

文章编号:1671-6833(2019)04-0036-06

# 基于 CNN 的非下采样剪切波域多聚焦图像融合

刘帅奇<sup>1,2</sup>, 王 洁<sup>1,2</sup>, 安彦玲<sup>1,2</sup>, 李子奇<sup>1,2</sup>, 胡绍海<sup>3</sup>, 王文峰<sup>4</sup>

(1.河北大学 电子信息工程学院,河北 保定 071000; 2.河北省机器视觉工程技术研究中心,河北 保定 071000; 3.北京交通大学 信息所,北京 100044; 4.中国科学院 新疆生态与地理研究所数字图像处理实验室,新疆 乌鲁木齐 830011)

**摘要:**结合非下采样剪切波变换的时频分离优良特性,提出了一种基于卷积神经网络(convolutional neural networks, CNN)的非下采样剪切波变换(non-subsampled shearlet transform, NSST)域图像融合算法.首先对源图像进行 NSST 分解,其次对分解的低频系数进行基于 CNN 的融合策略.最后对分解的高频系数进行基于向导滤波(guided filtering, GF)的改进加权的拉普拉斯能量和(improved weighted sum of Laplace energy, IWSML)模取大融合策略,然后将根据不同融合规则融合后的频率系数进行 NSST 反变化获取输出的清晰目标图像.实验结果表明,该方法不仅可以获得更利于人眼接受的视觉效果图,且有效地提高了融合图像的客观性能评价指标.

**关键词:**图像融合;多聚焦图像融合;非下采样剪切波变换;卷积神经网络;向导滤波

**中图分类号:** TU528.1; TJ01 **文献标志码:** A **doi:**10.13705/j.issn.1671-6833.2019.04.002

## 0 引言

一般将图像融合过程由低到高分为 3 个层次:像素级融合、特征级融合、决策级融合<sup>[1]</sup>.目前,空间域和变换域是主要的两种在像素级别基础上的融合方法.空间域算法主要包括基于区域能量和加权<sup>[2]</sup>、基于核范数最小化<sup>[3]</sup>等融合算法.而变换域算法主要采用多尺度几何变换进行图像融合<sup>[4]</sup>,例如文献[5]提出的基于轮廓波变换(contourlet transform, CT)的图像融合算法;文献[6]提出的基于非下采样轮廓波变换(non-subsampled contourlet transform, NSCT)与脉冲耦合神经网络(pulse coupled neural network, PCNN)图像融合算法;文献[7]提出的基于剪切波变换(shearlet transform, ST)与 PCNN 的图像融合算法.目前,处于主流位置的是变换域融合算法,其中应用最为广泛的为 CT 和 ST.与 CT 相比,ST 符合紧框架理论,可以有效地对图像的特征信息进行提取,同时其离散形式很容易实现<sup>[8]</sup>.因此,ST 在图像融合中得到了广泛的应用<sup>[9-10]</sup>.为了克服 ST 不具有平移不变性的缺点,也

为了更好地利用图像的空间连续性,文献[10]利用文献[11]提出的基于两尺度向导滤波的融合算法思想,在复剪切波变换(complex shearlet transform, CST)域中引入图像的空间连续性.虽然该算法明显地提升了融合后图像的视觉效果,但是在图像数据保留方面还可以提升,也就是低频部分的融合规则还可以进行改进.

近年来,由于深度学习理论的研究逐步深入并在相关领域取得了较好的研究成果<sup>[12-13]</sup>,因此 CNN 也被应用于多聚焦图像融合中,如文献[12]提出的一种采用 CNN 的多聚焦图像融合规则.虽然该算法获得很好的图像融合效果,但是在图像融合的过程中,该方法并没有对图像进行高低频的划分,这显然不符合人眼的视觉规律.因此,笔者结合该融合规则提出了基于 NSST 域的多聚焦图像融合算法.

## 1 非下采样剪切波变换原理

设 Shearlet 变换的基函数为:

$$\psi_{j,l,k}(x) = |\det \mathbf{A}|^{j/2} \psi(\mathbf{B}^l \mathbf{A}^j x - k), \quad (1)$$

收稿日期:2018-11-01;修订日期:2019-03-06

基金项目:国家自然科学基金资助项目(61572063,61401308);河北省自然科学基金资助项目(F2016201187, F2016201142);河北省高等学校科学技术研究项目(QN2016085);河北大学引进人才科研启动经费(2014-303)

通信作者:胡绍海(1964—),男,江西新余人,北京交通大学教授,博士,博士生导师,主要从事多维信号处理研究, E-mail:shhu@bjtu.edu.cn.

式中:  $j \geq 0; l = -2^j \sim 2^j - 1; k \in \mathbf{Z}^2; \psi \in L^2(\mathbf{R}^2); \mathbf{A}^j$  表示伸缩变换矩阵;  $\mathbf{B}^l$  表示保区域的几何变换矩阵.

因此函数  $f$  的 ST 可表示为:

$$S_\psi = \langle f, \psi_{j,l,k} \rangle. \quad (2)$$

NSST<sup>[8]</sup> 是 ST 的扩展, 能够很好地克服图像的下采样操作, 因而具备图像变换的平移不变性. 在进行图像融合时, 能够使融合图像中拥有更多源图像的信息特征, 使整体的融合图像效果得到极大的改善.

## 2 基于卷积神经网络的图像融合规则

在基于 CNN 的图像融合算法中, 图像融合中

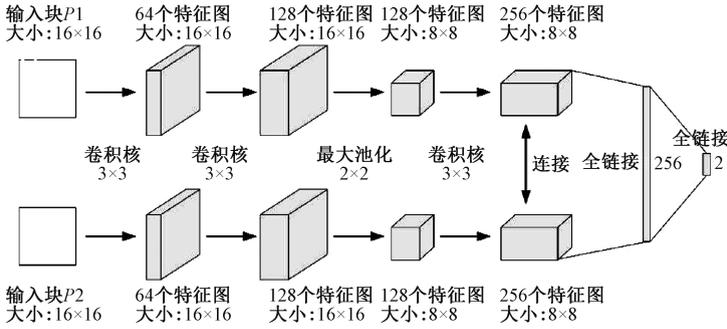


图 1 CNN 网络模型

Fig.1 The model of CNN

将 CNN 用于多聚焦图像融合算法的计算过程: 首先, 将两个源图像输送到预训练的 CNN 模型以输出得分图; 其次, 通过对重叠区域块求平均值, 从得分图中获得具有与源图像相同大小的焦点图; 再次, 将焦点图以 0.5 为阈值分割成二进制图; 然后, 使用小区域去除和向导滤波<sup>[15]</sup> 来精细化二进制分割图以生成最终决策图; 最后, 使用逐像素加权平均策略并利用最终决策图生成融合图像.

通过上述图像融合的步骤可知: 在图像融合过程中, 该方法并没有对图像进行高低频的划分, 这显然是不符合人眼的视觉规律的. 因此, 笔者结合该融合规则提出了一种基于 NSST 域的多聚焦图像融合方法.

## 3 向导滤波

向导滤波对图像的边缘保持具备很好的效果. 设向导图为  $I$ , 输入图像为  $p$ , 而输出图像为  $q$ <sup>[15]</sup>. 其中  $I$  与  $p$  可以是同一张图像. 向导滤波的关键部分就是假设一个线性模型, 即向导图  $I$  与输出图  $q$  之间存在局部线性模型:

$$q_i = a_k I_i + b_k, \forall i \in \omega_k, \quad (3)$$

式中:  $i$  为模型的像素索引;  $a, b$  为模型的线性系

决策图的生成通常被看作分类问题进行分析. 文献[12]给出了基于 CNN 的一种多聚焦图像融合算法, 将图像融合的主要任务转变成设计 CNN 的网络架构. 而活动水平测量和融合规则可以通过大量的图像数据进行 CNN 模型训练生成<sup>[12]</sup>.

文献[12]采用了文献[14]中提出的 Siamese 结构的 CNN 模型, 如图 1 所示. 由图 1 可知, 网络中的每个分支具有 3 个卷积层和 1 个最大池化层. 笔者设置图像块的大小为  $16 \times 16$ , 每个卷积核大小和步幅分别设置为  $3 \times 3$  和 1, 最大池化层的缩放因子和跨度分别设置为  $2 \times 2$  和 2.

数;  $\omega_k$  为以向导图  $I$  中一点  $k$  为中心, 长宽为  $(2r + 1)$  的正方形窗.

图像的边缘保持滤波问题则转换为如下的最优化问题:

$$E(a_k, b_k) = \sum_{i \in \omega_k} ((a_k I_i + b_k - p_i)^2 + \varepsilon a_k^2), \quad (4)$$

式中:  $\varepsilon$  表示归一化因子. 可以使用线性回归求解公式(4)的解<sup>[15]</sup>:

$$a_k = \frac{\frac{1}{|\omega|} \sum_{i \in \omega_k} I_i p_i - \mu_k \bar{p}_k}{\sigma_k^2 + \varepsilon}; b_k = \bar{p}_k - a_k \mu_k. \quad (5)$$

式中:  $\mu_k$  和  $\sigma_k^2$  表示向导图局部窗口中的均值和方差;  $\bar{p}_k$  表示  $p$  在窗口  $\omega_k$  中的均值;  $|\omega|$  表示窗口内的像素数. 为了更简单直观地表示, 在本文中使用  $G_{r,\varepsilon}(p, I)$  代表上述的向导滤波算法.

## 4 基于 CNN 的 NSST 域图像融合

笔者采用 NSST 进行图像分解, 该变换可以很好地抑制伪吉布斯效应. 文献[5]提出的改进的拉普拉斯能量和 (sum of Laplace energy, SML) 是一种能更好地表示图像边缘细节信息的区域能量函数, 在选择高频变换域系数中具有很好的效果, 笔者在此基础上对图像融合中的

拉普拉斯能量加上所对应的权值,并对高频选择系数图用向导滤波来进行平滑处理.设输入具有不同聚焦点的两幅源图像  $A$  和  $B$ ,经过融合后得到图像  $F$ .

首先,对两幅多聚焦的图像  $A$  和  $B$  采用相应 NSST 变换进行分解获得相应高频系数和低频系数为  $S_A^{l,d}(k)$  和  $S_B^{l,d}(k)$  (若没有下标则表示任一图像的分解系数),其中  $l$  和  $d$  分别表示 NSST 分解的尺度和方向,当  $l$  为 0 的时候表示低频系数,否则为其相应的高频系数.

其次,对于源图像分解得到的低频系数采用基于 CNN 的图像融合算法进行图像融合.利用文献[12]对不同的图像训练集训练得到的 CNN 网络模型获得图  $A$  和  $B$  的融合决策图  $m$ ,然后归一化并进行 NSST 来获取相应的低频系数决策图  $m_L$ ,则根据决策图获得的低频系数可由下式计算:

$$S_F^{0,d}(k) = m_L \cdot S_A^{0,d}(k) + (1 - m_L) \cdot S_B^{0,d}(k). \quad (6)$$

然后,对高频系数( $l$  大于 0)进行融合的规则为采用基于向导滤波的 IWSML 取大进行处理.

定义像素点的拉普拉斯能量如下:

$$M^{l,d}(x,y) = |2S^{l,d}(x,y) - S^{l,k}(x-1,y) - S^{l,k}(x+1,y)| + |2S^{l,k}(x,y) - S^{l,k}(x,y-1) - S^{l,k}(x,y+1)|. \quad (7)$$

由于高频图像的权值主要是由距离的大小和子带的方向所决定的,所以定义相应的权值公式:

$$\omega(x,y) = \omega^0(x,y) + \theta^d(x,y), \quad (8)$$

其中,  $\omega^0(x,y) =$

$$\begin{cases} \frac{1}{(x-x')^2 + (y-y')^2}, & (x,y) \neq (x',y') \\ 1, & (x,y) = (x',y') \end{cases},$$

而  $\theta^d$  中的  $\theta$  表示距离的大小,  $d$  表示 NSST 分解方向,其作用对  $\omega(x,y)$  的  $d$  方向数据进行翻倍.

因此, IWSML 的定义式为:

$$M_{\omega}^{l,d}(x,y) =$$

$$\sum_{\frac{n-1}{2}}^{\frac{n-1}{2}} \sum_{\frac{n-1}{2}}^{\frac{n-1}{2}} \omega(x,y) [M_{\omega_A}^{l,d}(x+a,y+b)]^2. \quad (9)$$

计算高频系数的 IWSML 值,可得  $M_{\omega_A}^{l,d}(k)$  和  $M_{\omega_B}^{l,d}(k)$  (其中  $k = (x,y)$ ),则计算高频决策图为:

$$\begin{cases} m_A(k) = \begin{cases} 1, & M_{\omega_A}^{l,d}(k) \geq M_{\omega_B}^{l,d}(k) \\ 0, & M_{\omega_A}^{l,d}(k) < M_{\omega_B}^{l,d}(k) \end{cases}; \\ m_B(k) = 1 - m_A(k). \end{cases} \quad (10)$$

然后,对  $m_A$  和  $m_B$  进行向导滤波来加强高频系数的空间联系性,如式(11)所示:

$$m_i = G_{r,\varepsilon}(m_i, S_i^{l,d}(k)), i = A, B. \quad (11)$$

随后,再将  $m_A$  和  $m_B$  进行数值的归一化规则处理,再通过式(12)能够获得最终所需要的高频部分系数  $S_F^{l,d}(k)$ :

$$S_F^{l,d}(k) = m_A \cdot S_A^{l,d}(k) + m_B \cdot S_B^{l,d}(k). \quad (12)$$

最终,将经过笔者提出的融合规则获得的低频系数  $S_F^{0,d}(k)$  和低频系数  $S_F^{l,d}(k)$  经过 NSST 重构获得最后所需要的目标图  $F$ . 经过上述的描述,所确定的本文的融合算法表示过程如图 2 所示.

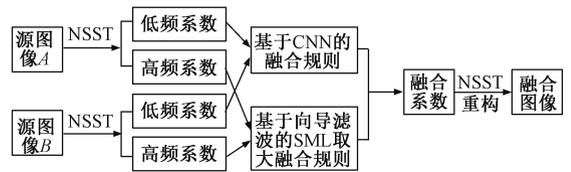


图 2 基于 CNN 的 NSST 域图像融合框架

Fig.2 The diagram of multi-focus image fusion algorithm based on CNN in NSST domain

## 5 实验结果

为了有效地评估该算法在不同聚焦程度图像中的融合性能,利用在不同图像融合中常用的融合测试图像所提出的算法,并且从主观和客观两方面与文献[5]中所提出的 CT-SML 图像融合方法、文献[6]提出 NSCT-PCNN 图像融合方法、文献[7]提出的 ST-PCNN 图像融合方法、文献[10]中提出的 CST-GF 图像融合方法和文献[12]提出基于 CNN 的图像融合方法进行对比.客观性能评价指标采用 5 种不同的方法,如图像方差 (standard deviation, Std)、图像平均梯度 (average gradient, Avg)<sup>[10]</sup>、图像  $Q^{AB/F}$  度量<sup>[5]</sup>、图像的互信息 (mutual information, MI)<sup>[10]</sup> 和图像空间频域度 (spatial frequency, SF)<sup>[10]</sup>. 这 5 个指标测试值越大,表明图像融合的评价指标越高,融合的图像越清晰,越符合人类的视觉系统.本文算法在实验过程中 NSST 的分解层数设置为 4 层,每层方向滤波器设置为有 6、10、10、18 个方向子带, CNN 参数选择与文献[12]相同,高频融合规则中的  $r=3, \varepsilon=1$ .

首先对多聚焦图像进行图像融合的测试,图 3(a)、(b)为经典的多聚焦图像融合测试图 Clock 的左侧部分聚焦和右侧部分聚焦的图像.然后采用所述的不同方法对多聚焦图像进行融

合,所得融合图像及与原测试图像的差图见图 3 (c) ~ (t).

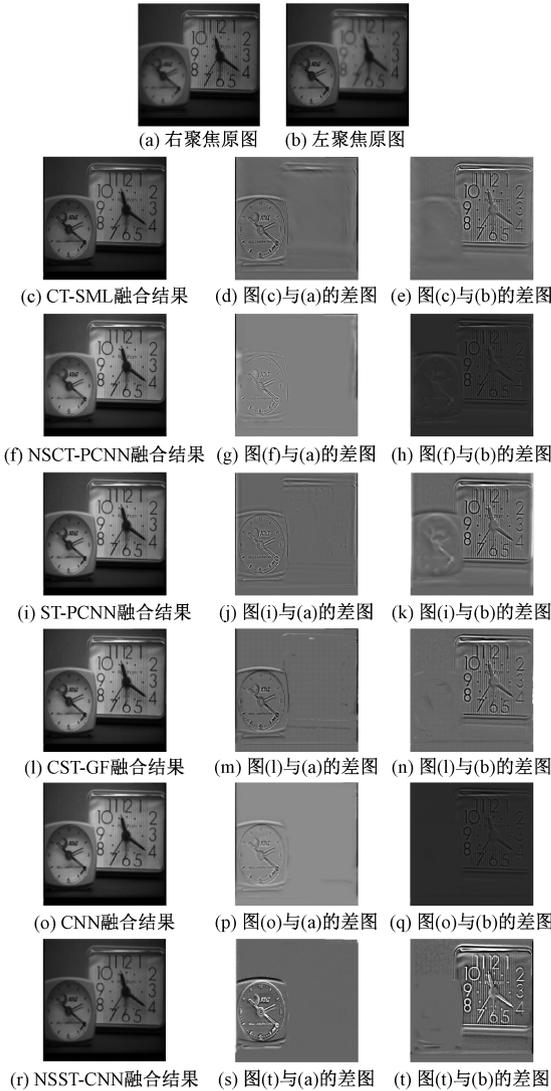


图 3 Clock 原图像和各融合算法的融合效果

Fig.3 Fusion effect of each fused method and Clock original image

在实验结果中,不同算法的融合结果不同,上图的融合结果中大表的左上角部分和大表中数字 8 附近的区域,能够发现本文的算法能够更清晰地表达图像,更符合人眼的视觉特性.与 CNN 相

比,本文的算法在多聚焦图像融合中这些区域稍差一些,但是本文的算法在图像的灰度层次呈现中更为清晰明显,这归功于 NSST 的应用.从上述各种融合图像实验结果与图 3(a) 的差图对比能够发现,本文算法得到融合图像的实验效果图是最好的,这说明了本文的算法在图像的细节信息保留方面是很有效的,这主要归功于 CNN 融合策略的引入.综合上述来看,从不同融合图像差图对比中可以明显地看出在视觉效果方面本文的算法是最好的,人造纹理产生的最少,这主要是由于引入向导滤波使图像的空间连续性得到很大的提升,从而有效地抑制了融合图像中人造纹理的产生.

表 1 给出了各融合算法相应的客观评价指标.由表 1 可知,本文所提出的融合算法在各项评价指标所取得的结果值都是最好的.这表明本文所提出的融合算法可以有效地提取源图像的聚焦信息并且较好地保留空间信息.最后从计算时间上看,虽然在数据的计算时间上本文的算法用时比较长,但是可以看到与其他融合算法相比,本文采用的 NSST 变换具有更好的时频特性,而与 CNN 相比,笔者提出的融合规则能充分地利用图像的统计信息.

然后对空拍图像进行融合测试.图 4(a)、(b) 为不同聚焦图像,各算法融合效果如图 5(c) ~ (t) 所示.

从实验结果中可以发现,融合结果中上下飞机前面和后面邻近的区域,与图 3 中的融合结果视觉效果类似,能够发现本文的算法在图像的融合质量和层次的清晰度表达上,更容易使人眼系统接受到更好视觉的图像信息特征,并且本文算法充分地利用了图像的空间连续性从而克服了其他融合算法在融合图像中引入的一些人造纹理.

同样,表 2 给出了各融合算法的客观评价指标.由表 2 可知,本文算法具有最好的客观评价标准.这也说明本文算法不但可以充分地提取源图

表 1 图 3 中各个融合算法融合效果的客观评价指标

Tab.1 Objective evaluation index of each fusion algorithm in Figure 3

融合方法	Std	Avg	$Q^{AB/F}$	MI	SF	时间/s
CT-SML	41.184 9	0.011 2	0.678 3	6.708 6	8.084 0	<b>43.177 4</b>
NSCT-PCNN	40.614 3	0.011 0	0.677 5	7.494 8	7.909 1	925.464 7
ST-PCNN	40.387 0	0.011 1	0.690 0	6.792 7	8.020 2	3 658.504 7
CST-GF	40.859 5	0.010 9	0.687 5	7.892 5	8.084 2	<b>216.255 0</b>
CNN	40.460 2	0.011 1	0.724 4	8.302 5	8.203 7	784.127 5
本文方法	<b>40.860 2</b>	<b>0.011 2</b>	<b>0.730 6</b>	<b>8.580 8</b>	<b>8.405 7</b>	984.127 5

表2 图4中各个融合算法融合效果的客观评价指标

Tab.2 Objective evaluation index of each fusion algorithm in Figure 4

融合方法	Std	Avg	$Q^{AB/F}$	MI	SF	时间/s
CT-SML	56.74	0.026	0.794	5.35	18.85	43.48
NSCT-PCNN	56.31	0.026	0.794	5.32	18.79	927.49
ST-PCNN	56.07	0.026	0.787	4.74	18.73	166.90
CST-GFF	54.86	0.026	0.779	5.04	18.44	3 602.31
CNN	56.87	0.026	0.798	5.42	18.85	204.66
本文方法	<b>56.88</b>	<b>0.026</b>	<b>0.799</b>	<b>5.57</b>	<b>18.86</b>	216.66

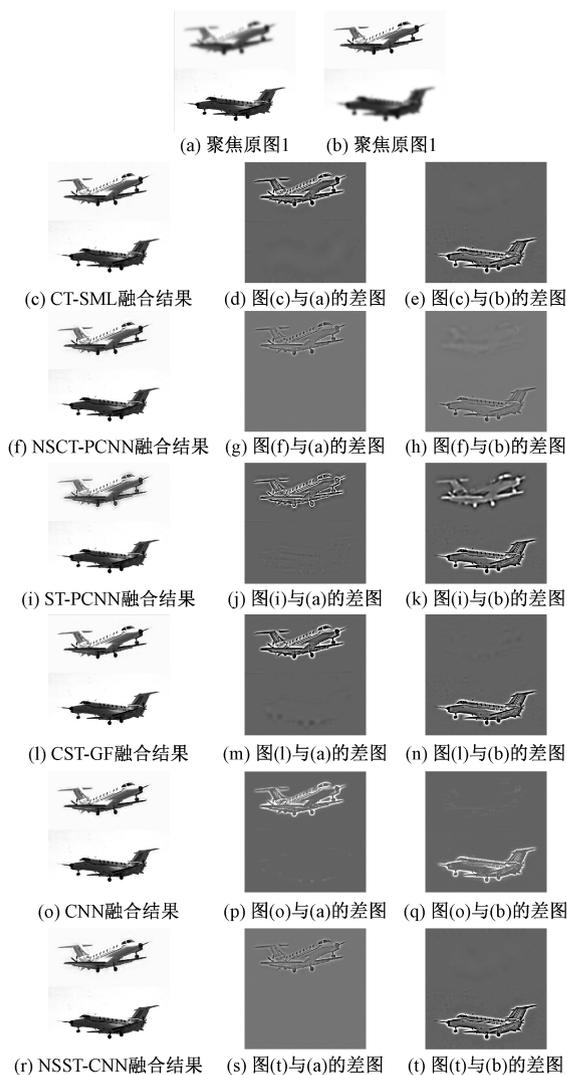


图4 飞机原图像和各融合算法的融合效果

Fig.4 Fusion effect of each fused method and Airplane original image

像中的细节信息,很好地抑制了最终融合图像中人造纹理的产生,并且对于不同种类的图像融合具有鲁棒性,所以综合来看,本文所提出融合规则是一种比较好的在图像融合中的方法。

## 6 结论

笔者提出了一种基于 CNN 的 NSST 变换域图像融合方法.该方法有效利用 NSST 变换域优良

的时频变换特性又充分应用了 CNN 图像融合规则较好的能量保持特点,并通过向导滤波将对融合过程中的高频图像进行平滑处理,进而更好地提高了图像的空间连续性,使图像的融合质量进一步增强.综合上述的实验表明,该方法具有更好的图像特征表达性能,不管是在图像的整体效果还是在客观性能指标上都要更优于文中所描述的目前的图像融合算法。

## 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] LIU S Q, ZHAO J, SHI M Z. Medical image fusion based on improved sum-modified-Laplacian [J]. International journal of imaging systems & technology, 2015, 25(3):206-212.
- [3] LIU S Q, ZHANG T, LI H L, et al. Medical image fusion based on nuclear norm minimization [J]. International journal of imaging systems & technology, 2015, 25(4):310-316.
- [4] GENG P, HUANG M, LIU S, et al. Multifocus image fusion method of ripplelet transform based on cycle spinning [J]. Multimedia tools & applications, 2016, 75(17):10583-10593.
- [5] 屈小波,闫敬文,杨贵德.改进拉普拉斯能量和的尖锐频率局部化 Contourlet 域多聚焦图像融合方法 [J].光学精密工程,2009,17(5):1203-1212.
- [6] QU X B, YAN J W, XIAO H Z, et al. Image fusion algorithm based on spatial frequency-motivated pulse coupled neural networks in nonsubsamped contourlet transform domain [J]. Acta automatica sinica, 2008, 34(12):1508-1514.
- [7] GENG P, WANG Z Y, ZHANG Z G, et al. Image fusion by pulse couple neural network with shearlet [J]. Optical engineering, 2012, 51(6):067005-1-067005-7.
- [8] LIM W Q. The discrete shearlet transform: a new directional transform and compactly supported shearlet

- frames [J]. IEEE transactions on image processing, 2010, 19(5): 1166-1180.
- [9] 石智,张卓,岳彦刚. 基于 Shearlet 变换的自适应图像融合算法[J]. 光子学报, 2013, 42(1): 115-120.
- [10] LIU S Q, SHI M Z, ZHU Z H, et al. Image fusion based on complex-shearlet domain with guided filtering [J]. Multidimensional systems & signal processing, 2017, 28(1):207-224.
- [11] 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.
- [12] 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.
- [13] 李章晓,宋微,田野. 基于深度学习和进化计算的外汇预测与投资组合优化[J]. 郑州大学学报(工学版), 2019, 40(1): 92-96.
- [14] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines[C]//Proceedings of the 27th International Conference on International Conference on Machine Learning. Haifa: Omnipress, 2010: 807-814.
- [15] HE K M, SUN J, TANG X O. Guided image filtering [J]. IEEE transactions on pattern analysis & machine intelligence, 2013, 35(6):1397-1409.

## Multi-focus Image Fusion Based on CNN in Non-sampled Shearlet Domain

LIU Shuaiqi<sup>1,2</sup>, WANG Jie<sup>1,2</sup>, AN Yanling<sup>1,2</sup>, LI Ziqi<sup>1,2</sup>, HU Shaohai<sup>3</sup>, WANG Wenfeng<sup>4</sup>

(1.College of Electronic and Information Engineering, Hebei University, Baoding 071000, China;2.Machine Vision Engineering Research Center of Hebei Province, Baoding 071000, China;3.Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China;4.Digital Image Processing Laboratory, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China)

**Abstract:** In this paper, a new multi-focus image fusion algorithm was proposed based on convolution neural network (CNN) in non-sampled Shearlet (NSST) domain by using the advantages of time-frequency of NSST. Firstly, the source image was decomposed by NSST. Secondly, the fusion strategy based on the convolution neural network was applied to the low frequency coefficients of the decomposition. Then, the improved weighted sum of Laplace energy (IWSML) based on the guided filtering (GF) were carried out to the high-frequency coefficients of the decomposition. Finally, the fused image could be gotten by inverse NSST transform. Experimental results showed that the fusion algorithm could not only achieve better visual effects, but also improve its objective evaluation index.

**Key words:** image fusion; multi-focus image fusion; NSST; CNN; guided filtering