Intelligent Automation & Soft Computing
DOI:10.32604/iasc.2022.020918
Article

Tech Science Press

Classification Similarity Network Model for Image Fusion Using Resnet50 and GoogLeNet

P. Siva Satya Sreedhar[1,*] and N. Nandhagopal[2]

[1]Faculty of Information and Communication Engineering, Anna University, Chennai, 600025, India
[2]Department of Electronics and Communication Engineering, Excel Engineering College, Namakkal, 637303, India
*Corresponding Author: P. Siva Satya Sreedhar. Email: sivasatyasreedhar@gmail.com

Received: 13 June 2021; Accepted: 14 July 2021

Abstract: The current trend in Image Fusion (IF) algorithms concentrate on the fusion process alone. However, pay less attention to critical issues such as the similarity between the two input images, features that participate in the Image Fusion. This paper addresses these two issues by deliberately attempting a new Image Fusion framework with Convolutional Neural Network (CNN). CNN has features like pre-training and similarity score, but functionalities are limited. A CNN model with classification prediction and similarity estimation are introduced as Classification Similarity Networks (CSN) to address these issues. ResNet50 and GoogLeNet are modified as the classification branches of CSN v1, CSN v2, respectively, to reduce feature dimensions. IF rules depend on the input dataset to fusion the extracted features. The output of the fusion process is fed into CSN v3 to improve the output image quality. The proposed CSN model is pre-trained and Fully Convolutional. At the time of IF, consider the similarities between the input images. This model applies to Multi-Focus, Multi-Modal Medical, Infrared-Visual and Multi-Exposure image datasets, and analyzed outcomes. The suggested model shows a significant improvement than the modern IF algorithms.

Keywords: Image fusion; CNN; CSN; GoogLeNet; ResNet50

1 Introduction

Digital Image Processing (DIP) transforms an image into digital form by doing some operations to get an enhanced image or extract features [1]. In an image processing system, signals are two-dimensional, and signal processing techniques are applied [2]. For Image Enhancement, Image Fusion techniques are convenient. However, image Enhancement is subjective, i.e., only required features are to be enhanced. As a result, unnecessary information may be padding to the image. Therefore, most researchers concentrate on enhancing the image and overlooking Image Restoration, which is objective [3].

The Image Fusion (IF) objective is to obtain the critical features from several input images and merge these features as one integrated image. Any IF model's outcome depends on the input image type, how these input images are processed, and the fusion rules applied [4]. IF perform operations at the pixel level. As a result, IF gives better results when compared to any other image enhancement technique. The quality of any IF technique depends on how features are extracting from input images and finding the similarity between the two input images. The proposed model mainly focus on these two factors while performing the pixel level fusion. IF has grown into our everyday lives. It has a significant role in Health Care, Agriculture, Disaster Management, Mobile Applications, and Remote Sensing. IF algorithms can be listed as two groups, i.e., spatial and transform domain algorithms. Machine Learning (ML) algorithms can process vast volumes of data and train the models [5]. Scholars have used these techniques in their research areas. Deep Learning (DL) is widely used in recent years to model the complicated relationship between data and features extracted from the inputs. DL techniques such as Convolutional Neural Network (CNN), Adversarial Network (AN) produce optimal results when compared to traditional Image Fusion techniques [6]. CNN models are more specific to the type of input image [7]. The main aim is to develop a CNN based IF model that can fuse most input image types without changing the framework.

2 Related Work

DL Models introduced CNN, which leads to a revolution in IF methods [8]. Ma et al. [9] applied CNN on Multi-Focus pictures. Liu Treated Multi-Focus IF as a classification job, and CNN is used for forecast the Focus Map (FM). DenseNet is used to improve the quality of the output image [10]. These two models post-processed the FM and recreated the fused images based on the refined Focus Maps. Vanmali et al. [11] addressed the thermal radiation problem in an Infrared–Visible (I-V) IF with a Hybrid Image Filtering technique derived from the Divide-and-Conquer strategy. Feng et al. [12] use Fully Convolutional Network (FCN) for fusing I-V pictures by applying "Local Non-Subsampled Shearlet Transform (LNSST)" and Average Gradient (AVG) as fusion rule and got the High-Quality visuals, objective assessments. Laplacian Pyramid (LP) and Max-Absolute as fusion rule to fuse I-V images [13]. They got the best results for some open-access datasets. Yin et al. [14] reformulate the Deep Neural Network (DNN) layers as Learning Residual (LeRU) functions and got the optimum image registration results. To fuse the Spatiotemporal Satellite Images, two CNN's are used [15]. These CNN are used to generate the Super-Resolution (S-R) pictures from the Low-Resolution (L-R) Landsat images. Feature extraction and weights are necessary to renovate the fused image. IF algorithms play a critical role in the detection of cancer genes [16]. Reddy et al. [17] explained the need for IF in real life. Sreedhar et al. [18] developed an embedded approach for Image Registration, Hyperspectral (HS) and Multispectral (MS) Image Fusion. They got optimum results compared to previous results but did not pay more attention to Image Registration. However, there is a space where a researcher can work on the issues.

3 Proposed Method

CNN have a modest pre-processing than other image classification algorithms. A CNN has an Input Layer (IL), number of Hidden Layers (HL) followed by an Output Layer (OL). These hidden layers of a CNN have several Convolutional layers (CLs) that convolve with dot product or multiplication. The ReLU layer acts as an activation function. Next, extra CL like Pooling Layers (PL) and Normalization Layers (NL) are present. Some pre-trained CNN's are available, but their training objective is different from image retrieval testing. The pre-trained CNNs ignore the similarities between the two images. As a result, the features learned for classification is not suitable for retrieval. CNN's also integrate similarity learning features. Throughout the training process, the training procedure needs to know whether the same class images are of the same class but do not care about their classes. Similarity Learning (SL) and Class Membership CM) prediction are complement to each other. By merging these two will generates additional features. In this paper, a new CNN model has proposed classification prediction and Similarity Estimation, known as Classification Similarity Network (CSN).

For Classification Branches (CB) of two CSNs, GoogLeNet and ResNet50 are changed. The outcome is to cut the dimension of Feature Vector, speedup retrieval. Total 5 FC layers are present between ResNet50 last PL and the Output Layer and treated as CSN v1. Fig. 1 shows CSN v1.

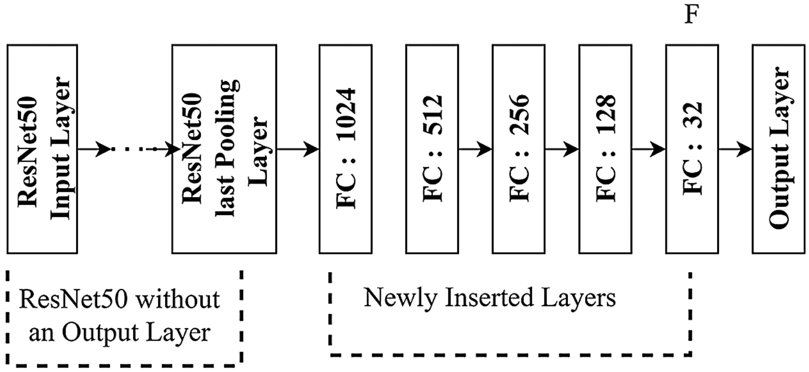

Figure 1: Classification branch architecture of CSN v1

Four Fully Connected (FC) Layers are present in between GoogLeNet Last PL and the Output Layer as shown in Fig. 2 and treated as CSN v2. CSN v3 designed the same as CSN v1. The objective of converting CNN into CSN is to diminish the dimensionality of the feature. At CSN v1, the feature dimension is 32, diminished from 2048. CSN v2 diminishes the feature dimension from 1024 to 32.

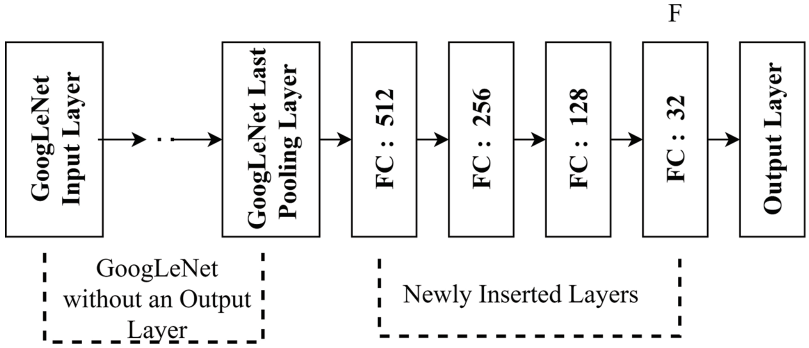

Figure 2: Classification branch architecture of CSN v2

CSN comprises two Classification Branches (CB), which is having the same model, weights. The CSN classifies the input images. Assume CB has n number of classes with Predicted Probabilities as P, P1. The two input images are having Feature Vectors as F, F1. CSN is having one Similarity Learning Network (SLN), and SLN has one Integration Layer (IL), one FC Layer, and one OL (which calculates the Similarity Score). The CSN architecture is in Fig. 3.

Eq. (1) gives the Activation Vector at the Integration Layer.

$$g = (F - F1) \otimes (F - F1) \tag{1}$$

where $\otimes$ performs the "Element-by-Element Multiplication and g is independent of the sequence order of input images.

CSN and FC layer in the SLN have 256 neurons. For each newly inserted layer in CSN, Batch Normalization and Rectified Linear Unit (ReLU) accelerate and regularize the learning. Next, CSN has to left CB and SLN. CB is used to obtain features, and SLN generates Similarity Scores. Later CSN is pre-trained to generate the network parameters for the newly suggested model. For training the proposed model, choose the pair of input images so that all classes must have the exact number of similar and dissimilar image duos. From all classes, these pairs of images are selected arbitrarily for optimum results. The suggested Image Fusion Architecture is in Fig. 4.

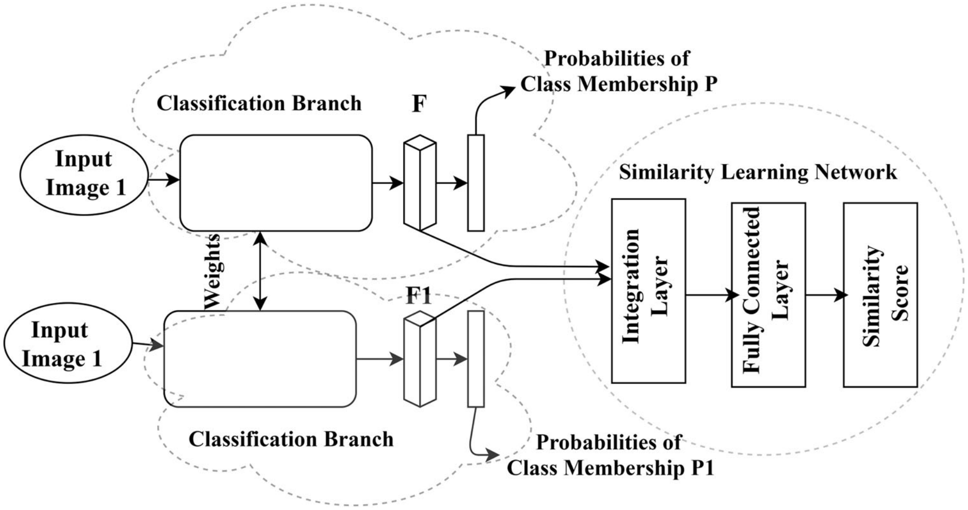

Figure 3: Architecture of CSN

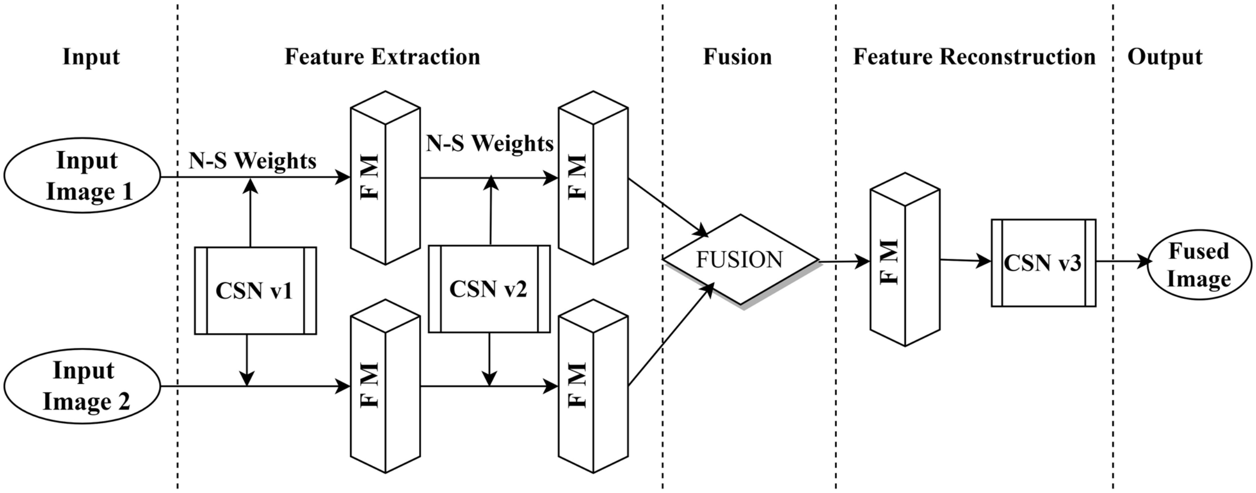

Figure 4: Architecture of suggested Image Fusion model

A minimum of one couple of images (similar or dissimilar) is necessary for the IF process. So, in the Input Phase, two images are shown. In the Feature Extraction Phase, two CSNs extracts the features from the input images. Training CNN for image datasets is rigid. CSN simplifies the training of the CNN. Both CSN v1 and CSN v2 are pre-trained on the input image Dataset. Both CSN v1, CSN v2 had 64 convolutional kernels of size 7 ×7, which is enough to extract extensive image features. CSN v1, CSN v2 extracts the effective image features. The obtained features are not used in the fusion, leading to the loss of features during fusion. Therefore, collected features from CSN v1, CSN v2 are passed once again to Feature Map (FM) before the fusion block for the fine-tuning features in input images. These fine-tuned features of input images have participated in the fusion, and Element-Wise Fusion takes place. It is independent of the number of input images in IF. The output of the fusion block is the input to FM and CSN v3 for fine-tuning of results. The mathematical representation of the suggested Image Fusion is in Eq. (2).

$$\check{R} = Ƒ_i\,(\tilde{N});\ \ 1 \leq I \leq N \tag{2}$$

$\check{R}$ = Fused Feature Maps of the j^{th} channel of image (x,y).

$Ƒ_i$ = Element-Wise Fusion rule on the i^{th} image.

$\tilde{N}$ = i^{th} input image (x,y) j^{th} Feature Map is extracted by CSN v1.

Before training the CSN, optimizing the model parameters with the appropriate loss function for precise outcomes is necessary. Here, Perceptual Loss Functions (P) regularizes the model to produces more structural likeness with the actual time image. Mean Square Error (MSE) of FMs of the expected fused picture and real-time fused picture extracted by the last CL of GoogLeNet [19] is the value of P. Perceptual loss is available in Eq. (3).

$$p = \frac{1}{\left[\hat{c}\hat{h}\hat{w}\right]} \sum_{i,x,y} \left[\dot{P} - \dot{T}\right]^2 \tag{3}$$

P = i^{th} channel Fusion Map of Predicted Image (x,y).

T = i^{th} channel Fusion Map of Ground-Truth Fusion Image.

$\hat{c}$ = Future Map channel number.

$\hat{h}$ = Future Map height.

$\hat{w}$= Future Map width.

P= Perceptual loss.

At initial, Basic Loss (B_l) is the MSE of the expected fused picture and the actual picture for pre-training the model. Then onwards, the sum of proposed Perceptual Loss (P) and the B_l to learn the model shown in Eq. (4).

$$Tl = (N - SW)\ Bl + (N - SW)\ P \tag{4}$$

Here, N-SW represents Nandha–Satya weights, whose value is 1.

4 Experimental Results

MATLAB R2020a and MatConvNet 1.0-beta25 are used to develop this model with Intel Core I7-10700K with 5.00 GHz Processor and NVIDIA TITAN X GPU. The suggested model evaluated on image datasets of M-F, I-V, MM-M, and M-E Image. The primary goal of any IF is, the fused image looks like a natural image. Key points to note down during the IF is in Tab. 1 [20].

The projected IF algorithm efficiency can be measured using the terms in Tab. 2 [21].

Multi-exposure images have more than two sources of images. The suggested model compares the results with the famous IF algorithm based on Guided Filtering (GF_IF) (it is IF generalized approach), Multi-Scale Transform Sparse Representation IF model (MSTSR_IF). In the suggested model has three image fusion rules (IF_Sum, IF_Mean, and IF_Max).

Table 1: Image types and their features to be monitor during the fusion

Image type	The features to be monitor during the fusion
Multi-Focus	Integrate sharp and clear features from source to fused image.
Infrared and Visual	From visual image visible appearance, from the Infrared image bright features.
Multi-Modal Medical	Combine typical features from various modalities into the fused image.
Multi-Exposure	Inject exact middle-exposure features from source pictures into fused pictures.

Table 2: Metrics and its purpose

Metrics	Its purpose
VIFF	To measure the ratio between visual information in the fused images to natural images.
ISSIM	To detect the structural identity of the processed picture with the actual picture.
SF and AG	To find the quantity of textual data that was present in the fused image from the input image.
NMI	How much data carried from the actual picture to the fused picture?

4.1 Multi-Focus Image Fusion (M-F IF)

The NYU-D2 dataset participated in the training of CSN's. The response of the proposed model, GF_IF and MSTSR_IF, are observed. Fusion images of GF_IF and MSTSR_IF got minor blurring around the fence, minimized in the proposed model with the three-fusion rules.

Fig. 5, the first row is the Far-Focus, Near-Focus images (from left to right). The second row is the fused images of Far-Focus and Near-Focus images. Second-row images are the first row's output images by applying the fusion rule as Mean, Max and Sum (from left to right), respectively. The image zone framed by green color in each sub-figure denotes the image spot close-up marked by a red color. The third row is output images of GF_IF and MSTSR_IF.

Tab. 3 IF_Max gives the best results for M-F IF. Next, IF_Mean, IF_Sum yield good results. "Element-Wise-Maxima" is the best fusion rule for M-F images. Fig. 6 is the char representation of Tab. 3.

Fig. 6 metrics are taking on X-axis, color bars are the values of the metric corresponding to the IF method. A total of five different IF methods have participated in the evolution. Each IF method is assigned a color to visualize better the result (The color box represents the particular method). From Fig. 6, IF_Max is the best for M-F IF.

4.2 Infrared - Visual Image Fusion

A person is standing on the mountain. The outdoor scene captured using the Infrared-Visual image format. Fig. 7, the first row has Infrared Image and Visual Image (from left to right). Second-row images are the first row's output images by applying the fusion rule as Mean, Max and Sum (from left to right), respectively. The image part framed by a green box in each subfigure denotes the close-up of the image part marked by a red rectangular box. The third row is output images of GF_IF and MSTSR_IF.

Tab. 4, IF_Max gives the best results for M-F IF. Next, IF_Mean, IF_Sum yield good results. "Element-Wise-Maxima" is the best fusion rule for I-V images. Fig. 8 is the char representation of Tab. 4. Fig. 9, the conclusion is that for I-V IF, IF_Max is good. Here also "Element-Wise-Maxima" fusion rule is good.

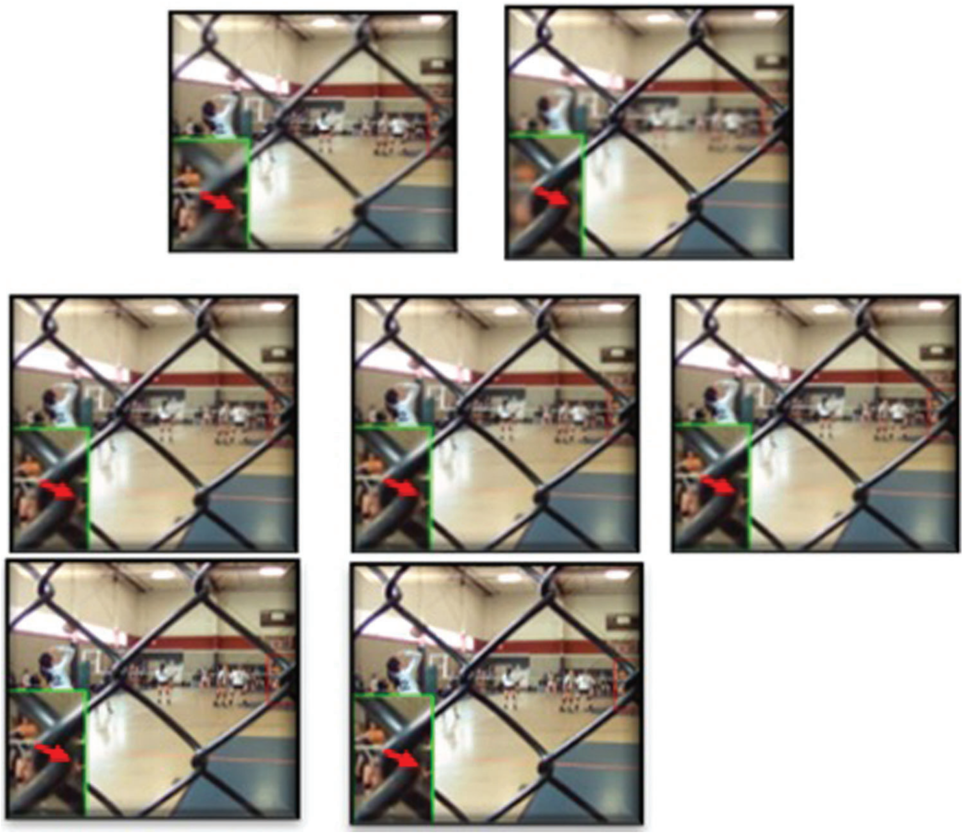

Figure 5: Fusion of multi-focus images results

Table 3: Multi-focus IF result analysis

	GF_IF	MSTSR_IF	IF_Mean	IF_Max	IF_Sum
VIFF	0.9806	0.993	0.9971	0.9981	0.9869
ISSIM	0.6187	0.6177	0.8939	0.7132	0.7913
NMI	1.044	0.9817	1.2198	1.3829	1.1825
SF	19.29	19.41	19.59	19.92	19.71
AG	2.864	2.891	3.7193	3.9701	3.5928

4.3 Multi-Modal Medical Image Fusion

CT and MR scanned images of the human brain are collected together. The fused image should have more skull information from the CT and textural tissue properties from MR. Fig. 9, the first row is the input images MR, CT scan images, respectively (From left to right). The second row is the fused images of the first row images. Mean, Max, and the sum is the Fusion rule applied during the fusion process (from left to right). The image part framed by a green box in each subfigure denotes the image part's close-up marked by a red box. The third row is output images of GF_IF and MSTSR_IF.

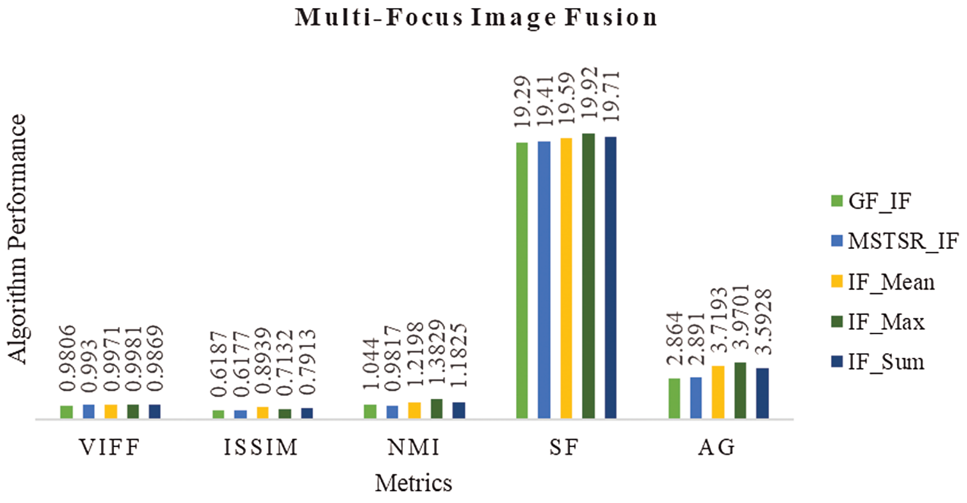

Figure 6: Result comparison of Multi-Focus IF

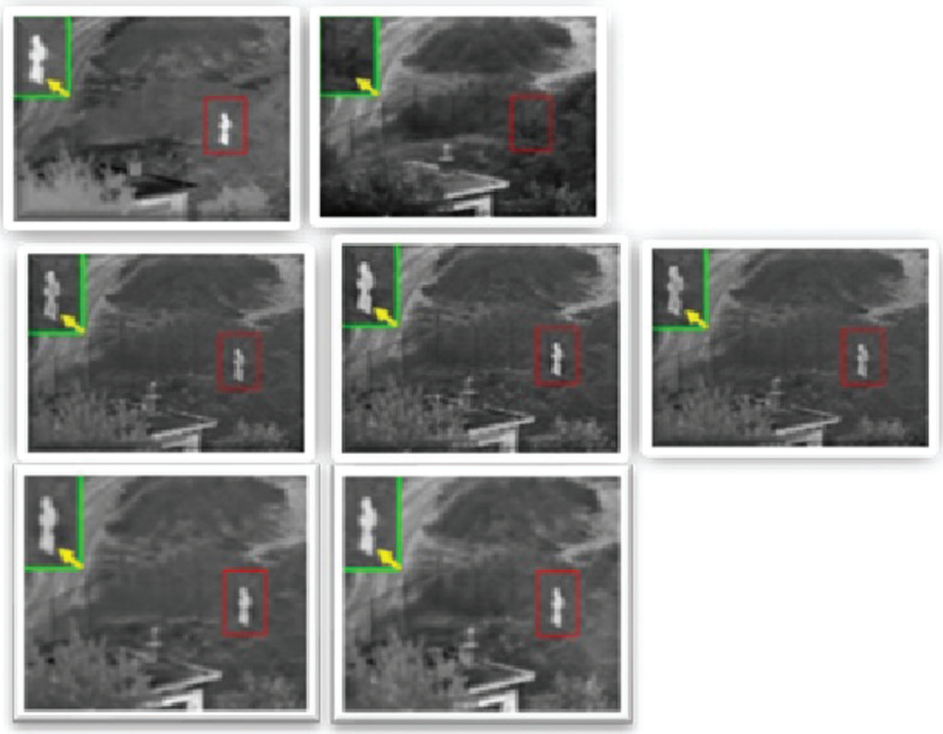

Figure 7: Infrared–visual Image Fusion

The experimental results are in Tab. 5.

Tab. 5 also concludes that for Medical IF also IF_Max gives the best response. Tab. 5 chart is in Fig. 10. Fig. 10 shows that IF_Max is the best suitable IF method over CT and MRI images.

Table 4: Infrared–visual image dataset result analysis

Metrics	GF_IF	MSTSR_IF	IF_Mean	IF_Max	IF_Sum
VIFF	0.5507	0.7327	0.7939	0.9781	0.8173
ISSIM	0.4272	0.4364	0.9714	0.8127	0.9486
NMI	0.3614	0.365	0.4189	0.4918	0.4012
SF	9.623	10.48	11.98	12.98	11.92
AG	1.505	1.678	1.7691	1.9872	1.8976

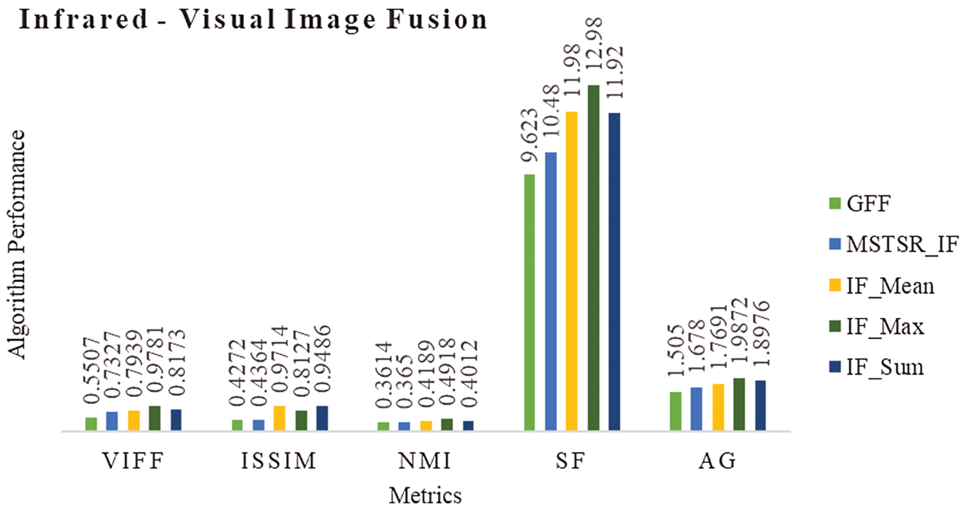

Figure 8: Result comparison of I-V IF

4.4 Multi-Exposure Image Fusion

Three input pictures from low exposure to high exposure and outdoor to indoor participate in the experiment. The fused image should integrate all features in both indoor and outdoor.

From, Fig. 11 The first row is the input images collected from different individual image capturing devices. The second row is the fused images of the first row images. Mean, Max, and the sum is the Fusion rule applied during the fusion process (from left to right). The image part framed by a green box in each subfigure denotes the image part's close-up marked by a red box. The third row is output images of GF_IF and MSTSR_IF. The results are available in Tab. 6.

The outcome from Fig. 12 is that IF_Max is suitable for M-E IF. Furthermore, experimental results conclude that IF_Max is the best fusion rule during IF. Tab. 7 represents the best fusion rule for various types of images used during the study.

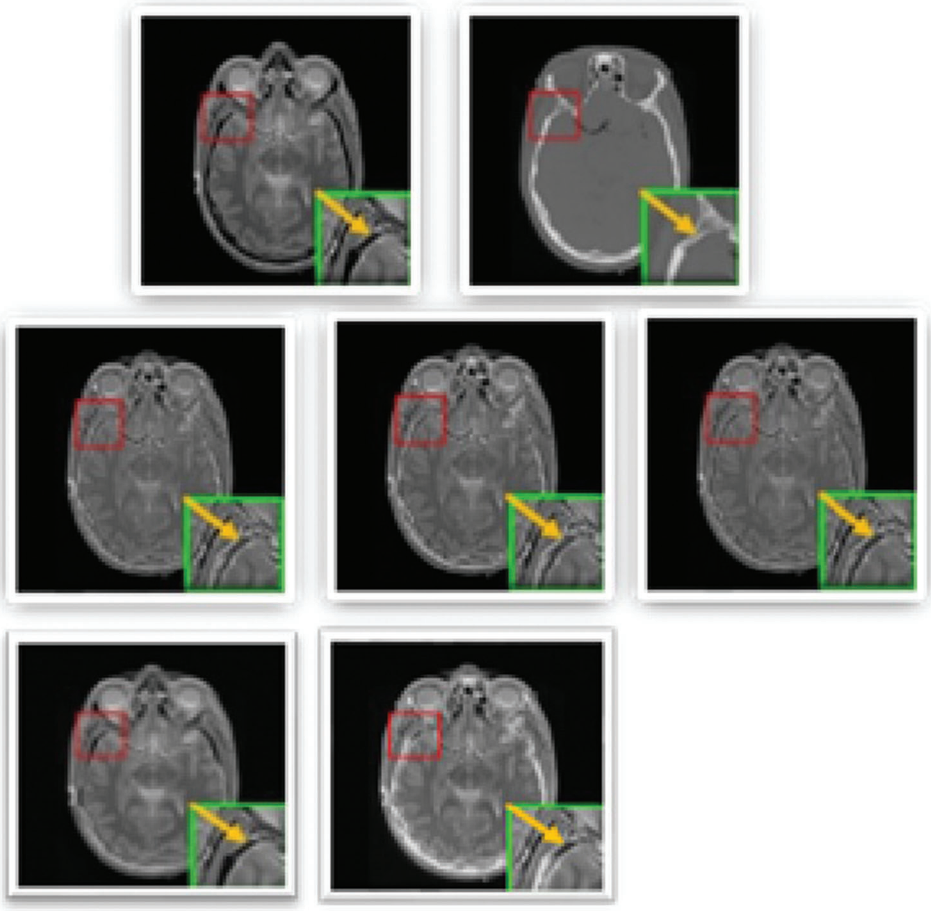

Figure 9: The medical image fusion

Table 5: Medical image dataset result analysis

Metrics	GF_IF	MSTSR_IF	IF_Mean	IF_Max	IF_Sum
VIFF	0.6714	0.7708	0.8976	0.9712	0.8971
ISSIM	0.415	0.4641	0.5871	0.5613	0.5129
NMI	0.6063	0.6883	0.8701	0.9182	0.8914
SF	21.73	24.13	24.83	25.81	24.79
AG	2.892	3.312	3.9812	4.8938	4.3986

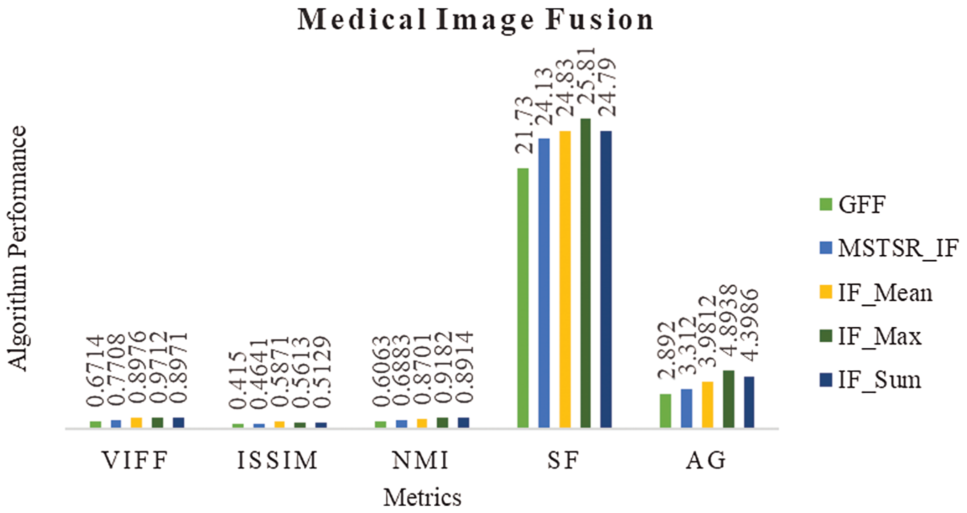

Figure 10: Result comparison of medical IF

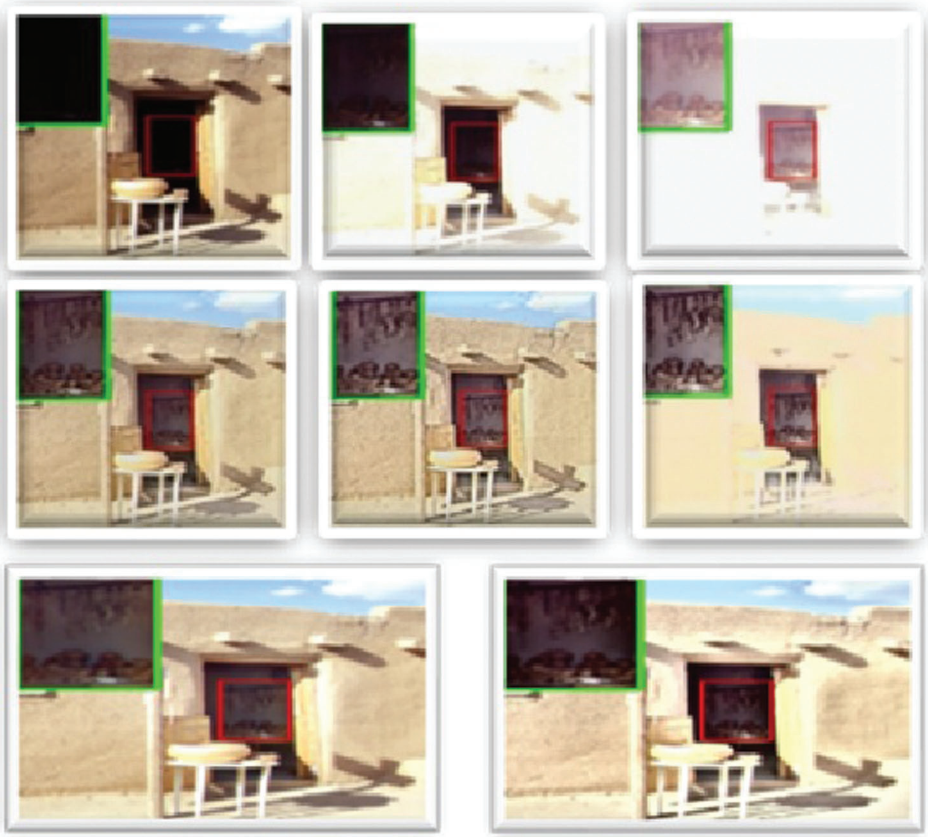

Figure 11: Multi-exposure image fusion

Table 6: Multi-exposure dataset result analysis

Metrics	GFF	MSTSR_IF	IF_Mean	IF_Max	IF_Sum
MESSIM	0.9204	0.8043	0.9986	0.9318	0.8598
SF	25.93	25.98	37.71	39.74	36.97
AG	3.645	3.495	8.697	9.6891	7.9817

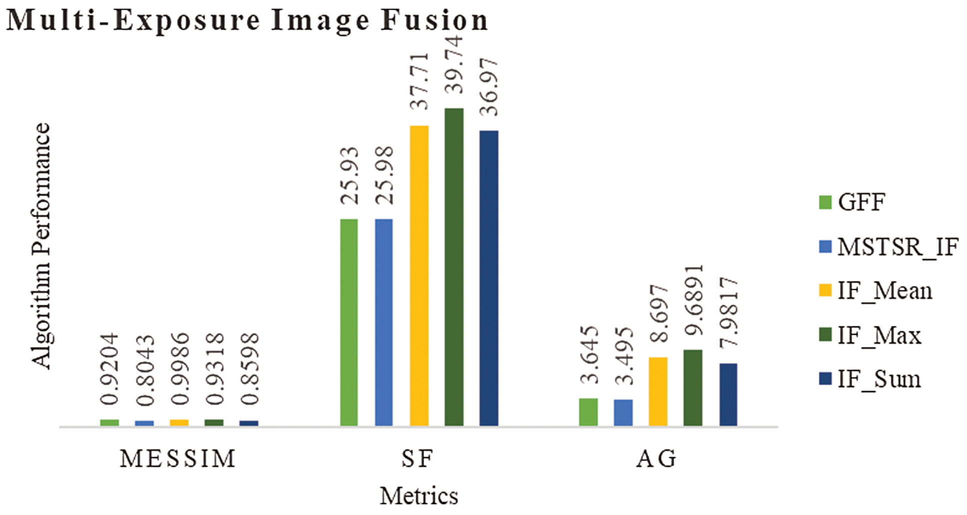

Figure 12: Result comparison of multi-exposure IF

Table 7: Image type and optimum fusion rule to be used

Image type	Optimum fusion rule
Multi-Focus images	Element-Wise maxima
Infrared and visual images	Element-Wise maxima
Medical images	Element-Wise maxima
Multi-Exposure images	Element-Wise maxima

5 Conclusion

The majority of the ML algorithms are focus on training and learning the CNN but less focused on Class Labels, which is the essential one in Image Processing. The suggested model is a fully pre-trained, End-to-End ML-based IF Framework focused on features like Similarity Learning and Class Labeling using CSN. As a result, the test results are more competitive with the current method results.

The suggested model left adequate room for researchers to work on this model further. The researchers go for DCNN instead of CSN. N-SW value is 1 in the present model. Fine-Tuning of N-SW is needed. Integration of Hash Function with CSN may yield competitive results to the suggested model.

Acknowledgement: We are thankful to T. K. K. Praneeth and K. Ashok Reddy for their contribution.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] C. Jaewan, Y. Kiyun and K. Yongil, “A new adaptive component-substitution-based satellite image fusion by using partial replacement,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 1, pp. 295–309, 2011.

[2] Y. Zhou, A. Rangarajan and P. D. Gader, “An integrated approach to registration and fusion of hyperspectral and multispectral images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3020–3033, 2020.

[3] H. Song, Q. Liu, G. Wang, R. Hang and B. Huang, “Spatiotemporal satellite image fusion using deep convolutional neural networks,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 821–829, 2018.

[4] L. Xuecan, H. Pengyu, L. Zhang and S. Jianguo, “MCFNet: Multi-Layer concatenation fusion network for medical images fusion,” *IEEE Sensors Journal*, vol. 19, no. 16, pp. 7107–7119, 2019.

[5] K. A. Reddy, R. Sathya and S. Narayana, “Machine learning algorithms for classification of genetic mutations arose during the cell’s rehabilitation to cancer tumor,” *Solid State Technology*, vol. 64, no. 2, pp. 2752–2758, 2021.

[6] S. Vinay and G. Sharma, “A hybrid approach of image fusion using modified DTCWT with high boost filter technique,” *International Journal of Computer Applications*, vol. 117, no. 5, pp. 22–27, 2015.

[7] R. Hashemzehi, S. J. S. Mahdavi, M. Kheirabadi and S. R. Kamel, “Detection of brain tumors from MRI images base on deep learning using hybrid model CNN and NADE,” *Biocybernetics and Biomedical Engineering*, vol. 40, no. 3, pp. 1225–1232, 2020.

[8] G. Zhang, S. Zhao, W. Li, Q. Du, Q. Ran *et al.,* “HTD-Net: A deep convolutional neural network for target detection in hyperspectral imagery,” *Remote Sensing*, vol. 12, no. 9, pp. 1–21, 2020.

[9] W. Ma, Q. Guo, W. Yue, W. Zhao, X. Zhang *et al.,* “A novel multi-model decision fusion network for object detection in remote sensing images,” *Remote Sensing*, vol. 11, no. 7, pp. 1–18, 2019.

[10] D. Xu and Y. Wu, “Improved YOLO-V3 with densenet for multi-scale remote sensing target detection,” *Infrared Physics & Technology*, vol. 20, no. 15, pp. 1–24, 2020.

[11] V. A. Vanmali and M. G. Vikram, “Visible and NIR image fusion using weight-map-guided Laplacian-Gaussian pyramid for improving scene visibility,” *Sadhana - Academy Proceedings in Engineering Sciences*, vol. 42, no. 7, pp. 1063–1082, 2017.

[12] Y. Feng, H. Lu, J. Bai, L. Cao and H. Yin, “Fully convolutional network-based infrared and visible image fusion,” *Multimedia Tools and Applications*, vol. 79, no. 21, pp. 15001–15014, 2020.

[13] N. Chang, K. Bai, S. Imen, C. Chen, W. Gao *et al.,* “Multisensor satellite image fusion and networking for all-weather environmental monitoring,” *IEEE Systems Journal*, vol. 12, no. 2, pp. 1341–1357, 2020.

[14] M. Yin, X. Liu, Y. Liu and X. Chen, “Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled Shearlet transform domain,” *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 1, pp. 49–64, 2019.

[15] M. A. Naji, A. Ali and E. Mehdi, “Ensemble of CNN for multi-focus image fusion,” *Information Fusion*, vol. 51, no. 11, pp. 201–214, 2019.

[16] Y. Sun, B. Xue, M. Zhang, G. G. Yen and J. Lv, “Automatically designing CNN architectures using the genetic algorithm for image classification,” *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 3840–3854, 2020.

[17] K. A. Reddy, R. Sathya and S. Narayana, “Performing uni-variate analysis on cancer gene mutation data using SGD optimized logistic regression,” *International Journal of Engineering Trends and Technology*, vol. 69, no. 2, pp. 59–67, 2021.

[18] P. S. S. Sreedhar and N. Nandhagopal, “Image fusion-the pioneering technique for real-time image processing applications,” *Journal of Computational and Theoretical Nanoscience*, vol. 18, no. 4, pp. 1208–1212, 2021.

[19] B. Lakshmipriya, N. Pavithra and D. Saraswathi, “Optimized convolutional neural network based color image fusion,” in *Proc. of the 2020 Int. Conf. on System, Computation, Automation and Networking (ICSCAN)*, Pondicherry, PY, India, pp. 1–4, 2020.

[20] P. Mhangara, W. Mapurisa and M. Naledzani, “Comparison of image fusion techniques using satellite pour l'Observation de la terre (SPOT) 6 satellite imagery,” *Applied Sciences (Switzerland)*, vol. 10, no. 5, pp. 1–13, 2020.

[21] A. Ilyas, M. S. Farid, M. H. Khan and M. Grzegorzek, “Exploiting superpixels for multi-focus image fusion,” *Entropy*, vol. 23, no. 2, pp. 247–269, 2021.