Industrial Engineering Journal
ISSN: 0970-2555
Volume : 49, Issue 7, No. 1, July : 2020

APPLYING DEEP LEARNING TECHNIQUES, MULTI-FOCUS FUSION INCLUDING ADAPTIVE AND GRADIENT JOINT CONSTRAINTS BEGAN

P.Rajesh, P.Bhargavi, T.Ajay Sagar, Dept of Electronics and Communication Engineering, Sree Venkateswara College Of Engineering, Nellore (Dt), Andhra Pradesh, India.
Mallishetty. Praveen Kumar, Dept of Computer Science and Engineering, Sree Venkateswara College Of Engineering, Nellore (Dt), Andhra Pradesh, India.

ABSTRACT
Image fusion aims to combining data by many images into a single image that, in theory, incorporates all the essential elements from each of the initial images. The imaging system's limited depth of field makes it challenging to retrieve all the useful information from a single image. Multi Focus Fusion (MFF-GAN), a generative adversarial network, is used in digital photography to combine images with very different focal points in order to reduce the Defocus Spread Effect (DSE)by making emphasis maps wherein the primary focus is correspondingly larger than the items. This framework includes an adaptive judgement block to decide when source pixels concentrate or not, depending on the difference of repeated blur. By extracting and reconstructing data, our technology enables multi-focus picture fusion, which almost eliminates blurring and feature loss near the boundary. The current approaches that make use of explicit and focused pictures are known as deep learning techniques. Numerous applications, including Multi Focus Image Fusion, use deep learning.

Keywords: Deep Learning, Generative Adversarial Network, Multi Focus Image fusion.

1. INTRODUCTION
Multi Focus Fusion is a method for fusing two pictures into one by emphasising the textures' fine details. Without removing any artefacts, it combines the key elements of many photos Within a single combined image. Multi-focus image fusion, the key step in the process of fusion that aims to increase the depth of field, is crucial by removing focused portions from several multi-focused pictures.
The two techniques employed in multifocus fusion are the spatial domain technique and the frequency domain technique. The spatial approach works by the pixel values of the input images, where the pixel values are changed to get a desirable result. Fusion methods like Weighted Averaging and the Selective Maximum Method are used in this sector. Each pixel in the source images is given a weight via the weighted averaging process, and each pixel value is added together to create the final image. To produce a fused image, the Selective Maximum Method chooses high intensity pixel values from images. The picture is initially moved into the frequency domain in frequency domain techniques, which implies that the image's fourier transform is computed first. All fusion procedures are conducted on the image's fourier transform before doing the inverse Fourier transform. This field includes techniques like Discrete Wavelet Transforms and Wavelet Base Methodology. We prefer to suggest a Generative Adversarial Network (GAN) architecture in this study. Our approach uses GANs, with the fundamental goal of learning from a collection of training data and creating new data having the same characteristics as the training data. Generative modelling is an unsupervised learning job in machine learning. The discriminator and the generator are the two primary parts of the GAN. The generator has been taught to create fictitious data from a random source. The differentiation is taught to differentiate between information from the generator and actual data.

2. RELATED WORK

The linked study involves the invention of generative adversarial networks (GAN) as well as the description of the history and current tactics used in this approach, which are all deep learning-based. Deep learning [1] has recently emerged as the most beneficial technique for research as well as a number of image processing applications, including medical ones. Deep learning is an instance of machine learning, which is a subset of artificial intelligence. Deep learning will generate clever generalization-capable fusion models from an excessive amount of data, making the fusion process more reliable. Supervised, semi-supervised, and unsupervised deep learning were the three categories.

The methodologies that were already in use emphasised the creation of precise decision maps. Convolution Neural Network (CNN) [2] manually creates the decision maps that categorise focussed and defocused areas, resulting in fusion pictures. A new approach is also suggested. In this technique, detection of the decision map process is used to distinguish among the focused and defocused portions of the input pictures using image segmentation depending multi focus image fusion [3]. Although both approaches improve the choice maps' accuracy, the output images would still lose some of their detail.

3. METHOD

3.1 GENERATIVE ADVERSARIAL NETWORK (GAN)

In order to create clear pictures from the source photos in this study without any loss, Generative Adversarial Network (GAN) [4] [5] is employed. Unsupervised learning is used by generative models; the input is used to train the kind, and it uses the training information to identify patterns and produce the output. Generator network creates a sample of data after taking a sample.

Deep convolutional GAN (DCGAN) and least squares GAN (LSGAN) are two GAN variations that are most similar to our approach. Convolution Neural Network (CNN) and GAN are used in our technique, DCGAN [6], to more effectively complete the picture fusion job. With the exception of the output layer, all layers of the generator and discriminator in DCGANs are activated by rectified linear units (ReLUs), Every of the against this layers are activated using leaky ReLUs, and fully interconnected layers are the last to be dropped in models with greater complexity. The DCGANs give the GAN access to the powerful CNN extraction of features abilities.

An addition to the GAN design that solves the issue of vanishing gradients and loss saturation is the Least Squares Generative Adversarial Network (LSGAN) [7]. It is a kind of generative adversarial network that uses the discriminator's least squares loss function.

The generative model examines the distribution of data in such a manner that the laplacian operator, using the maximum selection approach, discovers the joint gradient maps of the input data. Joint gradient is described here as genuine data, while gradient map is defined as false data of the fused picture.

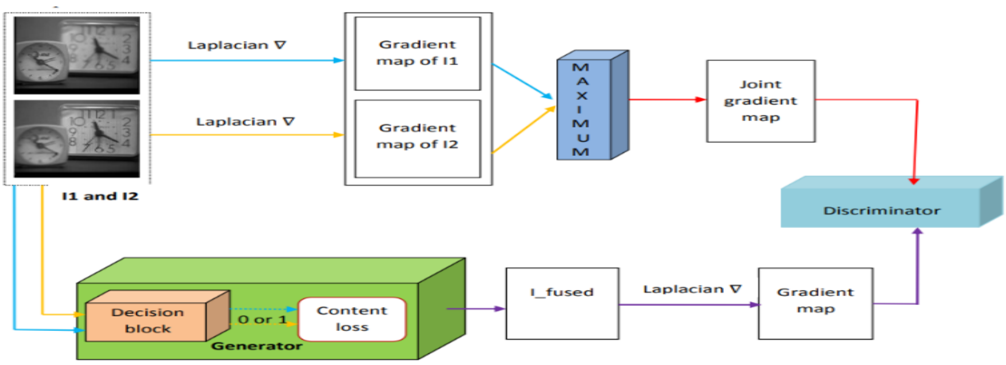


Figure 1: MFF-GAN Overall Fusion Frame

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 49, Issue 7, No. 1, July : 2020

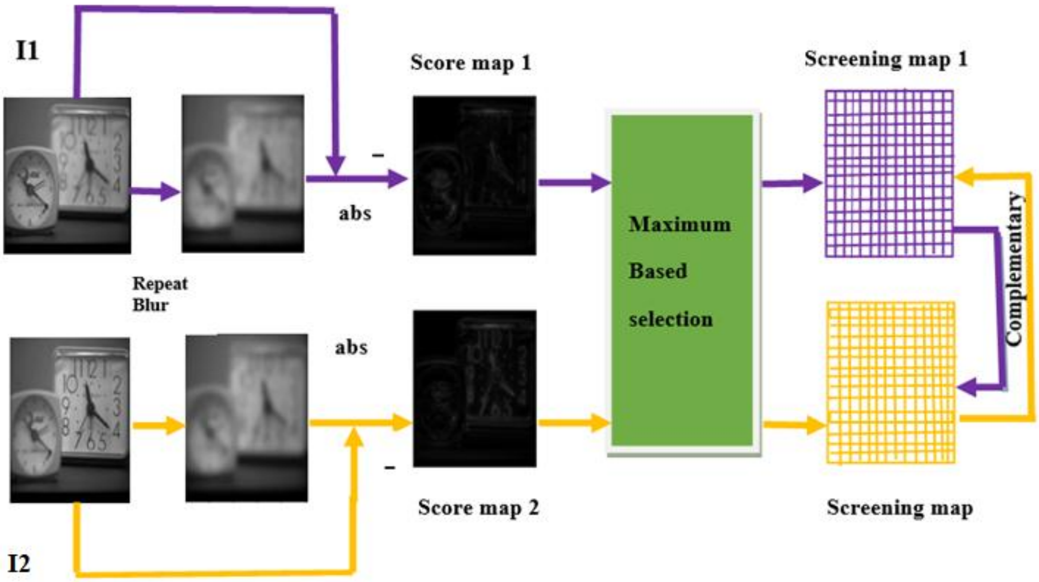


Figure 2: Decision Block

The two primary parts of GAN are the generator and discriminator. The laplacian operator from Figure 1 is applied to the input data to create gradient maps for the source pictures. The selected maximum approach is used to construct the joint gradient map. The decision block and contented loss make up the generator block, which generates score maps 1 and 2 based on the repeated blur theory. The screening map1 is produced using the maximum approach, and the screening map2 is its counterpart. These maps create fused images, and after applying Laplacian to the combined images, gradient maps are created. As the final step, the discriminator separates the joint gradient map from the gradient map to distinguish between fake and real data, allowing us to obtain the fused result with rich texture details.

3.2 EVALUATION METRICS

The Lytro data set and evaluation parameters were utilised to assess the quantitative analysis [8]. The indicators employed in this approach to calculate objective analysis are

3.2.1 Standard Deviation (SD): The fused picture contrast is estimated via standard deviation. This measure depicts how the image's pixel values are distributed, or how far apart they are from one another on average

$$SD = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(F(i,j)-\mu)^2}$$

Where M,N represents the sized of the image. μ represents the average value of the pixel. The larger opposition then more will be the SD.

3.2.2 Entropy: Entropy is an estimation of how much data a fusion image holds; the higher the value of EN, the more data it includes.

The EN reads this way,

3.2.3 Spatial Frequency: This metric evaluates the frequency in the fused image represents the whole activity level.

$$SF = \sqrt{RF^2 + CF^2}$$

here the column's frequency is CF and the row frequency is RF.

3.2.4 Edge information: The quantity of edge information [9] transmitted through the original image to the fused image is measured using this measure. Mathematically,

$$Q^{AB/F} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N} Q^{AF}(i,j)w^{A}(i,j) + Q^{BF}(i,j)w^{B}(i,j)}{\sum_{i=1}^{M}\sum_{j=1}^{N}(w^{A}(i,j) + w^{B}(i,j))}$$

Where edge capability and orientation values at coordinates (i, j) are indicated by QAF and QBF.The weights wA and wB represent the relative relevance of every source picture to the merged image. A high QAB/F indicates that the fused picture has received a significant amount of edge information.

3.2.5 Visual Information Fidelity (VIF): The content fidelity of the combined image is measured using this metric, That's in line with the way the visual system in humans works. VIF attempts to determine the distortion that exists among fused and source images using four steps. The source images are first filtered, and then the composite image is divided into numerous blocks. Second, with and without distortion are contrasted to every block's visual data. Next, the VIF for every sub band is calculated. The overall measure determined by VIF is then calculated.

3.2.6Sum of correlationsof differences (SCD): Correlation The correlation among the data sent to the fused image and the associated source image is measured by the coefficient.

$$r_{(X,F)} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(X(i,j) - \overline{X})(F(i,j) - \mu)}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(X(i,j) - \overline{X})^2 \sum_{i=1}^{M}\sum_{j=1}^{N}(F(i,j) - \mu)^2}}.$$

4. PROPOSED METHODOLOGY

The content of the image fusion in our research is to extract and fuse the most significant data from the sharp sections of the source pictures. We presented a GAN architecture. The data in the sharp sections should be ignored throughout the picture extraction procedure. Based on the aforementioned factors, we develop a generative adversarial network by adaptive and gradient joint restraint in order to alter the loss function during the optimisation process and minimise the smoothing impact while reinforcing the outcome.

4.1 Loss Functions

4.1.1 Loss function of Generator

The loss function of generator is Lg. It is the combination of contented loss and the adversarial loss. Mathematically it was shown that, $L_G = L_{G\text{adv}} + \alpha L_{G\text{con}}$

Whereα is utilized to alter the two loss terms for the same level of value.

4.1.2 Loss function of Differentiator

Ld is the discriminator's loss function. The differentiator can distinguish between bogus and authentic data because to its loss function. The false data in the suggested technique is a combined gradient map of the amalgamated picture and the true information is a gradient map made up of individual fused images. This procedure is carried out using the maximum selection approach.

5. ARCHITECTURE OF GAN

5.1 Generator Architecture

The generator is separated into two pathways in order to extract the information from the original photos. The pseudo-Siamese network is used in the generator network's architecture, and it is effective at handling two diverse inputs. The Pseudo-Siamese network is appropriate for photos with crisp or blurry pixels. Four convolutional layers are present on each of the generator network's two routes. Leaky ReLU is utilized as the stimulation function in each convolution layer to turn the layers on. Each convolution layer's input is the concatenation of the layers before it. The two routes are

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 49, Issue 7, No. 1, July : 2020

combined to create the final output, which uses a convolution layer to create the fused picture.

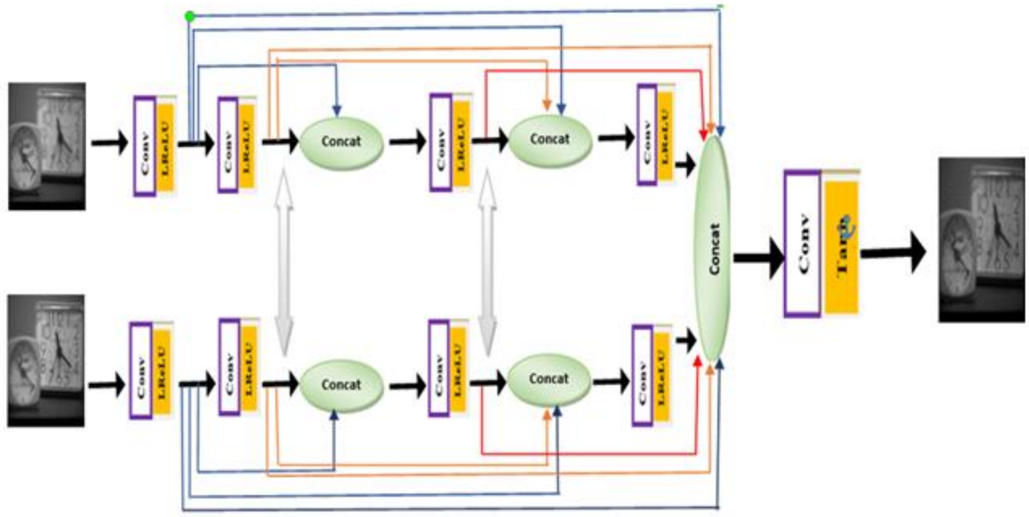


Figure3: Architecture of Generator Network

5.2 Discriminator Architecture

The gradient map of the source image and the joint gradient of the source image serve as the discriminator's two inputs. The joint gradient map uses the most effective technique to create the outcome. Leaky ReLU and four convolutional layers are used in the discriminator. The approximation of false data and genuine data is described in the probability block at the output.

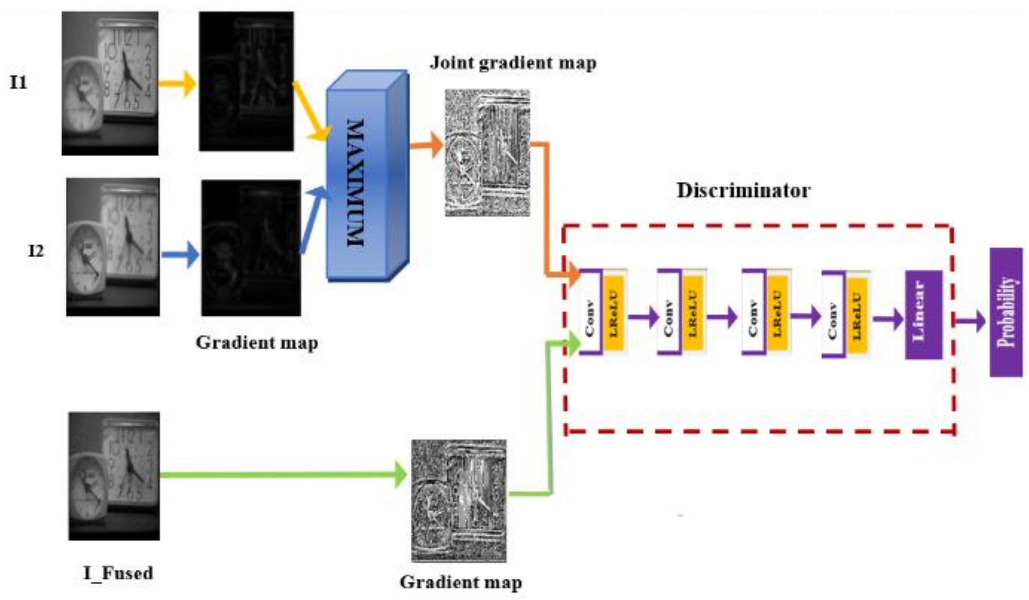


Figure 4: Architecture of Discriminator Network

6. RESULT

The result of the method can be discussed and evaluated using Qualitative analysis and Quantitative analysis for Lytro data set.

6.1 Qualitative Analysis

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 49, Issue 7, No. 1, July : 2020

Subjective human visual perception is a prerequisite for qualitative analysis. In this study, the qualitative analysis is evaluated using the Lytro data set. We may infer from the findings that MFF-GAN provides a number of benefits.The boundary lines of focussed and defocused parts of the source pictures may be reliably retained using our technique. MFF-GAN can better keep texture features, such as at boundary lines.

Figure 5: Weighted Average

Figure 6: Maximum Fused Image

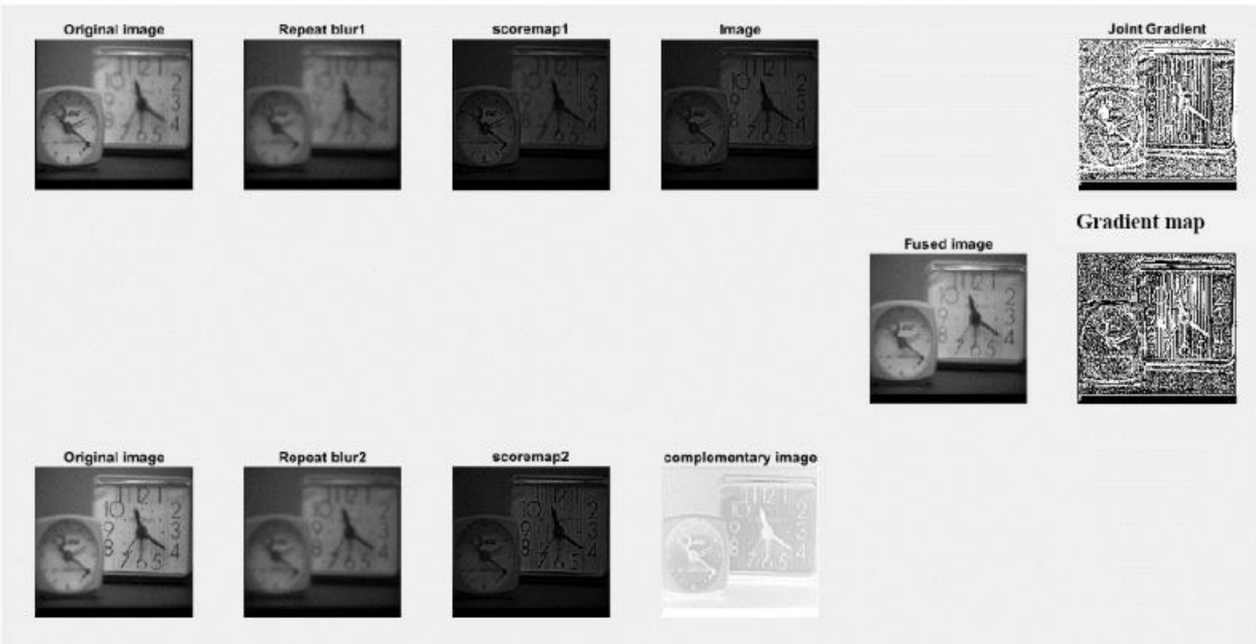


Figure 7: Step by step outputs of decision block for clock images

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 49, Issue 7, No. 1, July : 2020

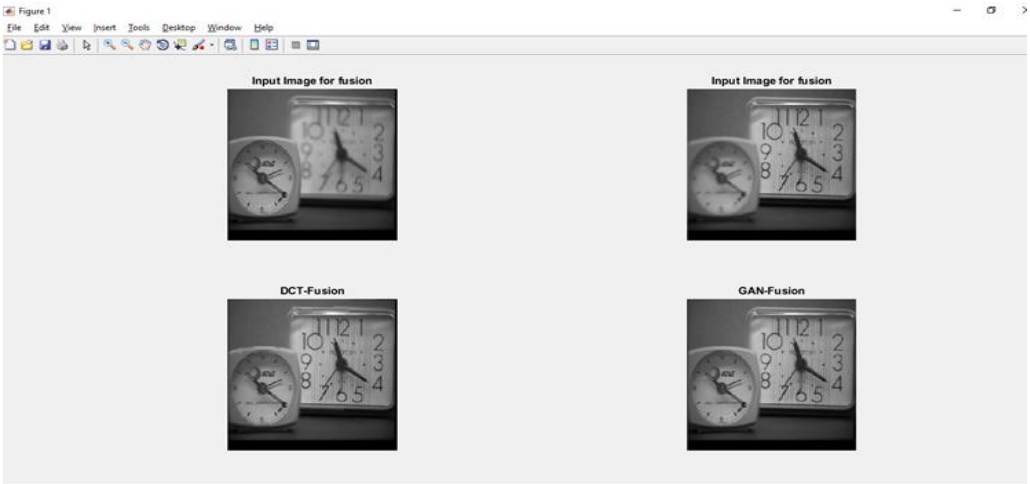
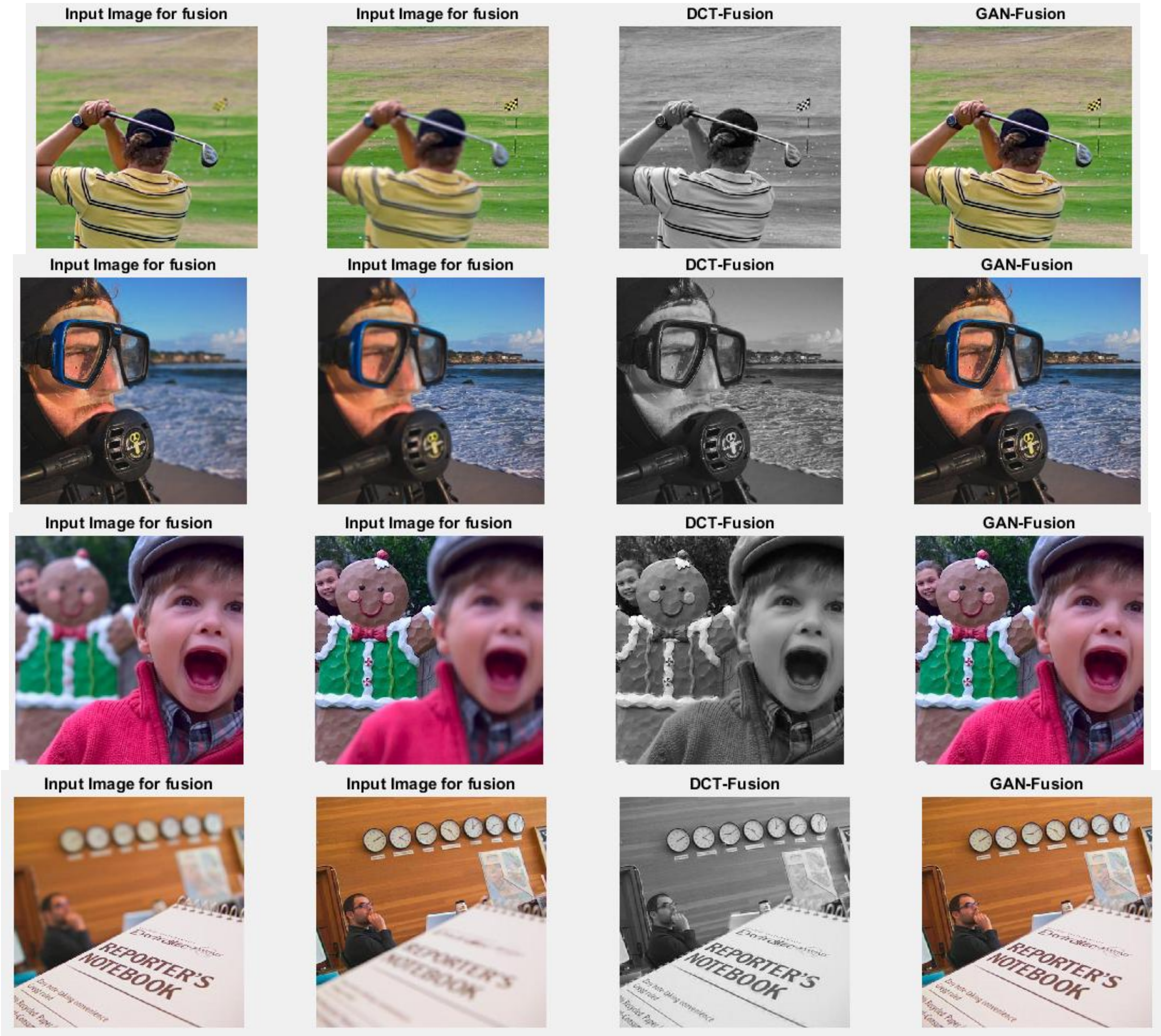


Figure 8: Fusion results of DCT and GAN for clock images



Industrial Engineering Journal
ISSN: 0970-2555
Volume : 49, Issue 7, No. 1, July : 2020

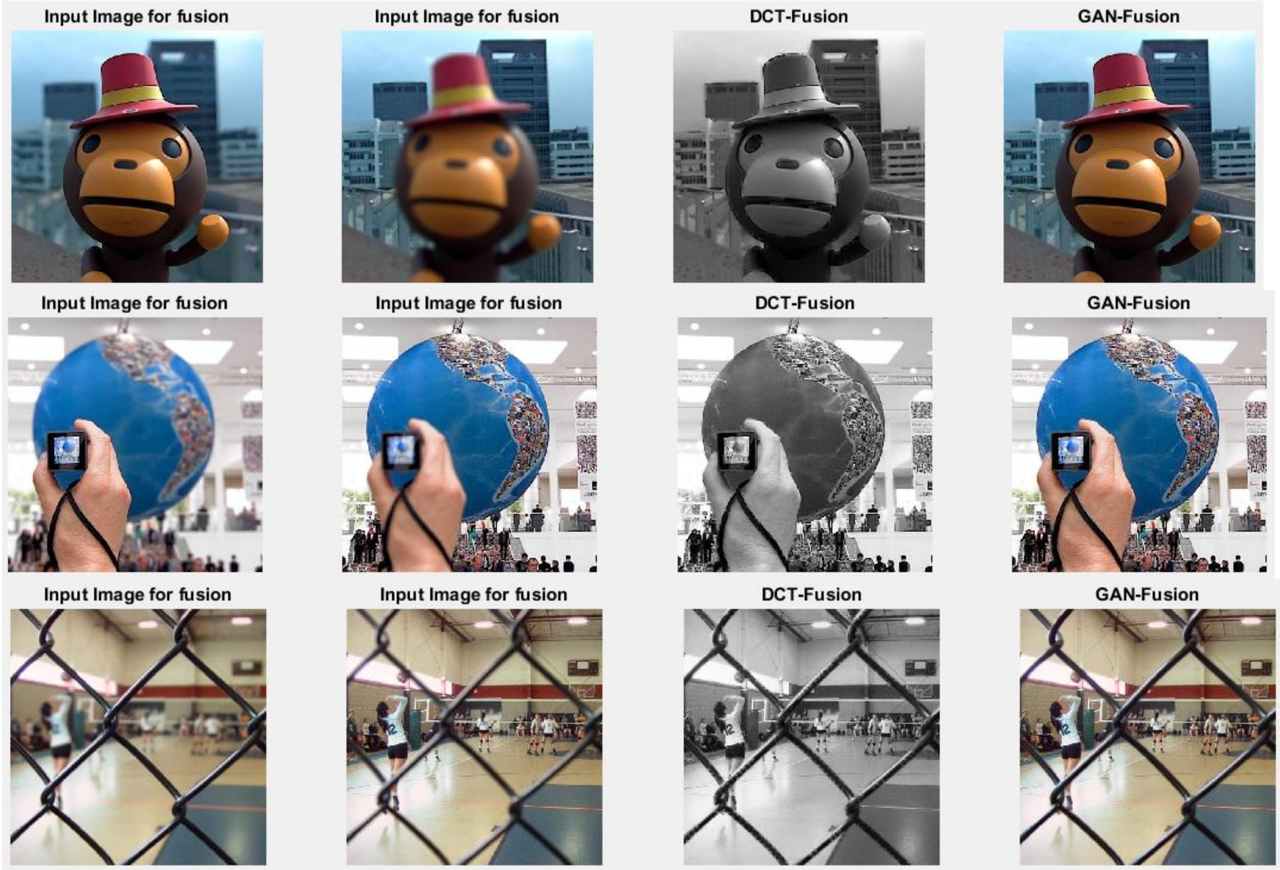


Figure 9: Fusion results of DCT and GAN for the few pairs of lytro dataset

Figures 5 and 6 demonstrate the maximum output from the source photos using the averaging and selected maximum methods, respectively. The selective maximum approach picks the highest pixel values from the source images and generates the output. The outputs of the decision block are shown in Figure 7 where the score maps are formed using the repeated blur concept, the screening map1 is made using the selected maximum approach, the screening map2 is produced by complementing the screening map1, and a fused picture is created. Results from the subjective analysis are described above.

6.2 Quantitative Analysis

In order to evaluate the quantitative analysis (Objective Analysis) we have used the Lytro data set and Evaluation metrics.

The six widely used statistics as objective metrics, which have already been described in related work to quantify the results of the merger of the DctVar and GAN, include: Standard Deviation (SD), Entropy (EN), Measures how much edge data is transmitted by source pictures to the fused image. Q(ABF), Spatial Frequency (SF), Visual Information Fidelity (VIF), and Sum of the Correlations of Differences (SCD) are some examples.

Metrics	Weighted average	Selective Maximum Method	DCT	GAN

Standard Deviation (SD)	0.3834	0.4587	0.4664	0.4668
Entropy(EN)	5.3328	7.2502	7.3119	7.3164
Spatial Frequency(SF)	0.0088	0.0088	0.0088	0.0088
Edge information(Q)	0.1131	0.1074	0.0449	0.1096
Visual Information Fidelity(VIF)	0.7813	0.7921	0.9158	0.8977
Sum of correlations of differences(SCD)	0.9641	0.9652	0.4895	0.9759

The outcomes of quantitative measurements show that GAN has a greater efficiency when compared to other approaches like weighted average, selected maximum, and DCT [10]. We may assess metrics for more picture pairs in the Lytro dataset in the same way.

CONCLUSION

An unsupervised generative adversarial network by gradient joint constraints and adaptive restraint is suggested in this effort to fuse multi-focus pictures. For focus detection in pixel units, an adaptive decision block depending on the continual blur concept has been used. Both qualitative and quantitative analyses have been done to determine how successful the fusion is. In order to validate the effectiveness of the GAN, the consequence are associatedby those from another approaches, such as the Discrete Cosine Transform (DCT). Our approach is roughly an order of magnitude quicker than the alternatives. The experimental findings demonstrate that this method's fusion output, the GAN, reports significant improvement in the subjective visual effects.

REFERENCES

[1] Liu, Yu, et al. "Deep learning for pixel-level image fusion: Recent advances and future prospects."
[2] Y. Liu "Multi-focus image fusion with a deep convolutional neural network Inf. Fusion"(2017).
[3] Du C, Gao S (2017) Image segmentation-based multi-focus image fusion through multi-scale convolutional neural network. IEEE Access 5:15750–15761.
[4] Jun Huang, Zhuliang Le, Yong Ma, Xiaoguang Mei & Fan Fan (2020) A generative adversarial network with adaptive constraints for multi-focus image fusion.
[5] X. Guo, R. Nie, J. Cao, D. Zhou, L. Mei, K. He, Fusegan: Learning to fuse multi-focus image via conditional generative adversarial network, IEEE Trans.Multimed. 21 (8) (2019) 1982–1996.
[6]TT Teoh, Z Rong - Artificial Intelligence with Python, 2022 - Springer. Deep convolution generative adversarial network (DCGAN).

[7] Xudong Mao, Qing Li, Haoran Xie, Raymond Y.K. Lau, Zhen Wang, Stephen Paul Smolley; Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2794-2802: Least Squares Generative Adversarial Networks.
[8] Jagalingam Pa, Arkal Vittal Hegdeb. A Review of Quality Metrics for Fused Image.
[9] Animesh Sengupta; Ayan Seal; Chinmaya Panigrahy; Ondrej Krejcar; Anis Yazidi -Edge Information Based Image Fusion Metrics Using Fractional Order Differentiation and Sigmoidal Functions.
[10] M.B.A. Haghighat, A. Aghagolzadeh, H. Seyedarabi, Multi-focus image fusion for visual sensor networks in dct domain, Comput. Electr. Eng. 37 (5) (2011) 789–797.