DOI:10.11721/cqnuj20220515

多尺度特征的深度学习及在多聚焦图像融合上的应用*

高文昌,余 磊
(重庆师范大学 计算机与信息科学学院,重庆 401331)

摘要:【目的】多聚焦图像融合指的是从同一场景下不同的图像中提取各自的聚焦区域,得到一幅全聚焦的图像,是近些年来图像处理领域一个热门的研究方向。传统的图像融合技术存在融合区域不清晰、失真、存在伪影等情况。针对这一现象,提出了一种基于深度学习的图像融合方法。【方法】所提算法整体使用了孪生网络来对图像中的聚焦区域进行分类,同时还引入了 GoogLeNet 中的 Inception 模块来提高网络的特征提取能力,取得了良好的性能。为了充分利用源图像中的特征信息,提出的算法中使用了不同大小的子块来提取源图像中不同尺度的特征信息,获得源图像多个尺度的特征。此外,提出的方法获得的二值图能够精确反映出源图像的聚焦区域和非聚焦区域,因此不需要应用后处理步骤来对二值图进行优化,降低了网络的复杂度。【结果】在 Lytro 多聚焦图像集和其他常用的多聚焦灰度图像集上的实验结果表明:相比于其他经典算法,提出算法的融合结果从主观和客观两个维度上都拥有显著的优势。【结论】提出的算法很好地融合了源图像中的细节特征,融合边缘自然平滑、无伪影产生,取得了较传统算法更好的融合效果。

关键词:孪生网络;Inception 模块;多尺度特征;多聚焦图像融合
中图分类号:TP301.6 **文献标志码**:A **文章编号**:1672-6693(2022)05-0118-09

在现代数字设备的成像中存在着一些问题,比如数码摄影拍出来的照片只有在关心的目标区域清晰可见,在其他区域则会模糊。这是因为数字设备原件会受到景深空间范围的影响,只有在景深的物体区域才会展现出清晰的效果。因此在数字图像处理领域,把多幅不同区域聚焦的图像融合得到一幅全聚焦的图像是一个热门的研究方向。近些年来,很多不同的融合方法被提出,这些方法可以从不同的角度实现多聚焦图像融合,根据这些方法的融合特点可以将它们分为两类[1],分别是基于空间域的融合方法和基于变换域的融合方法[2]。

基于空间域的方法指的是利用图像的强度来估计图像的聚焦特性。基于空间域的方法还可以进一步的分为基于区域的方法、基于像素的方法、基于块的方法。基于像素的方法获得的融合结果相比前两种方法来说要好,在研究中比较受欢迎,这些基于像素的方法包括梯度能量[3]、图像消光[4]等。早期的图像融合方法一般都是基于块的方法,其中融合结果直接受到块的大小的影响。基于块的融合方法的代表有基于形态学的焦点测量方法[5]、基于差分进化的算法[6]等。这些算法根据源图像中各种不同的特征,把源图像划分为很多块,然后对每一块进行聚焦测量筛选聚焦的块,最后利用这些块获得一幅聚焦清晰的图像。基于图像块的方法在聚焦和非聚焦的边界区域一般会产生伪影和模糊的效果。

基于变换域的方法的基本思想是将源图像转换到另外一个特征域中,然后在特征域中对图像进行处理,然后再转换为正常的图像。通常基于变换域的方法主要有 3 个过程:分解、融合、重建[7]。基于变换域的方法一般采用的是多尺度变换和离散余弦变换。这些方法有拉普拉斯金字塔(LP)[8]、基于离散小波变换(DWT)的方法[9]、基于非子采样轮廓波变换(NSCT)的方法[10]、基于双树复小波变换(DTCWT)的方法[11]等。这些方法在图像融合的细节和质量方面效果表现良好,但是这些方法的活动水平测量和融合规则都需要进行人工设计,而人工设计的难度很大,不能把所有的因素考虑在内,耗时耗力。

最近,深度学习在各个领域的发展十分快速,在数字图像领域,深度学习的作用越来越明显。CNN 是深度学习中常用的算法之一,能够学习图像中不同程度的信息和特征。Ma 等人[12]使用无监督的方式训练编码器-解

* **收稿日期**:2021-08-10 **修回日期**:2021-09-18 **网络出版时间**:2022-09-19 14:22
资助项目:国家自然科学基金(No. 72071019);重庆市自然科学基金(No. cstc2020jcyj-msxmX0068;No. cstc2021jcyj-msxm2791);成渝双城经济圈科技创新项目(No. KJCX2020024);重庆市自然科学基金面上项目(No. cstc2021jcyj-msxm2791)
第一作者简介:高文昌,男,研究方向为图像处理,E-mail:2019110516007@stu.cqnu.edu.cn;**通信作者**:余磊,男,教授,博士,E-mail:ylcqnu@163.com
网络出版地址:https://kns.cnki.net/kcms/detail/50.1165.n.20220916.1806.013.html

码器网络来获取图像的特征，然后用一种梯度的方法测量活动水平。Tang 等人[13]开发了一种高效 CNN 方法来识别聚焦和散焦区域，他们创建了基于公共图像数据库的综合训练图像集，生成的多焦点图像数据集使用了 12 种几何类型的遮罩考虑了许多不同的聚焦条件。但是，对于多聚焦图像融合，使用如此多的遮罩是不需要的，它增加了工作量。Zhang 等人[14]提出了一个通用的多曝光、多模态医学、红外视觉和多聚焦等图像融合模型。该模型有较好的泛化能力，但是该模型得到的融合图像相对专业领域内的结果并不是很好。Yang 等人[15]，提出了一个多级特征卷积神经网络，他们添加了一个 1×1 的卷积模块以减少冗余，融合的图像是利用带有决策图的权重方法生成的。Liu 等人[16]是最先把卷积神经网络应用到图像融合领域的研究人员，他们把确定源图像中的聚焦和非聚焦区域视为一个二分类问题，通过训练神经网络使神经网络能够区分聚焦像素和散焦像素。该算法很好地融合了图像，但是他们的算法使用子块的大小固定为 16×16，并不能准确的预测聚焦区域的分类。为了能够提高网络性能，本文提出一种多尺度卷积神经网络的方法：首先，提取不同尺度子块的特征信息，再对提取的不同子块的特征信息进行特征融合得到全局特征信息，这对提升网络的性能有很大的帮助。同时，我们把子块的最低大小缩减为 8×8，这使得子块中同时含有聚焦区域和离焦区域的概率变小，因此可以更精确的对聚焦区域进行分类，网络中还使用 GoogLeNet 中的结构来增强网络的性能。所提出的方法能够对像素精确分类，能够得到非常好的融合结果。

1 算法的提出

本文提出了一种基于深度卷积神经网络的图像融合方法。把多聚焦图像对输入到所提出的网络结构中得到焦点图，然后对焦点图做二值映射操作，处理后的结果和源图像采用像素加权平均规则计算融合图像。GoogLeNet 是一个深度网络，在分类方面具有出色的性能。对于深度学习来说，当网络结构变深时，性能会有效提升。同时，当网络变宽时也会提升性能。在深度方面，GoogLeNet 增加了两个 loss 函数防止梯度在传输的过程中消失。在宽度方面，采用了 Inception 结构增加网络的宽度。因此，为了提升网络的性能，提出的方法中并入了 Inception 结构，使整个网络宽度增加，提升网络在多尺度方面提取特征的能力。除了采用上述的 Inception 模块外，还使用了多尺度来提取图像中的高频和低频特征，通过对高频特征和低频特征的融合能够进一步提升网络的性能，实现像素的精确分类。

1.1 网络结构

本文设置的网络结构如图 1 所示。从整体上看，提出的网络结构是一个孪生网络，上下两个网络训练得到的权重一样。为了能够使网络能够精确分类聚焦像素和散焦像素，本文提出了多尺度的方法。在特征提取阶段，获取了图像多个尺度的信息，子块大小为：32×32，16×16 及 8×8，不同大小的子块作为网络的输入可以充分利用源图像中的多尺度信息来对源图像进行精确分类。此前，基于分类模型的算法把子块大小最低缩减为 16×16，这使得子块中同时含有聚焦区域和离焦区域的概率增大，不利于网络的分类，提出的算法把子块大小最小缩小为 8×8，这使得子块中同时包含聚焦区域和离焦区域的概率降低，增加分类的精度。此外，算法的网络结构中使用了 GoogLeNet 中的 Inception 模块来增加网络的性能(图 2)。在 Inception 模块中使用了 3 种卷积核：3×3，5×5 和 1×1，其中 3×3 卷积核和 5×5 卷积核也起到了多尺度提取图像特征的作用；1×1 的卷积核可以对特征进行降维，在解决计算瓶颈的同时也可以限制网络的参数大小，可以将网络做的更深。

对于 P_{11} 子块，在经过 3×3 卷积核进行卷积之后得到了 64 个特征图，然后经过 3×3 的卷积核得到了 128 个特征图，最后经过 3×3 的卷积核得到 256 个特征图，整个过程中使用了 padding 进行填充且未使用池化处理，所以特征图的大小没有改变。

对于 P_{12} 子块，在进行 Inception 模块的操作之后得到了大小为 16×16 的 480 个特征图，其中在 Inception 模块中也使用了 padding 进行填充。然后再经过大小为 2×2 的最大池化，得到 8×8 大小的 480 个特征图。

对于 P_{13} 子块，经过第一个 Inception 模块以及进行最大池化操作之后得到大小为 16×16 的 256 个特征图，然后再经过 Inception 模块和大小为 2×2 的最大池化操作得到大小为 8×8 的 480 个特征图，最后经过一次 Inception 模块得到大小为 8×8 的 512 个特征图。在网络的上半部分，把由 P_{11}，P_{12} 和 P_{13} 得到的特征图进行特征连接，最终得到大小为 8×8 的 1 248 个特征图。由于是孪生网络，网络结构的下半部分和上半部分的操作完全一样，把上下两个部分分别得到的特征图进行特征连接，作为全连接层的输入，网络的最后是利用 softmax 对特征进行二分类，计算每个子块的聚焦概率。

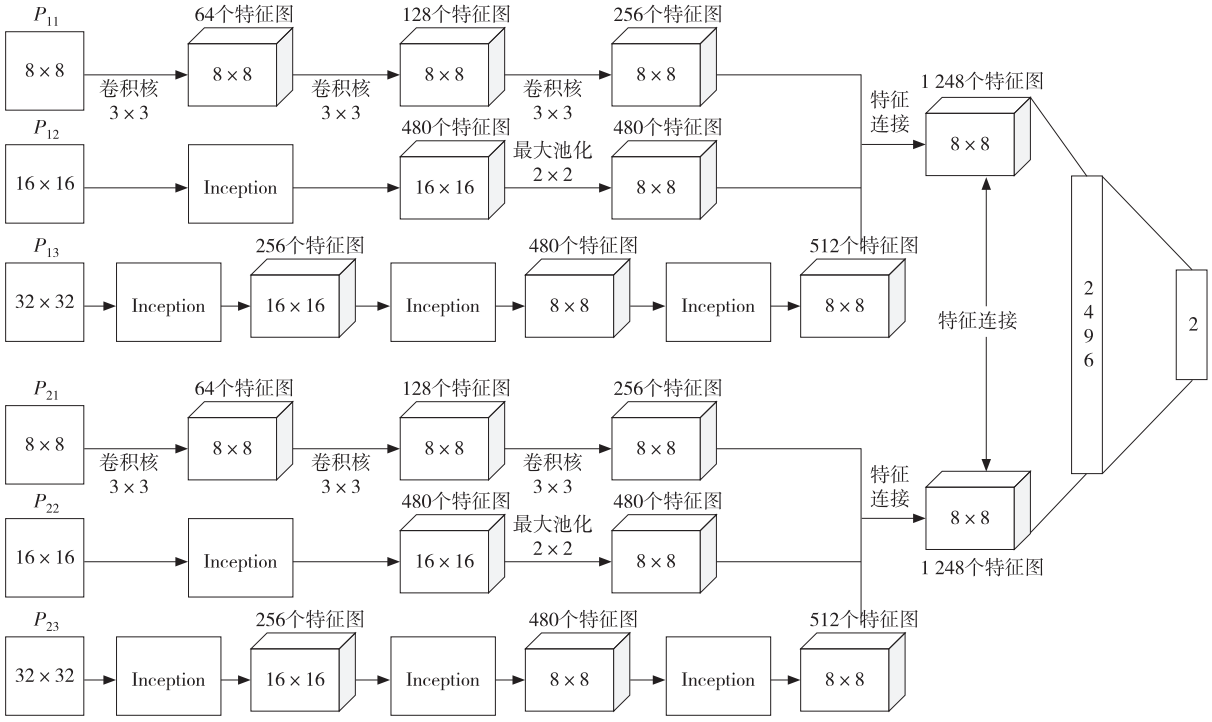


图1 整体网络结构

Fig. 1 The network architecture

1.2 训练

多聚焦图像融合可以看作为一个二分类问题，让 P_1 和 P_2 分别表示一对清晰和模糊的子块，当子块 P_1 比 P_2 清晰时该样本的标签定义为1，当子块 P_1 比子块 P_2 模糊时样本的标签定义为0。本文使用Cifar10数据集[17]生成训练集。这里选取Cifar10中的前45 000张图像生成所需要的训练集，后5 000张用来测试网络。首先将每一幅Cifar10图像转换成灰度图像，然后利用高斯模糊对每一幅图像进行5个不同程度的模糊，一共得到了225 000张32×32大小的模糊训练集，45 000张大小为32×32的清晰训练集。Cifar10源图像随机裁剪为16×16，8×8大小，然后执行与上述32×32子块一样的操作，分别得到225 000张大小为16×16，225 000张大小为8×8的模糊的训练集，以及45 000张大小为16×16，45 000张大小为8×8的清晰的训练集。在网络训练的过程中，使用了Adam作为优化器，批处理大小设置为64，动量设置为0.9，学习率设置为0.000 01，训练次数为1。

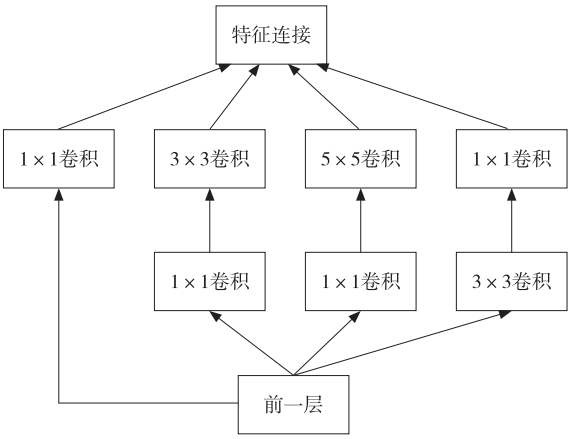


图2 Inception模块

Fig. 2 Inception module

1.3 算法步骤

图3为融合步骤(以Lytro图像集中的"child"为例子)。本文算法步骤如下：

输入：一对多聚焦图像；输出：一幅全聚焦图像。

步骤1：将一对多聚焦源图像分为多个32×32大小的图像块，然后将每个32×32子块对应的源图像区域随机裁剪为一个16×16大小的子块，一个8×8大小的子块。将两幅源图像分成的对应子块一起作为网络的输入，能够得到一个得分图。得分图中的每个数字能够说明一对图像块的聚焦程度。值接近1，说明来自源图像A的子块更聚焦。值接近0，说明来自源图像B的子块更聚焦。

步骤2：生成得分图的大小和源图像的大小是不一样的，为了能够得到和源图像大小一样的聚焦图M，将得分图中的每个值分配给对应M的全部像素，并将重叠像素进行平均。

步骤 3：将获得的聚焦图进行二值处理，阈值选择为 0.5。阈值处理公式为：

$$T(x,y)=\begin{cases}1, M(x,y)>0.5\\ 0, \text{否则}\end{cases}。$$

步骤 4：对得到的二值图 D 和源图像采用像素加权平均规则计算融合图 S，公式为：

$$S(x,y)=D(x,y)A(x,y)+(1-D(x,y))B(x,y)。$$

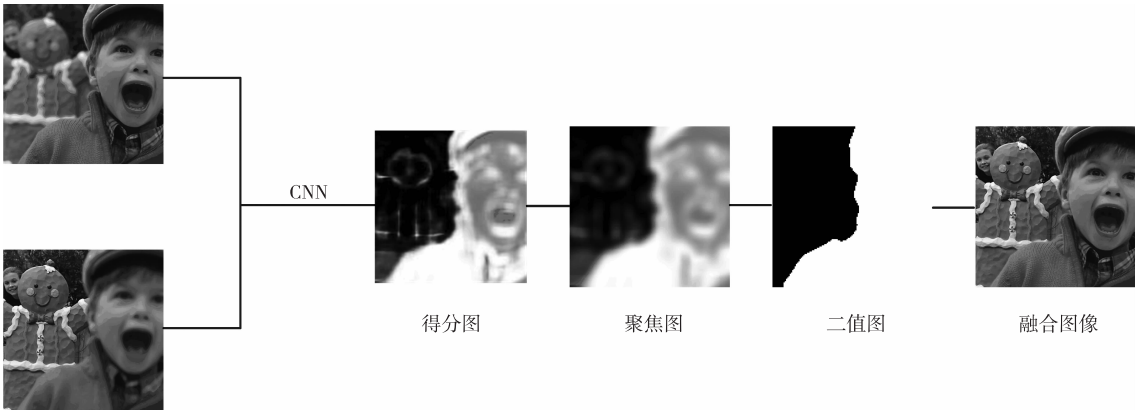


图 3　融合步骤

Fig. 3　Fusion step

2 实验结果与分析

2.1 客观评价指标

本文使用了两组不同的数据集来进行测试，其中第一组是 20 对 Lytro[18] 多聚焦彩色图像，第二组为 20 对其他常用的多聚焦灰度图像。图 4a，b 分别展示了两组图像集的部分源图像。本文选取 4 种经典的方法与本文提出的方法进行比较，分别是 NSCT[19]、IFCNN[14]、GFF[20] 和 CNN[16]。

在现实中，获得参考图像一般很困难，所以图像融合的定量评估并非易事。因此，许多指标被提出来评价图像融合的性能。目前，对于哪些度量可以完全的描述图像融合性能还没有共识。本文选用一些比较著名的融合度量进行评估，即：基于互信息量的度量 Q_{MI}[21] 和 Q_{Te}[22]、基于非线性相关信息熵的度量 Q_{NICE}[23]、基于图像相似度的度量 Q_{Y}[24] 和基于人类感知的度量 Q_{CB}[25]，这 5 个度量都是值越大代表融合的效果越好。表 1 给出了对各种评价指标的说明。

表 2 为 20 对 Lytro 图像在 5 种评价指标下的平均值，最高的值用粗体表示。由表 2 可以看出，本文提出的算法在 Q_{MI}，Q_{Te}，Q_{NICE}，Q_{Y} 等 4 个指标上均优于其他所对比的方法。对于 Q_{CB}，相比于 CNN，本文提出的算法获得的值略低，但是比其他的 3 种算法优秀。

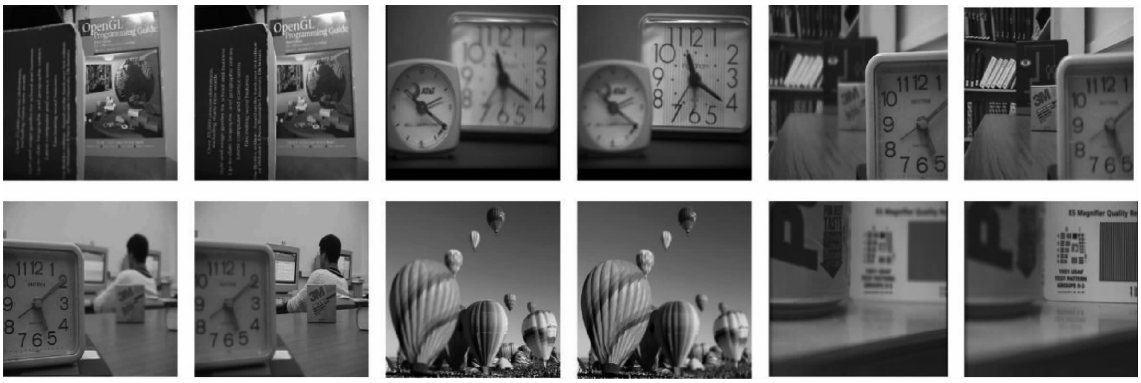
表 3 为其他常用 20 对多聚焦灰度图像在 5 种评价指标下的平均值，最高的值用粗体表示。由表 3 可以看出，提出的算法在 Q_{MI}，Q_{Te}，Q_{NICE}，Q_{Y} 等 4 个评价指标上表现优异，都要高于其他 4 种方法。对于 Q_{CB}，本文提出的方法与 CNN 的方法相差不大，且都明显高于其他 3 种方法。从结果可以看出传统算法得到的评价指标一般低于基于深度学习的算法。

2.2 主观效果评价

图 5 是具有代表性的灰度图像“desk”的实验结果，图 5a，b 为源图像，图 5c～g 为使用不同方法得到的融合图像，图 5h～l 是不同方法得到的融合图像与源图像 A 的差分图像。差分图在视觉上可以很好地表现出融合图像和源图像之间的差异，可以清楚直观地看出图像中的细节。图 5 中的差分图像，右侧时钟区域显示的信息越多，就说明融合的效果越差。从图 5h～k 可以看出图像右侧时钟上的数字，右侧时钟区域特征信息残留比较多，这说明融合的效果不是很好。图 5i 中右侧时钟区域融合的良好，但是时钟轮廓边缘残留的特征信息明显，这说明融合图像的边缘融合质量不是很好。从图 5j，l 可以看出，基于 CNN 的方法和提出的算法无论是在融合的边缘还是内容方面都展现出了良好的效果。为了进一步的比较基于 CNN 的算法和提出算法，时钟左上方红色矩形框内的内容被放大至图 5j，l 的左下角。从放大的区域可以看出，基于 CNN 的算法在时钟边缘有特征信息残

留,而提出的算法无残留的特征信息。同时,从时钟顶部的红色矩形框区域可以看出,提出的算法融合的图像更加的自然,边缘细节融合的更好。

a "Lytro"部分源图像



b 其他常用的多聚焦灰度图像

图4 部分测试图像

Fig. 4 Part of the test data in our experiment

表1 5种评价指标的说明

Tab. 1 Description of five evaluation indicators

评价指标	说明
Q_{MI}	计算融合图像和源图像之间的互信息量
Q_{Te}	计算融合图像和源图像之间的 Tsallis 熵
Q_{NICE}	计算融合图像和源图像之间的非线性相关信息熵
Q_{Y}	计算融合图像中保存源图像的结构信息量
Q_{CB}	强调人类视觉系统的主要特征

表2 20对 Lytro 图像融合指标的平均值

Tab. 2 The average objective assessment of Lytro images

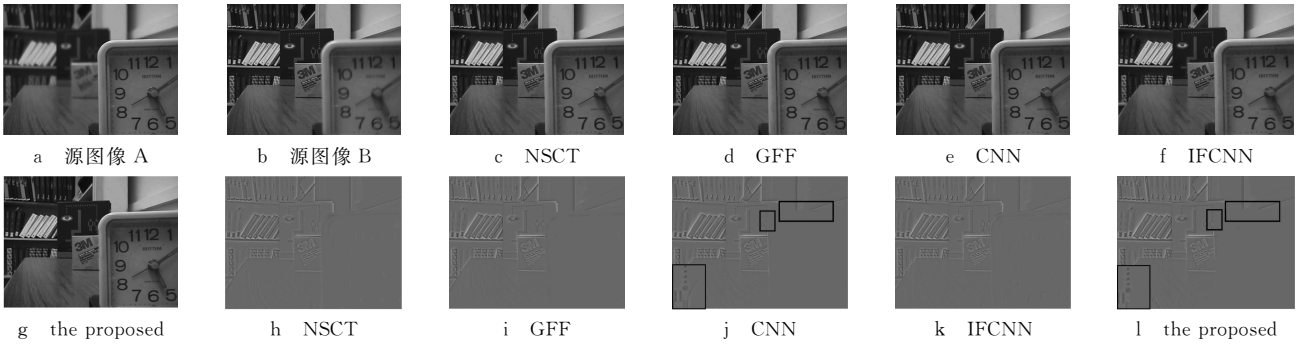
方法	Q_{MI}	Q_{Te}	Q_{NICE}	Q_{Y}	Q_{CB}
CNN	1.151 894	0.400 014	0.842 942	0.987 562	**0.808 448**
GFF	1.142 157	0.400 660	0.842 711	0.982 598	0.801 953
NSCT	0.952 551	0.385 656	0.830 751	0.961 644	0.747 797
IFCNN	0.937 691	0.384 978	0.829 759	0.952 212	0.729 189
本文方法	**1.184 761**	**0.412 261**	**0.865 199**	**0.998 258**	0.808 060

注:表中加粗的数字表示同指标下的最高值,下同

表 3　其他常用 20 对多聚焦灰度图像融合指标的平均值

Tab. 3　The average objective assessment of other commonly used 20 pairs of multi-focus gray-scale images

方法	Q_{MI}	Q_{Te}	Q_{NICE}	Q_{Y}	Q_{CB}
CNN	1.000 561	0.411 889	0.833 830	0.886 982	**0.744 002**
GFF	0.999 213	0.409 359	0.833 875	0.877 142	0.730 946
NSCT	0.926 431	0.395 060	0.829 587	0.863 441	0.722 552
IFCNN	0.945 648	0.397 140	0.829 750	0.885 724	0.703 829
本文方法	**1.033 504**	**0.443 796**	**0.839 719**	**0.922 495**	0.742 092



a　源图像 A　b　源图像 B　c　NSCT　d　GFF　e　CNN　f　IFCNN

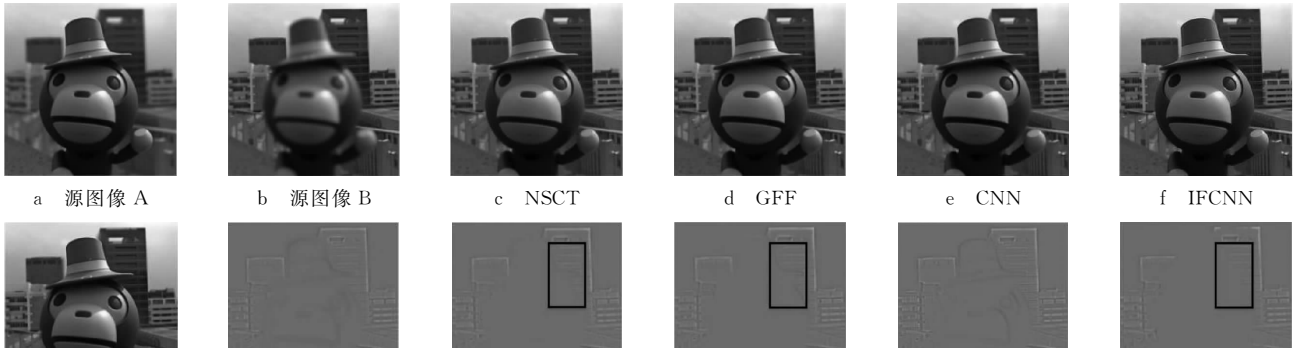
g　the proposed　h　NSCT　i　GFF　j　CNN　k　IFCNN　l　the proposed

注：a，b 为源图像，c～g 为使用不同方法得到的融合图像，h～l 是不同方法得到的融合图像与源图像 a 的差分图像

图 5　不同方法的“desk”融合图像以及对应的差分图

Fig. 5　The “desk” fusion images of different methods and the corresponding difference images

图 6 为 Lytro 图像集中 lytro-20 的实验结果。其中图 6a，b 为源图像，图 6c～g 为不同方法的融合图像，图 6h～l 分别为不同方法的融合图像和源图像 A 的差分图像。由图 6h 可以看出差分图像中物体的帽子、眼睛、鼻子、嘴巴区域特征信息较多，这说明基于 NSCT 融合算法得到的融合图像效果不好。由图 6k 可以看出融合图像中帽子和嘴巴区域特征信息较多，可以看出基于 IFCNN 的图像融合算法的效果也不是很好，但是比基于 NSCT 融合算法要好。由图 6i，j，l 可以看出基于 GFF，基于 CNN 以及提出的算法融合的效果都很好，但是由红色矩形区域可以看出提出的算法相比于基于 CNN 和基于 GFF 的算法而言，在边缘上，细节特征信息更少，边缘更加自然。



a　源图像 A　b　源图像 B　c　NSCT　d　GFF　e　CNN　f　IFCNN

g　the proposed　h　NSCT　i　GFF　j　CNN　k　IFCNN　l　the proposed

注：a，b 为源图像，c～g 为使用不同方法得到的融合图像，h～l 是不同方法得到的融合图像与源图像 a 的差分图像

图 6　不同方法的 Lytro-20 融合图像以及对应的差分图

Fig. 6　The lytro-20 fusion image and the corresponding difference image obtained by different methods

结合客观评价以及主观评价可知，本文所提出的算法很好地融合了多聚焦图像中的细节，融合边缘平滑自然，无伪影产生，融合之后的图像信息保存完好，融合结果从主观和客观两个维度上都拥有明显的优势。

2.3 参数讨论

2.3.1 阈值分割参数讨论 二值分割阈值会影响融合效果,在进行实验时发现当阈值大于0.5时,聚焦区域内存在大量误判标签,实验效果不理想。为了进一步选取有效的阈值,选取阈值0.4,0.45和0.5进行实验。对于20对Lytro图像和其他常用的20对多聚焦灰度图像数据集,实验结果如表4和表5所示,最高的值用粗体表示。从表5可以看出,当阈值选择为0.5时,得到的评价度量的值均大于阈值为0.4,0.45所得到的值。从表4可以看出,当阈值为0.5时,除了Q_{MI}的值小于阈值为0.4和0.45得到的Q_{MI}的值之外,其他4个度量均大于阈值为0.4和0.45得到的值。综上所述,阈值选为0.5最为合适。

表4 20对Lytro图像分割结果

Tab.4 20 pairs of Lytro images segmentation threshold results

评价指标	0.4	0.45	0.5
Q_{MI}	1.180 574	1.182 978	**1.184 761**
Q_{Te}	0.401 533	0.402 048	**0.412 261**
Q_{NICE}	0.844 745	0.844 989	**0.865 199**
Q_{Y}	0.986 768	0.987 859	**0.998 258**
Q_{CB}	0.801 660	0.806 731	**0.808 060**

表5 其他常用20对多聚焦图像

Tab.5 Other commonly used 20 pairs of multi-focus image segmentation threshold results

评价指标	0.4	0.45	0.5
Q_{MI}	**1.041 178**	1.020 705	1.033 504
Q_{Te}	0.411 836	0.411 687	**0.443 796**
Q_{NICE}	0.834 576	0.834 650	**0.839 719**
Q_{Y}	0.779 520	0.883 776	**0.922 495**
Q_{CB}	0.737 282	0.738 502	**0.742 092**

2.3.2 高斯模糊参数讨论 在基于深度学习的多聚焦图像融合算法研究中,通常采用高斯模糊对数据集进行增广,提出的算法采用高斯核为7×7,标准差为2的高斯模糊来生成训练集。为了比较不同模糊程度对提出算法结果的影响,使用高斯核均为7×7,但是标准差分别为1,2和3的高斯模糊对数据集进行增广。实验结果如表6和表7所示,最高的值用粗体表示。从表6和表7可以看出,标准差为2最合适。

表6 20对Lytro图像不同程度高斯模糊结果

Tab.6 20 pairs of Lytro images with different degrees of Gaussian blur results

评价指标	1	2	3
Q_{MI}	1.179 596	**1.184 761**	1.181 459
Q_{Te}	0.409 274	0.412 261	**0.416 958**
Q_{NICE}	0.862 596	**0.865 199**	0.864 395
Q_{Y}	0.994 495	**0.998 258**	0.986 473
Q_{CB}	0.803 597	**0.808 060**	0.805 537

表7 其他常用20对多聚焦图像不同程度高斯模糊结果

Tab.7 Other commonly used 20 pairs of multi-focus image with different degrees of Gaussian blur results

评价指标	1	2	3
Q_{MI}	**1.055 839**	1.033 504	1.014 968
Q_{Te}	0.439 843	**0.443 796**	0.441 016
Q_{NICE}	0.835 196	**0.839 719**	0.837 356
Q_{Y}	0.864 907	**0.922 495**	0.895 825
Q_{CB}	0.729 501	**0.742 092**	0.739 848

2.4 消融实验

为了体现提出算法多尺度的有效性,本文分别做了3种不同尺度的消融实验。在原始的网络结构的基础上,输入阶段分别全部使用32×32子块大小、16×16子块大小和8×8子块大小进行对比实验。在Lytro数据集和其他常用的多聚焦图像数据集上的实验结果如表8和表9所示。将源图像分成32×32子块进行聚焦分类时,由于子块较大,在聚焦和散焦的边界得到的子块可能同时包含聚焦区域和散焦区域,在进行聚焦分类时,网络不能对这一部分子块正确地分类,从而得到的效果较差。将源图像分成16×16子块进行聚焦分类时,效果较好。将源图像分成8×8子块进行聚焦分类时,子块中包含的图像信息较少,网络分类的精度无法保证,效果也比较差。提出的算法利用了多尺度的图像信息,可以保证网络的分类精度,相比于单一尺度大小的子块而言,效果更好。

3 结束语

本文提出了一种多尺度的卷积神经网络算法。在网络中并入了Inception模块,通过提升网络的宽度来增强网络的性能。同时还可以降低网络中参数的数量。提出的算法还通过图像多个不同尺度的细节特征增强像

素的分类精度。在 Lytro 图像集和其他常用的多聚焦灰度图像集上的实验结果表明，本文提出的算法很好地融合了源图像中的细节特征，融合边缘自然平滑、没有伪影，得到了较传统算法更好的融合效果。后期考虑将算法应用于多模态医学图像融合以及红外线可见光图像融合。

表 8　20 对 Lytro 图像多尺度对比结果

Tab. 8　Multi-scale comparison results of 20 pairs of Lytro images

评价指标	8×8	16×16	32×32	提出的算法
Q_{MI}	1.128 585	1.154 648	1.133 784	**1.184 761**
Q_{Te}	0.384 752	0.403 364	0.401 135	**0.412 261**
Q_{NICE}	0.842 973	0.847 297	0.850 061	**0.865 199**
Q_{Y}	0.932 856	0.966 843	0.912 093	**0.998 258**
Q_{CB}	0.738 364	0.791 735	0.748 573	**0.808 060**

表 9　其他常用 20 对多聚焦图像的多尺度对比结果

Tab. 9　Other commonly used 20 pairs of multi-focus image with multi-scale comparison results

评价指标	8×8	16×16	32×32	提出的算法
Q_{MI}	0.968 596	0.982 385	0.979 586	**1.033 504**
Q_{Te}	0.404 869	0.424 697	0.413 396	**0.443 796**
Q_{NICE}	0.801 945	0.815 963	0.795 993	**0.839 719**
Q_{Y}	0.879 475	0.895 836	0.904 616	**0.922 495**
Q_{CB}	0.719 586	**0.759 413**	0.706 011	0.742 092

参考文献：

[1] GOSHTASBY A A，NIKOLOV S. Image fusion：advances in the state of the art[J]. Information Fusion，2007，2(8)：114-118.

[2] 李娇，杨艳春，党建武，等. NSST 与引导滤波相结合的多聚焦图像融合方法[J]. 哈尔滨工业大学学报，2018，50(11)：145-152.
LI J，YANG Y C，DANG J W，et al. NSST and guided filtering for multi-focus image fusion algorithm[J]. Journal of Harbin Institute of Technology，2018，50 (11)：145-152.

[3] HUANG W，JING Z. Multi-focus image fusion using pulse coupled neural network[J]. Pattern Recognition Letters，2007，28(9)：1123-1132.

[4] LI S，KANG X，HU J，et al. Image matting for fusion of multi-focus images in dynamic scenes[J]. Information Fusion，2013，14(2)：147-162.

[5] DE I，CHANDA B. Multi-focus image fusion using a morphology-based focus measure in a quad-tree structure[J]. Information Fusion，2013，14(2)：136-146.

[6] ASLANTAS V，KURBAN R. Fusion of multi-focus images using differential evolution algorithm[J]. Expert Systems with Applications，2010，37(12)：8861-8870.

[7] PIELLA G. A general framework for multiresolution image fusion：from pixels to regions[J]. Information Fusion，2003，4(4)：259-280.

[8] BURT P，ADELSON E. The Laplacian pyramid as a compact image code[J]. IEEE Transactions on Communications，1983，31(4)：532-540.

[9] LI H，MANJUNATH B S，MITRA S K. Multisensor image fusion using the wavelet transform[J]. Graphical Models and Image Processing，1995，57(3)：235-245.

[10] ZHANG Q，GUO B. Multifocus image fusion using the nonsubsampled contourlet transform[J]. Signal Processing，2009，89(7)：1334-1346.

[11] LEWIS J J，O'CALLAGHAN R J，NIKOLOV S G，et al. Pixel-and region-based image fusion with complex wavelets[J]. Information Fusion，2007，8(2)：119-130.

[12] 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(11)：5793-5804.

[13] TANG H，XIAO B，LI W，et al. Pixel convolutional neural network for multi-focus image fusion[J]. Information Sciences，2018，433：125-141.

[14] 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.

[15] 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.

[16] 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.

[17] KRIZHEVSKY A，HINTON G. Learning multiple layers of features from tiny images[EB/OL]. (2009-01-01)[2021-08-10].

https://typeset.io/papers/learning-multiple-layers-of-featcues-from-tiny-images-seum9uf4g8.

[18] NEJATI M,SAMAVI S,SHIRANI S. Multi-focus image fusion using dictionary-based sparse representation[J]. Information Fusion,2015,25:72-84.

[19] ZHANG Q,GUO B. Multifocus image fusion using the nonsubsampled contourlet transform[J]. Signal Processing,2009,89(7):1334-1346.

[20] LI S,KANG X,HU J. Image fusion with guided filtering[J]. IEEE Transactions on Image Processing,2013,22(7):2864-2875.

[21] LIU Z,BLASCH E,XUE Z,et al. Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision:a comparative study[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2011,34(1):94-109.

[22] CVEJIC N,CANAGARAJAH C N,BULL D R. Image fusion metric based on mutual information and Tsallis entropy[J]. Electronics Letters,2006,42(11):626-627.

[23] WANG Q,SHEN Y,JIN J. Performance evaluation of image fusion techniques[J]. Image Fusion:Algorithms and Applications,2008,19:469-492.

[24] YANG C,ZHANG J Q,WANG X R,et al. A novel similarity based quality metric for image fusion[J]. Information Fusion,2008,9(2):156-160.

[25] CHEN Y,BLUM R S. A new automated quality assessment algorithm for image fusion[J]. Image and Vision Computing,2009,27(10):1421-1432.

Multi-scale Feature Based on Deep Learning and Its Application in Multi-focus Image Fusion

GAO Wenchang, YU Lei

(College of Computer and Information Science, Chongqing Normal University, Chongqing 401331, China)

Abstract: [**Purposes**]Multi-focus image fusion is to extract the respective focus areas from different images in the same scene to obtain an all-in-focus image, and is a popular research direction in the field of image processing in recent years. Traditional image fusion technology has some defects, such as unclear areas, distortions and artifacts. To overcome these disadvantages, an image fusion method based on deep learning is proposed. [**Methods**]The proposed network uses the siamese network as a whole to classify the focus area in the image, at the same time, the Inception module in GoogLeNet is also introduced to improve the feature extraction ability of the network and the proposed network has achieved good performance. To make the best of the feature information in the source images, the proposed network uses patches of different sizes to extract feature information of different scales in the source images, and obtain multiple scale features of the source image. In addition, the binary image obtained by the proposed method can accurately reflect the focus area and defocus area of the source images, so there is no need for post-processing steps to optimize the binary image, which reduces the complexity of the network. [**Findings**]According to the experimental results on the "Lytro" multi-focus image set and other commonly used multi-focus gray-scale image set, the fusion result of the proposed algorithm has distinct advantage in both subjective and objective dimensions, compared with other classic algorithms. [**Conclusions**] The proposed algorithm integrates the detailed features in the source image well, the fusion edge is naturally smooth, and no artifacts are generated, and it achieves a better fusion effect than the traditional algorithm.

Keywords: siamese network; inception module; multi-scale features; multi-focus image fusion

(责任编辑 许 甲)