计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2021,30(11):336−341 [doi: 10.15888/j.cnki.csa.008146]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

多尺度密集网络在红外和可见光图像融合应用①

盖 赟[1], 荆国栋[2]

[1](中国社会科学院大学 计算机教研部, 北京 102488)
[2](中国气象局气象干部培训学院, 北京 100081)
通讯作者: 荆国栋, E-mail: gyunsus@163.com

摘 要: 为了进一步提升红外和可见光图像的融合效果, 提出了一种基于多尺度卷积算子和密集连接网络的图像融合模型. 该模型首先使用多尺度卷积算子计算图像的直接多尺度特征, 然后使用密集连接网络计算图像的间接多尺度特征. 为了得到图像像素信息在不同尺度下的融合权重, 通过叠加的方式将各个尺度密集连接网络的输出进行融合, 并使用活动图方法计算两类图像的融合权重, 最后根据权重计算结果得到融合图像, 实验在 THO 数据集和 CMA 数据集获得较好的识别率.

关键词: 图像融合; 密集连接网络; 多尺度卷积; 特征提取; 特征融合

引用格式: 盖赟,荆国栋.多尺度密集网络在红外和可见光图像融合应用.计算机系统应用,2021,30(11):336–341. http://www.c-s-a.org.cn/1003-3254/8146.html

Application of Multi-Scale DenseNet in Image Fusion for Visual Image and Infrared Image

GE Yun[1], JING Guo-Dong[2]

[1](Department of Computer Teaching and Research, University of Chinese Academy of Social Sciences, Beijing 102488, China)
[2](China Meteorological Administration Training Centre, Beijing 100081, China)

Abstract: To further improve the fusion effect of visual and infrared images, this paper proposes an image fusion model based on multi-scale convolution operators and DenseNet. This model first uses multi-scale convolution operators to get the direct multi-scale features of images. Then, the DenseNet is used to calculate the indirect multi-scale features of images. To get the fusion weights of image pixel information on different scales, this paper fuses the DenseNet on different scales in a stacking manner, and the fusion weights of the two kinds of images can be derived by activity graphs. At last, the fused image is derived according to the fusion weights. The experimental results show that the recognition rate is high on the THO and CMA sets.

Key words: image fusion; DenseNet; multi-scale convolution; feature extraction; feature fusion

图像融合是将两幅或多幅图像中的重要信息合并为同一张图像的处理过程, 融合后的图像能够提供更多的场景信息, 这些信息对提高图像识别率和特征提取准确率都具有极大的推动作用. 红外传感器和可见光传感器是两种最常用的传感器: 红外传感器通过检测目标的热辐射信息完成图像成像, 这种图像能在可视条件差的情况下仍能捕获目标的位置、轮廓等信息. 但是红外图像的成像效果较差, 图像中的目标成像效果、且包含的噪声信息较多. 可见光传感器通过收集物体反射的光线来完成图像成像工作, 这类图像能够捕捉目标物体丰富的细节信息. 但是可见光图像在昏暗的条件下图像成像质量会大幅下降, 在光线条件特

① 基金项目: 中国社会科学院大学校级科研项目 (0016); 国家自然科学基金 (61602486)
Foundation item: Scientific Research Fund of University of Chinese Academy of Social Sciences (0016); National Natural Science Foundation of China (61602486)
收稿时间: 2021-01-18; 修改时间: 2021-02-23; 采用时间: 2021-03-03; csa 在线出版时间: 2021-10-22

别差的情况下可能完全无法成像. 如果能将两类图像中的重要信息融合在同一幅图像中, 一定可以提高图像的信息含量, 这对于提高图像的可用性具有极大的推动作用.

在过去的几十年里, 学者们在图像融合问题上做了大量的工作. 现有的融合方法大致可以分为 7 类: 多尺度变换方法、稀疏表示方法、神经网络方法、子空间方法、显著性方法、融合模型法和深度学习方法. 深度学习方法是近年来应用领域最广、效果最好的图像处理方法. 随着研究的深入, 学者提出了 VGG (Very Deep Convolution Network)[1]、AlextNet[2]、R-CNN (Region-CNN)[3] 等高效而准确的经典学习模型. 这些模型不仅可以自动学习图像的特征, 还可以根据学到的特征产生新的图像. 因此使用深度学习方法进行图像融合研究不仅可以提高图像特征提取的准确率, 还可以得到有效的融合规则.

目前基于深度学习的图像融合方法大致可以分为两类. 一类是使用现有的网络结构进行特征提取, 另一类是构建适合目标问题的网路结构. 第一类方法的代表性工作有: Li 等人[4] 通过多尺度分解将 VGG 模型应用在细节层, 他们还使用残差网计算出融合的权重图. 然而 VGG 模型是针对图像分类设计的, 所以直接将经典模型应用在图像融合问题的研究会存在适应性不足的问题. 第二类方法的代表性工作有: Liu 等人[5] 基于卷积神经网络方法构建领域了一个图像融合模型, 该方法使用行为水平图和权重分配来完成图像融合工作. 基于 Liu 的方法, Li 等人[6] 设计了一个包含卷积层、融合层和密集模块的学习模型, 并用在解码层完成了图像融合工作. Jiang 等人[7] 将残差网络和白化操作引入了模型的构建过程, 并基于卷积神经网络和残差网络提出了一个可以充分应用每个层输出的自编码模型. 但是该模型无法区分输入的信息是红外图像还是可见光图像. 除此之外, 这些方法还忽略低层特征对学习结果的影响, 仅仅是根据最后一层的输出做出网络决策. 针对这个问题, An 等人[8] 通过在编码层增加多个密集连接网络模块来提高模型对各层网络输出的感知, 但是数据在各个密集模块之间的流动仍是单向的, 后面的密集模块无法感知到前面密集模块的信息. Mustafa 等人[9] 对于每类图像的低层卷积输出应用了多组尺度不同的密集连接模块, 并通过将所有密集模块的输出融合来在一起完成多尺度特征的捕获. 图像的高层特征是对低层特征的进一步抽象, 所以如果低层特征没有改变, 高层特征使用何种尺度分析都无法让模型真正感知到图像在不同尺度下的特征.

为此本文提出了一个基于多尺度卷积算子, 融合密集连接网络和残差网络的图像融合方法. 该方法首先使用多个尺度的卷积算子分别对输入图像进行多尺度特征提取. 在得到多的尺度特征图之后, 再分别使用多个密集连接网络模块对低层特征进行计算. 每个密集连接模块中的卷积算子都是 的卷积算子. 为了让每个密集模块都能感知到低层卷积网络的输出, 本文使用了全局连接的机制, 即让低层特征图成为每个密集模块的输出. 在特征融合阶段, 所有尺度的特征图被整合在一起, 并形成了融合图像的权重系数, 最终通过加权的形式完成图像融合工作.

1 密集连接网络

随着深度学习网络的层数的增加, 梯度消失或梯度爆炸问题成为阻碍网络深度进一步增加的最重要因素. Gao 等人[10] 在 2017 年提出密集连接网络是近年来解决这一问题最好的一个网络结构. 该网络在短连接思路的影响下, 提出了包含密集连接的网络结构, 即: 将每个中间层神经元需要与之前所有层的输出进行连接, 也就是每一层的输入是由前面所有层的输出组成. 这样每个中间网络不仅可以感知到图像在高层网络的间接特征, 还可以感知到图像在低层网络的直接特征, 这对于各层网络都感知前面网络的输出是非常重要的.

在传统的网络结构中, 每个中间层网络的神经元只与前一层网络的输出相连接. 每一层都是上一层的局部特征再抽象, 高层特征是在低层特征基础之上提取出来的. 所以随着网络层数的增加, 传统网络的高层神经元再也无法感知到目标在低层网络的特征, 而低层网络的特征对于提高网络准确率具有重要的应用意义. 在密集网络的结构中, 每个密集模块的中间层与前驱层都建立了连接, 这样网络就可以综合分析目标在不同级别下的特征来提高模型学习的准确性.

2 模型结构

本节我们对本文提出的多尺度密集连接网络图像融合模型进行介绍. 本文提出的模型通过提取输入图像的多尺度特征, 并在密集网络模块的帮助下实现图像特征的提取和融合图像重建的工作. 图像融合模型

主要包含特征提取、特征融合、图像重建 3 个模块. 如图 1 所示, 首先将一组对齐的红外和可见光图像输入模型进行多尺度特征提取. 然后使用密集连接网络模块对每个尺度下的特征图进行深度特征计算, 并将每个尺度下的密集特征图进行融合得到输入图像的全局特征图. 最后对每个图像提取到的特征图进行融合, 并使用活动图完整图像融合工作.

2.1 特征提取

红外图像和可见光图像的融合实质是将红外图像中的信息根据一定的规则融合可见光图像. 融合时除了考虑亮度信息还需要考虑当前像素所在的连通区域面积, 如果连通区域只包含几个点就说明当前点是红外图像中包含的噪声. 如果能够充分考虑包含当前像素的多尺度区域特征, 就可以让提取关于目标对象更加准确的特征.

卷积神经网络是目前提取图像特征的常用方法, 通常网络的前几层提取的特征是图片的直接特征如: 边缘特征、区域特征等. 网络的高层卷积特征是以底层卷积特征为基础计算出来的. 如果在低层卷积只使用3×3的算子进行特征计算, 高层卷积是无法感知到图像在其他卷积尺度下的特征表现. 为了充分提取输入图像在不同尺度下的特征, 本文首先分别使用多个不同尺度的卷积算子分别对输入图像进行多尺度特征提取, 这些算子的尺度包含3×3、5×5和7×7.

网络中间卷积层提取的特征是图像直接特征的组合特征, 这些组合特征表达的信息可以被看作图像的间接特征. 深层卷积在计算特征时无法直接感知到浅层卷积的特征输出, 综合使用直接特征和间接特征能够进一步提升特征提取的准确性. 密集连接网络将每层卷积的计算结果输出至后续所有的卷积层, 这种方式可以让每个卷积层都能感知到前驱卷积层的计算结果. 为了能够充分利用图像各层卷积的特征输出, 本文在每个尺度的卷积后面都连接了 1 个密集连接网络模块, 该模块是由 4 个密集连接的卷积层组成. 密集连接模块中的卷积算子的尺度都是3×3.

在完成密集连接计算后, 本文将图像在各个尺度下的密集特征进行拼接形成一组多尺度的特征图. 因为在特征计算时所采用的补洞策略都是“SAME”, 所以不同尺度卷积得到的特征图尺度是一样的, 他们是可以直接通过堆叠方式拼接在一起. 为了防止特征图数量过多, 本文在完成特征图的堆叠后使用1×1的卷积操作对特征图数量进行缩减. 这样做一方面可以减少特征的数量, 另一方面也可以提高模型的稳定性.

2.2 特征融合和图像重建

融合模块包含局部特征融合和全局特征融合, 局部特征融合是指将单一图像的多尺度特征进行融合, 全局特征融合是指将多个图像的多尺度特征进行融合. 本文首先在各个图像范围内进行局部特征融合, 然后以此为基础进行全局特征融合. 全局融合模块需要将所有图像的特征图融合在一下, 并以此为基础完成融合图像重建.

在传统的框架中, 特征融合是通过将各个图像的特征图进行线性叠加来完成的. 令$\phi_i^m(i=1,2,\cdots,k;m\in\{1,2,\cdots,M\})$表示网络中训练到的特征图, 其中$m$表示特征图的索引, k代表图像的索引, M代表特征图的数量, f^m代表抽取出来的特征图, 文献 [5] 中给出的融合方法是:

$$f^m(x,y)=\sum_{i=1}^{k}\varphi_i^m(x,y) \tag{1}$$

这种融合策略过于简单, 为了对这些特征图进行有效的融合, 本文通过计算活动图的方式来完成融合工作. 活动图是一种可以将特征转化为融合权重的方法. 令A_i表示特征图, 特征图中对应位置的像素值累计可以用式 (2) 表示:

$$A_i(x,y)=\|\phi_1(x,y),\cdots,\phi_M(x,y)\|_1 \tag{2}$$

为了提高活动图的关于误匹配的鲁棒性, 本文使用块平均方法对初始映射和特征图A_i进行操作.

$$A_i(x,y)=\frac{\sum_{a=-r}^{r}\sum_{b=-r}^{r}A_i(x+a,y+b)}{(2r+1)^2} \tag{3}$$

其中, r表示块的尺寸, r的值越大算子鲁棒性越好. 但是当块尺寸过大时, 细节信息可能会丢失. 所以r的值一般设置为 3.

在获得了最终的活动图A_k^i之后, 权重映射W_k^i可以通过 Softmax 函数得到, W_k^i的计算公式如下所示:

$$W_k^i(x,y)=\frac{A_k^i(x,y)}{\sum_{n=1}^{K}A_n^i(x,y)} \tag{4}$$

其中, k代表图像的索引, $W_k^i(x,y)$的值被限制在 [0, 1] 范围内.

因为我们有很多不同的卷积组, 融合权重可以根

据初始权重在不同的权重图进行优化. 来自大尺度的权重映射可以表示结构图像, 来自小尺度的权重映射可以表示为细节特征. 为了充分使用这些权重, 我们对这些权重继续实施了 Softmax 操作.

$$W_{k,t}^{i}(x,y)=\frac{W_{k,t}^{i}(x,y)}{\sum_{t=1}^{G}W_{k,t}^{i}(x,y)} \tag{5}$$

其中, t表示卷积组的索引, 当最终的权重映射$W_{k,t}^{i}$被计算出来后, 候选融合图像就可以根据式 (6) 得出.

$$F_{d}^{i}(x,y)=\sum_{t=1}^{T}\sum_{k=1}^{K}W_{k,t}^{i}(x,y)\times I_{k}(x,y) \tag{6}$$

最终融合图像可以通过从候选融合图像中选择最大像素值来完成.

$$F_{d}(x,y)=\max\left(F_{d}^{i}(x,y),i\in\{1,\cdots,m\}\right) \tag{7}$$

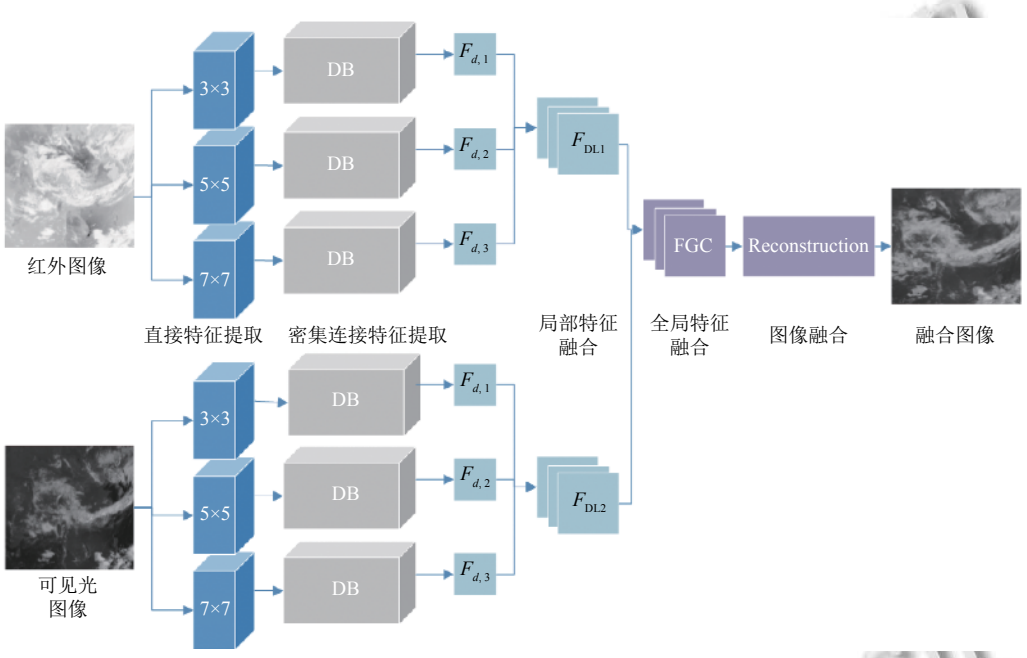


图 1 模型结构

3 实验结果和分析

为了验证本文提出方法的有效性, 我们在 20 组输入图片上进行了融合实验. 这些图片部分来自于 THO 图像数据库, 该数据库包含两类图像: 一类是红外图像和可见光图像. 另一组图像是从中国气象局 CMA 数据库获取的, CMA 数据库中包含的数据均是气象卫星云图, 这些图像资料也是由可见光图像和红外图像组成. CMA 中所有的图像都是由球状图像展开得到的平面图像, 所以图像形态呈扇形. 为了便于操作, 我们从中心区域裁剪下一块大小为544×544子图像作为研究对象. 本文所选用的部分样本图像如图 2 所示, 其中第 1 行是红外图像, 第 2 行是可见光图像.

3.1 损失函数

本文采用结构相似性函数 (Structure SIMilarity, *SSIM*)[11] 作为网络训练的损失函数, *SSIM* 被广泛地用于评估图像融合的质量, 该函数是基于输入图像的亮度、对比度和结构进行计算的, 函数的具体计算形式为:

$$SSIM(x,y)=\frac{\left(2\mu_{x}\mu_{y}+C_{1}\right)\left(2\sigma_{xy}+C_{2}\right)}{\left(\mu_{x}^{2}+\mu_{y}^{2}+C_{1}\right)\left(\sigma_{x}^{2}+\sigma_{y}^{2}+C_{2}\right)} \tag{8}$$

其中, C_1和C_2是 [0, 1] 之间的极小非零值, 用于确保分母部分不为零. 结构相似性指标重点关注图像中的关于物体的结构性信息, 图像融合工作的重点是将红外图像和可见光图像中的重要信息进行融合, 所以 *SSIM* 常用于图像融合的结果评估.

3.2 融合结果

本文选择了 CBF 方法[12] 和 Mustafa 等人提出的 MLDNet 方法[9] 作为比较对象, CBF 方法使用双线性滤波算子对两幅图像的显著性像素进行融合, MLDNet

方法选择多尺度密集连接算子进行特征计算和图像融合. CBF 方法是一种非深度学习方法法, MLDNet 是一种多尺度深度学习方法, 本文选择它们二者进行比较是为了评估两类方法和两种网络结构对融合结果的影响. 实验平台是 Matlab, 实验环境配置是: CPU: 2.6 GHz Intel(R) Core(TM) i7-8850H CPU; 内存: 32 GB RAM. 部分实验结果如图 3 所示.

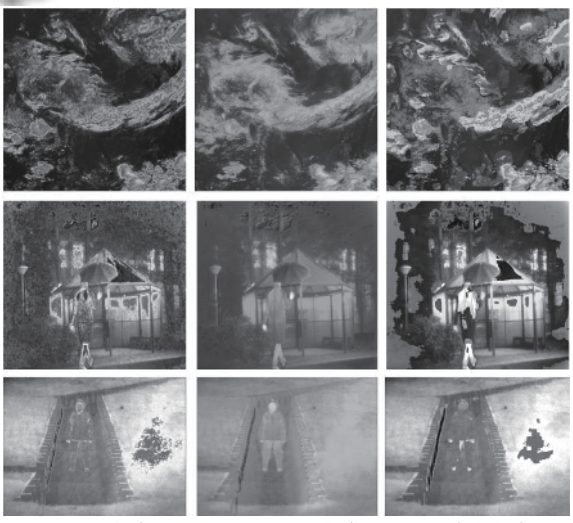
图 3(a) 是使用 CBF 方法得到的融合结果, 图 3(b) 是想使用 MLDNet 方法得到的融合, 图 3(c) 是使用本文方法得到的融合结果. 从图中可以很明显地看到, 图 3(c) 中包含的更丰富的细节信息和更清晰的结构信息. 从云图结果可以看出 CBF 方法只是将亮度信息融合如可见光图像, 而没有考虑云图的结构信息. MLDNet 的方法虽然考虑了结构信息, 但是图像清晰度不高, 很难分辨出那一部分是融合区域. 这是因为 MLDNet 方法的多尺度计算阶段停留在间接特征阶段, 不能真正的感知到图像多尺度直接特征. 要想让模型有效地感知到图像的多尺度特征, 必须在低层卷积层设置多尺度算子. 在云图分析领域, 研究者一方面希望看到清晰的结构信息, 另一方面也希望看到所有的亮度信息. 与前两种方法相比, 使用本文方法得到的图像更清晰, 被融合的红外信息更多. 这是因为我们在将直接特征和后续网络做了密集连接, 使得后续卷积模块都能感知到输入图像的直接信息, 所有得到的融合图像才能在结构和细节方面更加丰富.

(a) 气象云图 (b) 打伞的行人 (c) 道路上的人

图 2 三组输入图像

为了进一步比较本文方法与现有方法的差别, 本文使用 *SSIM* 函数对 3 种方法的融合结果进行了计算和比较, 计算结果如表 1 所示. *SSIM* 的计算结果是 0 和 1 之间的一个小数, 数值越大两个图像的相似性越高, 即图像融合质量越好. 当两幅图像完全一样时, *SSIM* 的值为 1.

表 1 中图 3(a)、图 3(b)、图 3(c) 分别代表气象云图、打伞的行人和道路上的人 3 类图像, 表中的 3 行数据分别代表着 3 种方法对 3 类图像的融合结果. 在从表 1 可以看出, 无论是在 3 个图像的比较看, 还是从平均结果看本文提出方法的评估结果和其他两个方法结果相比均有较好的表现. 值得注意的是本文的结果和 MLDNet 的评估结果相近, 这是因为这两类方法都是基于多尺度思想建立的, 不同之处在于 MLDNet 网络高层采用的是多分辨率密集连接模块, 而本文方法在网络底层采用的是多尺度卷积算子. 所以文本提出的方法能够更加直接地对输入图像进行多尺度特征捕获, 所以得到特征也更为准确.



(a) CBF 方法 (b) MLDNet 方法 (c) 本文方法

图 3 三组重建对比图像

表 1 不同方法在 *SSIM* 的结果

方法	图3(a)	图3(b)	图3(c)	AVG
CBF	0.6109	0.6198	0.6126	0.6144
MLDNet	0.8107	0.8046	0.8077	0.8077
Proposed	0.8182	0.8127	0.8231	0.8180

4 结论与展望

本文提出了一种有效的红外和可见光图像融合方法, 该方法首先使用多尺度卷积算子获得输入图像的多尺度特征, 然后使用密集网络计算图像的间接特征, 最后使用活动图的方法将卷积网络输出的特征图转化为融合权重, 并得到最终的融合结果. 该方法充分发掘了输入图像在多尺度直接特征和间接特征, 使用融合

模型对输入图像的特征表示更为准确. 但是多尺度算子融合的权重和结构相关性不明确, 下一步将重点研究多尺度算子的权重融合计算方法.

参考文献

1 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556v4, 2015.

2 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. International Conference on Neural Information Processing Systems, 2012. 1097–1105.

3 Girshick R, Donahue J, Darrell T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 580–587.

4 Li H, Wu XJ, Durrani TS. Infrared and visible image fusion with ResNet and zero-phase component analysis. Infrared Physics & Technology, 2019, 102: 103039.

5 Liu Y, Chen X, Cheng J, *et al*. Infrared and visible image fusion with convolutional neural networks. International Journal of Wavelets, Multiresolution and Information Processing, 2018, 16(3): 1850018. [doi: 10.1142/S0219691318500182]

6 Li H, Wu XJ. Dense fuse: A fusion approach to infrared and visible images. IEEE Transactions on Image Processing, 2019, 28(5): 2614–2623. [doi: 10.1109/TIP.2018.2887342]

7 Jiang ZT, He YT. Infrared and visible image fusion method based on convolutional auto-encoder and residual block. Acta Optica Sinica, 2019, 39(10): 1015001. [doi: 10.3788/AOS201939.1015001]

8 An WB, Wang HM. Infrared and visible image fusion with supervised convolutional neural network. Optik, 2020, 219: 165120. [doi: 10.1016/j.ijleo.2020.165120]

9 Mustafa HT, Zareapoor M, Yang J. MLDnet: Multi-level dense network for multi-focus image fusion. Signal Processing: Image Communication, 2020, 85: 115864. [doi: 10.1016/j.image.2020.115864]

10 Huang G, Liu Z, Maaten LVD, *et al*. Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE, 2017. 2261–2269.

11 Wang Z, Bovik AC, Sheikh HR, *et al*. Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing, 2004, 13(4): 600–612. [doi: 10.1109/TIP.2003.819861]

12 Kumar BKS. Image fusion based on pixel significance using cross bilateral filter. Signal, Image and Video Processing, 2015, 9(5): 1193–1204. [doi: 10.1007/s11760-013-0556-9]