

基于多尺度空洞卷积网络的多聚焦图像融合算法

尹海涛*, 周伟

南京邮电大学自动化学院、人工智能学院, 江苏 南京 210023

摘要 针对现有基于深度学习的图像融合算法中存在图像特征提取尺度单一、卷积核感受野小、不能有效突出显著特征等问题,提出了一种基于含注意力机制的多尺度空洞卷积网络的多聚焦图像融合算法。首先,构造一种多尺度空洞卷积模块,通过不同的扩张率改变卷积的感受野,从而提取源图像中的多尺度特征。此外,在多尺度空洞卷积模块中引入注意力机制,能自适应地选择显著性特征,进一步提高融合性能。所提融合网络包含特征提取、特征融合和图像重建等3个部分,其中特征提取部分主要由多尺度空洞卷积模块构成。相关实验结果表明,所提算法与现有基于深度学习的算法相比具有一定的竞争力。消融实验也验证了所提多尺度空洞卷积模块能强化网络的特征提取能力,提高图像融合质量。

关键词 图像处理; 多聚焦图像融合; 多尺度; 空洞卷积; 残差学习; 注意力机制

中图分类号 TP391

文献标志码 A

DOI: 10.3788/LOP212488

Multi-Scale Dilated Convolutional Neural Network Based Multi-Focus Image Fusion Algorithm

Yin Haitao*, Zhou Wei

College of Automation and College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing 210023, Jiangsu, China

Abstract According to the issues of single-scale image feature extraction, small receptive field, and cannot highlighting salient features in existing deep learning based image fusion algorithms, this paper proposes a multi-scale dilated convolution network with attention mechanism for multi-focus image fusion. First, a multi-scale dilated convolution block (MDB) is proposed. The MDB with different dilation rates can provide different receptive fields, and consequently it can extract the multi-scale features. Moreover, the attention mechanism is introduced into the MDB, which can adaptively select the salient features and improve the performance further. The proposed fusion network consists of three parts, including feature extraction, feature fusion, and image reconstruction. Specifically, the feature extraction part is composed of several MDBs. The experimental results demonstrate that the proposed method is competitive to some existing deep learning based methods. Some ablation studies also verify that the MDB can enhance the ability of feature extraction and improve the image fusion quality.

Key words image processing; multi-focus image fusion; multi-scale; dilated convolution; residual learning; attention mechanism

1 引言

由于光学镜头在景深上的限制,成像系统很难清晰地显示拍摄场景中所有目标。然而,部分不清晰的区域不利于后续的图像理解与应用,同时还可能造成一定的误差^[1]。多聚焦图像融合技术能有效地解决这个问题,通过集成同一场景的多聚焦图像生成一幅所

有目标都清晰的融合图像。目前,多聚焦图像融合技术已经在机器视觉、遥感监测和军事医学等领域发挥着至关重要的作用^[2-3]。

现有多聚焦图像融合算法大致可分为变换域方法、空间域方法和深度学习方法。变换域方法主要包括对源图像进行变换、对变换系数进行融合、对融合系数进行逆变换生成融合图像等3个步骤。常见变换域

收稿日期: 2021-09-10; 修回日期: 2021-10-25; 录用日期: 2021-11-10; 网络首发日期: 2021-11-20

基金项目: 国家自然科学基金(61971237)

通信作者: haitaoyin@njupt.edu.cn

方法包括拉普拉斯金字塔^[4]、小波变换^[5]、曲波变换^[6]、非下采样轮廓波变换(NSCT)^[7-8]及稀疏表示(SR)^[9]等。空间域方法则直接对源图像的像素或区域进行融合操作,包括基于主成分分析的融合算法^[10]、基于四叉树的融合算法^[11]、基于多尺度加权梯度的融合算法(MWGF)^[12]、基于密集尺度不变特征变换^[13]和空间频率^[14]等。变换域和空间域方法需要人工设计显著信息的活跃水平度量及融合规则,一定程度上限制了其普适性。

近些年,深度学习由于其强大的特征提取与数据表征能力,在语义分割^[15]、目标检测^[16]、图像分类^[17]、图像去噪^[18]和图像超分辨率^[19]等领域大幅提升了算法的性能。目前,研究者们也将深度学习引入了多聚焦图像融合领域。文献^[20]提出了一种基于卷积神经网络(CNN)的多聚焦图像融合算法,该算法将融合多聚焦图像视为分类问题,构建CNN模型对源图像中聚焦与非聚焦区域进行分类。文献^[21]提出了一种基于像素卷积神经网络的方法,该方法根据邻域信息来识别源图像中的聚焦和非聚焦像素点。文献^[22]提出了一种基于无监督深度学习的融合模型,该模型通过提取图像特征的空间频率来生成初始决策图。上述方法通过CNN提高了图像特征提取的准确性,但仍需要诸如形态学和引导滤波等后处理来提高融合决策的可靠性,而且采用的CNN模型并不是端到端的。近年来,针对这一问题,学者们提出了端到端的融合网络模型,如DeepFuse^[23]、IFCNN^[24]和UFAFuse^[25]。与传统融合算法相比,基于深度学习的融合算法在一定程度上提高了融合质量,但仍存在一些局限性:1)对源图像的特征提取尺度单一;2)对神经网络的特征利用不充分;3)不能有效地突出显著特征的重要性。

为解决上述问题,本文提出了一种基于含注意力机制的多尺度空洞卷积网络的多聚焦图像融合算法。该融合网络包含特征提取、特征融合和图像重建等3个子网络。特征提取部分主要由含注意力机制的多尺度空洞卷积模块(MDB)组成。利用不同扩张率的空间卷积构造多尺度空洞卷积模块,提取多聚焦图像的多尺度特征,能更加全面地对多聚焦图像进行表示。

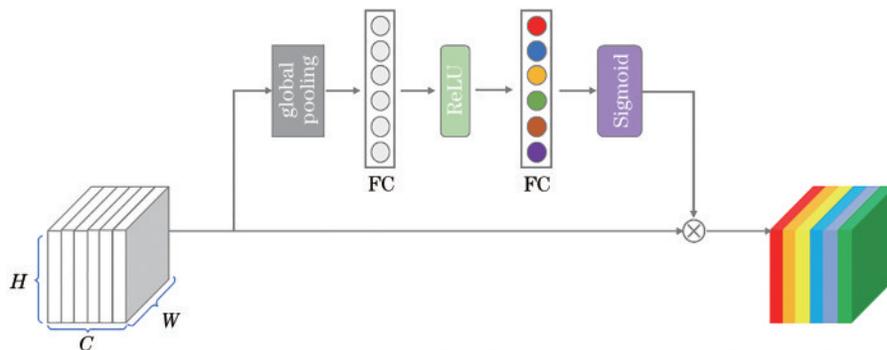


图2 SE注意力模块

Fig. 2 SE attention module

此外,为了进一步突出显著特征的重要性,在多尺度空洞卷积模块中引入了注意力机制对特征进行自适应地选择。特征融合和图像重建部分由常规的卷积模块组成。最后,通过端到端的监督学习对所提融合网络进行训练。在常用公开数据集上的实验结果验证了所提算法的有效性。

2 相关工作

2.1 残差网络

神经网络的网络层数对图像特征表示能力影响很大,较深的层数能提高神经网络的学习能力,但同时也容易导致梯度消失或爆炸现象、网络收敛困难等问题^[26]。为此,He等^[26]提出了残差网络,其核心思想是通过恒等连接将前面卷积层的信息传递到后面卷积层。图1为残差模块的结构图,对应的数学表达式为

$$H(\mathbf{x}) = F(\mathbf{x}) + \mathbf{x}, \quad (1)$$

式中: \mathbf{x} 表示输入; $H(\mathbf{x})$ 表示输出; $F(\mathbf{x})$ 为残差。残差网络使网络由原来的优化目标 $H(\mathbf{x})$ 转变为 $F(\mathbf{x})$ 。目前,残差模块作为基础结构被广泛地应用于各种神经网络的搭建,大幅提升了神经网络的性能。

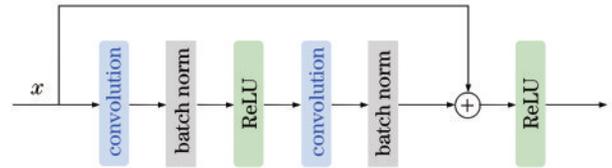


图1 残差模块

Fig. 1 Residual block

2.2 注意力机制

注意力机制源于人类视觉系统,能自动地注意神经网络中重要的特征,进一步优化神经网络、提升模型性能。Squeeze-and-Excitation(SE)^[27]是一种典型的注意力模块,核心思想是通过权重来自动地刻画不同特征通道的重要性。

SE注意力模块的结构图如图2所示,主要包括Squeeze与Excitation两个部分。假设 $\mathbf{x} \in \mathbf{R}^{W \times H \times C}$ 为SE模块的输入特征,其中 C 为输入特征的通道数, H

和 W 分别为特征的维度。Squeeze 部分对特征通道进行全局平均池化,即

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j), c = 1, 2, \dots, C, \quad (2)$$

式中: x_c 表示第 c 个特征通道。接下来,Excitation 部分将全局注意力信息 z 生成通道权重:

$$s = \sigma[W_2 \delta(W_1 z)], \quad (3)$$

式中: σ 为 Sigmoid 激活函数; δ 为 ReLU 激活函数; W_1 与 W_2 分别代表实现特征降维和升维的全连接(FC)操作; 向量 s 中的元素表示每个特征通道的重要性程度。最后,将 s 与 x 相乘实现特征通道权重赋予,从而得到 SE 模块的输出。

3 所提算法

3.1 网络结构

针对现有基于深度学习的图像融合算法对源图像特征提取的尺度单一及不能有效突出显著特征等问题,本研究提出了一种基于含注意力机制的多尺度空

洞卷积的多聚焦图像融合网络,简称 MDF-Net,网络整体结构如图 3 所示。所提 MDF-Net 包含 3 个部分:特征提取、特征融合和图像重建。

特征提取部分用于提取源图像的特征,该部分由两个网络分支组成,并共享参数以确保两个源图像特征的一致性和对应性。每条网络分支由 1 个 7×7 卷积层、3 个 1×1 卷积层和 3 个 MDB 构成。此外,将不同 MDB 的输出特征合并形成最终的特征,其中 3 个 MDB 输出特征的通道维度分别为 64、128 和 256, 1×1 卷积层用于调整特征通道维度。

特征融合部分主要实现特征的融合,由拼接操作和 1 个 1×1 卷积层组成,将特征提取部分得到的源图像特征进行拼接和 1×1 卷积操作,实现特征融合,其中 1×1 卷积层的输入通道和输出通道分别为 256 和 64。

图像重建部分主要将融合特征生成融合图像,该部分由 4 个 3×3 卷积层组成,输出特征通道数分别为 64、64、64 和 3。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。

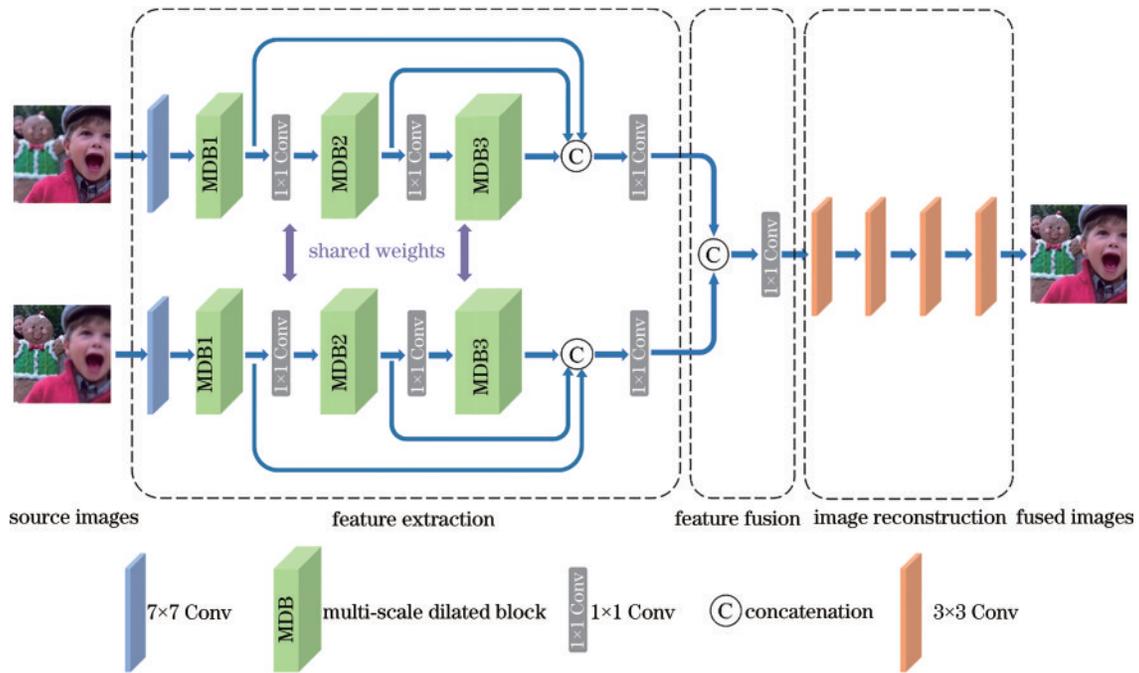


图 3 MDF-Net 结构图

Fig. 3 Architecture of MDF-Net

3.2 多尺度空洞卷积模块

MDB 模块是所提 MDF-Net 的核心模块。MDB 结构如图 4 所示,主要包括空洞卷积、SE 注意力模块及残差连接。

与传统卷积相比,空洞卷积通过扩张率使得同样尺寸的卷积核可以获得较大的感受野,能提取较大范围的图像特征,而且不增加参数量^[28]。为了能全面地提取多聚焦图像的特征,所提 MDB 由多个不同扩张率的空洞卷积组成。首先,将输入特征平均划分为

4 组。然后,采用扩张率分别为 1、2、3、4 的空洞卷积进行特征表示。空洞卷积示意图如图 5 所示。因此,图 4 显示的 MDB 含有 4 个不同扩张率的空洞卷积分支。此外,为了更好地突出多聚焦图像中显著特征的重要性,在每个空洞卷积分支中引入 SE 模块,自动地捕获多聚焦图像的显著特征,其中 SE 模块的降维比例设置为 16。接下来,4 个空洞卷积分支的输出通过拼接和 1×1 卷积层进行操作。最后,整个 MDB 模块通过残差形式连接。

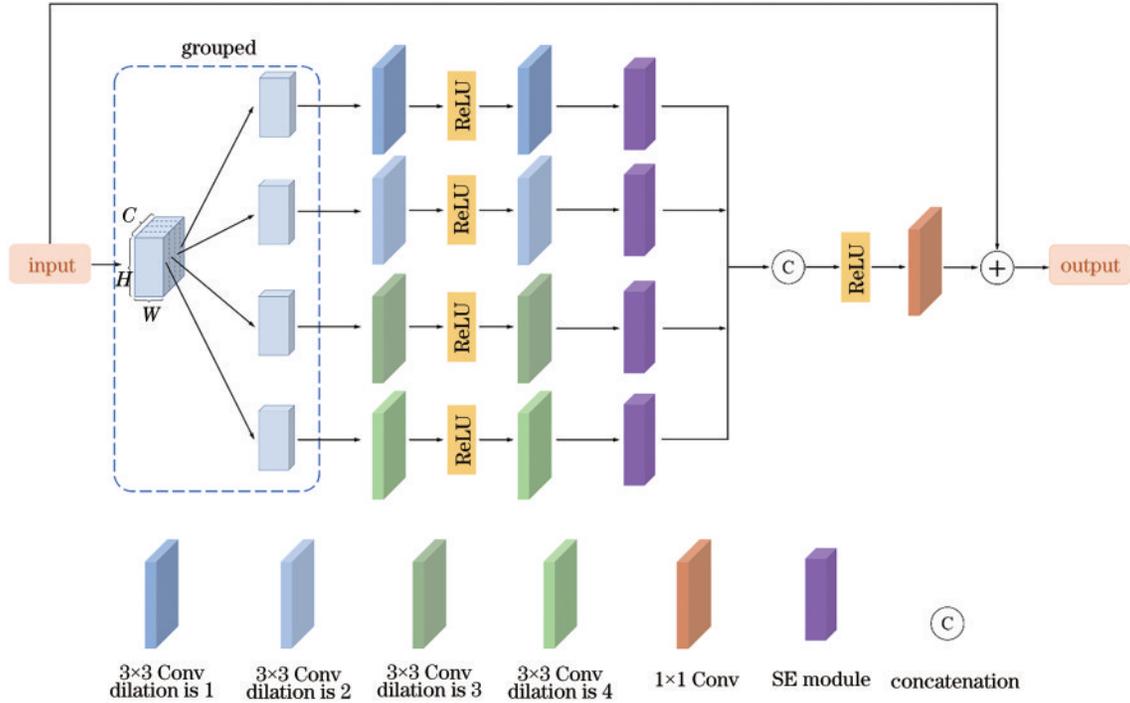


图 4 多尺度空洞卷积模块
Fig. 4 Multi-scale dilated block

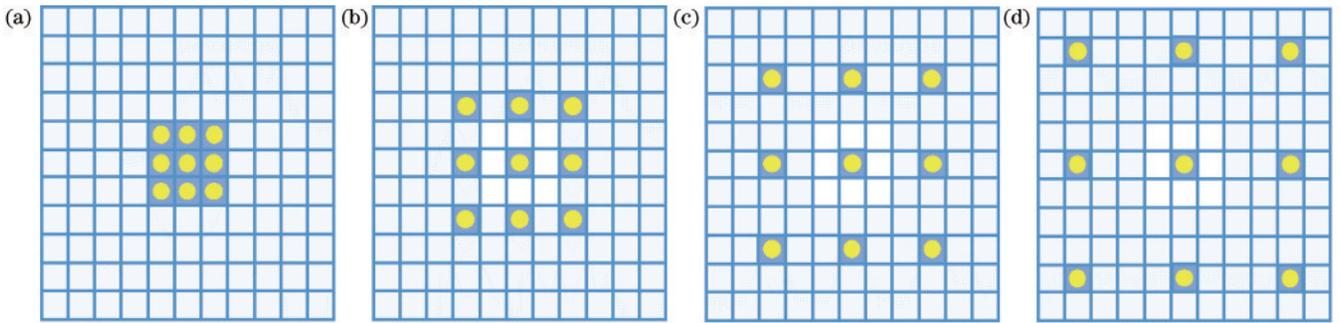


图 5 不同扩张率的空洞卷积示意图。(a)扩张率为 1;(b)扩张率为 2;(c)扩张率为 3;(d)扩张率为 4
Fig. 5 Dilated convolutional with different dilation rates. (a) Dilation rate is 1; (b) dilation rate is 2; (c) dilation rate is 3; (d) dilation rate is 4

3.3 网络训练

在实际应用中,很难获得多聚焦图像及其配对的全景深图像。为此,本研究构建了一个模拟多聚焦图像数据集。首先对清晰自然图像的不同区域进行高斯滤波处理,得到模拟多聚焦图像,然后将原始图像作为参考图像。选取 MS-COCO 数据集^[29]中 8000 张高清自然图像作为训练集。为了模拟不同景深产生的不同程度模糊,将选取的 8000 张图像分成 4 组,每组分别采用不同设置的高斯滤波器进行模糊处理。图 6 给出了合成多聚焦图像数据集的示例。

所提 MDF-Net 采用的损失函数为

$$L = \lambda L_1 + (1 - \lambda) L_{SSIM}, \quad (4)$$

$$L_1 = \| \mathbf{O} - \mathbf{I} \|_1, \quad (5)$$

$$L_{SSIM} = 1 - SSIM(\mathbf{O}, \mathbf{I}), \quad (6)$$

$$SSIM(\mathbf{O}, \mathbf{I}) = \frac{(2\mu_o\mu_i + C_1)(\sigma_{oi} + C_2)}{(\mu_o^2 + \mu_i^2 + C_1)(\mu_o^2 + \mu_i^2 + C_2)}, \quad (7)$$

式中: λ 为平衡参数,设为 0.2; L_1 表示 1-范数损失函数,用于计算融合图像与参考图像之间的绝对值误差,可以度量融合图像与参考图像之间的差异; \mathbf{O} 为 MDF-Net 网络的输出, \mathbf{I} 为参考图像; L_{SSIM} 表示结构相似度损失函数,用于计算网络的输出与参考图像之间的结构性差异; μ_o 和 μ_i 分别代表 \mathbf{O} 和 \mathbf{I} 的平均值; σ_o 和 σ_i 分别表示 \mathbf{O} 和 \mathbf{I} 的标准差; σ_{oi} 代表 \mathbf{O} 和 \mathbf{I} 的协方差;常数 C_1 和 C_2 是为了避免出现分母为 0 的情况。

采用 Pytorch 来实现 MDF-Net,程序运行的环境如下:RTX 3080/10 GB RAM;Intel Core i7-10700K@3.80 GHz。训练过程中的 batch 大小设为 8。采用 Adam 优化器对参数进行优化,优化器的初始学习率设为 0.001,并采用余弦退火衰减方式调整学习率。

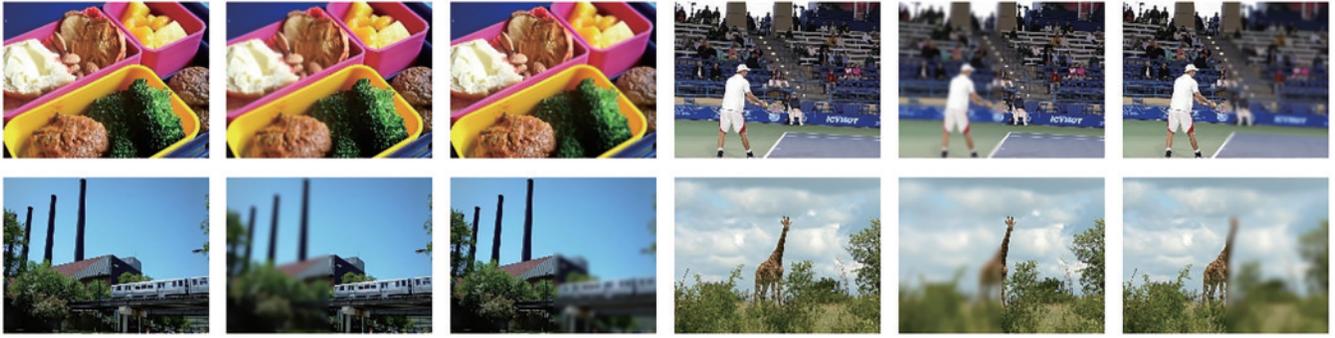


图 6 模拟多聚焦图像示例图

Fig. 6 Examples of simulated multi-focus images

4 实验结果与分析

4.1 实验设置

为评估 MDF-Net 的融合性能,选取了 NSCT 方法^[7]、SR 方法^[9]、IMF 方法^[30]、MWGF 方法^[12]、CNN 方法^[20]、DeepFuse 方法^[23]、DenseFuse-ADD 方法^[31]、DenseFuse-L1^[31]和 IFCNN-MAX 方法^[24]进行实验对比。所有对比方法均使用文献中提供的默认参数进行测试。采用 Lytro 数据集^[32]中 20 组彩色多聚焦图像作为测试图像。

此外,采用平均梯度(AG)^[33]、空间频率(SF)^[34]、视觉信息保真度(VIF)^[35]和 $Q^{AB/F}$ ^[36]等 4 种评价指标对融合图像质量进行客观评价。AG 指标能度量图像清晰程度,通过计算图像梯度的变化来反映图像细节轮廓的表征情况,值越大表示图像纹理边缘越清晰。SF 指标通过对图像的关联像素灰度值进行分析,计算图像的总活程度,值越高表示图像对微小细节的表达能力越强。VIF 指标从信息交互角度分析融合图像与源图像之间包含视觉信息的精准度,值越大表示融合图像的视觉信息越多,图像质量越佳。 $Q^{AB/F}$ 利用局部度量来估计来自输入的边缘信息在融合图像中的表现程度,值越高表示融合图像的质量越好。

4.2 对比实验结果

从 20 组融合结果中选取 4 组进行视觉效果对比,分别如图 7(“小孩”图像)、图 8(“猴子”图像)、图 9(“体育馆”图像)和图 10(“雕塑”图像)所示。此外,为了更加清晰地对比融合图像的细节效果,将融合图像局部区域放大,并置于融合图像的左下角。通过视觉对比可以看出,IMF 算法在边缘处存在不同程度的伪影,如图 7(e)、图 8(e)、图 9(e)和图 10(e)所示。图 7(d)、图 8(d)、图 9(d)和图 10(d)的局部放大区域表明,SR 算法产生了一定的模糊效应。NSCT 算法[图 7(c)、图 8(c)、图 9(c)和图 10(c)]和 MWGF 算法[图 7(f)、图 8(f)、图 9(f)和图 10(f)]虽然细节纹理特征清晰,但局部放大图像的对比度较低。DeepFuse 算法融合结果的整体亮度偏高,图像清晰度较低,如图 7(h)、图 8(h)、图 9(h)和图 10(h)所示;CNN 算法[图 7(g)、图 8(g)、图 9(g)和图 10(g)]

和 IFCNN-MAX 算法[图 7(k)、图 8(k)、图 9(k)和图 10(k)]的融合效果较佳,但在边缘保持和细节信息表征上整体效果较弱。DenseFuse-ADD 算法[图 7(i)、图 8(i)、图 9(i)和图 10(i)]和 DenseFuse-L1 算法[图 7(j)、图 8(j)、图 9(j)和图 10(j)]对于保留源图像中细节信息的能力较弱,图像清晰度较低。相比之下,所提 MDF-Net[图 7(l)、图 8(l)、图 9(l)和图 10(l)]融合结果的细节信息更加丰富、图像清晰度与对比度更高,图像视觉质量更优,具有更好的视觉效果。

表 1 为各种测试方法在“体育馆”单个图像上的性能指标。表 2 给出了各种测试方法在 Lytro 数据集上的平均指标值结果。从表 1 和表 2 可以看出,所提 MDF-Net 在 AG、SF 和 VIF 这 3 项指标上获得最优结果。对于 $Q^{AB/F}$ 指标,CNN 算法取得最好结果,略高于所提 MDF-Net。综合主观视觉效果和客观评价指标,所提 MDF-Net 能够有效地减少图像细节损失,更好地融合源图像中的细节信息,使融合图像的细节更清晰、色彩分布与源图像更接近,产生较好的融合效果。

4.3 消融实验

为了验证空洞卷积、注意力机制及多尺度空洞卷积间的拼接对网络性能的影响,分别测试了 MDF-Net 的 3 种变化模型:1)采用常规的 3×3 卷积代替了空洞卷积,记为 Without-Dilated;2)去掉 SE 模块,记为 Without-SE;3)去掉多尺度空洞卷积模块之间的拼接,记为 Without-Cat。表 3 列出了不同网络配置的 MDF-Net 在 Lytro 数据集上的平均客观指标。表 3 中的指标结果表明,所提 MDB 能更加有效地提取源图像特征,从而提高融合性能。

接下来,为了分析 MDB 模块数和空洞卷积分支数对 MDF-Net 的影响,MDB 模块数分别为 1、2、3、4 (其中 MDB 的输出特征通道数依次设置为“64”“64-128”“64-128-256”和“64-128-256-512”),空洞卷积分支数分别为 2、3、4、5 (其中每个空洞分支的输入特征通道数均为 16)。表 4 和表 5 分别给出了关于 MDB 模块数和空洞卷积分支数的消融实验结果。从表 4 和表 5 可以看出,含有 3 个 MDB 和 4 个不同扩张率空洞卷积分支的 MDF-Net 获得较好的融合效果。

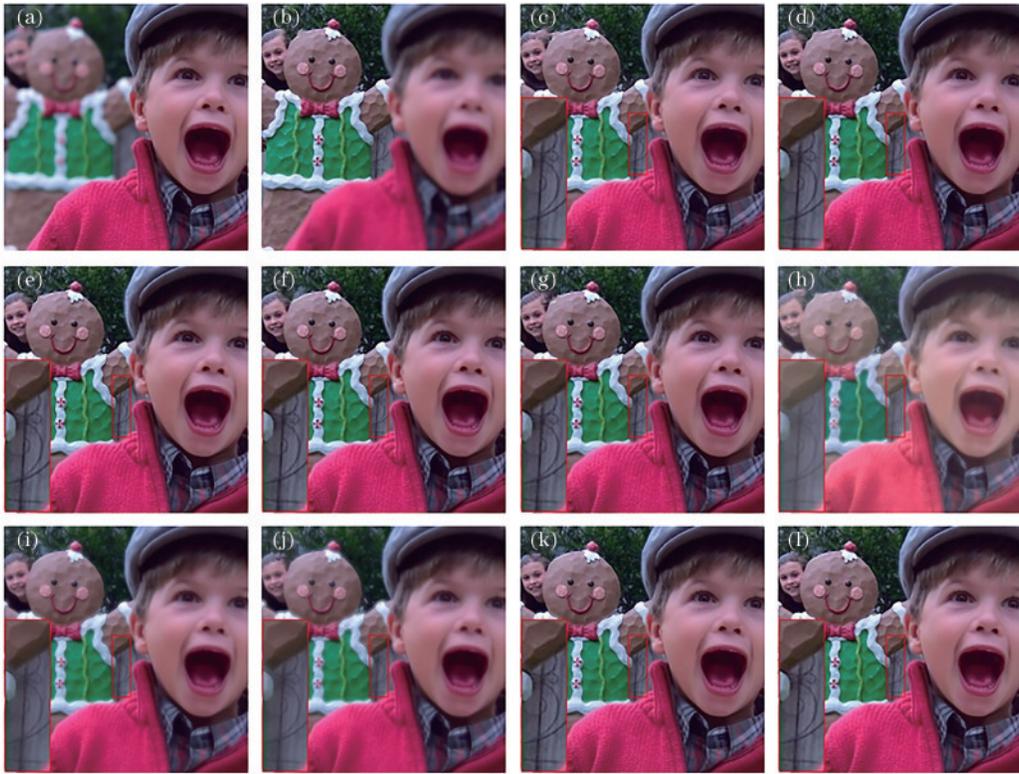


图 7 “小孩”图像的融合结果。(a)(b)源图像;(c) NSCT;(d) SR;(e) IMF;
 (f) MWGF;(g) CNN;(h) DeepFuse;(i) DenseFuse-ADD;(j) DenseFuse-L1;(k) IFCNN-MAX;(l) MDF-Net
 Fig. 7 Fused results of “children” image. (a)(b) Source images; (c) NSCT; (d) SR; (e) IMF; (f) MWGF; (g) CNN; (h) DeepFuse;
 (i) DenseFuse-ADD; (j) DenseFuse-L1; (k) IFCNN-MAX; (l) MDF-Net

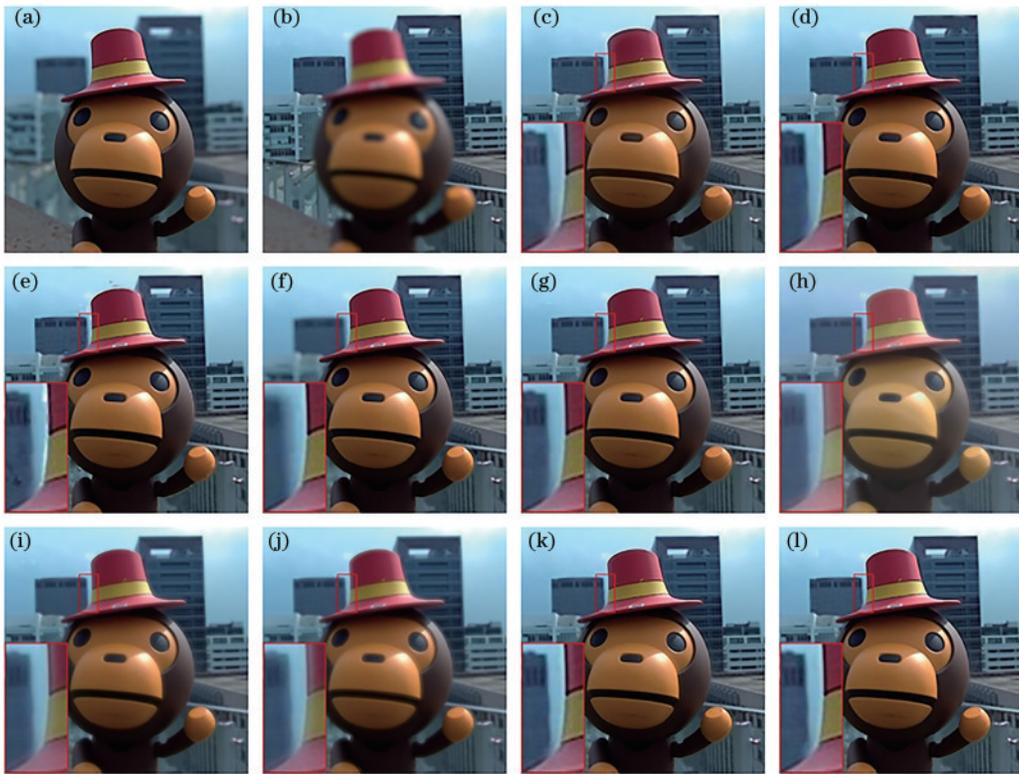


图 8 “猴子”图像的融合结果。(a)(b)源图像;(c) NSCT;(d) SR;(e) IMF;
 (f) MWGF;(g) CNN;(h) DeepFuse;(i) DenseFuse-ADD;(j) DenseFuse-L1;(k) IFCNN-MAX;(l) MDF-Net
 Fig. 8 Fusion results of “monkey” image. (a)(b) Source images; (c) NSCT; (d) SR; (e) IMF; (f) MWGF; (g) CNN; (h) DeepFuse;
 (i) DenseFuse-ADD; (j) DenseFuse-L1; (k) IFCNN-MAX; (l) MDF-Net

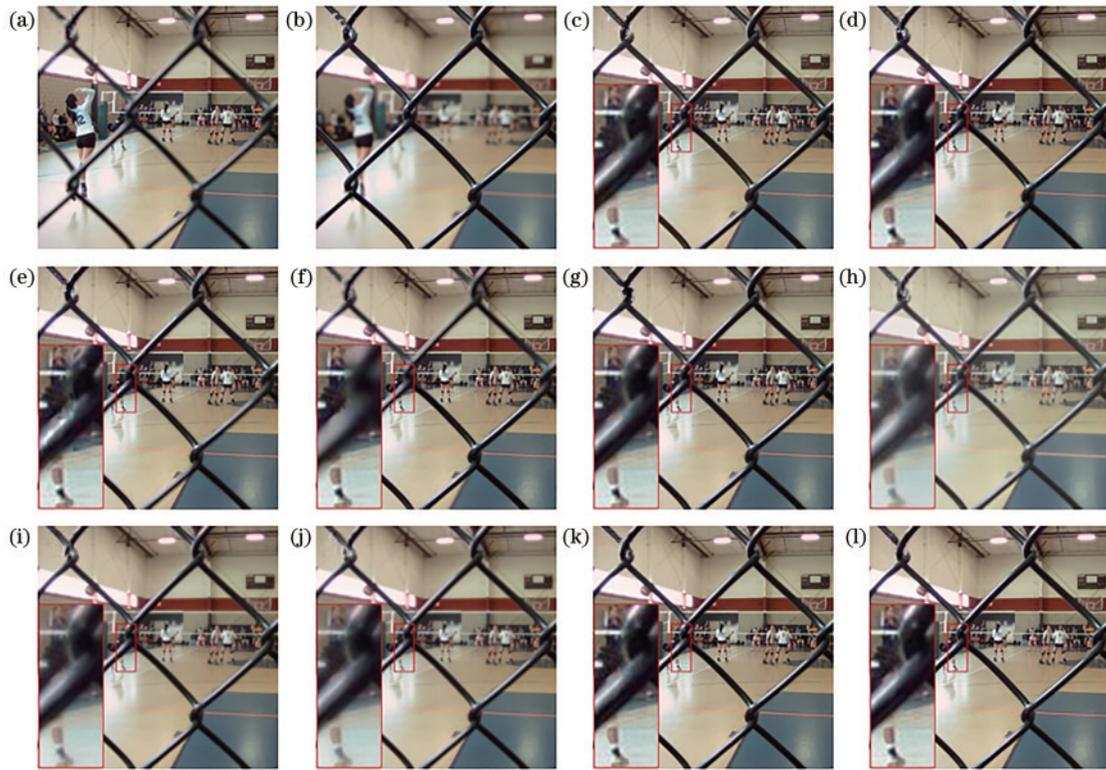


图 9 “体育馆”图像的融合结果。(a)(b)源图像;(c) NSCT;(d) SR;(e) IMF;(f) MWGF;(g) CNN;(h) DeepFuse;

(i) DenseFuse-ADD;(j) DenseFuse-L1;(k) IFCNN-MAX;(l) MDF-Net

Fig. 9 Fusion results of “gymnasium” image. (a)(b) Source images; (c) NSCT; (d) SR; (e) IMF; (f) MWGF; (g) CNN; (h) DeepFuse; (i) DenseFuse-ADD; (j) DenseFuse-L1; (k) IFCNN-MAX; (l) MDF-Net

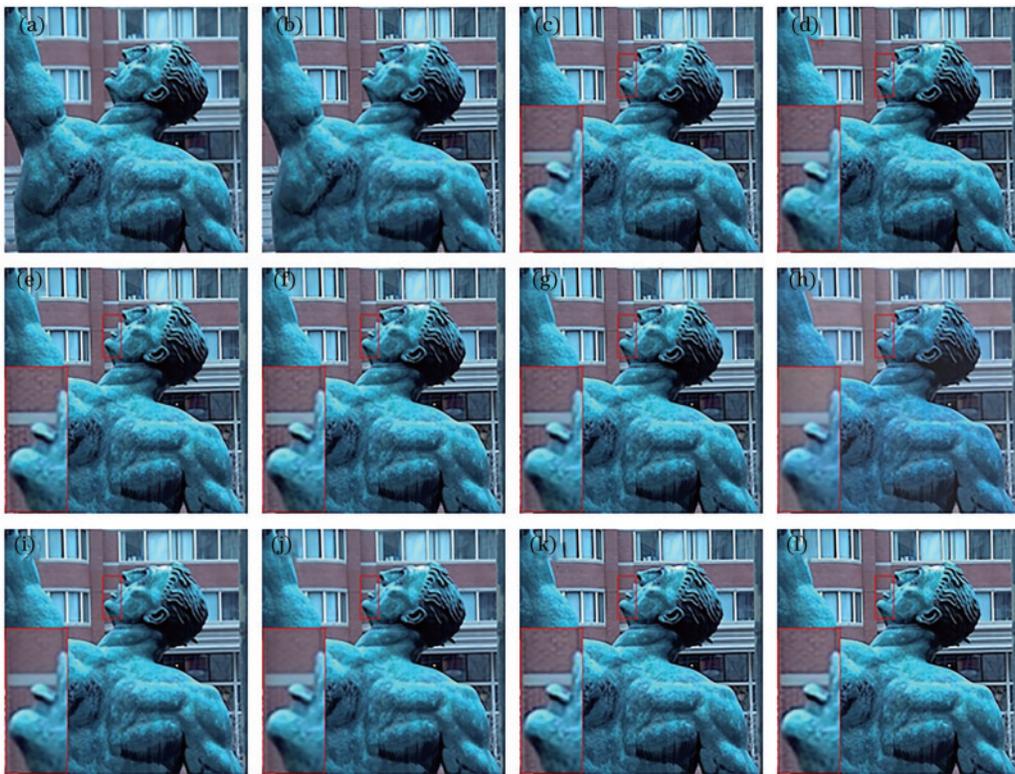


图 10 “雕塑”图像的融合结果。(a)(b)源图像;(c) NSCT;(d) SR;(e) IMF;

(f) MWGF;(g) CNN;(h) DeepFuse;(i) DenseFuse-ADD;(j) DenseFuse-L1;(k) IFCNN-MAX;(l) MDF-Net

Fig. 10 Fusion results of “statue” image. (a)(b) Source images; (c) NSCT; (d) SR; (e) IMF; (f) MWGF; (g) CNN; (h) DeepFuse; (i) DenseFuse-ADD; (j) DenseFuse-L1; (k) IFCNN-MAX; (l) MDF-Net

表 1 各种融合算法在“体育馆”图像上的指标值

Table 1 Indexes values of various fusion algorithms on “gymnasium” image

Algorithm	AG	SF	VIF	$Q^{AB/F}$
NSCT	6.2530	18.7075	0.9100	0.6685
SR	6.2463	18.8690	0.9156	0.7096
IMF	6.2806	18.7851	0.8795	0.6973
MWGF	6.0492	17.4456	0.8422	0.6906
CNN	6.1795	18.5287	0.9268	0.7165
DeepFuse	4.1720	10.9647	0.7062	0.5375
DenseFuse-ADD	4.5071	12.0923	0.8049	0.6110
DenseFuse-L1	4.4546	11.8492	0.7844	0.6084
IFCNN-MAX	6.4207	18.9627	0.9421	0.6876
MDF-Net	6.4541	19.3478	0.9424	0.7098

表 2 各种融合算法在 Lytro 数据集上的平均指标值

Table 2 Average indexes values of various fusion algorithms on Lytro dataset

Algorithm	AG	SF	VIF	$Q^{AB/F}$
NSCT	6.8506	18.8012	0.9187	0.7168
SR	6.8328	18.9940	0.9309	0.7437
IMF	6.9688	19.2298	0.9319	0.7406
MWGF	6.8446	19.0156	0.9308	0.7437
CNN	6.8814	19.0005	0.9373	0.7555
DeepFuse	4.0751	10.4187	0.6613	0.4860
DenseFuse-ADD	4.4676	11.5175	0.7727	0.5806
DenseFuse-L1	4.3616	11.0373	0.7510	0.5587
IFCNN-MAX	6.8742	18.9718	0.9327	0.7105
MDF-Net	6.9957	19.2976	0.9437	0.7488

表 3 MDF-Net 不同变体的客观指标结果

Table 3 Indexes results of different variants of MDF-Net

Variants of MDF-Net	AG	SF	VIF	$Q^{AB/F}$
Without-Dilated	6.8088	18.8046	0.9132	0.7485
Without-SE	6.9459	19.1778	0.9372	0.7472
Without-Cat	6.8248	18.8041	0.9213	0.7468
MDF-Net	6.9957	19.2976	0.9437	0.7488

表 4 关于 MDB 模块数的消融实验

Table 4 Ablation experiment on number of MDB modules

#MDB	AG	SF	VIF	$Q^{AB/F}$
1	6.7833	18.7125	0.9217	0.7465
2	6.9219	19.0769	0.9334	0.7473
3	6.9957	19.2976	0.9437	0.7488
4	6.7643	18.6264	0.9125	0.7443

最后,针对 MDB 中输出特征的通道数对 MDF-Net 性能的影响进行了实验分析。在第 4.2 节中, MDF-Net 的 3 个 MDB 的输出特征通道数分别设置为

表 5 关于 MDB 中空洞卷积分支数的消融实验

Table 5 Ablation experiment on number of dilated convolution branches in MDB

#Branch	AG	SF	VIF	$Q^{AB/F}$
2	6.7490	18.5700	0.9035	0.7484
3	6.8243	18.7955	0.9169	0.7446
4	6.9957	19.2976	0.9437	0.7488
5	6.9094	18.9495	0.9301	0.7403

“64-128-256”。为了验证特征通道参数设置的有效性,测试了其他几种关于 3 个 MDB 模块特征通道参数设置,分别为“64-64-64”“128-128-128”和“256-256-256”。表 6 给出了含不同特征通道数的 MDF-Net 在测试集上的平均性能指标。从表 6 可以看出, MDF-Net 中的 3 个 MDB 的特征通道数设置为“64-128-256”时获得较好的融合效果。

表 6 关于 MDB 中的特征通道数的消融实验

Table 6 Ablation experiment on number of feature channels in MDB

#Channel	AG	SF	VIF	$Q^{AB/F}$
64-64-64	6.8616	18.8961	0.9269	0.7457
128-128-128	6.8891	18.9124	0.9235	0.7455
256-256-256	6.8319	18.8076	0.9294	0.7401
64-128-256	6.9957	19.2976	0.9437	0.7488

5 结 论

提出了一种含注意力机制的多尺度空洞卷积模块,并基于该模块构建了一种多聚焦图像融合网络。所提多尺度空洞卷积模块通过引入不同扩张率空洞卷积来增大卷积核的感受野,从而更加全面地提取多聚焦图像中的多尺度特征,而且不增加参数量。此外,使用 SE 注意力模块进一步突出显著特征的重要性,能更准确地获取显著特征,提高融合图像的质量。客观和主观的对比实验结果验证了所提算法的有效性。在后续的工作中,将设计结合多聚焦图像融合特点的损失函数,进一步提升该算法的性能。

参 考 文 献

- [1] Li S T, Kang X D, Fang L Y, et al. Pixel-level image fusion: a survey of the state of the art[J]. Information Fusion, 2017, 33: 100-112.
- [2] 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.
- [3] Liu Y, Wang Z F. Simultaneous image fusion and denoising with adaptive sparse representation[J]. IET Image Processing, 2015, 9(5): 347-357.
- [4] Wang W C, Chang F L. A multi-focus image fusion method based on Laplacian pyramid[J]. Journal of Computers, 2011, 6(12): 2559-2566.

- [5] Li S T, Yang B. Multifocus image fusion by combining curvelet and wavelet transform[J]. Pattern Recognition Letters, 2008, 29(9): 1295-1301.
- [6] Nencini F, Garzelli A, Baronti S, et al. Remote sensing image fusion using the curvelet transform[J]. Information Fusion, 2007, 8(2): 143-156.
- [7] Zhang Q, Guo B L. Multifocus image fusion using the nonsubsampling contourlet transform[J]. Signal Processing, 2009, 89(7): 1334-1346.
- [8] 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.
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.
- [9] Yang B, Li S T. Multifocus image fusion and restoration with sparse representation[J]. IEEE Transactions on Instrumentation and Measurement, 2010, 59(4): 884-892.
- [10] Zhang Y X, Chen L, Zhao Z H, et al. Multi-focus image fusion based on robust principal component analysis and pulse-coupled neural network[J]. Optik, 2014, 125(17): 5002-5006.
- [11] Bai X Z, Zhang Y, Zhou F G, et al. Quadtree-based multi-focus image fusion using a weighted focus-measure[J]. Information Fusion, 2015, 22: 105-118.
- [12] Zhou Z Q, Li S, Wang B. Multi-scale weighted gradient-based fusion for multi-focus images[J]. Information Fusion, 2014, 20: 60-72.
- [13] Liu Y, Liu S P, Wang Z F. Multi-focus image fusion with dense SIFT[J]. Information Fusion, 2015, 23: 139-155.
- [14] Li S T, Kwok J T, Wang Y N. Combination of images with diverse focuses using the spatial frequency[J]. Information Fusion, 2001, 2(3): 169-176.
- [15] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 3431-3440.
- [16] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN [C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2980-2988.
- [17] 马永杰, 刘培培. 基于 DenseNet 进化的卷积神经网络图像分类算法[J]. 激光与光电子学进展, 2020, 57(24): 241001.
Ma Y J, Liu P P. Convolutional neural network based on DenseNet evolution for image classification algorithm[J]. Laser & Optoelectronics Progress, 2020, 57(24): 241001.
- [18] 甘建旺, 沙芸, 张国英. 基于非对称卷积神经网络的图像去噪[J]. 激光与光电子学进展, 2020, 57(22): 221018.
Gan J W, Sha Y, Zhang G Y. Image denoising based on asymmetric convolutional neural networks[J]. Laser & Optoelectronics Progress, 2020, 57(22): 221018.
- [19] 陈星宇, 张伟劲, 孙伟智, 等. 基于多尺度与多重残差网络的图像超分辨率重建[J]. 激光与光电子学进展, 2020, 57(18): 181009.
Chen X Y, Zhang W J, Sun W Z, et al. Super-resolution reconstruction of images based on multi-scale and multi-residual network[J]. Laser & Optoelectronics Progress, 2020, 57(18): 181009.
- [20] 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.
- [21] Tang H, Xiao B, Li W S, et al. Pixel convolutional neural network for multi-focus image fusion[J]. Information Sciences, 2018, 433/434: 125-141.
- [22] Ma B Y, Zhu Y, Yin X, et al. SESF-Fuse: an unsupervised deep model for multi-focus image fusion[J]. Neural Computing and Applications, 2021, 33(11): 5793-5804.
- [23] Prabhakar K R, Srikanth V S, Babu R V. DeepFuse: a deep unsupervised approach for exposure fusion with extreme exposure image pairs[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 4724-4732.
- [24] Zhang Y, Liu Y, Sun P, et al. IFCNN: a general image fusion framework based on convolutional neural network [J]. Information Fusion, 2020, 54: 99-118.
- [25] Zang Y S, Zhou D M, Wang C C, et al. UFA-FUSE: a novel deep supervised and hybrid model for multifocus image fusion[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-17.
- [26] 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, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [27] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.
- [28] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[EB/OL]. (2015-11-23)[2021-05-06]. <https://arxiv.org/abs/1511.07122>.
- [29] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context[M]//Fleet D, Pajdla T, Schiele B, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8693: 740-755.
- [30] 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.
- [31] Li H, Wu X J. DenseFuse: a fusion approach to infrared and visible images[J]. IEEE Transactions on Image Processing, 2019, 28(5): 2614-2623.
- [32] Nejadi M, Samavi S, Shirani S. Multi-focus image fusion using dictionary-based sparse representation[J]. Information Fusion, 2015, 25: 72-84.

- [33] Song H H, Yu S Y, Song L, et al. Fusion of multispectral and panchromatic satellite images based on contourlet transform and local average gradient[J]. *Optical Engineering*, 2007, 46(2): 020502.
- [34] Shah P, Merchant S N, Desai U B. Multifocus and multispectral image fusion based on pixel significance using multiresolution decomposition[J]. *Signal, Image and Video Processing*, 2013, 7(1): 95-109.
- [35] Han Y, Cai Y Z, Cao Y, et al. A new image fusion performance metric based on visual information fidelity [J]. *Information Fusion*, 2013, 14(2): 127-135.
- [36] Xydeas C S, Petrović V. Objective image fusion performance measure[J]. *Electronics Letters*, 2000, 36(4): 308-309.