Multi-Focus Fused Image using Inception-Resnet V2

M. Sobhitha, C. Shoba Bindu, E. Sudheer Kumar

Abstract**:** ***Multi-focus image fusion is the process of integration of pictures of the equivalent view and having various targets into one image. The direct capturing of a 3D scene image is challenging, many multi-focus image fusion techniques are involved in generating it from some images focusing at diverse depths. The two important factors for image fusion is activity level information and fusion rule. The necessity of designing local filters for extracting high-frequency details the activity level information is being implemented, and then by using various elaborated designed rules we consider clarity information of different source images which can obtain a clarity/focus map. However, earlier fusion algorithms will excerpt high-frequency facts by considering neighboring filters and by adopting various fusion conventions to achieve the fused image. However, the performance of the prevailing techniques is hardly adequate. Convolutional neural networks have recently used to solve the problem of multi-focus image fusion. By considering the deep neural network a two-stage boundary aware is proposed to address the issue in this paper. They are: (1) for extracting the entire defocus info of the two basis images deep network is suggested. (2) To handle the patches information extreme away from and close to the focused/defocused boundary, we use Inception ResNet v2. The results illustrate that the approach specified in this paper will result in an agreeable fusion image, which is superior to some of the advanced fusion algorithms in comparison with both the graphical and objective evaluations.***

Index Terms**:** ***Multi-Focus, Image Fusion, Fused Image, Convolution Neural Network.***

I. INTRODUCTION

The image fusion role in handling of images is vigorous by extracting the best and balancing features from 2 or more images and incorporating that information by using proper algorithm in order to provide better recognition characteristics. Image fusion plays a vital role in many applications. Revealing of the focused region is the major issue for the multi-focus image fusion algorithm.

When capturing a picture of a 3D view, it is challenging to obtain a picture where every object is centered, while the intensity of the field is inadequate. However, it is further appropriate and efficient to use all-in-focus images as key in for various image processing tasks. For decades, enormous multi-focus image fusion algorithms have been suggested. Best of them fall into these both gatherings, i.e., transform domain-based algorithms [1] and spatial domain-based algorithms [2].

The determination of studies on multi-focus image fusion is to attain a collected image where all objects are to be captured in a focus. Paralleled to the original images, the novel one is more precious info and enhanced graphical performance. Recently experts are very much attracted in using deep learning algorithms in processing the image data. The suggested fusion method exploits the capabilities of artificial neural networks. Furthermore, the ability of the neural network to learn to customize the image fusion process.

A decision map consists of thorough and precise info about the image which needs to be fused, which is essential for several image fusion concerns, particularly in multi-focus image fusion. Still, to get an adequate image fusion effect, the accomplishment of a decision map is very much necessary and typically challenging to finish. Here, the boundary aware multi-focus image fusion method with the deep neural network is suggested to overcome the shortfall of information. The contributions are as follows: as a first step the two basis images were considered from the same position and they are converted into grayscale images for normalization purpose, then initial score map is generated, and based on the decision map of an image the fused image is created. Results on different test images endorse that this method produces better superiority fused images than that of the state-of-the-art image fusion methods.

II. RELATED WORK

In [3], the first CNN is used for retrieving the defocus level descriptors and assessment rules in a sequential way. However, the results of earlier neural network related approaches [3] [4] are still unconvinced, specifically areas close to the focused/defocused boundary (FDB), hence, ample post-processing is being deliberated to overcome this issue.

There have been some prior attempts performed to use DNNs for image fusion. DNA-based techniques are also being used which can process distant images from multispectral (MS) and panchromatic (PAN) images to incorporate corresponding spectral and spatial features into one product. In [5], PAN image and MS image is considered by using an enhanced Small Denoising Autoencoder (MSDA). Though, the images are treated as a multi-channel image and distributed straight away to the initial convolutional layer of the network. Li et al. [6], considers a classification problem for the multi-focus image fusion and suggested a fusion

Revised Manuscript Received on August 05, 2019.

M. Sobhitha, Department of CSE, JNTUA, Ananthapuramu, Andhra Pradesh, India.

C. Shoba Bindu, Department of CSE, JNTUA, Ananthapuramu, Andhra Pradesh, India.

E. Sudheer Kumar, Department of CSE, JNTUA, Ananthapuramu, Andhra Pradesh, India.

technique which relies on ANN.

Tang et al. [4] popularized pixel CNN (p-CNN) to handle multi-focus image fusion. Source images are considered to create the score matrix, then a decision map is generated by correlating with the score matrix values, the final output is fused image. Du et.al [7] introduced a segmentation related multi-focus image fusion network which can be used at multi-level input. Prabhakar et al. [8] given a CNN-based unsupervised method for Multi-Exposure Fusion (MEF), wherein a very modest CNN with multi convolutional layers is proposed for encoding the input images and for generating feature map progression and finally for fused maps.

Zhao et al. [9] proposed the concept of a multilevel deep CNN based on supervised learning for image fusion and they established an end-to-end network for learning various aspects.

In [10], for low-dose X-ray CT restoration, a residual network established on the directional wavelet transforms domain is proposed. The direction wavelet is recycled to syndicate the key in dataset and the labeled dataset into elevated-dimensional feature space and absorb its scaling.

In [11], Wavelet-based CNN has proposed a multi-level super-resolution network for faces, resulting in fine detailing of high-resolution images. In another method, multi-focus image fusion, a mixture of super-resolution and CNN, can be used directly to generate super-resolution and full-focus output images, resulting in improved fusion images. The solutions discussed above focuses not only on boundary level information but also on the focused/defocused boundary area, which is very important for image fusion work.

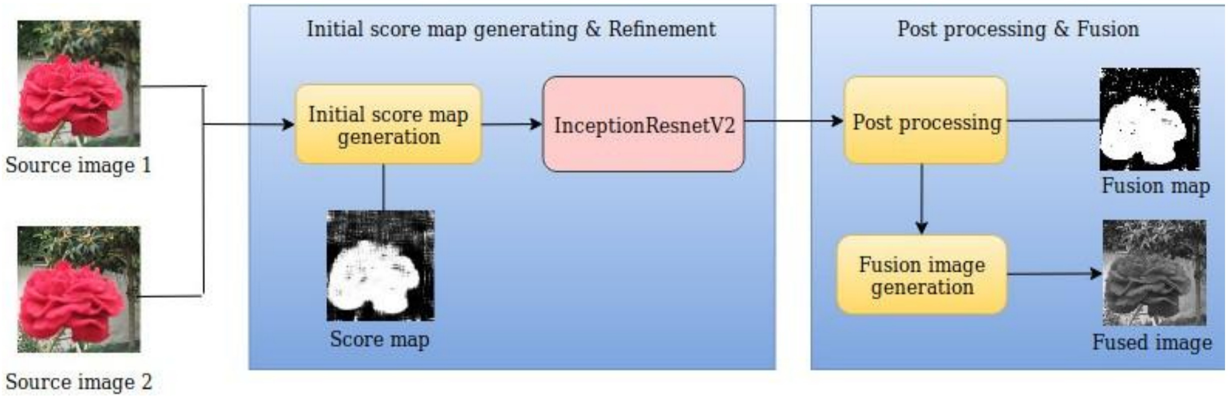

Fig 1: The process flow diagram of the suggested approach

III. PROPOSED FUSION METHOD

This section discusses our approach consisting of 2 steps they are: first generating a score map, refining the score map, followed by a post-processing method, and fusing the image. As a base for this approach, we have considered [20]. The process flow diagram of this approach is in Figure 1.

A. First Generating a Score Map

CNN is the common model for pixel-wise multi-focus image fusion classification problem. However, when the network is dependent, the problem of degradation arises because optimization is not equally easy for all systems. It is a good selection to use two source images as input.

Fig 2: Score Map generation

Moreover, in the process of multi-focus image fusion, 2 images were picked up from a similar position. The same result has been anticipated when you fused those two images into one. Henceforth, for pixel-wise image fusion, the two-channel representation is additional capable and simpler than the model in [12]. The initial score map is generated (i.e, Fig 2) for the gray scale image which is then fed in to the network for generating fusion map.

After generating score map for the image, it is fed in to the convolution neural network Inception-ResNet v2 for generating fusion map. This model recognizes the 1000 different classes of objects in the ImageNet 2012 Large Scale Visual Recognition Challenge. The model consists of a deep convolutional net using the Inception-ResNet-v2 architecture that was trained on the ImageNet-2012 data set. The input to the model is a 299×299 image, and the output is a list of estimated class probabilities. The Inception network was an important milestone in the development of CNN classifiers. Inception-ResNet v2 has a computational cost that is similar to that of Inception v4.

B. Post-Processing and Fused Image

When refining, a minimal post-processing step is required, such as binarization and small-area removal [3]. In particular, to obtain a decision map, the binarization is smeared to the focus score map. Small areas within the Focused or De-Focused area can also be cleared.

After attaining the decision map, the next step of fusion image (ImgFu) will be obtained from the decision map (DeMa) and two source images (ImgSo) is as follows:

$$ImgFu = DeMa.\ ImgSo_A + (1 - DM).\ ImgSo_B.$$

After the post-processing phase, the fusion map is generated as shown in Fig 3.

Fig 3: Fusion Map after post-processing

IV. EXPERIMENTAL RESULTS

Here we specify about the network used in our work, configurations that are used for implementation and finally the metrics that are considered to evaluate the proposed system.

A.Network Details and Configuration Settings

In our execution, typical Inception-ResNet v2 network is used. The effects of depth of the network and size of the patches are examined. If a deeper network or maximum patches used, in turn the performance will enhance. However, there is a conflict between performance at the expense of time and memory. In order to perform training and testing of this approach we have used Intel Core i7-8700 CPU @3.20 GHz ×12, memory 16GB, Radeon RX 550 series GPU, 64-bit processor with Ubuntu 18.04 LTS. This paper is implemented in Keras framework with Tensorflow backend and the programming language used is Python 3.6.

The dataset used for training is Alpha Matting dataset [13], which has 27 images along with ground truth information is used as foreground dataset, and 700 COCO 2017 dataset [14] background images are used. The sample images which are in the database has shown in the Fig 4. These three images were selected randomly for validation purpose.

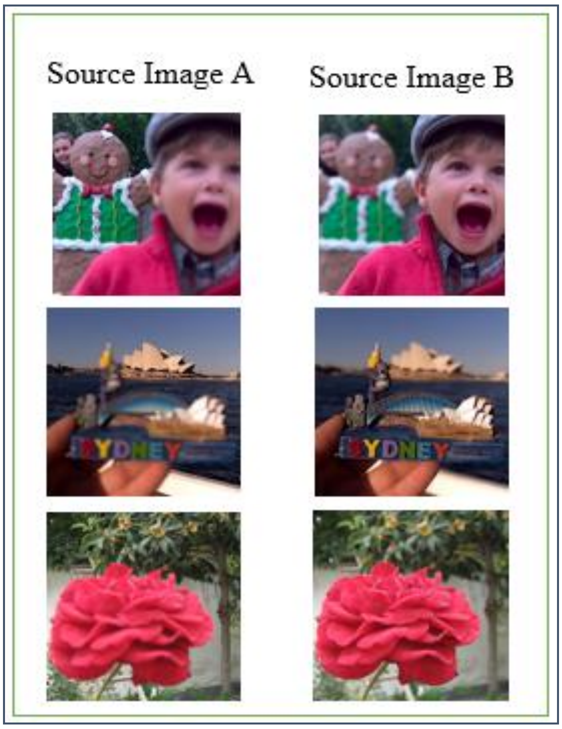

Fig 4: Sample images for testing

B.Quality Evaluation Metrics

In this paper for evaluating the image quality generated from the network we have used Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) no-reference image quality score metric [15] and NIMA: Neural Image Assessment metric [16]. The results generated for these two metrics are specified in table 1.

Table 1: Image Quality Assessment by using two metrics for proposed approach.

S. No	Image Name	BRISQUE	NIMA Score
1	Boy	9.641	5.068 +- (1.382)
2	Rose	5.736	4.847 +- (1.437)
3	Sydney	6.012	6.151 +- (1.406)

V. COMPARISON RESULTS

By comparing the suggested method with 3 other multi-focus fusion methods, namely NSCT [17], SR [18], DSIFT [19]. The evaluation has performed on 10 combinations of multi-focus images taken from "Lytro" dataset and considering the average of all results for each method, Lytro is the commonly cited database for image fusion, wherein others have considered "clock", "book", "lab" and "flower", which are most widely used. The results of various methods on the above three images are mentioned in table 2 and also plotted in graphs shown in Figure 5 & 6.

Table 2: Comparison results of different techniques with the proposed approach

S. No	Method Name	BRISQUE	NIMA Score
1	NSCT	5.157	3.369
2	SR	5.469	3.947
3	DSIFT	6.397	4.699
4	Our Method	6.832	5.125

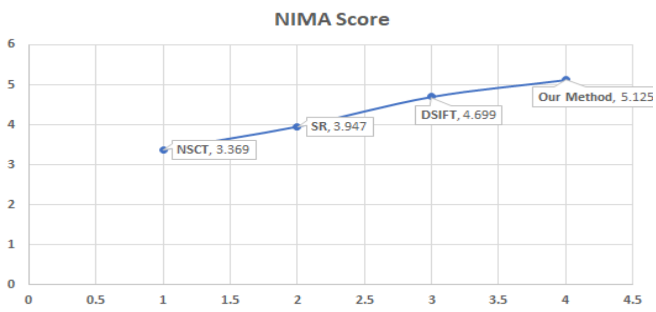

Figure 5: Comparison result by using NIMA Score

The below Fig 7 depicts the original source images we considered for testing and their corresponding gray scale image, its initial score map, fusion map and fusion result.

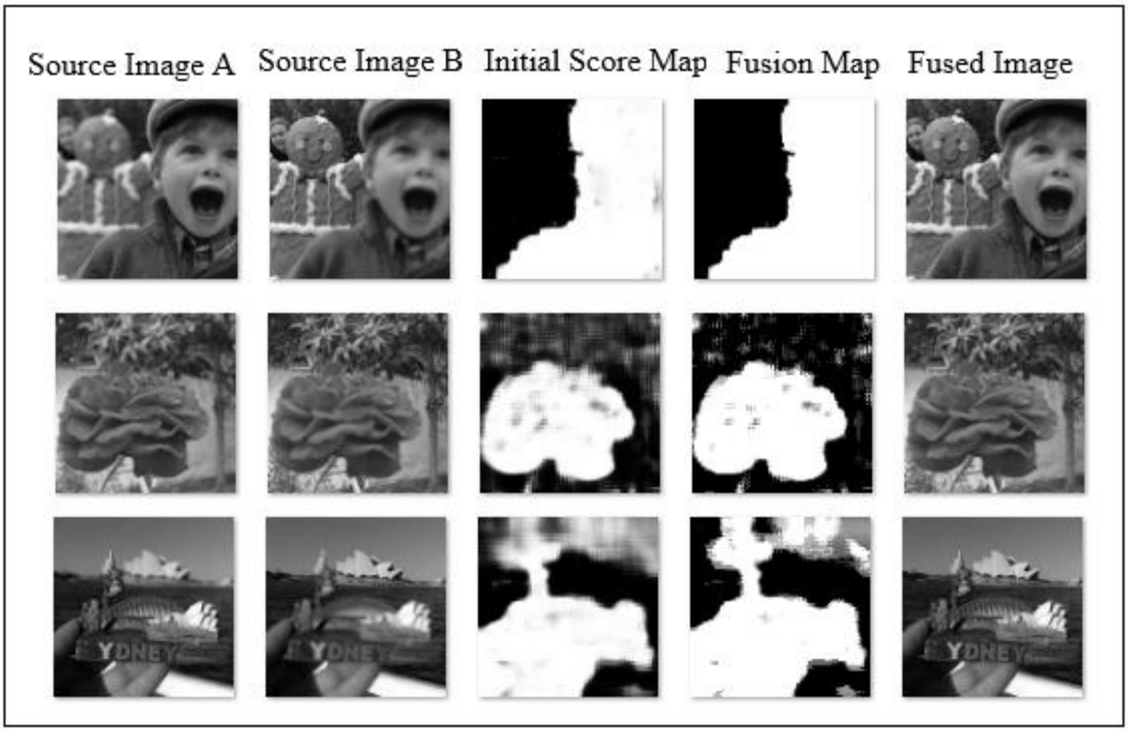

Figure 7 : This figure depicts the original gray scale source images we considered for testing, its initial score map, fusion map and fusion result.

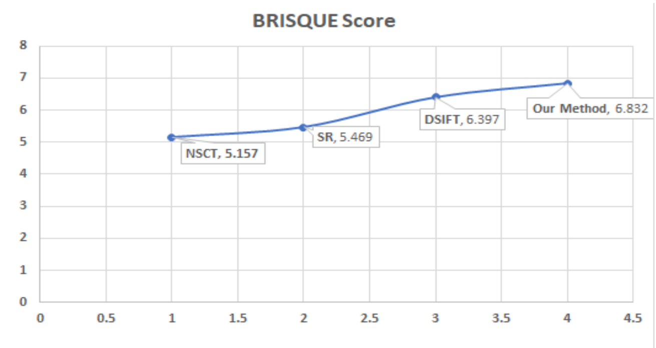

Figure 6: Comparison result by using BRISQUE Score

VI. CONCLUSION

Multi-focus image fusion is an all-in-focus image that combines dissimilar depth-of-field pictures with various focus levels of the similar scene. One of the basic need for image fusion is that, the overall details should be taken from various source images and conserved in the final fusion image. In this paper, we discourse this issue with a deep learning-based approach, intended at understanding the straight mapping between source images and focus map. The pixel-level map is that which comprises resolution information later comparing with the activity level of the source images. The 2-channel structure in this paper is designed to well extract the absolute diffuse information of the two source images and advance the fusion results. The suggested deep architecture delivers advanced performance in terms of individual and objective measurements.

REFERENCES

1. J. J. Lewis et.al, “Pixel- and region-based image fusion with complex wavelets,” Information Fusion, vol. 8, pp. 119–130, 2007.
2. M. Li,W. Cai, and Z. Tan, “A region-based multi-sensor image fusion scheme using pulse-coupled neural network,” Pattern Recognition Letters, vol. 27, no. 16, pp. 1948–1956, 2006.
3. Yu Liu et.al, “Multi-focus image fusion with a deep convolutional neural network,” Information Fusion, vol. 36, 2017.
4. Han Tang et.al, “Pixel convolutional neural network for multi-focus image fusion,” Inf. Sciences, vol. 433-434, pp. 125–141, 2018.
5. W. Huang et.al, “A new pan-sharpening method with deep neural networks,” IEEE Geosc and Rem Sens Lettrs, vol. 12, pp. 1037–1041, 2015.
6. S. Li, J. Kwok, Y. Wang, Multifocus image fusion using artificial neural networks, Pattern Recognit. Lett. 23 (8) (2002) 985–997.
7. C. Du et.al, Image segmentation-based multi-focus image fusion through multi-scale CNN, IEEE Access 2017.
8. K.R. Prabhakar et.al, DeepFuse: a deep unsupervised approach for exposure fusion with extreme exposure image pairs, IEEE International Conference on Computer Vision 2017, pp. 4714–4722.
9. L. Zhao et.al, “Multiple description convolutional neural networks for image compression,” IEEE Transactions on Circuits and Systems for Video Technology, 2018.
10. E. Kang and J. C. Ye, “Wavelet domain residual network (WavResNet) for low-dose X-ray CT reconstruction,” 2017.
11. H. Huang, R. He, Z. Sun, and T. Tan, “Wavelet-srnet: a wavelet-based cnn for multi-scale face super resolution,” in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1689–1697, Honolulu, HI, USA, July 2017.
12. S. Zagoruyko and N. Komodakis, “Learning to compare image patches via convolutional neural networks,” in CVPR, 2015, pp. 4353–4361.
13. C. Rhemann et.al, “A perceptually motivated online benchmark for image matting,” in CVPR, 2009, pp. 1826–1833.
14. Tsung-Yi Lin et.al, “Microsoft coco: Common objects in context,” in ECCV, 2014, pp. 740–755.
15. A. Mittal, A. K. Moorthy and A. C. Bovik, "No-Reference Image

Quality Assessment in the Spatial Domain," in IEEE Transactions on Image Processing, vol. 21, no. 12, pp. 4695-4708, Dec. 2012. doi: 10.1109/TIP.2012.2214050.
16. Talebi, H., & Milanfar, P. (2018). NIMA: Neural Image Assessment. IEEE Transactions on Image Processing, 27, 3998-4011.
17. Qiang Zhang and Bao long Guo, "Multifocus image fusion using the nonsubsampled contourlet transform," Sig Proc, vol. 89, no. 7, pp. 1334 – 1346, 2009.
18. B. Yang and S. T. Li, "Multifocus image fusion and restoration with sparse representation," IEEE Transactions on Instrumentation and Measurement, vol. 59, no. 4, pp. 884–892, 2010.
19. Y. Liu et.al, "Multi-focus image fusion with dense sift," Inf. Fusion, vol. 23, pp. 139–155, 2015.
20. Ma, Haoyu et.al. (2019). Boundary Aware Multi-Focus Image Fusion Using Deep Neural Network. arXiv:1904.00198.

AUTHORS PROFILE

M. Sobhitha, studying M.Tech in Computer Science, she completed her B.Tech from JNTUA University. Her Interests include Data Science, Image Processing. She can be reached at sobhithamallapu009@gmail.com.

C. Shoba Bindu, Ph.D in CSE from JNTUA, Anantapuramu, Andhra Pradesh. She is currently working as Professor in department of CSE, JNTUA. Her research areas include Network Security, Data Mining , Data Analytics and Cloud Computing. She can be reached at shobabindhu@gmail.com.

E. Sudheer Kumar, pursuing Ph.D from JNTUA University, Anantapur, received B.Tech and M.Tech in Information Technology from JNTUA, Ananthapuramu, India. His research interests include Data Science, Software Engineering, Software Architecture, and other latest trends in technology. He can be reached at sudheerkumar.e@gmail.com.