

文章编号: 1002-2082 (2020) 03-0500-08

# 改进 VGG 网络的多聚焦图像的融合方法

陈清江, 汪泽百, 柴昱洲

(1. 西安建筑科技大学 理学院, 陕西 西安 710055;  
2. 中国空间技术研究院西安分院, 陕西 西安 710100)

**摘 要:** 为更好地保留原图像信息, 提高图像融合性能, 提出一种改进 VGG 卷积神经网络与边缘像素统计特征相结合的融合算法。首先, 该算法将完整图像拆分成图像块, 以图像块的预处理来获取较高的图像分类, 精度达到 0.985 以上, 以改进的 VGG 卷积神经网络来加快模型收敛速度, 当图像块输入到网络当中, 可以初步得到二分类的权值矩阵。其次, 在高频细节部分, 对于左聚焦图像和右聚焦图像的清晰模糊模块分别进行模糊化处理, 根据像素点之间的统计特征经阈值分割后得到有明显边界的权值矩阵。最后, 结合两次分割的权值矩阵, 通过加权求和的融合策略, 得到处处清晰的聚焦图像。为说明算法有效性, 在实验部分展示其融合主观视觉效果图与信息熵等客观评价, 该算法对比其他算法表现突出, 可较好地保留原图像的信息。

**关键词:** 图像融合; VGG; 图像分块; 统计特征; 多聚焦图像

中图分类号: TN201; TP391.4

文献标志码: A

DOI: 10.5768/JAO202041.0302003

## Multi-focus image fusion method based on improved VGG network

CHEN Qingjiang, WANG Zebai, CHAI Yuzhou

(1. College of science, Xi'an University of Architecture and Technology, Xi'an 710055, China;  
2. Cast-Xi'an, Institute of Space Radio Technology, Xi'an 710100, China)

**Abstract:** In order to better retain the original image information and improve the image fusion performance, an improved fusion algorithm combining visual geometry group(VGG) convolutional neural network with edge pixel statistical features was proposed. Firstly, this algorithm divided the complete image into image blocks, preprocessed the image blocks to obtain a higher image classification accuracy of 0.985 or more, and used the improved VGG convolutional neural network to accelerate the convergence of the model. When the image blocks were input into the network, the weight matrix of binary classification could be preliminarily obtained. Secondly, in the high-frequency detail part, the clear blur modules of left-focus image and right-focus image were respectively subjected to blurring processing, and the weights matrix with obvious boundaries was obtained by the threshold segmentation based on the statistical feature between the pixel points. Finally, combined with the two-step weight matrix, the fusion strategy of weighted sum was used to obtain the clear focus image. In order to illustrate the effectiveness of the algorithm, the experimental part shows the objective evaluation of its fusion subjective visual effect diagram and information entropy, which is outstanding compared with other algorithms and can better retain the information of the original image.

**Key words:** image fusion; VGG; image segmentation; statistical features; multi-focus image

## 引言

图像融合是图像处理中的重要部分, 能够协

同利用同一场景的多种传感器图像信息, 输出一幅更适用于人类视觉感知或计算机进一步处理

收稿日期: 2019-10-12; 修回日期: 2019-12-03

基金项目: 国家自然科学基金 (61403298); 陕西省自然科学基金 (2015JM1024); 陕西省教育厅专项科研计划 (2013JK0586)

作者简介: 陈清江 (1966-), 男, 博士, 教授, 主要从事小波分析及其在图像处理中的应用等方面的研究。

E-mail: chen66xauat@126.com

与分析的融合图像。它可明显改善单一传感器的不足, 提高结果图像的清晰度及信息含量, 有利于更为准确、可靠、全面地获取目标或场景信息。

在可见光图像中成像设备捕获目标图像, 对于捕获目标部分的有效聚焦图像是清晰可见的, 而非目标区域的其他物体则为模糊。因此, 对于光学透镜捕捉的成像图很难做到所有物体处处聚焦, 提出多种多焦点图像融合算法<sup>[1]</sup>。总的来说, 这些方法可以分为两类: 变换域方法和空间域方法<sup>[2]</sup>, 多尺度变换(MST)是最常用的变换域方法之一。传统的融合方法包括基于金字塔的图像融合<sup>[3]</sup>、基于剪切变换图像融合和基于非子采样轮廓变换(NSCT)<sup>[4]</sup>的图像融合算法等。最近提出的融合方法包括基于像素的融合方法, 主要有引导滤波(DSIFT)<sup>[2]</sup>、基于多尺度加权梯度的图像融合算法(MWGF)<sup>[5]</sup>、基于低秩矩阵(LRR)<sup>[6]</sup>的多聚焦噪声图像融合算法、基于离散小波变换的多聚焦图像融合算法<sup>[7]</sup>, 随着深度学习的发展, 结合深度学习解决多聚焦图像融合<sup>[8]</sup>的方法得到广泛推广, 如基于卷积神经网络(CNN)的多聚焦图像融合<sup>[9-10]</sup>、基于全卷积网络的多聚焦图像融合<sup>[11-12]</sup>。对于一个简单的二分类问题, 本文优先选择精度相对较高的 VGGNet 网络<sup>[13]</sup>进行修改。输入子块的多特征, 保证了图像的分类精度。全图处理容易损失较多高频信息, 于是本文只对左右图像的混合聚焦散焦部分进行处理, 在信息熵、互信息等方面得到较好的融合效果。

## 1 网络结构模型

### 1.1 VGG16 模型介绍

VGGNet<sup>[13]</sup>是由牛津大学的 K.Simonyan 和 A. Zisserman 提出的卷积神经网络模型, 该模型诠释了隐层深度对于预测精度的影响, 训练时间和 AlexNet 相比大大减少。该模型在 ImageNet 中达到了 92.7% 的 top5 测试精度, VGGNet 结构根据层数的不同分为不同的版本, 常用的结构是 VGG16 和 VGG19, 在 VGG16 中所有的卷积层都有相同的配置, 卷积层中卷积核大小为  $3 \times 3$ , 步长大小为 1, 最大池化层共 5 个, 其核大小都为  $2 \times 2$ , 步长为 2; 全连接层共 3 个, 前两层共有 4 096 个通道, 第 3 层共 1 000 个标签类别; 最后一层为 Softmax 层; 所有隐层后都带有 ReLU 非线性激活函数, 总体结构如图 1 所示。

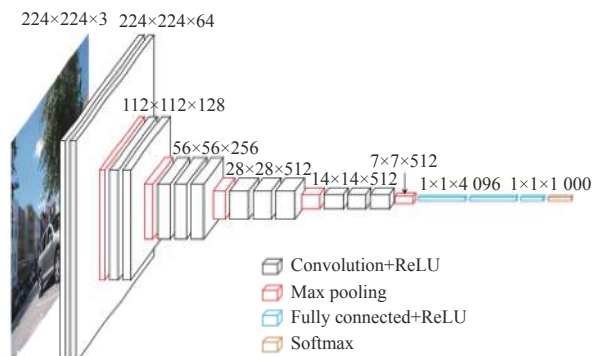


图 1 VGG16 网络模型

Fig. 1 VGG16 network model

### 1.2 Crop-VGG 网络结构

本文只需要处理聚焦图和散焦图的简单二分类问题, 并不需要过深的隐层, 多个残余的隐层影响了训练速率, 通过逐个隐层删除, 部分隐层添加, 最终设计为本文的 Crop-VGG 网络, 在保证精度的同时提高了训练速率。

如图 2 所示, Crop-VGG 是基于 VGG16 对二分类问题的更改, 并不需要过深的网络层, 于是裁剪了一部分网络, 在保留原有分类效果的同时, 提高分类速度。与原网络相比, Crop-VGG 卷积层卷积核大小、池化层大小、步长都与原 VGG 网络保持一致, 区别在于卷积层数由原来的 13 变为 8, 由 2-2-3-3-3 变为了 2-3-3 结构, 最大池化层由 5 个变为 3 个, 卷积层中卷积核大小都为  $3 \times 3$ , 步长为 1, 最大池化层核大小  $2 \times 2$ , 步长为 2; 全连接层共 3 层, 前两层通道数由原来的 4 096 变为 2 048, 输出类别变为 2, 每个隐层都跟有非线性激活函数 ReLU, 最后一层为 Softmax 层。基于 caffe 框架下, 将预处理数据集网络训练, 随着迭代次数增加, loss 的收敛速度有明显提升, 随着迭代次数增加, 当 loss 值达到收敛后, 精确率保持在 0.985 以上 (如图 3)。

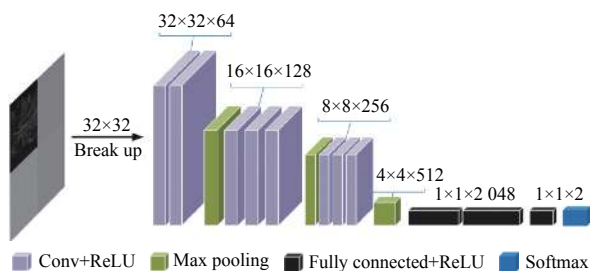


图 2 Crop-VGG 网络模型

Fig. 2 Crop-VGG network model

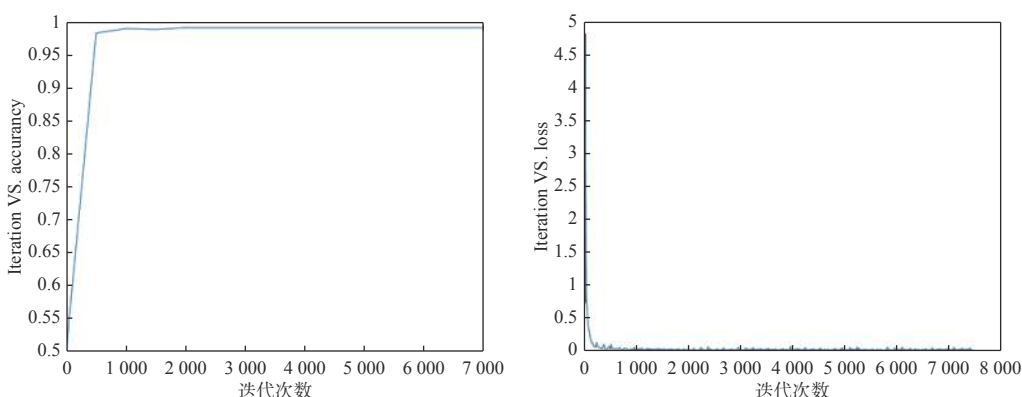


图3 epoc图与loss图

Fig. 3 Diagram of epoc and loss

## 2 使用网络进行统计特征分类

### 2.1 数据预处理

步骤1) 选取纹理多样的一组清晰图, 使用点扩散(point spread function)方法进行模糊处理。因为点源在经过任何光学系统后都会由于衍射而形成扩大的像点, 通过测量系统的点扩展函数, 能够更准确地提取图像信息, 这里使用点扩散函数更加符合光学散焦过程。对于分类训练, 需要扩大清晰和模糊的区分度, 于是设置扩散半径  $r=10$ , 获取一组聚焦和散焦图像。

步骤2) 将两组图像分别分割为  $32 \times 32$  像素的小块。首先使用二维离散小波<sup>[14]</sup>变换对图像进行分解。设原始图像为  $C^0 = (c_{mn}^0)$ , 则

$$\begin{cases} c_{m,n}^{-i} = \frac{1}{2} \sum_{k,l \in \mathbb{Z}} c_{k,l}^{-i+1} \tilde{h}_{k-2n} \tilde{h}_{l-2m}, \\ d_{m,n}^{-i1} = \frac{1}{2} \sum_{k,l \in \mathbb{Z}} c_{k,l}^{-i+1} \tilde{h}_{k-2n} \tilde{g}_{l-2m} \\ d_{m,n}^{-i2} = \frac{1}{2} \sum_{k,l \in \mathbb{Z}} c_{k,l}^{-i+1} \tilde{g}_{k-2n} \tilde{h}_{l-2m}, \\ d_{m,n}^{-i3} = \frac{1}{2} \sum_{k,l \in \mathbb{Z}} c_{k,l}^{-i+1} \tilde{g}_{k-2n} \tilde{g}_{l-2m} \end{cases} \quad (1)$$

式中:  $L = (h_n)_{n \in \mathbb{Z}}$  表示分解过程的低通滤波器;  $H = (g_n)_{n \in \mathbb{Z}}$  表示高通滤波器,  $g_n = (-1)^n h_{1-n}$ ,  $n \in \mathbb{Z}$ ; 而  $\tilde{L} = (\tilde{h}_n)_{n \in \mathbb{Z}}$ ,  $\tilde{H} = (\tilde{g}_n)_{n \in \mathbb{Z}}$  分别表示  $L$ 、 $H$  的镜像滤波器, 于是获取其横向  $HL$ 、纵向  $LH$  以及对角  $HH$  的高频细节分别表示为

$$\begin{cases} D_{-j}^{LH} = (d_{m,n}^{-j1})_{(m,n) \in \mathbb{Z}^2} \\ D_{-j}^{HL} = (d_{m,n}^{-j2})_{(m,n) \in \mathbb{Z}^2} \\ D_{-j}^{HH} = (d_{m,n}^{-j3})_{(m,n) \in \mathbb{Z}^2} \end{cases} \quad (2)$$

再使用 Sobel 算子对图像进行卷积运算, 提取轮廓纹理特征, 并进行  $\frac{1}{2}$  的下采样。

步骤3) 将所有对应位置的 Sobel 算子处理后的图像与步骤2的  $HL$ 、 $LH$ 、 $HH$  拼接成一幅图像组作为输入训练数据输入目标网络中, 完整步骤如图4所示。

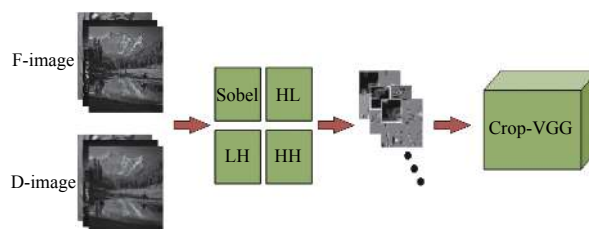


图4 训练集预处理模型

Fig. 4 Training set preprocessing model

### 2.2 分类训练

对于训练好的二分类网络模型逐块进行模糊评测<sup>[15]</sup>, 进行阈值为0.5的分割, 其中, 聚焦区块概率接近于1, 散焦区块概率接近于0。对于介于中间聚焦散焦混合区块做如下判别:

$$\begin{cases} y(i, j) = 1, \sum a \geq 2; \text{ else } y = 0 \\ x(i, j) = 1, \sum a \geq 3; \text{ else } x = 0 \\ c(i, j) = 1, \sum a \geq 5; \text{ else } c = 0 \end{cases} \quad (3)$$

式中  $y$ 、 $x$ 、 $c$  为分别为四角区块、四边区域方块、中间区域方块, 它们是否为1取决于周围方块  $a$  的聚焦块数量, 这种方法可以将图像中空洞区域块有效去除, 如图5所示。

### 2.3 基于统计特征的像素分割

通过分类训练后对目标图A和目标图B进行了分类处理, 如图6所示, 依据形态学分割<sup>[16]</sup>为3个权重矩阵, 分别为  $\alpha$ 、 $\beta$ 、 $\gamma$ 。其中  $\alpha$ 、 $\beta$  为已确定的聚焦区域的权值矩阵, 只需要对  $\gamma$  矩阵的细节处

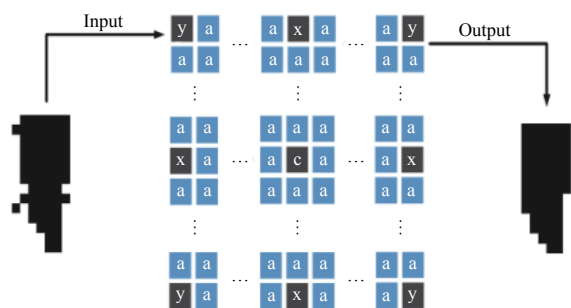


图5 处理误判的区块的矫正矩阵

Fig. 5 Correction matrix for handling misjudged blocks

理。 $T1$ 、 $T2$ 是由两幅原聚焦图像通过 $\gamma$ 权值矩阵分割出来的聚焦散焦混合部分,即:

$$\begin{cases} T1 = A \times \gamma \\ T2 = B \times \gamma \end{cases} \quad (4)$$

分别对  $T1$  和  $T2$  进行点扩散函数  $\partial$  的散焦处理,  $\varphi$  函数用来求目标像素点与  $3 \times 3$  区域邻近像素

点方差,通过统计像素点之间的模糊变化关系,经过阈值分割后能够提高在边界部分像素点区分度,如(5)式所示:

$$\begin{cases} GA = \varphi(T1 - \partial(T2))^2 \\ GB = \varphi(T2 - \partial(T1))^2 \end{cases} \quad (5)$$

由于只判断中间混合区域,因此本文方法实现时受全局误判点干扰较少,边界部分信息明确,生成初步边界区域明显的分割权值图像矩阵  $\tau1$ :

$$\begin{cases} \tau1(i, j) = 1, GA \geq GB \\ \tau1(i, j) = 0, GA < GB \end{cases} \quad (6)$$

对  $\tau1$  中的空洞点再进行一次形态学空洞补全的修复处理,得到最终分割区域明显的  $\tau2$  权值矩阵。最终的融合图像  $AB$  为

$$AB = A \times \alpha + B \times \beta + T1 \times \tau2 + T2 \times (1 - \tau2) \quad (7)$$

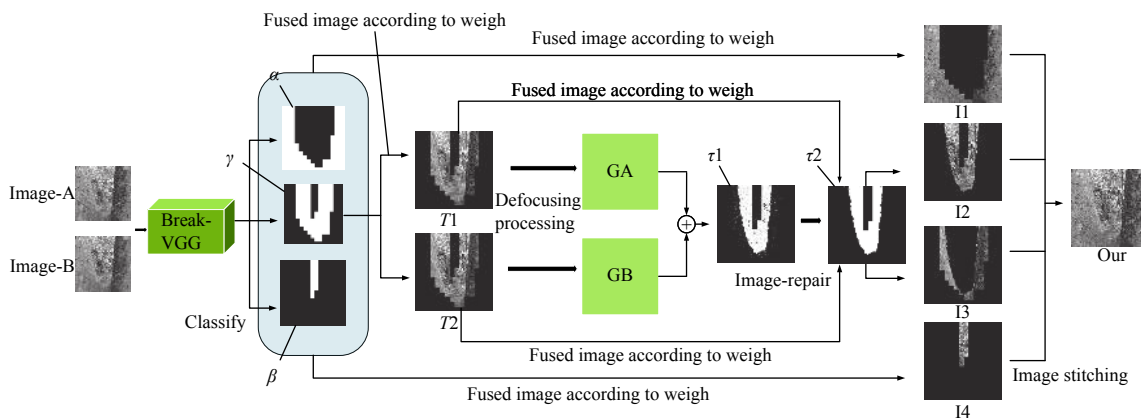


图6 像素级形态融合

Fig. 6 Pixel level morphological fusion

### 3 实验结果

在实验中,我们使用了几对输入图像作为测试图像,验证了该方法的可行性。将本文提出的多焦点图像融合算法与最新的多聚焦图像融合算法如 MWGF<sup>[4]</sup>、DCTvarcv<sup>[17]</sup>、DSIFT<sup>[2]</sup> 等进行了比较,下面介绍图像融合的详细讨论分析结果。

图7~图10展示了图像融合视觉质量的主观评价,为了对比不同的多聚焦图像算法的融合效果,选取4组多聚焦图像。Image A 与 Image B 为需要融合的左右聚焦图,其他图像为每个算法对应的融合图像。例如,在图7中图片整体清晰度较高,在图8中,树木边缘参差的纹理特征比较其他算法接近于原始图像,除去边缘

的其他细节部分完全保留原始图像信息,而在图9中手背部分与图10瓶子边界部分清晰程度较高。

表1~表5为 Vegetables、Tree、A globe、The bottle 4类多聚焦融合图像在几种算法中的表现,通过数据结果可以看出本文算法在信息熵、平均梯度、图像清晰度、边缘信息保持度、互信息上的得分情况基本优于其他算法。这说明本文算法能够减少细节与边缘信息的损失,对于图像整体信息量保存较完整,边缘层次比较多,清晰度也明显高于其他算法。改进的 VGG 网络能够更好地保存高频信息,抽选两组表的数据在各类算法的比较如图11所示。



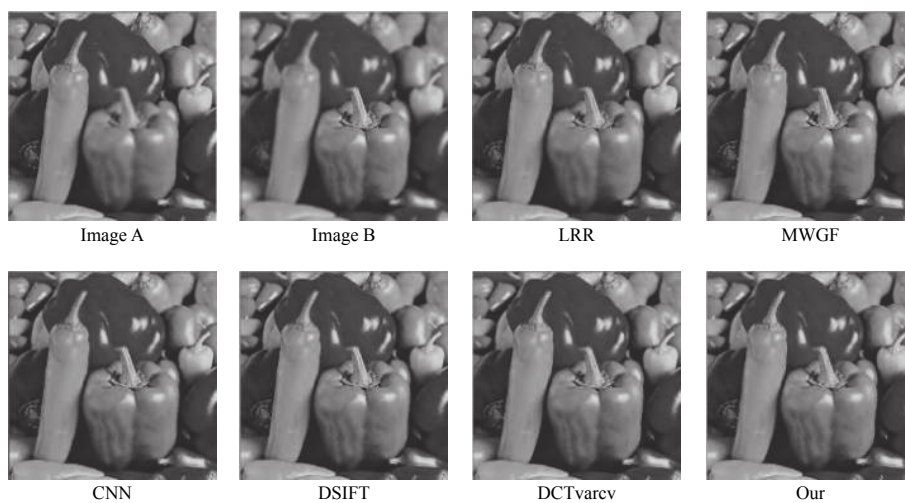


图 7 Vegetables 在不同算法的融合表现

Fig. 7 Fusion performance of Vegetables in different algorithms

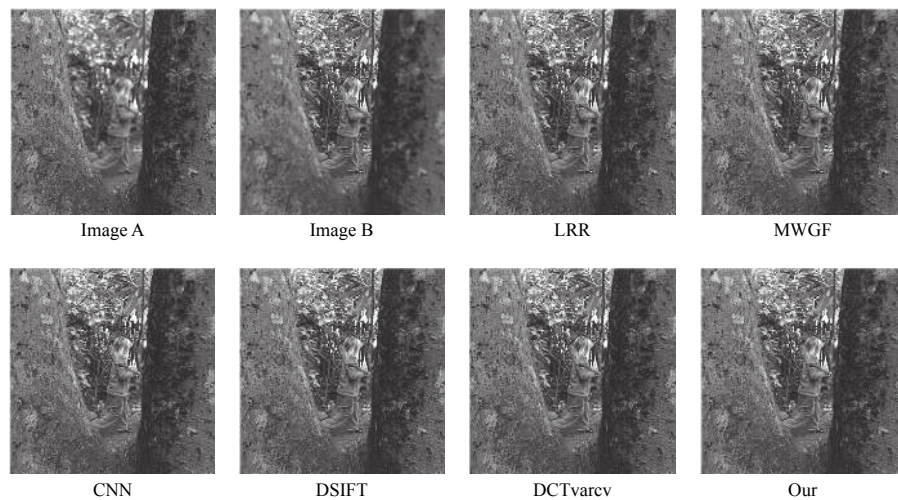


图 8 Tree 在不同算法的融合表现

Fig. 8 Fusion performance of Tree in different algorithms

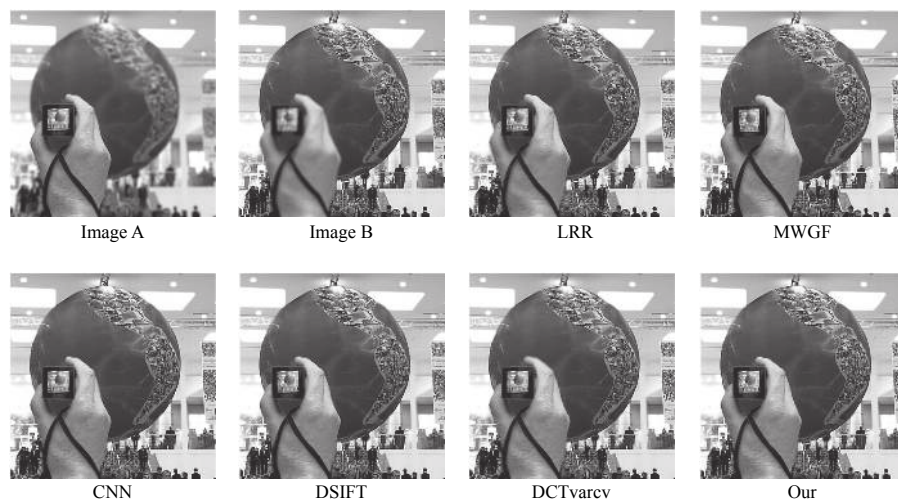


图 9 A globe 在不同算法的融合表现

Fig. 9 Fusion performance of A globe in different algorithms

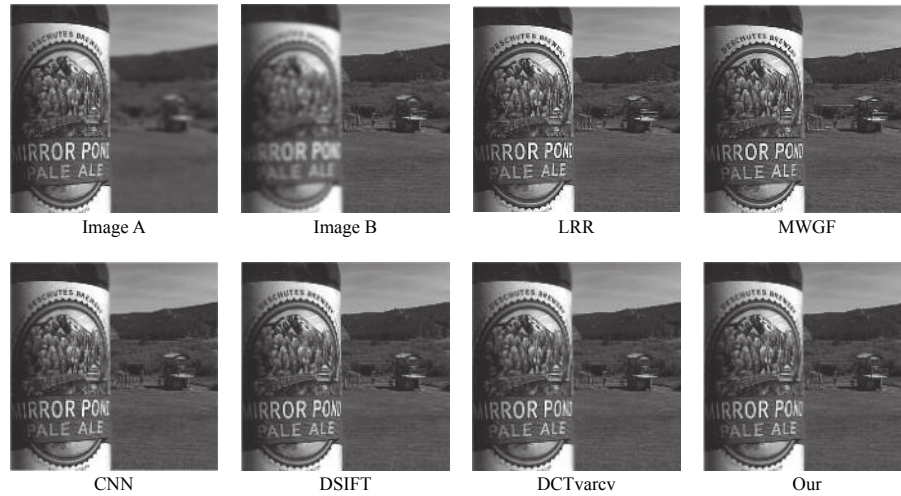


图 10 The bottle 在不同算法的融合表现

Fig. 10 Fusion performance of The bottle in different algorithms

表 1 各类算法在信息熵的对比

Table 1 Comparison of various algorithms in information entropy

$Q_E$	LRR	MWGF	DCTvarecv	DSIFT	CNN	Our
Vegetables	7.602 2	7.599 3	7.602 8	7.590 9	<b>7.603 7</b>	7.599 4
Tree	7.629 9	7.634 3	7.632 8	7.642 1	7.632 2	<b>7.635 5</b>
A globe	7.688 6	7.685 6	7.683 6	7.677 9	7.677 4	<b>7.689 1</b>
The bottle	7.617 6	<b>7.628 9</b>	7.628 3	7.625 2	7.618 1	7.623 5

表 2 各类算法在平均梯度的对比

Table 2 Comparison of various algorithms in average gradient

$Q_{AG}$	LRR	MWGF	DCTvarecv	DSIFT	CNN	Our
Vegetables	7.084 2	7.434 3	7.422 1	7.292 9	7.442 8	<b>7.487 3</b>
Tree	13.597 2	14.551 9	14.546	14.554	14.533 4	<b>14.589 6</b>
A globe	8.473 3	8.875 8	8.850 7	8.847 5	8.865 9	<b>8.884 1</b>
The bottle	8.623 5	9.174 8	9.176 6	9.142 7	9.180 5	<b>9.186 6</b>

表 3 各类算法在图像清晰度的对比

Table 3 Comparison of various algorithms in image clarity

$Q_{Fd}$	LRR	MWGF	DCTvarecv	DSIFT	CNN	Our
Vegetables	7.938 5	8.540 7	8.510 1	8.123 1	8.531 3	<b>8.592 9</b>
Tree	16.423 7	17.691 2	17.679 6	17.660 2	17.669	<b>17.736 9</b>
A globe	10.428 5	10.961 4	10.973 7	10.897 5	10.956 9	<b>10.981 5</b>
The bottle	10.156 7	10.915 9	<b>10.946 4</b>	10.834 4	10.923 5	10.928 9

表 4 各类算法在边缘信息保持度的对比

Table 4 Comparison of various algorithms in edge information retention

$Q_{AB/F}$	LRR	MWGF	DCTvarecv	DSIFT	CNN	Our
Vegetables	0.672 4	0.694 0	0.693 4	<b>0.771 7</b>	0.696 4	0.697 0
Tree	0.678 1	0.741 6	0.742 4	0.743 0	0.741 4	<b>0.743 7</b>
A globe	0.687 7	0.767 6	0.757 9	0.768 1	0.759 9	<b>0.770 8</b>
The bottle	0.693 3	0.801 6	0.793 2	0.8	0.795 6	<b>0.803 8</b>

表 5 各类算法在互信息的对比

Table 5 Comparison of various algorithms in mutual information

$Q_{MI}$	LRR	MWGF	DCTvarev	DSIFT	CNN	Our
Vegetables	4.663	4.728 7	4.756 5	4.709 6	<b>4.770 4</b>	4.758 6
Tree	3.327 4	5.021 8	<b>5.418 4</b>	5.215 6	4.563 9	5.401 8
A globe	4.967	6.277 1	6.41	6.334 8	5.640 9	<b>6.468 5</b>
The bottle	3.956 4	5.358	5.652 7	5.404 6	4.856 7	<b>5.698 2</b>

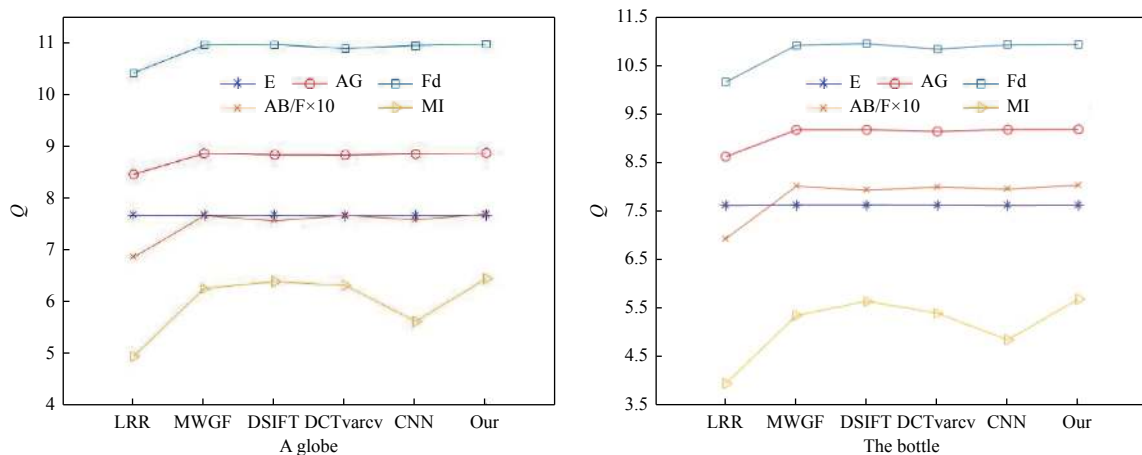


图 11 A globe 和 The bottle 在各类算法的融合表现

Fig. 11 Fusion performance of A globe and The bottle in different algorithms

## 4 结论

本文提出的结合深度学习与统计特征的方法解决复杂的实际问题是可行的, Crop-VGG 网络中能较少运算速率保留更多的原始图像信息,同时结合统计特征的融合算法在各类指标上具有良好的表现,所以针对不同的问题复杂度设计相应的深度网络能更加便捷地处理问题。

### 参考文献:

- [1] LIU S, CHEN J, RAHARDJA S. A new multi-focus image fusion algorithm and its efficient implementation[C]// IEEE Transactions on Circuits and Systems for Video Technology. USA: IEEE, 2019.
- [2] LIU Yu, LIU Shuping, WANG Zengfu. Multi-focus image fusion with dense SIFT[J]. *Information Fusion*, 2015, 23: 139-155.
- [3] SAHU A, BHATEJA V, KRISHN A. Medical image fusion with Laplacian pyramids[C]//2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom). USA:IEEE, 2014: 448-453.
- [4] CHAI Yi, LI Huafeng, ZHANG Xiaoyang. Multifocus image fusion based on features contrast of multiscale products in nonsubsamped contourlet transform domain[J]. *Optik-International Journal for Light and Electron Optics*, 2012, 123(7): 569-581.
- [5] ZHOU Zhiqiang, LI Sun, WANG Bo. Multi-scale weighted gradient-based fusion for multi-focus images[J]. *Information Fusion*, 2014, 20: 60-72.
- [6] LI Hui, WU Xiaojun, DURRANI T. Multi-focus noisy image fusion using low-rank representation[C]. USA: arXiv.2018: 1804. 09325.
- [7] JOSHI K, KIROLA M, CHAUDHARY S, et al. Multi-focus image fusion using discrete wavelet transform method[J/OL]. SSRN, 2009: 3383141 [2019-10-12]. <https://ssrn.com/abstract=3383141>.
- [8] DU C B, GAO S S, LIU Y, et al. Multi-focus image fusion using deep support value convolutional neural network[J]. *Optik*, 2019, 176: 567-578.
- [9] LI H, NIE R, ZHOU D, et al. Proceedings of the 2018 2nd International Conference on Algorithms, Computing and Systems, ICACS '18Convolutional Neural Network Based Multi-Focus Image Fusion[C]. [S.l.]: [s.n.] 2018: 148-154.

- [ 10 ] AMIN-NAJI M, AGHAGOLZADEH A, EZOJI M. Ensemble of CNN for multi-focus image fusion[J]. *Information Fusion*, 2019, 51: 201-214.
- [ 11 ] GUO Xiaopeng, NIE Rencan, CAO Jinde, et al. Fully convolutional network-based multifocus image fusion[J]. *Neural Computation*, 2014, 30(7): 1-26.
- [ 12 ] CHEN Qingjiang, ZHANG Xue. Image defogging algorithms based on multiscale convolution neural network[J]. *Journal of Applied Optics*, 2019, 40(4): 596-602.  
陈清江, 张雪. 基于全卷积神经网络的图像去雾算法[J]. *应用光学*, 2019, 40(4): 596-602.
- [ 13 ] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. *arXiv preprint arXiv*, 2014:1409-1556.
- [ 14 ] LI J, YUAN G, FAN H. Multifocus image fusion using wavelet-domain-based deep CNN[J]. *Computational Intelligence and Neuroscience*, 2019, 2019: 1-23.
- [ 15 ] CHEN Qingjiang, LI Yi, CHAI Yuzhou. A multi-focus image fusion algorithm based on deep learning[J]. *Laser & Optoelectronics Progress*, 2018, 55(7): 246-254.  
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): 246-254.
- [ 16 ] QIU X, LI M, ZHANG L, et al. Guided filter-based multi-focus image fusion through focus region detection[J]. *Signal Processing: Image Communication*, 2019, 72: 35-46.
- [ 17 ] HAGHIGHAT M B A, AGHAGOLZADEH A, SEYEDARABI H. Multi-focus image fusion for visual sensor networks in DCT domain[J]. *Computers & Electrical Engineering*, 2011, 37(5): 789-797.