

# 深度密集卷积神经网络协同检测的多聚焦图像融合方法

杨威<sup>1</sup>, 梅礼晔<sup>2</sup>, 徐川<sup>3</sup>, 张欢<sup>3</sup>, 胡传文<sup>4\*</sup>, 邓騫闯<sup>1</sup>

<sup>1</sup>武昌首义学院信息科学与工程学院, 湖北 武汉 430064;

<sup>2</sup>武汉大学工业科学研究院, 湖北 武汉 430072;

<sup>3</sup>湖北工业大学计算机学院, 湖北 武汉 430068;

<sup>4</sup>浙江省测绘科学技术研究院, 浙江 杭州 311100

**摘要** 针对传统方法不能充分挖掘图像聚焦关联信息导致融合细节失真的问题,提出了一种基于深度密集卷积神经网络协同检测的多聚焦图像融合方法。将多聚焦源图像进行集成实现协同聚焦特征检测,利用深度密集卷积神经网络的特征复用、低级特征与高级特征相结合等特点,来加强多聚焦图像特征表达能力,可以更好地挖掘图像语义信息。采用多尺度金字塔池化策略聚合不同聚焦区域的全局上下文信息,增强聚焦与离焦的区分能力,得到粗略融合概率决策图。进一步采用卷积条件随机场对其进行优化,获得精细化概率决策图,最终得到细节保持的融合图像。将一对多聚焦图像合并为6通道送入网络进行训练,保证了训练时聚焦图像相关性。利用公开数据集对提出的融合方法进行主观与客观评价,实验结果表明该方法具有较好的融合效果,能够充分挖掘聚焦关联信息、保留足够的图像细节。

**关键词** 图像处理; 多聚焦图像; 图像融合; 密集卷积神经网络; 金字塔池化; 协同检测

中图分类号 TP391.9

文献标志码

DOI: 10.3788/LOP202259.2410004

## Multi-Focus Image Fusion Method Based on Cooperative Detection via a Deep Dense Convolutional Neural Network

Yang Wei<sup>1</sup>, Mei Liye<sup>2</sup>, Xu Chuan<sup>3</sup>, Zhang Huan<sup>3</sup>, Hu Chuanwen<sup>4\*</sup>, Deng Qianchuang<sup>1</sup>

<sup>1</sup>School of Information Science and Engineering, Wuchang Shouyi University, Wuhan 430064, Hubei, China;

<sup>2</sup>The Institute of Technological Sciences, Wuhan University, Wuhan 430072, Hubei, China;

<sup>3</sup>School of Computer Science, Hubei University of Technology, Wuhan 430068, Hubei, China;

<sup>4</sup>Zhejiang Academy of Surveying and Mapping, Hangzhou 311100, Zhejiang, China

**Abstract** Traditional methods cannot fully mine image-focus association information, thereby leading to the distortion of fusion details. In this study, a multi-focus image fusion method based on collaborative detection via densely connected convolutional neural networks is proposed to address this issue. Multi-focused source images are integrated to detect focused features collaboratively, and the features of deep dense convolutional networks, such as feature reuse, and the combination of low-level and high-level features are used to enhance the multi-focused image feature representation, which better mine the images' semantic information. By leveraging feature reuse, multi-focus source images are integrated to achieve collaborative focus feature detection. The multi-scale pyramid pooling strategy is used to aggregate the global context information of different focus regions to enhance the ability to distinguish the focused and defocused areas and obtain a rough fusion-probability decision graph. Furthermore, a convolution conditional random field (CRF) is adopted to optimize the decision graph and the refined probabilistic decision graph is obtained. Finally, the fused image is obtained with its details preserved. A pair of multi-focused images are combined into six channels and fed into the network for training, thus ensuring that the focused areas are correlated. The proposed method is evaluated subjectively and objectively using public data sets. The experimental results show that the proposed method produces effective fusion and

收稿日期: 2021-09-13; 修回日期: 2021-10-15; 录用日期: 2021-10-27

基金项目: 国家重点研发计划(2016YFB0502600)、国家自然科学基金(41601443)、湖北省教育厅科学技术研究项目(B2021351)

通信作者: \*Chuanwenhu@qq.com

fully mines the focused association information and retains sufficient image detail.

**Key words** image processing; multi-focus image; image fusion; densely connected convolutional neural networks; pyramid pooling; cooperative detection

## 1 引言

在光学成像中,由于镜头景深问题导致图像中只有局部区域可以聚焦,难以获得覆盖全场景的清晰图像。而多聚焦图像融合技术通过提取多幅局部对焦图像的互补信息,将其融合成一幅全聚焦清晰图像,从而使图像质量增强,便于视觉理解、提升图像信息利用率。目前,多聚焦图像融合广泛应用于医学显微成像、机器视觉测量、机器识别、军事安防等领域<sup>[1]</sup>。

一般来说,多聚焦图像融合方法可分为3种:基于变换域的融合方法、基于空间域的融合方法以及基于深度学习的融合方法。其中,基于多尺度变换(MST)域的图像融合方法,包括基于拉普拉斯的算法、基于小波变换的算法、基于非子采样轮廓变换(NSCT)的算法等<sup>[2]</sup>。其融合过程主要包括3个步骤:1)根据图像的多尺度特征,将源图像分解为高频成分和低频成分;2)选择不同的融合规则得到高融合和低融合映射;3)通过逆MST得到最终的融合映射。然而基于MST的方法在变换融合过程中会出现空间不一致性问题,易导致不同程度的失真。基于空间域的图像融合方法,主要通过线性组合进行图像融合,通常可分为3类:基于像素的、基于块的和基于目标区域的方法<sup>[3-4]</sup>。然而,利用像素的梯度相关信息或图像块进行图像融合,易在融合结果中引入伪影块,导致效果不佳。典型的空域融合方法,包括基于引导滤波图像融合(GFF)<sup>[5]</sup>和图像抠图(IFM)<sup>[6]</sup>的融合方法等,虽然在图像的特征提取和细节表达方面较好,但是难以人工设定理想的融合规则。近年来出现了基于深度学习的多聚焦图像融合方法,可充分发挥其学习能力强、泛化能力强、可移植性好的优点。如Liu等<sup>[7]</sup>基于卷积神经网络(CNN)的方法,梅礼晔等<sup>[8]</sup>基于空间金字塔池化的方法,这些方法通过图像块来进行融合,导致运算较为复杂,且图像边缘会存在块效应。此外,Guo等<sup>[9]</sup>提出了基于全卷积神经网络的方法,虽然较好地解决了图像分块问题,但由于未充分考虑上下文信息之间的关联,使得适合全局特征的图像块被忽视省略。李恒等<sup>[10]</sup>提出了一种全卷积神经网络的聚焦图像融合算法,但其主要是对灰度图像进行融合,没有考虑到彩色图像信息,限制了多聚焦图像融合能力。

针对传统方法不能充分挖掘图像聚焦关联信息导致融合细节失真的问题,本文将多聚焦图像融合当作上下文关联约束的二分类分割问题,即区分聚焦与非聚焦区域,提出了一种基于深度密集卷积神经网络协同检测的多聚焦图像融合方法。该方法利用深度密集

卷积神经<sup>[11]</sup>特征复用、低级特征与高级特征相结合等特点,来加强多聚焦图像特征表达能力,更好地挖掘图像语义信息。利用金字塔池化网络来提取多尺度信息,将不同聚焦区域的上下文信息聚合起来,使其散聚焦的区分能力得到优化,可以较好地检测聚焦边界。进而,在区分非聚焦与聚焦区域的处理过程中引入卷积条件随机场(ConvCRFs)<sup>[12]</sup>,可对网络概率预测图的准确度进行优化,进一步增强融合效果。

此外,由于真实场景中的镜头模糊情况通常与高斯模糊不同,近似呈平均模糊。本文利用点扩展函数(圆盘disk模糊)来模拟真实多聚焦情况<sup>[6]</sup>,可以更加符合镜头模糊情况,提升数据集模拟真实散聚焦情况,从而获得更好的训练模型。在训练中,本文将多聚焦图像A和多聚焦图像B合并为6通道送入网络进行训练,保证了训练时聚焦图像的相关性,即左聚焦图像与右聚焦图像相结合,能更好地提升训练效果。最后,对本文方法与7种主流融合方法进行实验对比,从主观视觉效果评价和客观对比评价两方面验证了本文方法的高效性与优越性。

## 2 深度密集卷积神经网络协同聚焦检测的图像融合方法

考虑到两张经过配准的多聚焦源图像 $O_A$ 与 $O_B$ 存在关联信息,因此,本文将其集成为一张多通道图像,在此基础上进行基于密集卷积神经网络的协同检测操作。此后,利用金字塔池化进行多尺度特征信息提取,利用二值交叉熵作为损失函数,以获得粗略的二分类概率决策图。利用卷积条件随机场实现概率决策图精化,最后根据融合计算规则进行融合以获得最终的多聚焦图像融合结果。本文融合方法的总体框架如图1所示。

假设经过概率决策图精化后的二值图像矩阵为 $W_A$ ,其另一半二值图为 $W_B$ ,即 $W_B = 1 - W_A$ 。源图像为 $I_A$ 和 $I_B$ ,因此最终融合图像 $F$ 的计算规则为

$$F = W_A \cdot I_A + I_B \cdot (1 - W_A) \quad (1)$$

### 2.1 密集网络协同聚焦检测

密集网络协同聚焦检测的核心是深度密集卷积神经网络<sup>[11]</sup>,其基于卷积神经网络,并在理论上有了进一步的探索,引入了深度密集的概念。如图2所示,为一个深度密集模块结构图,其最为主要的构成是:密集模块(dense block)和过渡层(transition layer)。其中密集块彼此相互连接,既能减轻梯度消失又能够使网络特征信息的利用率得到较好提升。如图2所示,该模块有着4个密集块。一般来说深度学习网络中,网络的

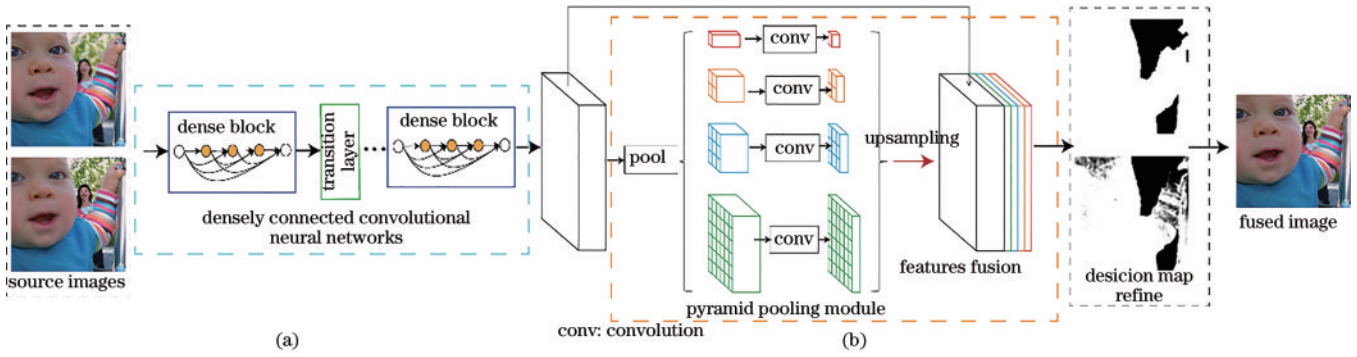


图 1 提出的融合方法总体框架图。(a)深度密集卷积神经网络协同检测;(b)金字塔池化网络多尺度信息提取

Fig. 1 Overall framework of proposed fusion method. (a) Collaborative detection via densely connected convolutional neural networks; (b) multi-scale information extraction for pyramid pooling network

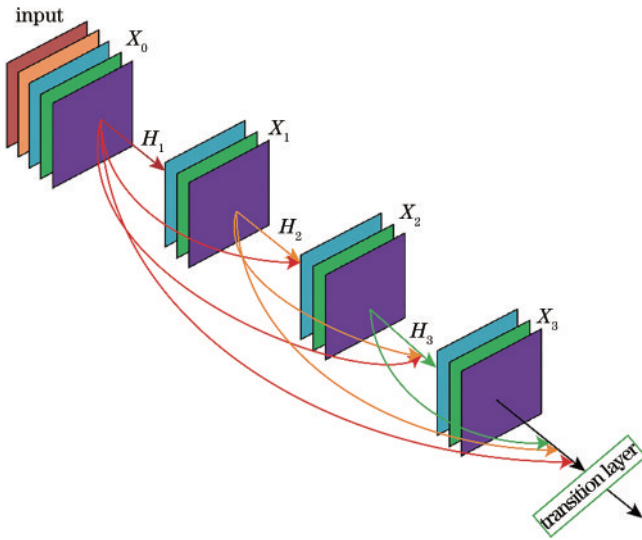


图 2 深度密集模块结构图

Fig. 2 Structure of deep dense block

深度越深,其梯度消失的情况就越发严重,因而引入密集块来进行缓解。具体来说,每个 dense block 包含  $L$  层卷积模块组成,  $X_0$  为输入层,其输入图像经过一个  $3 \times 3$  卷积层和批归一化层(BN)后输出一个特征图,  $X_l$  为第  $l$  层的输出 ( $l = 0, 1, 2, \dots, L - 1$ ), 每一层通过均非线性变换  $H_l(\cdot)$  激活,  $H_l(\cdot)$  由卷积、修正性线性单元 ReLU 激活函数、dropout 层组成。密集卷积神经网络则采用了串联的方式,  $X_l$  层输入等于前面所有层特征融合, 即,  $X_l = H_l([X_0, X_1, \dots, X_{l-1}])$ , 其中  $[X_0, X_1, \dots, X_{l-1}]$  表示将 0 到  $l-1$  层的输出特征图在通道维度上进行相加, 融合了所有较早层特征, 具有特征重用的特点, 使网络参数更加高效, 强化了特征复用能力。为了压缩网络参数, 密集卷积神经网络在两个 dense block 之间使用了过渡层, 过渡层由一个  $1 \times 1$  的卷积层、BN 层和 dropout 层组成,  $1 \times 1$  的卷积层起到降维的作用。总的来说, 在残差网络的基础上, 深度密集卷积神经网络的结构更加复杂, 且在加入了密集连接之后使各层都有着先前的所有层特征, 可有效实现特征复用, 使得不同层次的多聚焦图像特征信息的传

输得到了优化和利用。

## 2.2 金字塔池化多尺度信息提取

考虑到多聚焦图像最难检测点在聚焦区域和非聚焦区域, 而高层卷积神经网络中未充分获得重要的全局先验知识, 且高层特征包含了更多的语义和较少的位置信息, 为进一步减少不同子区域间上下文信息的丢失, 如图 1 所示, 本文引入金字塔池化模型<sup>[13]</sup>。其为一种有效的全局上下文先验模型, 包含不同尺度、不同子区域间的信息, 可有效提升网络利用全局上下文信息的能力, 使网络充分挖掘到聚焦与非聚焦区域的边界信息, 嵌入困难融合场景的上下文特征, 提升融合效果。具体来说金字塔池化模型主要池化层和上采样层构成, 在经过深度密集卷积模块后, 对其输出特征图进行自适应平均池化操作, 即使用一个平均池化层, 将特征图按照指定尺寸进行池化操作, 在本文中生成了  $1 \times 1, 2 \times 2, 3 \times 3, 6 \times 6$  不同尺度的特征图, 以此构成了多尺度的特征图。同时为了保证全局特征的权重, 如果金字塔共有  $N$  个级别, 则在每个级别后使用  $1 \times 1$  的卷积将级别通道降为原本的  $1/N$ 。再通过双线性插值得到未池化前的大小, 最终将所有层级特征图在通道维度上进行合并, 以得到全局信息的特征图。在合并多尺度特征图后, 对合并后特征图进行  $3 \times 3$  卷积操作, 随后使用了连续 3 个上采样层将特征图插值到源图像尺寸。

## 2.3 卷积条件随机场概率决策图精化

虽然卷积神经网络在源图像全局上下文信息提取上具有较好效果, 然而概率图中仍存在被误分类的像素, 如图 3 所示。因而, 为获得更准确更优秀的分割能力, 利用 ConvCRFs<sup>[12]</sup> 对概率决策图进行优化。ConvCRFs 是在全连接条件随机场(FullCRFs)<sup>[14]</sup> 基础上进行优化。输入图像  $\mathbf{O}$  可以经过  $\arg \max P(\mathbf{K}|\mathbf{O})$  来进行求解, 其具体解析式为

$$P(\mathbf{K} = \hat{k}|\hat{\mathbf{O}} = \mathbf{O}) = \frac{1}{Z(\mathbf{O})} \exp(-E(\hat{k}|\mathbf{O})), \quad (2)$$

式中,  $\mathbf{K} = \{K_1, \dots, K_n\}$  表示随机场。





图 3 多尺度金字塔池化协同检测的概率决策图。(a)源图像;(b)粗分割;(c)精细分割

Fig. 3 Probabilistic decision diagram for multi-scale pyramid pooling collaborative detection. (a) Source image; (b) rough segmentation; (c) refined segmentation

能量函数  $E(\hat{k}|\mathbf{O})$  表达式为

$$E(\hat{k}|\mathbf{O}) = \sum_{i \leq N} \psi_{\mu}(\hat{k}_i|\mathbf{O}) + \sum_{i \neq j \leq N} \psi_{\rho}(\hat{k}_i, \hat{k}_j|\mathbf{O}) \quad (3)$$

在 FullCRFs 的框架中添加条件独立性因素,该操作使得在图形处理器(GPU)上能够利用 ConvCRFs 的卷积运算进行相应的有效推测,可高效结合 CNN 的特征提取能力以及随机字段的建模能力,使 ConvCRFs 可以有效地对信息进行传递。

为实现对目标区域的精确显著性检测,通过 ConvCRFs 来整合概率图的全局、局部以及边界信息,可有效获得精化概率决策图。在优化过程中,以卷积运算为基础对概率图的多个特征信息进行计算,通过随机条件场(CRFs)把这些特征融合到精化图中,最终得到经过 ConvCRFs 优化后的精确图。

## 2.4 网络训练

### 2.4.1 数据集制作

本文所使用的数据集为 VOC2012 数据集,总共分为 4 个大类:车辆、住户、动物、人,该数据集分类为 20 个类别(加上背景共有 21 类),总计包含 17125 张图像。为了模拟多聚集图像,本文采用 Guo 等<sup>[9]</sup>的图像生成方法,利用点扩展函数(圆盘 disk 模糊)来模拟真

实多聚焦情况,可以更符合镜头模糊情况,提升数据集模拟真实散聚焦情况,从而获得更好的训练模型。总计通过 5 步获得合成后的多焦点图像,分别为高斯模糊、图像转化、图片反转、逐像素相乘,以及逐像素相加。

### 2.4.2 训练参数

本文实验是基于深度学习框架 Pytorch,在 NVIDIA 1080Ti 下进行训练。本文将多聚焦图像 A 和多聚焦图像 B 合并为 6 通道送入网络进行训练,训练阶段图像大小为 256 pixel × 256 pixel,采用 Adam 作为梯度优化器,其学习速率为 0.001,正则化项为 0.9,损失函数采用二值交叉熵。在实验中每次送入网络 1 张图像,总共训练次数为 50,当训练损失连续 10 次不下降时,保存模型。在测试阶段,输入图像为图像原尺寸。

## 3 实验结果及分析

### 3.1 对比方法

为证明本文所提融合方法的优越性与有效性,选择 Lytro 多聚焦的彩色图像数据集<sup>[15]</sup>进行实验。选择了 7 种主流的图像融合方法作为对比方法,分别

为:非下采样轮廓波变换(NSCT)融合方法<sup>[3]</sup>、基于导引滤波(GFF)的融合方法<sup>[5]</sup>、基于图像抠图的多聚焦图像(IFM)融合方法<sup>[6]</sup>、基于双边过滤器(CBF)<sup>[16]</sup>的融合方法、基于离散余弦谐波小波变换(DCHWT)的融合方法<sup>[17]</sup>、基于 CNN 的融合方法<sup>[18]</sup>、基于金字塔池化网络(PSPF)的融合方法<sup>[19]</sup>。

### 3.2 定性分析

为了展示本文所提算法的优越性,本文展示了代表性的两例图像融合结果,并与不同方法进行了对比,如图 4 所示。图 4A 为不同融合方法对 Lytro-3 的视觉融合结果,图 4B 展示了每幅融合图像减去源图像 A 得到的伪彩色差异图像。首先从融合结果中可以看出在大多数对比方法中“男孩的耳朵边缘”变得较为模糊,但本文提出方法所求出的融合结果图的边缘结构相对

清晰,可以证明提出方法具有更好的边缘信息提取能力。同样地,从差异图像可以看出,提出方法的图像融合结果在细节位置较好,尤其从图中标注的矩形框中也可以看出,相比于其他的 CNN、PSPF、IFM 方法,提出方法在边缘结构上更为清晰,保留了更多细节,也验证了图 4A 中融合结果视觉表现。这也说明提出方法通过深度卷积神经网络协同检测,可以更好地融合图像特征,学习到图像更多的语义信息,实现深度信息挖掘,从而可以较好地进行聚焦区域检测,以获得高质量的融合图像。

为了进一步验证提出方法的效果,图 5 给出了 Lytro-17 的融合结果,以及融合图像与源图像 A 的伪彩色差异图。从中容易看出,在不同图中,其留下的聚焦区域的痕迹越少,则说明聚焦部分的图像信息被提

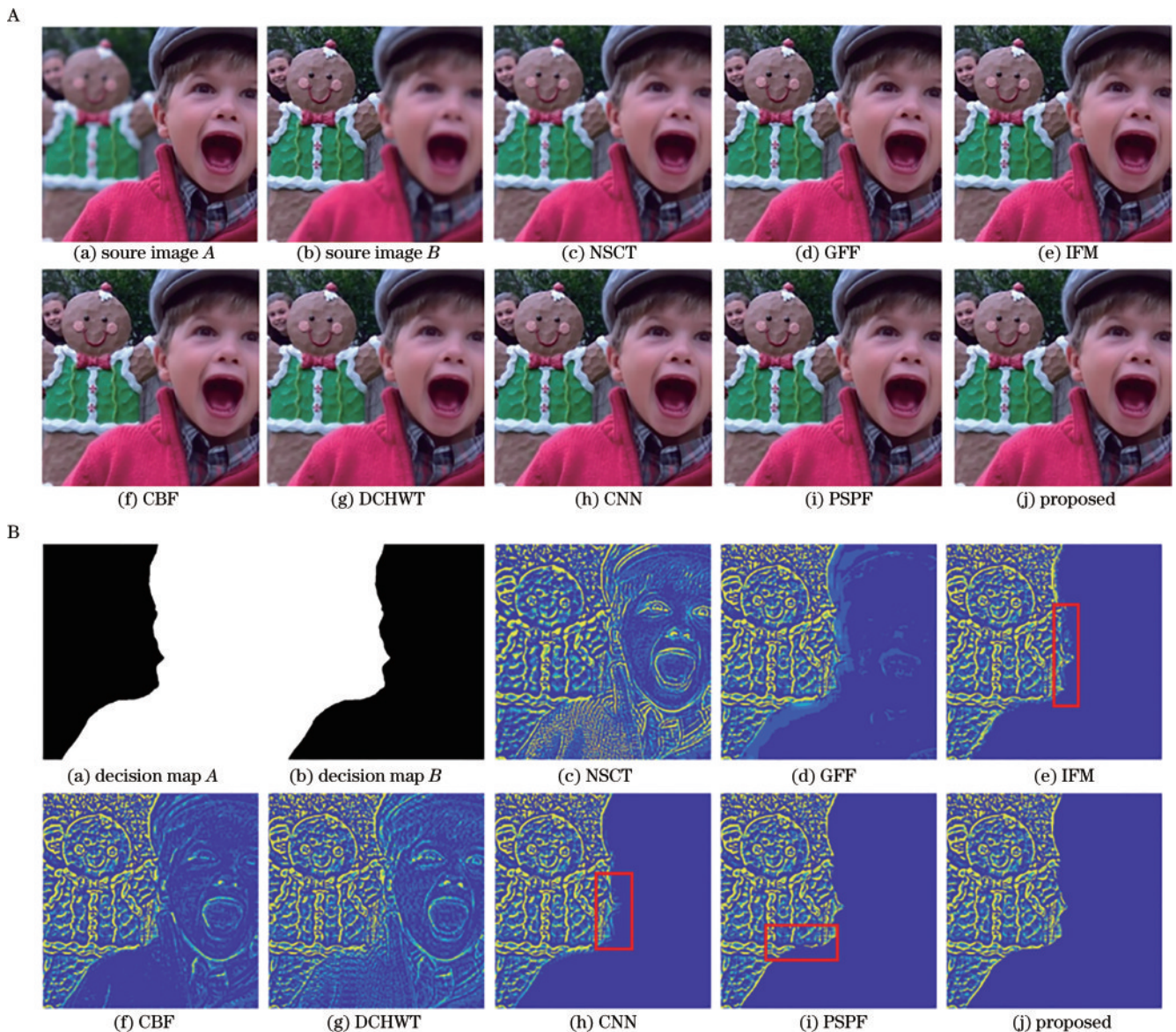


图 4 Lytro-3 图像融合结果。A 为不同方法的融合结果对比, B 为每幅融合图像减去源图像 A 得到的伪彩色差异图像  
Fig. 4 Lytro-3 image fusion results. A is corresponding fusion results obtained by various methods, and B is difference images in pseudo color map form obtained by subtracting source image A from each fused image



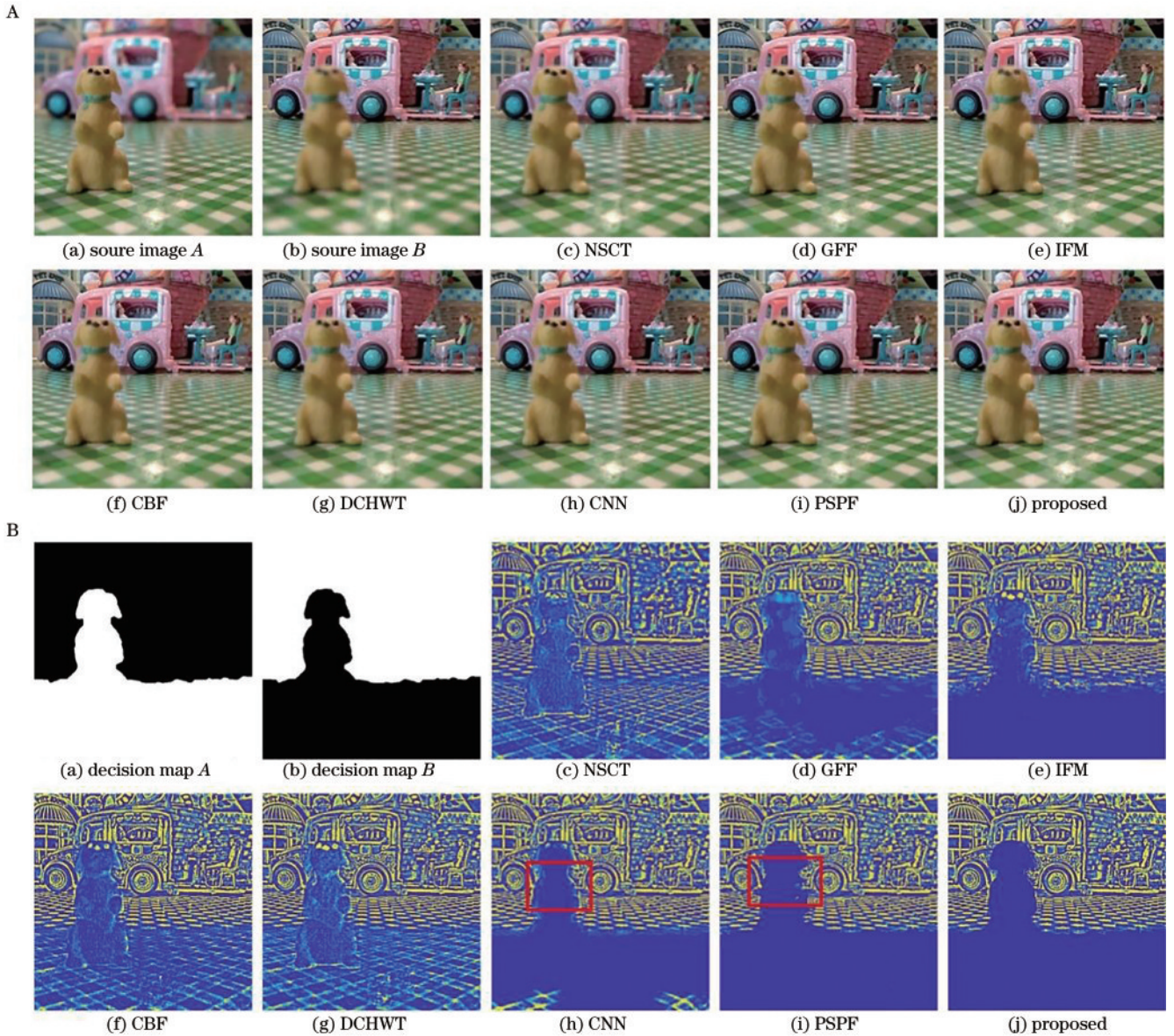


图 5 Lytro-17 图像融合结果。A 为不同方法的融合结果对比, B 为每幅融合图像减去源图像 A 得到的伪彩色差异图像  
Fig. 5 Lytro-17 image fusion results. A is corresponding fusion results obtained by various methods, and B is difference images in pseudo color map form obtained by subtracting source image A from each fused image

取到融合图中更多,也就意味着融合性能更好。从图中可以看出,提出算法在图像下方信息较少,很明显优于其他方法,而 CNN 以及 PSPF 方法都在边界区域有所不足,说明提出的融合方法边缘区域较好,充分利用图像上下文信息,较好地检测了聚焦边界,在主观视觉效果评价上较对比方法而言有着较为良好的融合效果。

剩下 18 张图像的最终融合结果,如图 6 所示,第 1 列到第 4 列、第 5 列到第 8 列分别展示了 18 幅多聚焦图像融合结果。无论是融合决策图,还是最终融合图像,都可以看出本文提出算法具有较好的融合效果。

### 3.3 定量比较

为了客观证明本文方法的有效性,采用 4 种主流的多聚焦图像评价指标,作为量化指标,分别为互信息

( $Q_{MI}$ )、非线性相关信息熵( $Q_{NCIE}$ )、边缘保持度( $Q_{AB/F}$ )、视觉保真度( $Q_{VIF}$ )<sup>[17]</sup>,并与主流的 7 种方法进行了对比。各评价指标具体描述如下。

#### 1) 互信息( $Q_{MI}$ )

互信息能够计算融合图像与源图像之间的相似程度,其值的大小能够直观地反映出融合图像中所包含源图像特征信息量的多少。其计算公式为

$$Q_{MI} = 2 \left[ \frac{MI(A, F)}{H(A) + H(F)} + \frac{MI(B, F)}{H(B) + H(F)} \right], \quad (4)$$

$$MI(A, F) = H(A) + H(F) - H(A, F), \quad (5)$$

$$MI(B, F) = H(B) + H(F) - H(B, F), \quad (6)$$

式中: $MI(A, F)$ 与是源图像 A 与融合图像 F 之间的互信息; $MI(B, F)$ 是源图像 B 与融合图像 F 之间的互信息; $H(A, F)$ 、 $H(B, F)$ 分别是源图像 A、源图像 B





图 6 提出方法更多结果展示。第 1 列到第 4 列, 第 5 列到第 8 列分别展示了 18 幅多聚焦图像融合结果  
 Fig. 6 More results of proposed method. The first to fourth and fifth to eighth columns show fusion results of 18 pairs of multi-focus images, respectively

与融合图像  $F$  之间的联合信息熵;  $H(A)$ 、 $H(B)$  和  $H(F)$  分别为源图像  $A$ 、源图像  $B$ 、与融合图像  $F$  的边际熵。 $Q_{MI}$  越大, 表示融合图像包含源图像中的信息量越多。

2) 非线性相关信息熵 ( $Q_{NCIE}$ )

在多变量情况下, 每两个变量之间的一般关系可

以根据归一化互相关(NCC)的定义获得, 因此, 相关  $K$  个变量的非线性相关矩阵可以定义为

$$R = \{NCC_{ij} \mid 1 \leq i \leq K, 1 \leq j \leq K\}, \quad (7)$$

式中,  $NCC_{ij}$  表示第  $i$  个和第  $j$  个变量的非线性相关系数。在此基础上非线性联合熵  $H_R$  被定义为

$$H_R = - \sum_{i=1}^K \frac{\lambda_i^R}{K} \log_b \frac{\lambda_i^R}{K}, \quad (8)$$

式中,  $\lambda_i^R, i = 1, \dots, K$  是非线性相关矩阵的特征值。根据矩阵特征值理论, 可以推导出  $0 \leq \lambda_i^R \leq K, i = 1, \dots, K$  和  $\sum_{i=1}^K \lambda_i^R = K$ 。非线性相关信息熵  $Q_{\text{NCIE}}$  作为有关变量的非线性相关度量, 被定义为

$$Q_{\text{NCIE}} = 1 - HR = 1 + \sum_{i=1}^K \frac{\lambda_i^R}{K} \log_b \frac{\lambda_i^R}{K}, \quad (9)$$

具体来说,  $Q_{\text{NCIE}}$  越大, 融合效果越好。

### 3) 边缘保持度 ( $Q_{\text{AB/F}}$ )

边缘保持度用于描述源图像边缘在融合结果中的保留情况, 如下式所示:

$$Q_{\text{AB/F}} = \frac{\sum_{i=1}^M \sum_{j=1}^N [Q^{\text{AF}}(i, j) \times \omega_A(i, j) + Q^{\text{BF}}(i, j) \times \omega_B(i, j)]}{\sum_{i=1}^M \sum_{j=1}^N [\omega_A(i, j) + \omega_B(i, j)]}, \quad (10)$$

$$Q^{\text{AF}}(i, j) = Q_g^{\text{AF}}(i, j) Q_o^{\text{AF}}(i, j), \quad (11)$$

$$Q^{\text{BF}}(i, j) = Q_g^{\text{BF}}(i, j) Q_o^{\text{BF}}(i, j), \quad (12)$$

式中:  $Q_g^{\text{AF}}(i, j)$  和  $Q_o^{\text{AF}}(i, j)$  表示像素点  $(i, j)$  处的边缘强度和方向保留值;  $\omega_A$  表示权重系数。  $Q_{\text{AB/F}}$  值越大, 说明源图像转移到融合图像中的边缘信息量越多。

### 4) 视觉保真度 ( $Q_{\text{VIF}}$ )

视觉信息保真度是一个基于自然场景统计和人类视觉系统提取的图像信息概念的图像融合质量评估指数。  $Q_{\text{VIF}}$  指数定义为

$$Q_{\text{VIF}} = \frac{\sum_{j \in \text{subbands}} I(\bar{C}^{N,j}; \bar{F}^{N,j} | S^{N,j} = s^{N,j})}{\sum_{j \in \text{subbands}} I(\bar{C}^{N,j}; \bar{E}^{N,j} | S^{N,j} = s^{N,j})}, \quad (13)$$

式中:  $\sum_{j \in \text{subbands}} I(\bar{C}^{N,j}; \bar{F}^{N,j} | S^{N,j} = s^{N,j})$  代表从融合图像中

提取的信息量;  $\sum_{j \in \text{subbands}} I(\bar{C}^{N,j}; \bar{E}^{N,j} | S^{N,j} = s^{N,j})$  表示从源

图像中提取的信息量。  $Q_{\text{VIF}}$  越大, 表示融合质量越高。

为了进一步验证提出算法的可行性和融合效果, 如表 1 所示, 展示了 5 幅图像融合结果的客观评价对比情况。其中, 加粗数值表示最优结果, 斜体数值表示次优结果。

从表 1 可以看出, 提出方法在 4 种客观评价指标中均取得了较好结果。从互信息、非线性相关信息熵、边缘保持度、视觉信息保真度等指标的得分来看, 提出方法大部分都达到最优, 说明采用所提的密集卷积神经

表 1 融合结果客观评价对比

Table 1 Comparison of objective evaluation of fusion results

Image	Metrics	Fusion method							
		NSCT	GFF	IFM	CBF	DCHWT	CNN	PSPF	Proposed
Lytro-1	$Q_{\text{MI}}$	2.8165	3.6591	3.7414	3.3332	2.9543	3.7935	3.9901	<b>4.0177</b>
	$Q_{\text{NCIE}}$	0.8193	0.8301	0.8313	0.8255	0.8208	0.8318	0.8349	<b>0.8354</b>
	$Q_{\text{AB/F}}$	0.6004	0.7531	0.7405	0.7418	0.6945	0.7539	0.7560	<b>0.7575</b>
	$Q_{\text{VIF}}$	0.5432	0.7461	0.7037	0.6863	0.6476	<b>0.7476</b>	0.7452	0.7455
Lytro-3	$Q_{\text{MI}}$	2.8810	3.8132	4.0430	3.4832	3.1244	4.0501	4.1158	<b>4.1351</b>
	$Q_{\text{NCIE}}$	0.8203	0.8320	0.8355	0.8273	0.8229	0.8356	0.8368	<b>0.8370</b>
	$Q_{\text{AB/F}}$	0.5497	0.7437	0.7432	0.7311	0.6964	<b>0.7466</b>	0.7415	0.7455
	$Q_{\text{VIF}}$	0.4804	0.7546	0.7518	0.6936	0.6484	<b>0.7593</b>	0.7512	0.7571
Lytro-9	$Q_{\text{MI}}$	2.5304	3.8522	4.0807	3.6374	2.9118	4.0713	4.1162	<b>4.1549</b>
	$Q_{\text{NCIE}}$	0.8160	0.8319	0.8356	0.8288	0.8199	0.8355	0.8363	<b>0.8367</b>
	$Q_{\text{AB/F}}$	0.4625	0.8004	0.7992	0.7919	0.7448	0.8015	0.8010	<b>0.8017</b>
	$Q_{\text{VIF}}$	0.4193	0.7760	0.7703	0.7337	0.6546	0.7766	<b>0.7793</b>	0.7753
Lytro-17	$Q_{\text{MI}}$	2.4034	3.4225	3.6484	3.0606	2.7818	3.7062	3.8529	<b>3.8604</b>
	$Q_{\text{NCIE}}$	0.8143	0.8275	0.8301	0.8229	0.8191	0.8313	0.8328	<b>0.8337</b>
	$Q_{\text{AB/F}}$	0.5400	0.7660	0.7547	0.7529	0.7297	<b>0.7695</b>	0.7677	0.7693
	$Q_{\text{VIF}}$	0.4701	0.7123	0.6971	0.6496	0.6148	0.7227	0.7250	<b>0.7270</b>
Lytro-18	$Q_{\text{MI}}$	4.0753	4.8021	4.8594	4.5228	4.2013	4.9014	4.8296	<b>4.9635</b>
	$Q_{\text{NCIE}}$	0.8403	0.8525	0.8537	0.8476	0.8422	0.8545	0.8531	<b>0.8557</b>
	$Q_{\text{AB/F}}$	0.6282	0.7397	0.7352	0.7285	0.6877	0.7414	0.7332	<b>0.7425</b>
	$Q_{\text{VIF}}$	0.6226	0.8049	0.7916	0.7441	0.7232	<b>0.8086</b>	0.7891	0.7982



网络协同检测方法可以很好地挖掘多聚焦图像深度特征,从而进行聚焦估计,而采用的空间金字塔池化模块更好地约束了上下文信息,从而使其图像边缘聚焦检测较好,因此,该算法的整体融合质量较高,能更多地保留源图像聚焦信息。

为了进一步地综合验证本文提出算法的综合图像

效果,对整个测试集进行量化评估,表 2 为 20 幅图像客观评价的平均结果,从中可以看出,提出算法在 3 个指标上均获得了最优效果。在视觉信息保真度评价指标上,也获得了仅次于 CNN 方法的次优结果。这也表明,在综合情况下,提出方法与其他方法相比有着更好的融合效果。

表 2 融合结果平均客观评价对比

Table 2 Comparison of average objective evaluation of fusion results

Metrics	Fusion method							
	NSCT	GFF	IFM	CBF	DCHWT	CNN	PSPF	Proposed
$Q_{MI}$	3.1473	4.1211	4.2879	3.8211	3.3649	4.3210	4.4357	<b>4.4656</b>
$Q_{NCIE}$	0.8250	0.8395	0.8425	0.8344	0.8278	0.8430	0.8449	<b>0.8459</b>
$Q_{AB/F}$	0.5799	0.7601	0.7534	0.7528	0.7124	0.7618	0.7590	<b>0.7669</b>
$Q_{VIF}$	0.5132	0.7430	0.7233	0.6870	0.6465	<b>0.7465</b>	0.7406	0.7435

## 4 结 论

提出了一种基于深度密集卷积神经网络协同检测的多聚焦图像融合方法。该融合方法利用密集卷积神经网络中的特征复用的优势,通过对多聚焦的源图像进行集成,以此来使其达成协同聚焦的特征检测,然后利用金字塔池化网络来进行多尺度的全局信息提取,将不同聚焦区域的上下文信息聚合起来,使其散聚焦的区分能力得到优化,并获得粗略的融合概率决策图。接着使用 ConvCRFs 进行进一步的优化,由粗略融合概率决策图得到精确概率决策图,最后产生细节保持良好的融合图像。实验结果表明,所提方法在视觉上获得了较好结果,在 4 个定量指标上,也获得了最好结果,充分证明该方法的有效性,可有效应用于自动化成像视觉任务中。

## 参 考 文 献

- [1] 杨艳春, 李娇, 党建武, 等. 基于引导滤波与改进 PCNN 的多聚焦图像融合算法[J]. 光学学报, 2018, 38(5): 0510001.  
Yang Y C, Li J, Dang J W, et al. Multi-focus image fusion based on guided filtering and improved PCNN[J]. Acta Optica Sinica, 2018, 38(5): 0510001.
- [2] Zhang Q, Guo B L. Multifocus image fusion using the nonsubsampling contourlet transform[J]. Signal Processing, 2009, 89(7): 1334-1346.
- [3] 唐超影, 浦世亮, 叶鹏钊, 等. 基于卷积神经网络的低照度可见光与近红外图像融合[J]. 光学学报, 2020, 40(16): 1610001.  
Tang C Y, Pu S L, Ye P Z, et al. Fusion of low-illumination visible and near-infrared images based on convolutional neural networks[J]. Acta Optica Sinica, 2020, 40(16): 1610001.
- [4] 吴一全, 王志来. 基于目标提取与引导滤波增强的红外与可见光图像融合[J]. 光学学报, 2017, 37(8): 0810001.  
Wu Y Q, Wang Z L. Infrared and visible image fusion based on target extraction and guided filtering enhancement[J]. Acta Optica Sinica, 2017, 37(8): 0810001.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 梅礼晔, 郭晓鹏, 张俊华, 等. 基于空间金字塔池化的深度卷积神经网络多聚焦图像融合[J]. 云南大学学报(自然科学版), 2019, 41(1): 18-27.  
Mei L Y, Guo X P, Zhang J H, et al. Spatial pyramid pooling in deep convolutional networks for multi-focus image fusion[J]. Journal of Yunnan University (Natural Sciences Edition), 2019, 41(1): 18-27.
- [9] Guo X P, Nie R C, Cao J D, et al. Fully convolutional network-based multifocus image fusion[J]. Neural Computation, 2018, 30(7): 1775-1800.
- [10] 李恒, 张黎明, 蒋美容, 等. 基于监督学习的全卷积神经网络多聚焦图像融合算法[J]. 激光与光电子学进展, 2020, 57(8): 081015.  
Li H, Zhang L M, Jiang M R, et al. Multi-focus image fusion algorithm based on supervised learning for fully convolutional neural networks[J]. Laser & Optoelectronics Progress, 2020, 57(8): 081015.
- [11] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2261-2269.
- [12] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [13] Veksler O. Efficient graph cut optimization for full CRFs

- with quantized edges[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(4): 1005-1012.
- [14] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6230-6239.
- [15] Nejati M, Samavi S, Shirani S. Multi-focus image fusion using dictionary-based sparse representation[J]. Information Fusion, 2015, 25: 72-84.
- [16] Kumar B K S. Multifocus and multispectral image fusion based on pixel significance using discrete cosine harmonic wavelet transform[J]. Signal, Image and Video Processing, 2013, 7(6): 1125-1143.
- [17] Kumar B K S. Image fusion based on pixel significance using cross bilateral filter[J]. Signal, Image and Video Processing, 2015, 9(5): 1193-1204.
- [18] 郝一帆, 蹇毅. 基于多尺度局域加权方差的多聚焦融合方法[J]. 光子学报, 2021, 50(12): 1210001.  
Hao Y F, Jian Y. Multi-focus fusion method based on multi-scale local weighted variance[J]. Acta Photonica Sinica, 2021, 50(12): 1210001.
- [19] Zhou J C, Hao M L, Zhang D H, et al. Fusion PSPnet image segmentation based method for multi-focus image fusion[J]. IEEE Photonics Journal, 2019, 11(6): 1-12.