学习。
肖 斌: 男, 1982年生, 教授, 研究方向为图像处理、模式识别和数字水印。
毕秀丽: 女, 1982年生, 副教授, 研究方向为图像处理、多媒体安全和图像取证。
李伟生: 男, 1975年生, 教授, 研究方向为智能信息处理与模式识别。
- 责任编辑: 陈 倩