

DOI:10.19651/j.cnki.emt.2313916

# 基于区域聚焦特性的多聚焦图像融合算法<sup>\*</sup>

林 妙 李伟彤

(广东工业大学信息工程学院 广州 510006)

**摘要:**针对现有多聚焦图像融合算法存在的伪影和信息残留问题,提出了一种依据图像聚焦特性,最大程度保留各区域信息和清晰度的算法。首先,通过区域检测得到聚焦区域决策图,利用该决策图进行初始融合和边界提取,得到边界区域决策图;其次,利用ACS网络学习多聚焦图像的融合规则,生成网络融合图;最后,根据边界区域决策图对初始融合图和网络融合图进行加权求和,得到最终的融合图像。实验结果表明:该算法在聚焦区域和边界区域都优于其它比较算法,各项评估指标分别提高4.8%和1.5%以上;同时主观效果更符合HVS。实验证明了在保留源图像的细节信息和避免各个区域的视觉伪影上,该算法都能取得很好的效果。

**关键词:**图像处理;空间域;回归网络;残差空洞卷积金字塔

**中图分类号:** TP391 **文献标识码:** A **国家标准学科分类代码:** 520.6040

## Multi-focus image fusion algorithm based on region focus property

Lin Miao Li Weitong

(School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China)

**Abstract:** To address the problem of artifacts and information residuals in the existing multi-focus image fusion algorithms, an algorithm is proposed to maximize the retention of information and clarity of each region based on the focusing characteristics of the image. Firstly, the focus region decision map is obtained by region detection, which is then used for initial fusion and boundary extraction to obtain the boundary region decision map; secondly, the ACS Network is used to learn the fusion rules of multi-focus images and generate the network fusion image; finally, the initial fusion image and the network fusion image are weighted and summed according to the boundary region decision map to obtain the final fusion image. The experimental results demonstrate that the algorithm outperforms other comparable algorithms in both the focus region and the boundary region, and the evaluation indexes are improved by more than 4.8% and 1.5%, respectively; Meanwhile, the subjective effect is more in line with HVS. The experiments have proved that the algorithm achieves good results in retaining the detailed information of the source image and avoiding visual artifacts in various regions.

**Keywords:** image processing; spatial domain; regression network; residual atrous convolution pyramid

## 0 引 言

由于光学镜头景深的限制,只有处于景深范围内的物体才能呈现聚焦效果,所以采用单一的固定光学镜头无法获取同一场景所有区域的清晰聚焦成像。为了将不同景深的图像融合成一张全聚焦图像,并获取更完整的场景信息,研究者们提出了多聚焦图像融合算法。这些算法主要可以归纳成3类:基于空间域、基于变换域和基于深度学习。

基于空间域的算法可以分为基于块的方法、基于区域的方法和基于像素的方法。基于块的方法<sup>[1]</sup>受块划分大小的影响,容易产生块效应;基于区域的方法<sup>[2]</sup>利用分割技术

对源图像进行区域划分,其性能受到分割技术的限制;基于像素的方法<sup>[3]</sup>则通过聚焦度量评估源图像中像素的显著性,生成逐像素权重图。空间域方法<sup>[4]</sup>能保留大量的空间信息,但图像的活动水平测量和融合规则方面需要人工设计,难以适应所有情况。

基于变换域的多聚焦图像融合算法<sup>[5]</sup>主要根据图像的多尺度特征将其分解为高频和低频成分;然后,针对高频和低频成分分别选择不同的融合规则进行融合;最后,通过逆变换得到最终的融合图像。这类算法没有充分考虑源图像处理过程中可能出现的空间一致性问题,因此,所得到的融

收稿日期:2023-06-16

<sup>\*</sup> 基金项目:广东省科技计划项目(2017A010101016)资助

合图像可能会丢失部分空间信息,并产生不同程度的失真。

深度学习方法一般分为基于分类模型<sup>[6]</sup>和基于回归模型<sup>[7]</sup>,前者将聚焦区域的检测视为一个分类或分割问题,以生成图像融合的决策图;后者采用卷积网络实现从源图像到融合图像的端到端映射。Amin 等<sup>[8]</sup>通过训练网络,确定所输入图像块是聚焦还是散焦,该方法避免了人为活动水平测量的设计,但依赖于图像块的选取,很难对包含部分聚焦区域和散焦区域的边界进行分类。2020 年 Xu 等<sup>[9]</sup>提出了一种无监督的端到端图像融合网络,该算法在聚散焦边界无伪影产生,但在聚焦区域存在信息残留。与传统方法不同,深度学习算法<sup>[10]</sup>通过网络学习形成融合规则,提高了算法的普适性,但存在特征提取尺度单一、特征没有得到合理充分利用等问题。

从光学镜头成像特点出发,本文提出了一种针对区域聚焦特性的多聚焦图像融合算法,该算法将融合进一步细化为聚焦区域融合、空洞卷积孪生(atrous convolutional siamese, ACS)网络融合和最终图像融合。聚焦区域融合通过区域检测生成决策图,进而得到初始融合图。ACS 网络是基于回归模型的端到端映射网络,利用不同膨胀系数的空洞卷积构造残差空洞卷积金字塔模块,以提取多聚焦图像的多尺度特征,同时引入注意力机制对特征进行自适应选择,得到 ACS 网络融合图像。最终,将初始融合图和 ACS 网络融合图合成一张全聚焦图像。通过验证,该方法在多聚焦图像的各区域都取得较好融合效果,解决了空间域方法融合规则人为设计的局限性和深度学习方法容易造成信息残留等问题。

## 1 相关工作

### 1.1 算法分析

通常将多聚焦图像融合算法分为传统方法和深度学习方法,两者并不是完全独立的,存在类比关系:空间域类比分类网络算法,变换域类比回归网络算法。

在空间域<sup>[11]</sup>中依靠测量活动水平生成决策图来提取特征,通常利用最大选择规则进行多聚焦图像融合。在大多数空间域方法中,首先通过聚焦度量得到初始决策图,然后利用一致性验证做进一步处理,以获得更加精确的聚焦区域决策图。深度学习分类方法<sup>[12]</sup>利用网络对源图像聚焦区域进行判定,得到决策图。由于融合图像中的聚焦区域是根据决策图直接从对应的源图像复制而来的,故空间域方法和分类网络方法可以最大程度地保留原始聚焦信息,这有助于保持融合图像的清晰度。然而,直接复制聚焦区域可能在聚焦边界区域引入伪影,对于未精确校准的图像缺乏鲁棒性。

变换域方法<sup>[13]</sup>的处理过程通常包括 3 个步骤:分解、融合和重建,其基本思想是将原始图像的空间信息转换为其它域中的信息进行融合处理。在回归网络中,图像融合<sup>[14]</sup>包括特征提取、特征融合和特征重建 3 个步骤,回归

网络方法与变换域方法具有类似框架。由于变换域方法与人类视觉系统(human vision system, HVS)具有一致的信息处理机制,因此在融合图像时,尤其是处理聚焦/散焦边界区域,变换域方法可以获得更加自然的视觉效果。然而,与源图像相比,变换域方法的融合图像清晰度容易降低,这是因为在分解、融合和重构过程中可能会引入像素强度误差,同样的现象也存在于回归网络模型中。多聚焦图像融合各算法优劣分析如图 1 所示。

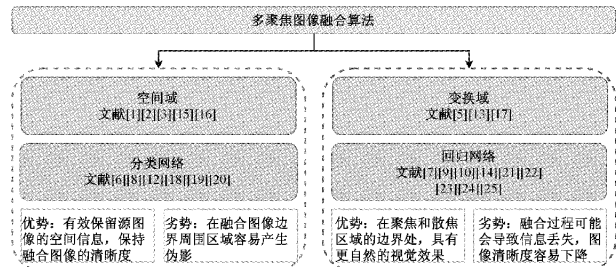


图 1 多聚焦图像融合方法研究现状

### 1.2 聚散焦边界

现有空间域方法和深度学习分类方法难以获得真实、清晰的融合结果,特别是在聚焦/散焦边界(the focus/defocus boundary, FDB)附近区域<sup>[15]</sup>。当前景聚焦而背景散焦时,前景不受到背景散焦影响,如图 2(a)所示,前景和背景在虚线处明确分开。相反,当前景散焦时,散焦扩散效应(the defocus spread effect, DSE)将导致原始模糊区域扩大,如图 2(b)所示,图中虚线显示前景对象的边界,实线显示散焦扩散的实际范围。可以看出:在 FDB 附近,聚焦和散焦区域共存,甚至存在两幅源图像都模糊的区域。

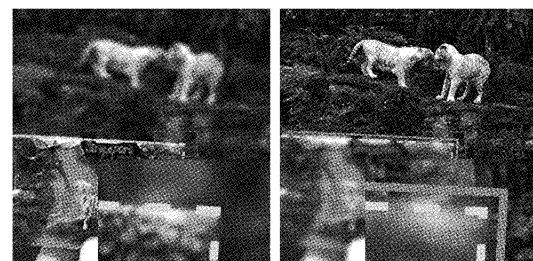


图 2 聚焦不同区域的扩散现象

## 2 本文方法

综合考虑了空间域方法和回归网络在聚焦区域和 FDB 的优劣,本文提出一个空间域与回归网络相结合的图像融合框架,旨在对图像不同区域采用不同的精细处理,以达到最佳融合效果。具体算法为:对于聚焦区域使用决策图进行图像信息提取,对于 FDB 区域采用 ACS 网络进行处理。图 3 为本文方法框架图,整个过程包括 3 个融合部分:1)聚焦区域融合部分:首先利用空间频率实现聚焦区域检测,通过小区域去除生成聚焦区域决策图;然后,利用聚焦区域决策图与源图像对进行初始融合,并采用边界提取

算法生成边界区域决策图;2) ACS 网络融合部分:通过输入源图像对到网络中,生成网络融合图;3)最终图像融合部

分:根据边界区域决策图,将初始融合图和网络融合图进行加权求和,得到最终的融合图像。

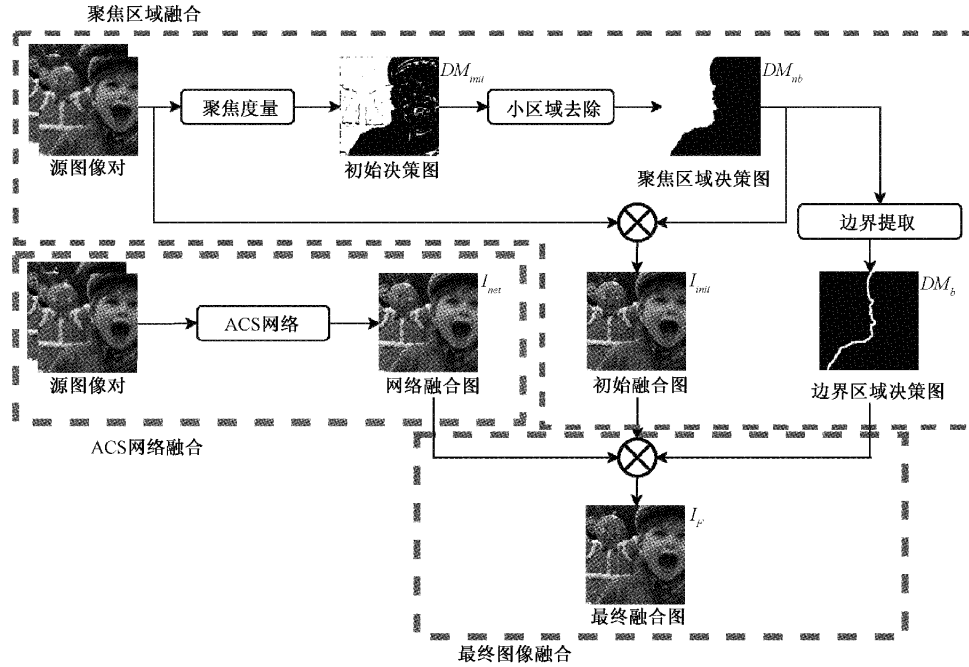


图 3 本文方法框架图

## 2.1 聚焦区域融合

聚焦区域融合包含 4 个步骤:聚焦度量、小区域去除、初始图像融合和边界提取。

步骤 1) 聚焦度量。图像的空间频率 (spatial frequency, SF) 是表征图像中灰度变化剧烈程度的指标,故对多聚焦图像应用 SF 计算像素点  $(m, n)$  的聚焦度量值,记为  $SF(m, n)$ , 并选择最大聚焦度量值的像素作为聚焦像素,具体表示为:

$$RF(m, n) = \sqrt{\sum_{-r \leq a, b \leq r} [F_{m+a, n-b} - F_{m+a, n-b-1}]^2} \quad (1)$$

$$CF(m, n) = \sqrt{\sum_{-r \leq a, b \leq r} [F_{m+a, n-b} - F_{m+a-1, n+b}]^2} \quad (2)$$

$$SF(m, n) = \sqrt{\frac{(RF(m, n))^2 + (CF(m, n))^2}{(2r+1)^2}} \quad (3)$$

其中,  $F$  为输入图像,  $(m, n)$  为输入图像坐标;  $RF$  和  $CF$  分别是行和列的空间频率;  $r$  是核半径,  $r = 5$ 。原始的空间频率是基于块的,而本文运用基于像素的方法,在源图像的四周做了填充,以得到精准的逐像素决策图  $DM_{init}$ , 具体公式如下:

$$DM_{init}(m, n) = SF_A(m, n) - SF_B(m, n) \quad (4)$$

其中,  $SF_A(m, n)$  和  $SF_B(m, n)$  为源图像 A、B 的聚焦度量值。

步骤 2) 小区域去除。初始决策图已经对源图像中明显聚焦区域进行了标记,但在局部区域存在误判。这里采用小区域去除消除误判区域,即通过获取二值图像的连通域个数,计算各连通域的面积,当连通域的面积小于某一

阈值时,对该连通域的聚焦属性值进行修正,得到聚焦区域决策图  $DM_{nb}$ 。

步骤 3) 初始图像融合。根据聚焦区域决策图选取源图像中的聚焦部分,获得初始融合图,具体表示如下:

$$I_{init}(m, n) = \begin{cases} I_A(m, n), & DM_{nb}(m, n) > 0 \\ I_B(m, n), & \text{其他} \end{cases} \quad (5)$$

其中,  $I_{init}$  是初始融合图像,  $DM_{nb}$  是聚焦区域决策图,  $I_A$  和  $I_B$  分别是源图像,  $(m, n)$  为像素位置。

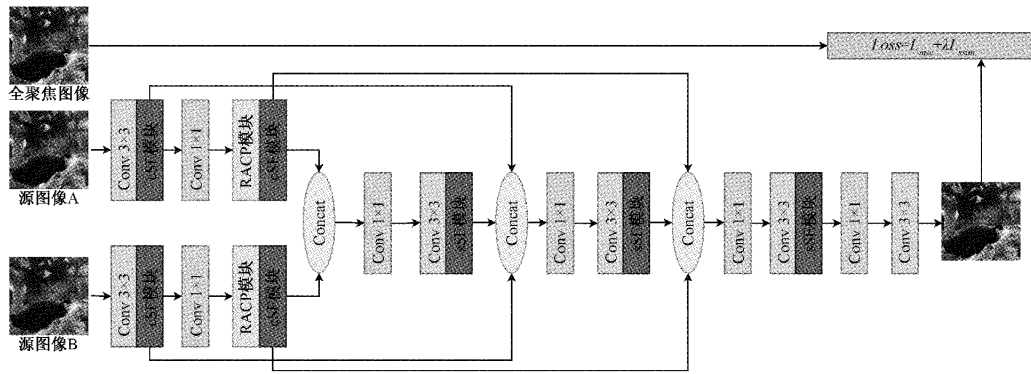
步骤 4) 边界提取。通过对聚焦区域决策图做 Canny 边缘检测,取得边界;再对边界进行扩展和引导滤波得到边界区域决策图  $DM_b$ 。

## 2.2 ACS 网络融合

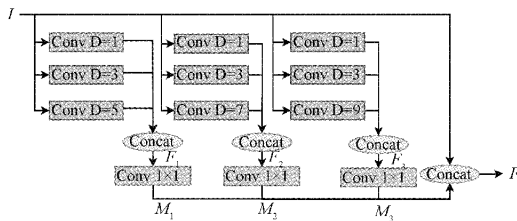
### 1) 网络结构

ACS 网络结构如图 4(a) 所示。从整体上看,该网络为孪生网络结构,上下两个分支共享权重。该网络未使用全连接层,因此对输入图像没有尺寸的限制。除此之外,为尽量保留源图像的信息,整个网络均未使用池化层。为了解决去除池化层带来的参数增加和训练难度增大的问题,在网络中添加  $1 \times 1$  卷积。这是由于  $1 \times 1$  卷积不仅可以减少参数数量,还能增加网络的非线性特性,从而提高网络的理解能力。

该网络由三部分组成:特征提取、特征融合和特征重建。在特征提取部分采用了 cSE (spatial squeeze and channel excitation) 模块、残差空洞卷积金字塔 (residual atrous convolution pyramid, RACP) 模块、 $3 \times 3$  和  $1 \times 1$  两



(a)主要网络



(c)cSE模块

图 4 ACS 网络详解

种卷积。在卷积神经网络所提取的特征图中,不同通道提取不同特征信息,其通道特征的重要程度不一致。cSE 模块可以对多个特征的差异性进行更精确的描述,并传递到后续网络,这有助于不同细节特征的重建,改善图像质量。对于受 DSE 影响的边界区域,RACP 模块能够在更大感受野上获取图像的特征,通过参考边界和聚焦区域的特征信息,摆脱局部特征的限制,从而实现边界区域更加精准的预测。特征融合部分是将两个分支所提取的特征在通道上进行堆叠,从而得到融合特征。特征重建使用 sSE (channel squeeze and spatial excitation)模块将通道信息进行整合,通过参数学习为每个空间特征生成权重,以有效构建局部特征。我们使用 cSE 和 sSE 模块以增强深度特征的鲁棒性和网络的表达能力。此外,本文采用了密集连接架构,分别连接特征提取部分的不同层,加强了特征传播并减少了参数量。在网络训练过程中,通过比较融合图像和标签图像,不断更新参数权重,逐步形成融合规则。

2)RACP 模块

在深度学习网络中,为了增加感受野和降低计算量,通常需要进行池化操作,但这也导致空间分辨率的损失,而空洞卷积能够在不损失信息的情况下扩大感受野<sup>[21]</sup>。因此,本文引入了 RACP 模块,以更好地提取图像特征,结构如图 4(b)所示。

RACP 模块包括 3 个卷积组,每个卷积组由不同膨胀系数的空洞卷积构成,用于提取同一输入下不同感受野的图像特征。模块首先将不同感受野的图像特征级联降维,然后将 3 个卷积组降维后的特征与输入进行级联。具体公式表示如下:

$$\begin{aligned}
 F_1 &= Concat(I * Conv_{D=1}, I * Conv_{D=3}, I * Conv_{D=5}) \\
 F_2 &= Concat(I * Conv_{D=1}, I * Conv_{D=3}, I * Conv_{D=7}) \\
 F_3 &= Concat(I * Conv_{D=1}, I * Conv_{D=3}, I * Conv_{D=9})
 \end{aligned} \tag{6}$$

$$M_i = F_i * Conv_{1 \times 1} \tag{7}$$

式 (6)中  $F_i$  是  $I$  经过第  $i$  组空洞卷积提取的所有特征,  $i = \{1, 2, 3\}$ ,  $I$  为输入特征集,  $Conv_{D=d}$  是膨胀系数为  $d$  的空洞卷积,  $*$  表示卷积运算,  $Concat(\cdot)$  表示级联运算。式 (7)中  $M_i$  是  $F_i$  进行  $1 \times 1$  卷积降维后的特征。最终输出特征  $F$  如下:

$$F = Concat(M_1, M_2, M_3, I) \tag{8}$$

3)cSE 和 sSE 模块

SE(squeeze and excitation)<sup>[26]</sup>中的 Squeeze 是指利用全局平均池化操作来提取全局感受野的信息,将每个特征通道都抽象为一个点。Excitation 通过利用两层的多层感知机网络来进行非线性的特征变换,构建特征图之间的相关性。

如图 4(c)所示,在 cSE 中,首先令输入特征图  $U = \{u_k(m, n) | k = 1, 2, \dots, C; m = 1, 2, \dots, H; n = 1, 2, \dots, W\}$ , 将其进行全局平均池化为  $z, z = (z_1, z_2, \dots, z_C)^T$ ; 然后使用两个全连接层 ( $W_1$  和  $W_2$  为权重)和 ReLU 激活函数  $\delta(\cdot)$  进行信息处理,得到  $C$  维向量  $\hat{z} = (\hat{z}_1, \hat{z}_2, \dots, \hat{z}_C)^T$ ; 对其使用 sigmoid 函数  $\sigma(\cdot)$  进行归一化,最后通过相乘,得到经过信息校准过的特征图  $\hat{U}_{cse}$ , 相关公式如式(9)所示。

$$z_k = \frac{1}{H \times W} \sum_{m=1}^H \sum_{n=1}^W u_k(m, n)$$

$$\hat{z} = W_1(\delta(W_2 z))$$

$$\hat{U}_{cSE} = F_{cSE}(U) = \{\sigma(\hat{z}_k)u_k(m, n) \mid k = 1, 2, \dots, C\} \quad (9)$$

在 sSE 中,首先输入特征图  $U$ , 然后通过  $1 \times 1 \times C$  卷积进行通道压缩生成空间映射张量  $Q = \{q(m, n) \mid m = 1, 2, \dots, H; n = 1, 2, \dots, W\}$ ; 并使用 sigmoid 函数  $\sigma(\cdot)$  激活得到空间权重图,最后通过特征图与权重图逐点相乘,得到经过信息校准过的特征图,如图 4(d)。公式具体表示为:

$$\hat{U}_{sSE} = F_{sSE}(U) = \{\sigma(q(m, n))u_k(m, n) \mid m = 1, 2, \dots, H; n = 1, 2, \dots, W\} \quad (10)$$

### 2.3 最终图像融合

利用初始融合图像  $I_{init}$  和网络融合图像  $I_{net}$ , 根据边界

区域决策图进行加权求和,得到最终融合图像  $I_F$ , 具体表示为:

$$I_F = DM_b \times I_{net} + (1 - DM_b) \times I_{init} \quad (11)$$

其中,  $DM_b$  是边界区域决策图。

## 3 实验与讨论

### 3.1 训练数据集

有监督网络训练需要大量数据,但目前缺少带标记的、真实的数据集,为此构建了一个多聚焦图像数据集。首先,从 VOC2007 中选取 2 000 张图像,将其裁剪成  $256 \times 256$  的标签图像;然后对每张标签图像划分为  $8 \times 8$  块(每块大小为  $32 \times 32$ );其次,对划分块进行梅花采样高斯模糊,模糊半径为 3;最后,对互补区域进行同样的处理<sup>[23]</sup>。数据集效果如图 5,其中第 1 行为标签图像,第 2、3 行为模糊区域互补的多聚焦图像。



图 5 训练数据集

### 3.2 网络设置

考虑到输出图像应符合 HVS 且接近真实图像,我们使用结构相似性损失  $L_{SSIM}$  和均方误差损失  $L_{MSE}$  相结合来构造损失函数  $L$ , 如式 (12), 其中  $\lambda$  是 MSE 的权重, 在实验中  $\lambda = 10$ 。

$$L = L_{SSIM} + \lambda L_{MSE} \quad (12)$$

$$L_{SSIM} = 1 - SSIM$$

$$SSIM = \frac{2\mu_O\mu_T + C_1}{\mu_O^2 + \mu_T^2 + C_1} \cdot \frac{2\sigma_O\sigma_T + C_2}{\sigma_O^2 + \sigma_T^2 + C_2} \cdot$$

$$\frac{\sigma_{O,T} + C_3}{\sigma_O^2\sigma_T^2 + C_3} \quad (13)$$

$$L_{MSE} = \frac{1}{W \times H} \sum_{m=1}^W \sum_{n=1}^H [O(m, n) - T(m, n)]^2 \quad (14)$$

式中:  $O$  和  $T$  分别代表网络融合图像和目标图像,  $\mu_O$  和  $\mu_T$  代表图像的均值,  $\sigma_O$  和  $\sigma_T$  代表图像方差,  $\sigma_{O,T}$  为两图像的协方差,  $C_1, C_2, C_3$  为常数, 用于避免分母为 0。式 (14) 中  $W$  和  $H$  为图像的宽和长,  $(m, n)$  为像素位置。在网络的训练过程中, 本文使用 Adam 作为优化函数, 学习率为 0.000 5, 批处理大小为 4。

### 3.3 实验设置

为了验证本文模型的有效性, 将所提出的融合方法与

8 种代表性方法进行了比较, 这 8 种方法是基于寻找边界的方法<sup>[15]</sup> (BF)、基于引导滤波的方法<sup>[16]</sup> (GFDF)、基于卷积神经网络的方法<sup>[19]</sup> (CNN)、基于分类模型的方法<sup>[20]</sup> (ECNN)、基于回归模型的方法<sup>[22]</sup> (IFCNN)、基于密集连接和 Siamese 结构结合的方法<sup>[23]</sup> (SF\_MFIF)、基于无监督网络的方法<sup>[24]</sup> (SESF) 和基于端到端网络的方法<sup>[25]</sup> (GACN), 并从定性和定量两方面对算法进行了评估。为了评估所提出方法的性能, 使用 28 对公共多聚焦图像作为测试图像, 其中 20 对取自 Lytro 数据集<sup>[17]</sup>, 其余 8 对是被广泛应用的多聚焦图像, 图 6 为实验所用到的部分多聚焦测试图像。

### 3.4 主观评估

不同方法对“Book”图像融合结果如图 7 所示, 融合结果中存在差异的位置用矩形标出, 并进行了局部放大。图 7(a) 中, 书角模糊且存在 DSE, 但远离边界的背景部分是清晰的; 图 7(b) 中, 书角清晰而背景模糊; 图 7(c)~(e) 背景标题清晰, 但书角模糊, 没有很好地保留图 7(b) 中的信息; 在图 7(f)、(g)、(i)、(j) 中, 书角和背景之间的边界并不清晰, 甚至在图 7(f) 中出现小块的错误识别。相比其它算法结果, 图 7(h) 和 (k) 具有更加清晰的边界, 书角和背景部分都很清晰。



图 6 部分多聚焦测试图像

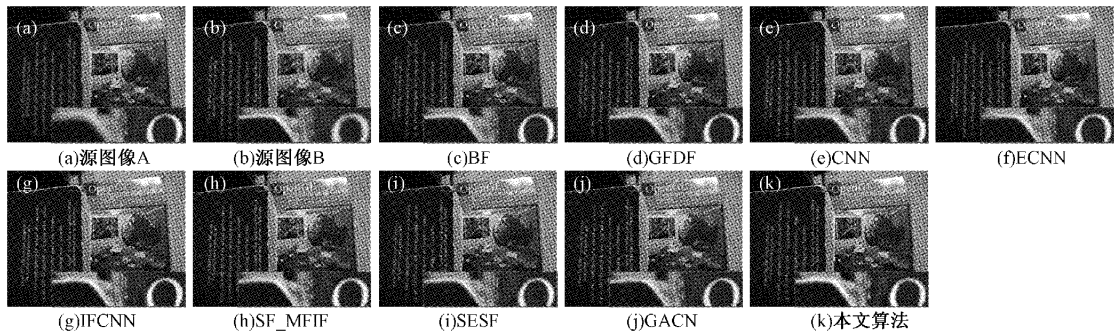


图 7 不同方法对“Book”图像融合结果

图 8 为“Book”融合结果与源图像的差分图像,可以看到:图 8(e)、(f)中存在残留的英文字母信息,而其它方法在聚焦区域上表现良好。通过观察图 7 和 8 的 FDB 区域,可以看出本文方法的差分图像边界更明显,

在融合图上能更容易分辨书角边缘,且在边界区域前景和背景都清晰。这是因为本文方法对源图像的边界区域采用针对性的预测方法,而非直接对源图像进行加权求和。

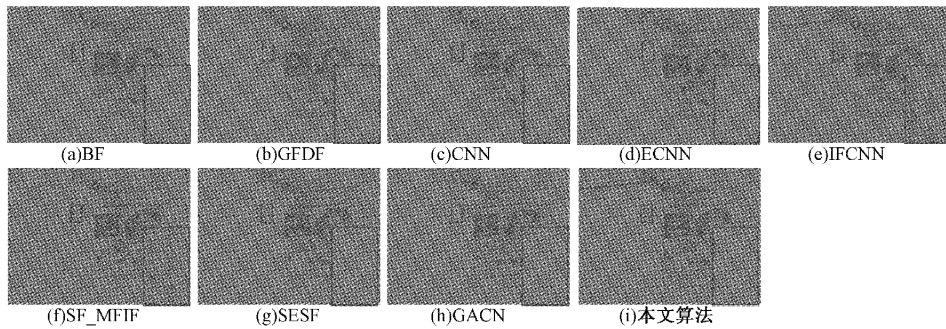


图 8 “Book”融合结果与源图像的差分图像

图 9 是 Lytro 图像集中“Children”的融合结果,图 10 是图 9 融合结果与源图像的差分图像。从小男孩的帽子部分可以看出,图 9(c)、(d)、(i)融合图像的边界模糊严重,图 9(e)、(j)在融合边界出现视觉伪影,图 9(f)在边界区域存在误判导致融合边界畸变,图 9(g)、(h)、(k)在边界处清晰且不存在伪影。通过观察差分图像,可以发现:在图 10(e)、(f)中,图像背景部分和源图像存在明显细节特征的差异,在相应位置的融合图像中可以观察到玩偶衣襟部分的伪影和像素扩散;图 10(a)~(d)、(g)~(i)在聚焦的平坦区域无残留信息,即融合图像完整保留了源图像该区域的纹理细节。总体而言,本文方法所获得的融合图像具有较好的视觉效果,能够保留细节信息,且在边界处没有

出现伪影和畸变等现象。

### 3.5 客观评估

为了对融合图像进行全面评估,采用客观评价指标对融合结果进行定量分析。不同的多聚焦图像融合算法在不同的评估指标上可能存在较大差异,为了减少单个指标的片面性,采用基于信息论的度量 EN 和 MI、基于图像特征的度量 STD、基于图像结构相似性的度量  $Q_Y$  和基于人类感知的度量 VIF 多种指标评价融合图像的性能。表 1~4 中每一项指标中加粗值表示最优指标,加下划线值表示次优指标。

表 1 为所选 20 幅 Lytro 融合图像各评价指标的平均值。由表 1 可知,对于  $Q_Y$ , 算法 BF、GFDF 优于本文方法;

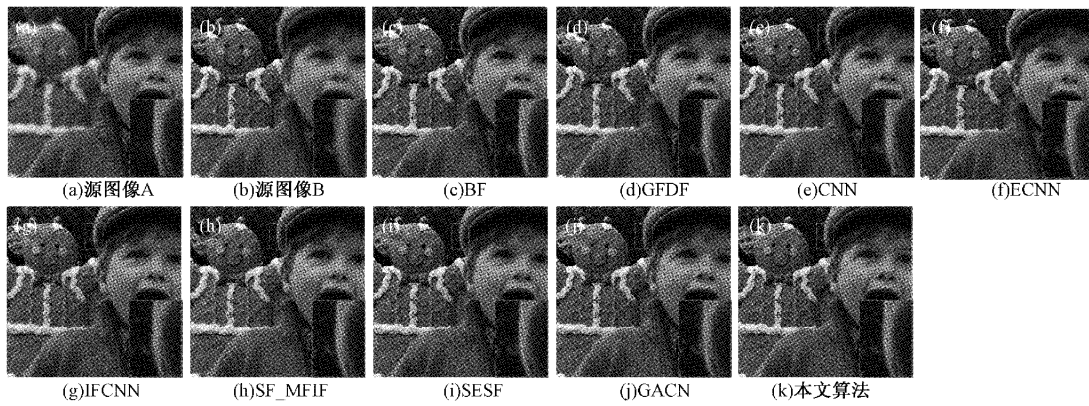


图 9 不同方法对“Children”图像融合结果

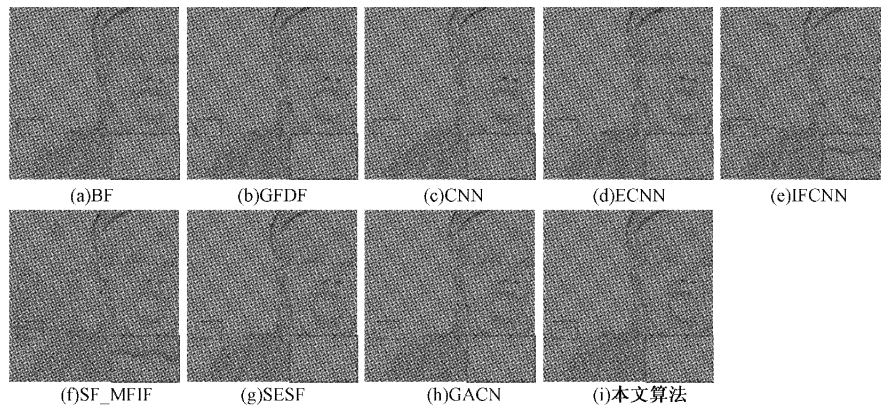


图 10 “Children”融合结果与源图像的差分图像

表 1 Lytro 图像集各评价指标的平均值

Algorithm	EN	MI	STD	$Q_Y$	VIF
BF	7.528 6	15.057 1	57.304 6	<b>0.989 3</b>	1.118 0
GFDF	7.530 9	15.061 9	57.532 6	0.988 3	1.135 3
CNN	7.530 4	15.060 7	57.468 7	0.987 1	1.128 9
ECNN	7.531 2	15.062 4	57.551 5	0.987 5	1.135 7
IFCNN	7.531 9	15.063 9	57.550 3	0.951 1	1.133 9
SF_MFIF	7.528 4	15.056 9	57.646 1	0.933 0	<b>1.174 9</b>
SESF	7.532 3	15.064 6	57.593 5	0.986 8	1.141 5
GACN	7.531 3	15.062 6	57.486 7	0.986 7	1.133 0
Proposed	<b>7.539 8</b>	<b>15.079 6</b>	<b>57.705 9</b>	0.981 6	1.163 1

表 3 Lytro 图像集在聚焦区域的评价指标

Focused region	FMI <sub>w</sub>	PSNR	SSIM	VIF
BF	0.554 6	47.443 8	0.812 0	0.973 5
GFDF	0.554 9	47.427 3	0.812 0	0.973 8
CNN	0.554 1	47.466 4	0.812 1	0.973 6
ECNN	0.560 3	47.321 6	0.812 0	0.973 8
IFCNN	0.436 8	46.825 2	0.815 4	0.967 5
SF_MFIF	0.444 4	47.926 8	0.814 3	0.990 0
SESF	0.560 0	47.305 4	0.812 1	0.973 7
GACN	0.553 0	47.344 2	0.812 2	0.973 9
Proposed	<b>0.636 7</b>	<b>49.903 0</b>	<b>0.855 3</b>	<b>1.082 7</b>

表 2 其余 8 幅测试图像各评价指标的平均值

Algorithm	EN	MI	STD	$Q_Y$	VIF
BF	7.161 2	14.322 5	50.004 0	0.922 8	0.731 8
GFDF	7.164 6	14.329 2	49.900 0	0.972 8	0.837 7
CNN	7.155 2	14.310 4	49.980 3	0.913 1	0.729 1
ECNN	7.165 8	14.331 7	50.197 4	<b>0.973 9</b>	0.843 1
IFCNN	7.159 2	14.318 3	49.890 3	0.878 8	0.874 6
SF_MFIF	7.182 1	14.364 2	50.540 2	0.878 5	0.893 1
SESF	7.171 2	14.342 4	50.370 5	0.968 9	0.854 2
GACN	7.166 2	14.332 4	<b>51.179 2</b>	0.966 8	0.867 5
Proposed	<b>7.188 4</b>	<b>14.376 8</b>	51.176 8	0.959 9	<b>0.914 5</b>

表 4 Lytro 图像集在边界区域的评价指标

Boundary region	AG	EI	EAV	LIF
BF	1.310 3	14.060 4	9.741 6	0.041 0
GFDF	1.319 8	14.148 8	9.819 8	0.040 1
CNN	1.297 8	13.933 5	9.622 3	0.040 4
ECNN	1.341 2	14.336 7	10.047 2	0.040 1
IFCNN	1.371 8	14.666 2	10.295 1	0.039 9
SF_MFIF	1.392 6	14.879 8	10.446 1	0.039 4
SESF	1.355 1	14.490 0	10.140 8	0.039 6
GACN	1.338 0	14.318 9	10.006 1	0.040 3
Proposed	<b>1.417 0</b>	<b>15.103 6</b>	<b>10.644 4</b>	<b>0.037 4</b>

对于 VIF,算法 SF\_MFIF 略优于提出方法。除此之外,对于其他指标,本文的方法均优于其他对比算法。表 2 为其余 8 幅融合图像每个指标的平均值,可以看出:对于 STD,本文方法仅劣于 GACN;对于  $Q_V$ ,本文方法处于中等水平,其余指标均最优。虽然在结构相似性指标上表现一般,但对于基于信息论、图像特征、人类特征感知等指标,本文的方法都基本优于其他算法。

为了进一步对图像清晰度进行定量分析,利用图 3 中边界区域决策图  $DM$ ,和各算法融合结果加权求和,获取聚焦区域和边界区域,并计算相应区域的评价指标。对于聚焦区域,本文采用基于小波变换的特征互信息  $FMI_w$ 、峰值信噪比 PSNR、结构相似性 SSIM 和基于人类感知的度量 VIF 指标,通过评估融合图像从源图像中获取的信息量来衡量两者之间的相似程度。对于边界区域存在源图像对都不清晰的情况,采用无参考的评价指标,包括平均梯度 AG、边缘强度 EI、点锐度算法 EAV 和线性模糊指数 LIF 来衡量边界区域的融合效果。表 3 和 4 分别是 Lytro 数据集在聚焦区域和边界区域下的客观性能比较,可以看出:本文所提出的算法在聚焦区域和边界区域,其客观指标均优于其他算法。因此本文方法的总体性能优越。

综合以上主、客观评价可知,本文方法在保留多聚焦图像细节和纹理信息方面表现出色,同时在边界区域融合更精准,总体上提升了图像的清晰度。

#### 4 结 论

针对多聚焦图像聚焦区域信息残留和边界区域散焦扩散的问题,本文提出了一种针对区域聚焦特性的多聚焦图像融合算法。通过对空间域方法和回归网络方法进行分析,综合两种方法在聚焦区域和边界区域的优势,生成高质量的融合图像。实验结果验证了本文方法能够有效保留源图像中的细节和纹理信息,且能在聚焦区域和边界区域同时避免视觉伪影的产生,从而进一步提高多聚焦融合图像的清晰度。边界的准确性对最终融合结果有很大影响,由于本文是根据聚焦度量结果得到边界,没有对源图像对的边界分别进行细化提取,而源图像对的边界未必完全重合,所以对多聚焦源图像边界不一致情况的处理是未来要解决的问题。

#### 参考文献

- [1] KAUSAR N, MAJID A, JAVED G S. A novel ensemble approach using individual features for multi-focus image fusion[J]. Computers and Electrical Engineering, 2016, 54: 393-405. DOI: 10.1016/j.compeleceng.2016.01.013.
- [2] HUANG Y, LI W, GAO M, et al. Algebraic multi-grid based multi-focus image fusion using watershed algorithm[J]. IEEE Access, 2018, 6:47082-47091.
- [3] CHEN Y, GUAN J, CHAM W K. Robust multi-focus image fusion using edge model and multi-matting[J]. IEEE Transactions on Image Processing, 2017, 27(3): 1526-1541.
- [4] ZHANG X. Deep learning-based multi-focus image fusion: A survey and a comparative study[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(9):4819-4838.
- [5] 姚海, 秦华旺. 基于 NSCT 的焦点度量优化和 ISML 多聚焦图像融合[J]. 电子测量技术, 2022, 45(4): 85-90.
- [6] ZHAO W, YANG H, WANG J, et al. Region- and pixel-level multi-focus image fusion through convolutional neural networks[J]. Mobile Networks and Applications, 2021, 26(1):40-56.
- [7] MUSTAFA H T, YANG J, ZAREAPOOR M. Multi-scale convolutional neural network for multi-focus image fusion[J]. Image and Vision Computing, 2019, 85:26-35.
- [8] AMIN-NAJI M, AGHAGOLZADEH A, EZOJI M. CNNs hard voting for multi-focus image fusion[J]. Journal of Ambient Intelligence and Humanized Computing, 2020, 11(4):1749-1769.
- [9] XU H, MA J, JIANG J, et al. U2Fusion: A unified unsupervised image fusion network [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(1):502-518.
- [10] 尹海涛, 周伟. 基于多尺度空洞卷积网络的多聚焦图像融合算法[J]. 激光与光电子学进展, 2023, 60(2): 80-89.
- [11] LI S, KANG X, FANG L, et al. Pixel-level image fusion: A survey of the state of the art [J]. Information Fusion, 2017, 33:100-112.
- [12] YANG Y, NIE Z, HUANG S, et al. Multilevel features convolutional neural network for multifocus image fusion[J]. IEEE Transactions on Computational Imaging, 2019, 5(2):262-273.
- [13] 杨建翠, 马庆功. 基于非下采样 Shearlet 变换耦合导向法则的多聚焦图像融合算法[J]. 电子测量与仪器学报, 2020, 34(3):36-42.
- [14] 李春艳, 孙韬, 谢俊峰. EMF 深度学习可见光/红外图像融合算法[J]. 国外电子测量技术, 2020, 39(10): 25-32.
- [15] ZHANG Y, BAI X, WANG T. Boundary finding based multi-focus image fusion through multi-scale morphological focus-measure[J]. Information fusion, 2017, 35:81-101.
- [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, DOI: 10.1016/j.image.2018.12.004.
- [17] NEJATI M, SAMAVI S, SHIRANI S. Multi-focus image fusion using dictionary-based sparse representation[J]. *Information Fusion*, 2015, 25:72-84, DOI: 10.1016/j.inffus.2014.10.004.
- [18] MA H, LIAO Q, ZHANG J, et al. An  $\alpha$ -matte boundary defocus model-based cascaded network for multi-focus image fusion[J]. *IEEE Transactions on Image Processing*, 2020, 29:8668-8679.
- [19] 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.
- [20] AMIN-NAJI M, AGHAGOLZADEH A, EZOJI M. Ensemble of CNN for multi-focus image fusion[J]. *Information fusion*, 2019, 51:201-214.
- [21] JIANG L, FAN H, LI J, et al. Pseudo-Siamese residual atrous pyramid network for multi-focus image fusion[J]. *IET Image Processing*, 2021, 15(13): 3304-3317.
- [22] 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.
- [23] LI H, ZHANG L, JIANG M, et al. Multi-focus image fusion algorithm based on supervised learning for fully convolutional neural network[J]. *Pattern Recognition Letters*, 2021, 141:45-53.
- [24] MA B, ZHU Y, YIN X, et al. SESF-Fuse: an unsupervised deep model for multi-focus image fusion[J]. *Neural Computing and Applications*, 2021, 33:5793-5804, DOI: 10.1007/s00521-020-05358-9.
- [25] MA B, YIN X, WU D, et al. End-to-end learning for simultaneously generating decision map and multi-focus image fusion result[J]. *Neurocomputing*, 2022, 470:204-216.
- [26] HU J, SHEN L, SUN G. Squeeze-and-Excitation Networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 42(8):7132-7141.

### 作者简介

林妙, 硕士研究生, 主要从事图像融合方面的研究。

E-mail: linmiao1701@163.com

李伟彤(通信作者), 博士, 副教授, 硕士生导师, 主要从事图像融合、图像质量评估等方面的研究。

E-mail: liweitong@gdut.edu.cn