International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys.13, No.1 (Apr-23)

http://dx.doi.org/10.12785/ijcds/130161

Improved Image Fusion Technique Using Convolutional Neural Networks and The Hybrid PCA-Guided Filter

Nalini S. Jagtap[1,2] and Sudeep D. Thepade[3]

[1]*Pimpri Chinchwad College of Engineering,*
[2]*Dr D Y Patil Institute of Engineering Management and Research, nalinisjagtap@gmail.com*
[3]*Pimpri Chinchwad college of Engineering, sudeepdthepade@gmail.com*

Received 10 Mar. 2022, Revised 26 Dec. 2022, Accepted 6 Feb. 2023, Published 16 Apr. 2023

Abstract: In order to better accurately forecast the diagnostic features, medical image fusion attempts to combine multi-focus and multimodal medical data into an original image. The outcomes of deep learning-based image processing can be visually beautiful. The incomprehensibility of the outcomes in Medical Imaging is a critical topic. This paper offers a feature-level multi-focus, multi-exposure, and multimodal image fusion using a hybrid layer of Principal Component Analysis (PCA) and Guided Filter (GF) to maximize the anatomical details and eliminate significant noise and artefacts. The proposed method utilizes a Convolutional Neural Network (CNN) based network for feature extraction. The original image is initially decomposed using Principal Component Analysis (PCA). A PCA decreases its dimensionality while preserving all of the essential information in the picture in the first stage and produces a revised weight map. A Guided Filter is used at the PCA output to uphold the edges and further augment the features, reducing the ringing and blurring effects. In the third step, a pre-trained CNN network creates a new weight map by extracting critical characteristics from pictures in the input dataset. The output feature map combines the weight maps produced by the GF and CNN, which is further fused with the reference picture to create the fused image output. The contribution of developed method:

- To improvise image quality features by removing noise, ringing and blurring.
- Increase the quality of the image by using the hybrid mechanism for extracting more underlying critical features of images [1].

The estimation is based on three multimodal imaging datasets, including CT-MRI, MRI-PET, and MRI-SPECT. Furthermore, the proposed method excels in existing state-of-the-art techniques in terms of fusion quality. .

Keywords: image fusion, multimodal image, PCA

1. INTRODUCTION

Many medical imaging techniques have been widely employed in recent years to see complementary information. Medical image fusion enhances diagnosis accuracy by combining the information acquired from various multimodal pictures into a single composite image [2], - [3], [1], [4], [5]. Because no one modality is capable of capturing all of a tissue's information, researchers are often faced with a data overload problem. The goal is to retain as much diagnostic data as possible from the input pictures while maintaining observable visual quality and avoiding spatial and phase distortion. Medical image fusion is a subset of biomedical Imaging that uses radiography to identify illnesses.

The joint image filtering technique is used in [6] provides guidance image as the prior image transfers structural details from the guidance image to the [7] target image. Unlike existing methods that solely consider the guidance image, the proposed algorithm can transmit salient components that are compatible with both the guidance and target images selectively. Biomedical Multimodal image fusion plays an essential role since images are captured and combined using various sensors to integrate additional image information [8]. Fused pictures can be generated by integrating data from several modalities - 1. MRI (offers the most profound view, showing the tiniest anomalies), 2. SPECT (integration of CT-PET - provides information about both functional and anatomical areas), 3. CT (produces cross-section of the human body's anatomy), 4. PET (utilizes a small amount of radioactive FDG (fluorodeoxyglucose), tracer reacts to positrons and creates radiation within cells detected) [9].

Many studies on multimodal medical picture fusion have been published in recent years. Fusion approaches are commonly divided into three levels: pixel, feature, and decision level [10], [11]. The transform and spatial domains are often used to build these image fusion abstraction layers. On the other hand, spatial domain approaches have certain disadvantages, like poor spatial contrast and

E-mail address: nalinisjagtap@gmail.com, sudeepdthepade@gmail.com

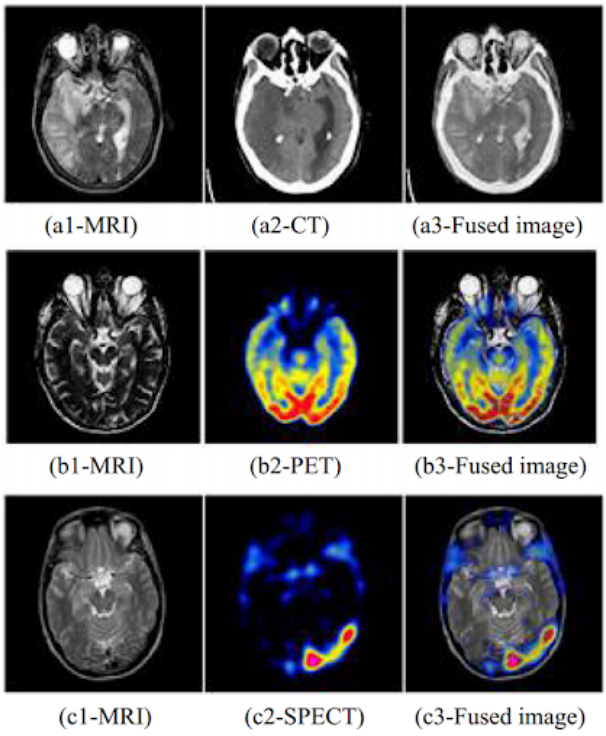

Figure 1. Example of Multi-Modal Biomedical Image Fusion a (1,2). MRI + CT, b (1,2). MRI + PET, c (1,2). MRI + SPECT

localization [12]. To provide a useful platform for better localization of image contour and texture aspects, many proposals have been put forward, such as different multiscale decompositions in the transform domain such as stationary wavelet transform, discrete wavelet transform [13], dual-tree complex WT [14], [15], [16], nonsubsampled shearlet transform [17], [18], curvelet transform [17], nonsubsampled contourlet transforms [19], and, contourlet transform, etc. Under various combinations of fusion rules, the NSCT-based methods produce fairly satisfactory fusion outcomes. Picture fusion using guided filters [20], [21], [22] has also gained popularity since it maintains good spatial consistency, reduces the curved edge, but suffers from staircase effects at image boundaries. Recent developments in neuroscience have demonstrated that deep feature representations can be studied hierarchically, beginning with simplistic ideas like oriented edges and progressing to higher-level complicated designs like segments, parts, textures, and objects. This finding sparked the development of so-called "deep learning" methods, which are now state-of-the-art in several fields. A well-established deep learning technique, the CNN (convolutional neural network) [22], [23] has produced outstanding accomplishments in the field of pattern recognition, which is an advantage for medical imaging and fusion techniques. A CNN-based (convolutional neural network) biomedical image fusion approach performs a critical function in the fusion process from the pixel level to the decision level. Such methods have several drawbacks, chief among them being their implementation complexity and the need for large amounts of data for model training. But the performance of functional CNN can still be made better. This can be done by combining a CNN classification network with different image fusion techniques to make the system work better. In this research, a novel fusion model for multimodal medical pictures is proposed, employing CNN and a combination of PCA-GF. The proposed system captures spatial information effectively, maintains consistency in the spatial domain, and suppresses noise present in the image. The adaptable system presented here can also be used for multi-focus and multi-exposed pictures. The multifunctioning of the system shows the technique's versatility.

1.1 Key Contribution of proposed system

The following are the key contributions of the proposed system:

- At the beginning of the fusion process, Principal Component Analysis (PCA) is used to decrease noise and preserve crucial information in the image.
- To retain the required edges and create a comprehensive weighted map, a guided filter is used. This further reduces the blurring and ringing effects caused by uneven focus.
- To make medical pictures more accurate, pixel activity and fusion weight maps are made using a CNN-based approach and the above hybrid technique.
- The final weight map has enhanced feature preservation since it is a weighted sum of the weight maps created by GF and CNN. This map will be added to the original image to make the final fused image.

The remaining paper is planned as follows: Section (II) gives a Related work of hybrid techniques in image fusion. Section (III) Technical background. Section (IV) outlines the anticipated fusion model along with crucial execution stages, while Section (V) depicts the experimental analysis and outcomes, summarizes the conclusions and future work.

2. Related Work

Creating radiology reports is time-consuming and necessitates extensive competence in the field. As a result, reliable automatic radiology report generation is critical for reducing load. Despite the development of deep learning algorithms for image classification and captioning, radiology report preparation remains challenging in terms of grasping and integrating complicated medical visual contents with accurate natural language descriptions. Additionally, open-access datasets, including medical photos and reports, have very small data sizes. A fresh method has been developed to deal with this problem [24]. Maintenance of gradient and intensity plays an important role in the extraction of image quality features. In order to deal with this aspect, we developed a fast unified fusion network based on proportional maintainance of gradient and intensity (PMGI)

which realizes a variety of tasks including infrared and visible image fusion, multi-focus image fusion, and pan-sharpening [25]. In a number of fusion challenges, both in terms of visual effect and quantitative metric, PMGI did better than the state of the art. Furthermore, as compared to the above available method, our proposed system is more expedited. Due to low scar-to-background contrast and picture quality, quantifying myocardial scarring in late gadolinium-enhanced (LGE) cardiac magnetic resonance imaging can be difficult. With the development of deep learning models for combining LGE and cine images to improve robustness and accuracy of scar quantification [26]. Deep learning has been used to come up with an image fusion method that can keep both structural and functional information. MGMDcGAN not only shows better aesthetic effects, but also preserves the maximum or approximate maximum amount of information in source pictures, according to adequate experimental results on the fusion of MRI-PET, MRI-SPECT, and CT-SPECT [27]. Multiscale DenseNet is proposed by adding a multi-scale mechanism into DenseFlow and applying the improvised method for medical image fusion because DenseFlow only works in a single scale model. A similar approach has proposed multi-scale DenseNet for medical image fusion [30]. A guided image filter is proposed [28] as a way to deal with the problem of low contrast, detail-efficient image fusion algorithms that use multiple silent features. When it comes to analysing image features, it involves a large computational cost and takes a lot of time. To address this problem, an adolescent identity search algorithm is proposed [29]. The real-time image fusion technique has been proposed in [30] which uses pre-trained neural networks to generate a single image containing features from multi-modal sources. The images are merged using a novel strategy based on deep feature maps extracted from a convolutional neural network. These feature maps [31] are compared to generate fusion weights that drive the multi-modal image fusion process. With recent advancements in medical technology, medical picture quality requirements are becoming increasingly demanding. A medical image enhancement method based on convolutional neural networks (CNNs) and frequency band broadening (FBB) is presented to suit clinical diagnostic needs [32].

Motivation for the proposed system encompasses a detailed study of related works in the past and its drawbacks, such as:

1) Absence of multisource image feature extraction in previous studies.

2) Guided filter based feature extraction used in [32], [20] did not ensure image output for reduced noise and blurring effects, as a hybrid approach of PCA-GF with CNN assures a more detailed feature extraction exercise.

3) There have been no previous methods or research areas that have covered the combined approach of image fusion using CNN and a hybrid PCA-GF guided filter.

Implementation of the CNN approach with PCA-GF defines the robustness of the image along with better medical predictions. With the help of PCA-GF, images from different sources can be changed in a good way. This gives people hope when try

3. TECHNICAL BACKGROUND

3.1 PCA

is a vector space transform that is frequently used to reduce multidimensional data sets to fewer dimensions for analysis. Because its purpose is to reveal the fundamental structure of data in an unbiased fashion, PCA is the simplest and most useful of the genuine eigenvector-based multivariate analyses [24], [5]. Covariance is a technique for calculating correlations between random variables. The covariance [33] of two real-valued random variables (A, B) is defined in equitation 1:

$$Cov(A, B) \equiv \langle (A - \langle A \rangle)(B - \langle B \rangle) \rangle \tag{1}$$

3.2 Guided Filter

Like a bilateral filter, the guided image filtering has strong edge-preserving and smoothing qualities. Many functions, such as image augmentation, noise reduction, picture dehazing, and image fusion, benefit significantly from the GF. The filtering result is a direct transformation of the guiding image that is applied locally [[34]

3.2.1 Generalized Concept

- A filtering input image p, a guiding picture I, and an output image q are used in a generic linear translation-variant filtering method. Both I and p are supplied at the input and are assumed to be identical. At pixel I, the filtering output is given as a weighted average as given in equation 2 :

$$q_i = \sum_j W_i j(I) p_j \tag{2}$$

Where pixel indexes are i and j. The filter kernel Wij is independent of p and is a function of the guiding picture I.

3.2.2 Linear Models

- A linear correlation between the guiding image I and the filtering outcome q is the fundamental hypothesis of the guided filter. Equation 3 depicts a window wk with the pixel k as its centre, q is a linear transform of I:

$$q_i = a_k I_i + b_k, \forall i \epsilon w_k \tag{3}$$

a_k, b_k = some Linear Constant Coefficients in window wk. In the local linear model, a square window having a radius of r is utilized. According to the model, q will have an edge only if the guiding image I have an advantage proven by the equation: $\nabla q = a \nabla I$.

This model is beneficial in super-resolutionof the image,dehazing, image matting, etc.
Unwanted noise components n can be subtracted from the output using the expression 4:

$$q_i = p_i - n_i \tag{4}$$

The linear coefficients (a_k, b_k) can be calculated by using the linear ridge regression model of equation 5 and 6:

$$= \frac{\frac{1}{|w|}\sum_{i \epsilon w_k} I_i p_i - \mu_k \bar{p}_k}{\sigma_k^2 + \epsilon} \tag{5}$$

$$b_k = \bar{p}_k - a_k \mu_k \tag{6}$$

μ_k = Mean of I inw_k

σ_k^2 = Variance of I inw_k

$|w|$ = Number of pixels

ϵ = Regularization parameterw_k

$\bar{p}_k = \frac{1}{|w|} p_i$ – Mean of p inw_k

Lastly, using these values of linear coefficient, the final filtering output can be obtained as in equation 7:

$$q_i = \frac{1}{|w|}\sum_{i \epsilon w_k}(a_k I_i + b_k) \tag{7}$$

3.3 Edge Preserving Filter

- The guided filter's edge-preserving filtering feature can be simply stated as follows: Suppose that the guided image I and the input are the same.

$$I = p$$

In this case as given in equation 8,

$$a_k = \frac{\sigma_k^2}{\sigma_k^2 + \epsilon}, b_k = (1 - a_k)\mu_k \tag{8}$$

$$If\ \epsilon > 0 :$$

1) Case I: *High Variance* -
The guide image vary a lot inside the window

$$w_k.\ \text{And}\ \sigma_k^2 >> \epsilon,$$

therefore

$$a_k \approx 1, b_k \approx 1.$$

2) Case II: *Flat Patch* -
Guide I is almost invariable in window

$$w_k.\text{And}\ \sigma_k^2 << \epsilon,$$

therefore

$$a_k \approx 0,\ b_k \approx \mu_k.$$

More crucially, parameter determines whether a patch is flat or has a significantly high variance. The patches having values of σ_k^2 much more minor than ϵ are smoothened and the patches having variance values higher than are preserved.

3.4 CNN (Convolutional Neural Network)

CNN stands for convolutional neural network, and it is a sort of deep learning neural network. Consider CNN to be a machine learning system that can take an input image and assign importance (learnable weights and biases) to various aspects/objects in the image, as well as distinguish between them. CNN works by extraction of various features of image, Any CNN consists of:

a) An input layer is an image with multiple altered aspects such as varied exposure, scale, and source.

b) Output layer which is a binary or multi-class label.

c) Hidden layers consist of convolutional layers, ReLu (Rectified linear unit) layers, the pooling layers, and a fully connected neural network.

4. Proposed Convolutional-Neural-Networks And The Hybrid Pca-Guided Filter Image Fusion

The proposed system employs Convolution Neural Networks in conjunction with a hybrid image fusion technique containing Guided Filters, and PCA. The CNN outperforms conventional ML (machine learning) methods such as ANN, SVM, and Naive Bayes [5]. CNN-based systems accuracy may be increased by intensifying the quality of input data, giving better training, or both. The CNN Network described also plays a significant influence in enhancing outcomes. More layers of CNN simply mean more exceptional training and a more comprehensive calculation of the image's permutations of pixel densities and related characteristics. Images are pre-processed before being input into the convolution network to improve image quality and to locate-retrieve the Region-of-Interest [35], [36]. The proposed architecture shown in figure (2) (3) presents a CNN and Hybrid PCA-Guided Filter Image Fusion approach applied to multimodal and multi-exposed medical images. The system is built to enhance picture quality at every level by using fusion techniques such as:

1) PCA,
2) Guided filter
3) CNN

Using the fused weight maps produced from the GF-CNN layers and the original image, the ultimate output delivers a fused image with all the necessary features without compromising the [3] quality of the pictures.

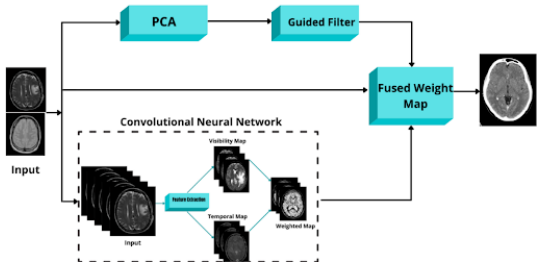

Figure 2. The proposed system

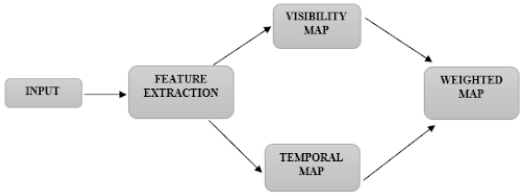

Figure 3. Role of CNN in PCA-GF model

The eigenvalues and the characteristic equation is used to compute the Eigen vector-matrix using the covariance-matrix stated as in equation 9,

$$(\lambda_i - EA) = 0 \tag{9}$$

The column-vector used to represent the weight values in a weight map should be normalized

$$W_T$$

By selecting the column vector with the most significant eigenvalues that correspond to the major components. The PCA technique averages the weighted pixels' average values at each pixel position for all pixels in the reference images. PCA is primarily used to improve the low-frequency components or smooth parts of an image to generate a detailed outcome.

The weight maps derived from various high-level components can be merged as [37] as shown in equation 10.

$$W = \frac{1}{3}\sum_{j}^{3} W_T \tag{10}$$

For Horizontal, Vertical and Diagonal j is stated as

$$j = 1, 2, 3$$

The Pixel's weight at the location (x, y) of an image is determined by the strength of vertical and horizontal edges. Weight map is built using image statistics for guided filtered approximation. This weighted map generated by PCA is given to the guided filter, which reduces the curved edges of the images and makes them smooth. The redefined weight maps (shown in equation 11) can be upsampled via bicubic interpolation and then processed using a guided filter as,

$$W_{GF} \underleftarrow{GF}\ (W, I, \gamma, \epsilon) \tag{11}$$

Where gamma epsilon are smoothing parameters utilized in the guided filter process, the guided filter's output is subsequently used to simultaneously calculate a weighted sum of weight maps generated by GF and PCA.

In the next step, CNN is utilized for increasing temporal consistency and accuracy. The CNN is indeed a hierarchical modal that ultimately generates an FC layer that executes the imaging process. For feature map extraction, the same input images fed into the PCA-GF system are fed into the CNN in a sequence according to a pattern specified by N. Dense features at shallow layers are directed by sparse features at deeper layers with unique high-level spatial information in CNN. The output feature map of a convolution is created when the feature maps of the preceding layer are transformed to learnable kernels using the activation function. Convolutions from several input maps will be combined with each output map. Neurons, number of layers, biases, filter sizes, stride, learning rate, activation function, and weights are just a few of the hyper-parameters and factors that make up a CNN. Because the combinatorial process analyses the immediate area of the input pixels, other correlation values can be examined based on other filters. Filter sizes represent different levels of complexity; tiny filters extract fine-grained data, while big filters retrieve coarse-grained data. As a result, in the early 2000s, researchers used spatial filters to improve quality and investigated the relationship between spatial filters and network learning. When changing filters, a variety of studies released during this period demonstrated that CNN might perform well on coarse and fine-grained data. For

$$I_i(i = 1, 2, 3, ...N)$$

is the multimodal input images, then the feature maps can be derived as given in equation 12:

$$F_i(x, y) = CNN(I_i)(x, y) \tag{12}$$

Where CNN(.) is a deep network [38]. We can acquire a feature vector employed in that convolutional layer for one pixel (x,y) whose dimension is the number of filters. A pixel's response [39] to CNN can reveal if the pixel is essential or valuable for visual representation. As a result, the strength of the vector

$$V_i(x, y)$$

(shown in equation 13) can be used to

calculate the visibility (luminosity) of pixel $I_i(x, y)$ by using L_1 norm of the feature vector,

$$V_i(x, y) = \|F_i(x, y)\| \tag{13}$$

The value of weight applied to this pixel will be affected by its visibility. Larger weights are often assigned to pixels with high local contrast. Motion recognition is required to avoid the ghosting effect in circumstances when the original images were taken in motion. Motion detection can be discovered by computing the Euclidean distance between two normalized feature vectors (given in expression 14):

$$\bar{F} = \frac{F}{\|F\|} \tag{14}$$

The following formula of 15 is used to compute how far two feature vectors in two photos are apart at a single pixel:

$$s_{ij}(x, y)^2 = \|\bar{F}_i(x, y) - \bar{F}_j(x, y)\|_2^2 \tag{15}$$

The consistency is lost due to movement that results in a low score of similarity. The similarity (shown in equation 16) between

$$F_i(x, y) and F_j(x, y)$$

$F_j(x, y)$ is mapped onto the range [0, 1] using a Gaussian kernel:

$$s_i(x, y) = \sum_{j=1}^{N} exp \frac{-s_{ij}(x, y)^2}{2\sigma^2} \tag{16}$$

Where is a constant that may be changed depending on the function that has to be performed, the temporally constant pixels should be given more weight.

With the help of visibility map and similarity map, the resultant feature weight map can be given as given in equation 17:

$$W_{CNN}(x, y) = \frac{V_i(x, y) x S_i(x, y) x M_i(x, y)}{\sum_{j=1}^{N} V_i(x, y) x S_i(x, y) x M_i(x, y) + \alpha} \tag{17}$$

We add a tiny coefficient to prevent division by zero. A widely used mask function (expression 18) is also used for weight map calculation given by:

$$M_{thei}(x, y) = \begin{cases} 1 & \beta < I_i(x, y) < 1 - \beta \\ 0 & \text{else} \end{cases} \tag{18}$$

Where $\beta \epsilon [0, 1]$is a parameter that controls the quality of exposure after normalizing the input images, it is capable of successfully removing pixels with low exposure. The CNN layers classify and amplify the input images to generate an even more precise weight map W_{CNN}

In the final procedure of fusion, the W_{CNN} is combined with the redefined weight mapW_{GF} generated from the PCA-GF hybrid system to form the concluding fused weight map W_f.
The weighted sum can be expressed as shown in equation 19:

$$W_f = W_{CNN} + W_{GF} \tag{19}$$

Lastly, for generating an all-featured fused image $I_f(x, y)$consisting of detailed data from both the medical modalities, the fused weight map W_f is multiplied with the original image $I_i(x, y)$ expressed as in given in equation 20:

$$I_f(x, y) = \sum_{i=1}^{N} (I_i)(x, y) X W_f(x, y) \tag{20}$$

When compared to other traditional approaches, the fused image will contain more information since the weighted sum of maps

$$(W_{CNN} + W_{GF})$$

reflects more precise details by retaining virtually all of the essential data at each level in the hybrid system.

5. EXPERIMENTAL ANALYSIS

5.1 Objective Analysis

(i) *Comparative Analysis of developed CNN and Hybrid PCA-GF Approach with Various Conventional Fusion Techniques*

The proposed technique is contrasted to the eight traditional image fusion algorithms, which are PCA[40] (Principal Component Analysis), DWT [41](Discrete Wavelet Transform), GIF (Graphics Interchange Format)[39], PCNN [42] (Pulse Coupled NN), NSCT (Non-Subsampled Contourlet Transform)[41], CNN (Convolution Neural Network)[22], NSCT-GIF[39], Guided Filters[28], Multi - focus image fusion technique [25], MSDNet for medical image fusion [30], Zero Learning fast medical image fusion [30], Multi-exposure fusion [38], Multi spectral image fusion [43]. Various performance indicators are chosen to provide a comparison assessment in order to assess the effectiveness of the targeted approach.

To arrive at a subjective conclusion, researchers compared the visual impressions of the original and merged images, while the quantitative study employed multiple quality criteria to demonstrate the perceived results' utility. Tables 1 provide a thorough examination of the quality metrics requirements for traditional fusion methods and proposed hybrid fusion methods as measured by the metrics IQI, mSSIM, and PSNR. The ultimate goal for the performance metrics values

shown in Table 1 is that for a successful image fusion technique, it should be the highest value, with the value closest to 1 indicating a higher quality fused picture. The greater the PSNR number expressed in decibels, the better the image quality. According to the data in Table 1 suggested hybrid methodology's fusion output beats all currently used fusion procedures in theoretical and practical analyses.

Table 1: Comparative Analysis for Quality Metrics for Various Fusion Methods (set1, set2 and set 3, set 4, set 5, set 6)

Sets	Algorithm	IQI	mSSIM	PSNR
1	PCA[40]	0.43	0.42	35.38
	DWT[41]	0.44	0.44	37.08
	GIF[39]	0.55	0.52	39.64
	PCNN[42]	0.59	0.55	38.30
	NSCT[41]	0.60	0.59	34.69
	CNN[21]	0.91	0.89	34.70
	NSCT-GIF[39]	0.94	0.82	30.06
	Guided Filter[27]	1.00	1.00	42.62
	Proposed System	**1.00**	**1.00**	**96.88**

Sets	Algorithm	IQI	mSSIM	PSNR
2	PCA[40]	0.46	0.42	29.05
	DWT[41]	0.56	0.55	32.39
	GIF[39]	0.54	0.60	35.55
	PCNN[42]	0.53	0.58	37.40
	NSCT[41]	0.70	0.67	39.58
	CNN[21]	0.92	0.94	39.72
	NSCT-GIF[39]	0.94	0.94	36.62
	Guided Filter[27]	0.99	1.00	42.99
	Proposed System	**0.99**	**1.00**	**90.50**

Sets	Algorithm	IQI	mSSIM	PSNR
3	PCA[40]	0.49	0.45	32.68
	DWT[41]	0.52	0.56	33.76
	GIF[39]	0.59	0.58	38.56
	PCNN[42]	0.63	0.59	35.80
	NSCT[41]	0.68	0.69	39.67
	CNN[21]	0.97	0.94	33.65
	NSCT-GIF[39]	1.00	0.99	45.68
	Guided Filter[27]	1.00	1.00	45.79
	Proposed System	**1.00**	**1.00**	**94.36**

The above mentioned six sets of data depicts that the proposed methodology gives more effective results in terms of IQI, mSSIM and PSNR values. Moreover, if we compare the proposed method with the previously used guided filter method based on PSNR value it shows marginally high accuracy with better image quality.

The bar graphs are shown in figure (4), (5), and (6) further represent the performance level of the proposed hybrid system as compared to traditional methods.

Sets	Algorithm	IQI	mSSIM	PSNR
4	PCA[40]	0.43	0.44	28.60
	DWT[41]	0.53	0.47	29.45
	GIF[39]	0.61	0.55	32.56
	PCNN[42]	0.65	0.59	35.32
	NSCT[41]	0.80	0.65	39.78
	CNN[21]	0.99	0.93	23.80
	NSCT-GIF[39]	0.97	0.95	40.67
	Guided Filter[27]	1.00	1.00	44.07
	Proposed System	**0.99**	**1.00**	**90.54**

Sets	Algorithm	IQI	mSSIM	PSNR
5	PCA[40]	0.41	0.45	34.78
	DWT[41]	0.53	0.52	36.40
	GIF[39]	0.57	0.60	43.01
	PCNN[42]	0.60	0.65	40.23
	NSCT[41]	0.66	0.70	40.30
	CNN[21]	0.89	0.85	40.20
	NSCT-GIF[39]	0.93	0.91	32.30
	Guided Filter[27]	1.00	1.00	42.86
	Proposed System	**1.00**	**1.00**	**91.81**

Sets	Algorithm	IQI	mSSIM	PSNR
6	PCA[40]	0.57	0.53	30.56
	DWT[41]	0.57	0.55	32.58
	GIF[39]	0.62	0.63	34.56
	PCNN[42]	0.60	0.61	33.30
	NSCT[41]	0.77	0.70	38.40
	CNN[21]	0.90	0.90	37.46
	NSCT-GIF[39]	0.93	0.98	38.40
	Guided Filter[27]	0.99	1.00	43.28
	Proposed System	**1.00**	**1.00**	**93.89**

Compared to techniques such as PCA, DWT, GIF, and NSCT in the Image Quality Index figure (4), the proposed system excels by raising the quality index by about 50 percent, with practical simulated values of 0.99, which is very close to 1. Similarly, in figure (5), mSSIM is used for measuring Mean Structural Similarity. A resulting mSSIM index ranges between 0 and 1 in decimal numbers, with the value "1" indicating a complete resemblance in unique parallel sets of images. A score of 0 implies that there is no structural resemblance.The proposed hybrid technique achieves mSSIM accuracy with an output value of 0.99 (almost 1).

PSNR is most often used to evaluate the quality of the reconstruction of pictures and is expressed on a logarithmic scale. In general, a greater PSNR suggests that the restoration is of higher quality, typical PSNR values in image compression range between 30 and 50 decibels. In figure (6), the PSNR values of the preceding eight approaches are in the 30-45 dB range, but the proposed system contributes values in the 90-96 dB range. This signifies that the reconstructed, fused picture has a superior resolution and quality.

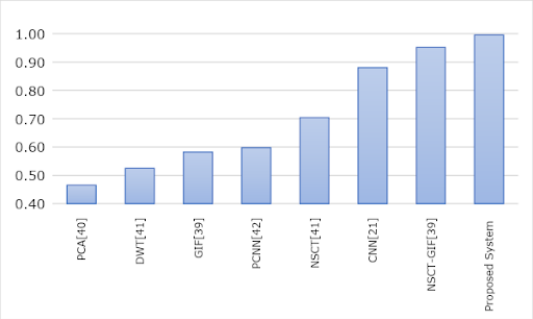

Figure 4. Comparative analysis for Image Quality Index (IQI)

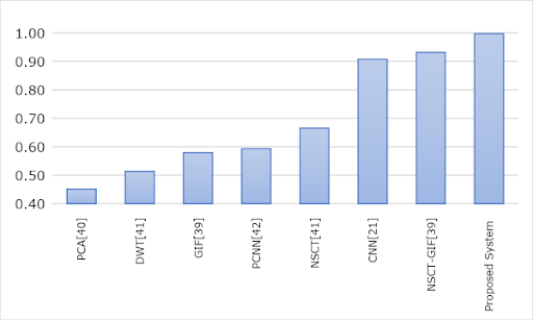

Figure 5. Comparative analysis for mSSIM

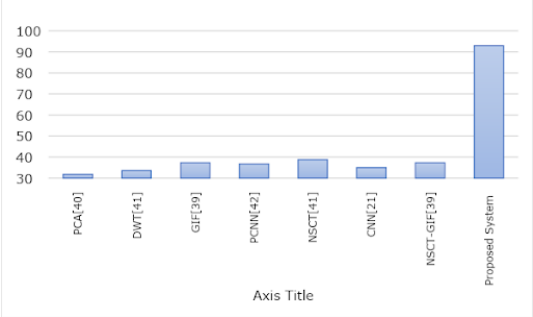

Figure 6. Comparative analysis for Peak-Signal-to-Noise Ratio (PSNR)

(ii) *Comparative Analysis of Proposed CNN and Hybrid PCA-GF Approach with base Guided Filter Technique[28]*

When contrasted to our earlier work, both the IQI and the mSSIM have a maximum value of one. Denoising in an image can be improved by lowering the percentage of pixels with a comparable degree of similarity and increasing the levels of pixels with a comparable degree of similarity. A good parameter setting is required to analyze a noisy image while maintaining an appropriate balance between retaining relevant structural features and reducing noise. Denoised pictures can be analyzed using two measurements: noise reduction and structure preservation. Both of these metrics are derived using the SSIM metric's resemblance comparison. Compared to PSNR and MSE, the structural similarity (mSSIM) metric [4] shows more accurately how much noise has been cut down by using the proposed hybrid technique.

Compared to traditional state-of-the-art techniques, the earlier work (Guided Filter) obtained outstanding results for IQI and SSIM. Still, the new strategy improved the results much more, as seen in figure (7) bar graphs. The graph in figure (8) compares the PSNR values of the earlier and now presented technique. The hybrid design that uses neural networks surpasses the guided filter methodology by producing PSNR values of more than 90 dB. In contrast, the prior approach computed values in the 40-50 decibel region. Peak Signal-to-Noise Ratio is the ratio of a signal's highest possible power to the power of corrupting noise, and it impacts the quality of its representation. As previously stated, PSNR is an approximation showing how well humans see a reconstruction.

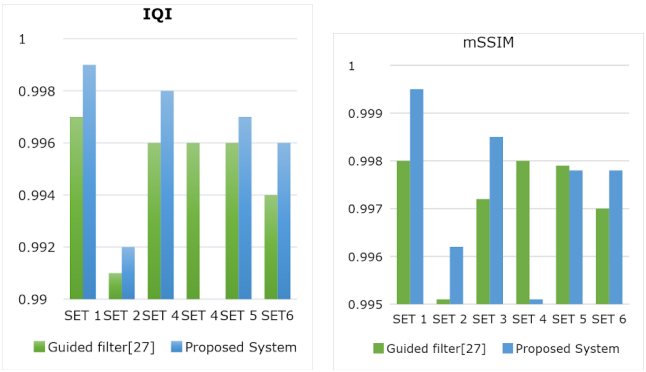

Figure 7. IQI and mSSIM Parameter Analysis for Guided Filter and Proposed (CNN and PCA - GF) System (Higher the Value better the performance)

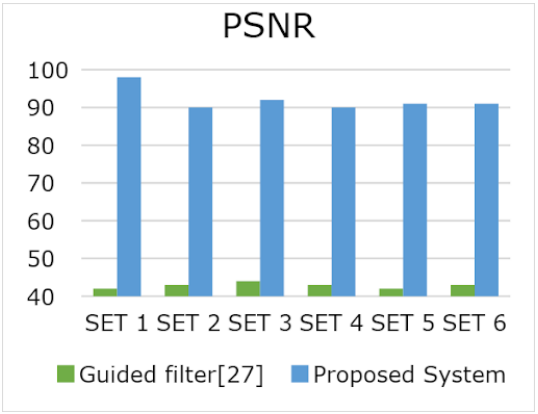

Figure 8. PSNR Comparison between Guided Filter and Proposed CNN and PCA-GF System (Higher the Value better the performance)

(iii) *Analysis with more performance metrics*

Further, the experimentation is extended with more performance metrics using existing Guided Filter [48] methods and proposed system of CNN and Hybrid PCA-GF. Signal to Noise Ratio (SNR), Mean absolute error (MAE), Mean square error (MSE), Root mean square error (RMSE), Naturalness Image Quality Evaluator (NIQE), Entropy and Correlation coefficient (CC) are the performance metrices against which the performance of proposed system is evaluated as shown in Table 3 for all 6 sets of input images. The proposed system computed a comprehensive collection of additional parameters important for evaluating image quality to offer a clearer picture, as shown in Table (2). The MSE and RMSE of any fusion technique should be close to zero for it to be effective. This condition is fulfilled by the suggested approach since the resulting MSE values are zero and the RMSE values are incredibly close to zero. Entropy, NIQE, Correlation coefficient (CC), and MAE are also computed to illustrate the proposed system's remarkable efficiency.

Table 2: Comparision of Prior Guided Filter Technique Quality Metrics Obtained by the Proposed CNN and PCA-GF and Prior Guided Filter Technique

Parameters	SET 1		SET 2		SET 3	
	GF[27]	Proposed System	GF[27]	Proposed System	GF[27]	Proposed System
SNR	0.0018 dB	-0.00099 dB	-0.00026 dB	-0.00098 dB	-0.00007 dB	-0.00093 dB
MAE	0.05408	0.00006	0.04266	0.00019	0.02689	0.00012
MSE	0.10764	0	0.09883	0	0.05188	0
MAE	0.32808	0.00064	0.31437	0.00132	0.22777	0.00085
NIQE	5.2601	6.2727	3.9176	4.5375	4.494	3.5369
ENTROPY	6.5836	6.3469	4.847	5.0362	4.4893	6.2884
CC	0.9998	1	0.9998	0.9999	0.9998	0.9999
PARAMETERS	SET 4		SET 5		SET 6	
	GF[27]	Proposed System	GF[27]	Proposed System	GF[27]	Proposed System
SNR	0.00017 dB	-0.00140 dB	0.00019 dB	-0.00145 dB	-0.00009 dB	-0.007 dB
MAE	0.03951	0.0002	0.05099	0.00018	0.0484	0.00009
MSE	0.07703	0	0.10176	0	0.09242	0
RMSE	0.27755	0.00132	0.319	0.00114	0.30401	0.0009
NIQE	12.1482	2.9098	22.0537	6.389	2.3514	4.4983
ENTROPY	5.954	4.6698	6.882	6.8704	4.6216	4.7506
CC	0.998	0.9999	0.9998	0.9999	0.9999	0.9999

All the parameters are shown in Table (2) are also represented graphically for six sets in figure (9).

- Higher values of **Entropy** give more comprehensive information. Entropy is a measure of picture information content that may be understood as the average unpredictability of a source of information. Entropy is defined in images as the matching states of intensity level to which individual pixels may adapt. It is utilized in quantitative analysis and assessment of picture details; the entropy value is employed because it allows for a more accurate comparison of image features. Sets (2, 3, 5, 6) have a higher entropy value for the proposed technique when compared to previous work, indicating a more accurate fused picture.
- The **Mean Square Error** (MSE) is the squared average difference between the filtered and source images. The smaller the Mean Square Error, the closer the pixels in the predicted picture are to the pixels in the targeted image. The proposed system returns a "0" MSE value for each set, indicating strong prediction capacity, whereas the Guided Filter method returns non-zero values.
- The MAE **(Mean Absolute Error)** value represents the average amount of error in the predicted picture. According to the MAE graph in figure (9)(b), the values derived from distinct sets are extremely near to zero, implying a greater accuracy rate. The values of MAE in the prior method indicate more amount of error since the values are non-zero.
- Lower Naturalness **Image Quality Evaluator** (NIQE) scores indicate a larger information richness in the fused image. More miniature scores of the Naturalness Image Quality index are achieved with the lowest at set 3, set 4, and set 5 compared to Guided Filter, indicating a better image quality.
- The **Root Mean Squared Error** evaluates the difference between the fused and sample image. Low RMSE values suggest that the test picture is very similar to the reference image. The proposed system also achieves lower RMSE values than the previously proposed method showing augmented predictions.
- The **Correlation Coefficient Analysis** is an indicator for strength measurement of the link between relative movements of two variables. The CC values lie between a range of [-1.0, 1.0]. A computed value less than -1 or greater than 1 signifies that the correlation coefficient has errors. The proposed technique yields greater values of CC as compared to the prior proposed approach.
- The **Signal-to-Noise Ratio** shows the ratio of signal power present in an image to the noise power. The lower levels of SNR obtained using the proposed technique show a significant reduction in the noise present in the final fused image.

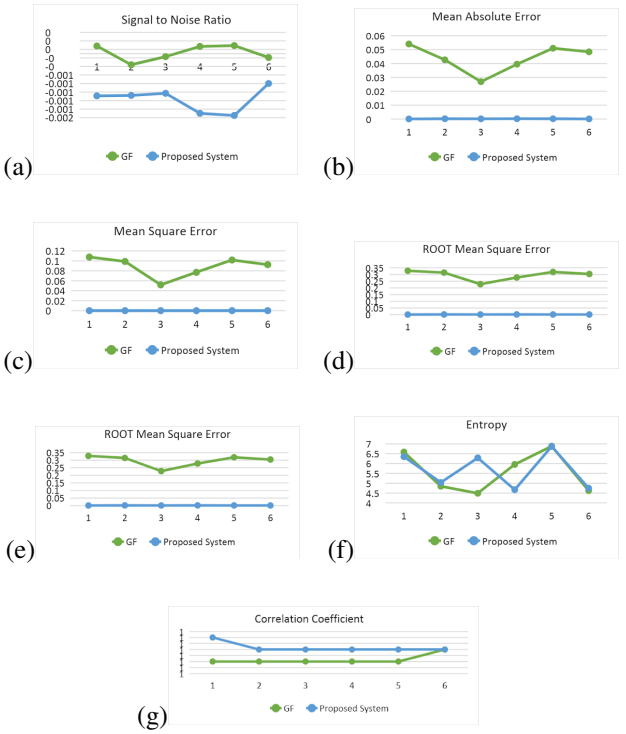

Figure 9. Comparative Analysis of Image Quality Parameters [(a)-(g)] for Different sets of Guided Filter [48] and Proposed CNN and PCA-GF System

5.2 Subjective Analysis

As shown in Figure 10, we perform a paired fusion of various source pictures produced with CT, MR-Gadolinium contrast media, MR-Proton density, MR-T2 weighted images, PET, and SPECT imaging techniques, respectively, to demonstrate the efficacy of the proposed approach for combining pictures with different modalities. These pictures were obtained from Harvard Medical School's entire brain atlas project (https://www.med.harvard.edu/aanlib/). To generate the final fused image, the weight maps and original photos are merged. The developed system produces promising results in enhancing the details on every pair-wise fusion [44], as highlighted by the boxes shown in Figure 10. Similarly, Figure 11 shows the successful implementation of the proposed system for multi-exposure images and yields fused output images with better image quality metrics.

Dataset: Images were sourced from Harvard whole-brain atlas, atlas is unique in its terms, comprehensive and fulfilled our objective as it contains a huge compilation of modern cross-sectional imaging, including CT, MRI, and SPECT in health and disease.

5.3 Statistical Analysis

To demonstrate that the proposed system is statistically significant, a statistical study was performed. The analysis of variance test was used to compare traditional and proposed approaches for three distinct image

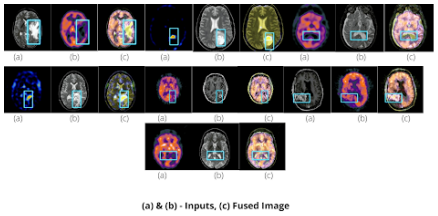

Figure 10. Multimodal Input datasets and Fused Outputs of the Hybrid System

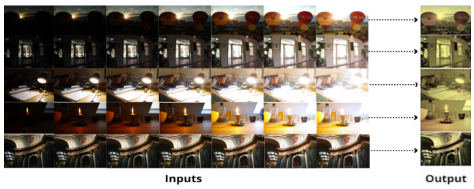

Figure 11. Multi-exposure Input Datasets and Fused Outputs of the Hybrid System

quality metrics. Table (4) shows that the presented CNN and hybrid PCA-GF System have higher average values than any conventional technique, indicating that the created system outperforms existing techniques. The dataset consisting of different techniques yields p-values of less than 0.05, suggesting that the null hypothesis can be discarded and the comparison can be concluded as statistically significant. Tables (4) indicate that

$$F_{crit} << F,$$

which confirms the significance, further supports the hypothesis and verifies the analysis.

Table 4: Table for PSNR for respective Sum, Average and Variance values

Source of Variation	SS	df	MS	F	P-value	F crit
Rows	17420.14	8	2175.267	179.3299	7.61E-29	2.18017
Error	485.199	40	12.12998			
Total	17989.48	53				

Thus the proposed system outperforms the existing various methods in objective, subjective and statistical analysis.

6. Conclusions and Future Work

Proposed system of multimodal image processing that merges the benefits of CNN and the guided image filter and yields comprehensive results in picture quality and resolution. A neural network learns the optimal mapping of multimodal, multi-focus, and multi-exposed inputs to build the feature map directly from the data. The proposed system based on many tasks such as high resolution, denoising, edge-preserving [15], elimination of ringing and blurring effects, and image smoothening, reveals that the fused

images from the CNN and hybrid model of a PCA-guided filter are numerically and qualitatively similar to those obtained from the most recent cutting-edge techniques. Concretely, the presented method makes image fusion with deep learning understandable in domains such as medical imaging, which requires limited tolerance for mistakes or unknowns. Furthermore, the investigation of several imaging parameter metrics verifies and satisfies all of the claims made by the proposed hybrid method. The suggested system improved IQI by 50 percent and has a great peak-to-noise ratio, which means it has better resolution and picture quality [21]. Regarding structural resemblance, the proposed technique reaches higher accuracy with values closer to the maximum value, i.e., one. The mSSIM measure more appropriately reflects the significant amount of noise that has been decreased. The system displays a mean square error of zero, indicating that the model makes accurate predictions. The technique has many uses, such as getting rid of noise, dealing with limited lighting, and getting rid of ringing and blurring effects. The following are the major challenges that this hybrid system still faces and has the potential to overcome as a future scope.

- The sizes and resolutions of multiple modalities and multi-exposure pictures fluctuate. The system rejects input datasets of varying sizes and formats.
- Analysis can be further improved with the help of Xydeas-Petrovic index (Qab/f) , mutual information (MI) and normalized mutual information (NMI) metrics as it further scope for evaluating the amount of information transferred from source image to fused image.
- The developed method can only fuse images taken from a single viewpoint. It is inapplicable to input [45] datasets with changing angles, such as Sagittal, Coronal, and Axial, in the case of MRI.
- The investigation can be done to overcome these obstacles by making the proposed system more flexible in future work. It can accept input datasets of varying resolutions, sizes, and functions from various perspectives on medical imaging modalities.

References

[1] B. V. Dasarathy, "A special issue on image fusion: advances in the state of the art," *Information Fusion*, vol. 8, no. 2, p. 113, 2007.

[2] A. Galande and R. Patil, "The art of medical image fusion: A survey," in *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2013, pp. 400–405.

[3] B. Meher, S. Agrawal, R. Panda, and A. Abraham, "A survey on region based image fusion methods," *Information Fusion*, vol. 48, pp. 119–132, 2019.

[4] A. de Juan, A. Gowen, L. Duponchel, and C. Ruckebusch, "Image fusion," in *Data handling in science and technology*. Elsevier, 2019, vol. 31, pp. 311–344.

[5] P. D. Vora and M. N. Chudasama, "Different image fusion techniques and parameters: a review," *Int J Comput Sci Inf Technol*, vol. 6, no. 1, pp. 889–892, 2015.

[6] Y. Li, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Joint image filtering with deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 8, pp. 1909–1923, 2019.

[7] H. Li, X.-J. Wu, and J. Kittler, "Mdlatlrr: A novel decomposition method for infrared and visible image fusion," *IEEE Transactions on Image Processing*, vol. 29, pp. 4733–4746, 2020.

[8] L. Chandrashekar and A. Sreedevi, "Advances in biomedical imaging and image fusion," *International Journal of Computer Applications*, vol. 179, no. 24, pp. 1–9, 2018.

[9] A. M. Opportunity, "Academic success breeds mediocrity."

[10] A. P. James, S. Thiruvenkadam, J. S. Paul, and M. Braun, "Special issue on medical image computing and systems," *Information Fusion*, 2014.

[11] V. Aslantas and A. N. Toprak, "A pixel based multi-focus image fusion method," *Optics Communications*, vol. 332, pp. 350–358, 2014.

[12] B. Balachander and D. Dhanasekaran, "Comparative study of image fusion techniques in spatial and transform domain," *ARPN Journal of Engineering and Applied Sciences*, vol. 11, May 2016.

[13] S. Thakral and P. Manhas, "Image processing by using different types of discrete wavelet transform," in *International Conference on Advanced Informatics for Computing Research*. Springer, 2018, pp. 499–507.

[14] P. R. Hill, N. Anantrasirichai, A. Achim, M. E. Al-Mualla, and D. R. Bull, "Undecimated dual-tree complex wavelet transforms," *Signal Processing: Image Communication*, vol. 35, pp. 61–70, 2015.

[15] S. Ghofrani, "Comparing nonsubsampled wavelet, contourlet and shearlet transforms for ultrasound image despeckling," *International Journal of Image, Graphics and Signal Processing*, vol. 7, no. 2, p. 15, 2015.

[16] Z. Ding, D. Zhou, R. Nie, R. Hou, and Y. Liu, "Brain medical image fusion based on dual-branch cnns in nsst domain," *BioMed Research International*, vol. 2020, 2020.

[17] P. R. Muduli, "Development and implementation of image fusion algorithms based on wavelets," Ph.D. dissertation, 2013.

[18] M. Diwakar and M. Kumar, "A hybrid method based ct image denoising using nonsubsampled contourlet and curvelet transforms," in *Proceedings of International Conference on Computer Vision and Image Processing*. Springer, 2017, pp. 571–580.

[19] K.-m. HE and J. Sun, "Tang xiao-ou. guided image filtering," in *Proceedings of the llth European Conference on Computer Vision. Heraklion, Crete, Greece*, 2010, pp. 1–14.

[20] H. Wu, S. Zheng, J. Zhang, and K. Huang, "Fast end-to-end trainable guided filter," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1838–1847.

[21] H. Yang, M. Zhu, Z. Zhang, and H. Huang, "Guided filter based edge-preserving image non-blind deconvolution," *arXiv preprint arXiv:1609.01839*, 2016.

[22] Y. Li, J. Zhao, Z. Lv, and Z. Pan, "Multimodal medical supervised image fusion method by cnn," *Frontiers in Neuroscience*, p. 303, 2021.

[23] J. Yuan, H. Liao, R. Luo, and J. Luo, "Automatic radiology report generation based on multi-view image fusion and medical concept enrichment," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 721–729.

[24] H. Zhang, H. Xu, Y. Xiao, X. Guo, and J. Ma, "Rethinking the image fusion: A fast unified image fusion network based on proportional maintenance of gradient and intensity," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 12 797–12 804.

[25] S. Maqsood, U. Javed, M. M. Riaz, M. Muzammil, F. Muhammad, and S. Kim, "Multiscale image matting based multi-focus image fusion technique," *Electronics*, vol. 9, no. 3, p. 472, 2020.

[26] A. S. Fahmy, E. J. Rowin, R. H. Chan, W. J. Manning, M. S. Maron, and R. Nezafat, "Improved quantification of myocardium scar in late gadolinium enhancement images: deep learning based image fusion approach," *Journal of Magnetic Resonance Imaging*, vol. 54, no. 1, pp. 303–312, 2021.

[27] J. Huang, Z. Le, Y. Ma, F. Fan, H. Zhang, and L. Yang, "Mgmdcgan: Medical image fusion using multi-generator multi-discriminator conditional generative adversarial network," *IEEE Access*, vol. 8, pp. 55 145–55 157, 2020.

[28] N. Mehta, S. Budhiraja, and S. Goel, "Multimodal medical image fusion using modified weight maps and guided filter," in *2017 2nd International Conference on Telecommunication and Networks (TEL-NET)*. IEEE, 2017, pp. 1–6.

[29] J. Jose, N. Gautam, M. Tiwari, T. Tiwari, A. Suresh, V. Sundararaj, and M. Rejeesh, "An image quality enhancement scheme employing adolescent identity search algorithm in the nsst domain for multi-modal medical image fusion," *Biomedical Signal Processing and Control*, vol. 66, p. 102480, 2021.

[30] F. Lahoud and S. Süsstrunk, "Zero-learning fast medical image fusion," in *2019 22th International Conference on Information Fusion (FUSION)*. IEEE, 2019, pp. 1–8.

[31] M. Suganya, R. Sabitha, and J. Jasmine, "Deep learning technique for brain tumor detection using medical image fusion," *International Journal of Innovative Technology and Exploring Engineering*, 2019.

[32] T. Qiu, C. Wen, K. Xie, F.-Q. Wen, G.-Q. Sheng, and X.-G. Tang, "Efficient medical image enhancement based on cnn-fbb model," *IET Image Processing*, vol. 13, no. 10, pp. 1736–1744, 2019.

[33] M. G. Borgognone, J. Bussi, and G. Hough, "Principal component analysis in sensory analysis: covariance or correlation matrix?" *Food quality and preference*, vol. 12, no. 5-7, pp. 323–326, 2001.

[34] S. Hao, D. Pan, Y. Guo, R. Hong, and M. Wang, "Image detail enhancement with spatially guided filters," *Signal processing*, vol. 120, pp. 789–796, 2016.

[35] R. P. Desale and S. V. Verma, "Study and analysis of pca, dct & dwt based image fusion techniques," in *2013 International Conference on Signal Processing, Image Processing & Pattern Recognition*. IEEE, 2013, pp. 66–69.

[36] S. P. R. Borra, R. K. Panakala, and P. R. Kumar, "Vlsi implementation of image fusion using dwt-pca algorithm with maximum selection rule," *International Journal of Intelligent Engineering and Systems*, vol. 12, no. 5, 2019.

[37] M. Ch, M. M. Riaz, N. Iltaf, A. Ghafoor, and S. S. Ali, "A multifocus image fusion using highlevel dwt components and guided filter," *Multimedia Tools and Applications*, vol. 79, no. 19, pp. 12 817–12 828, 2020.

[38] H. Li and L. Zhang, "Multi-exposure fusion with cnn features," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 1723–1727.

[39] B. Rajalingam, F. Al-Turjman, R. Santhoshkumar, and M. Rajesh, "Intelligent multimodal medical image fusion with deep guided filtering," *Multimedia Systems*, pp. 1–15, 2020.

[40] K. Padmavathi, M. V. Karki, and M. Bhat, "Medical image fusion of different modalities using dual tree complex wavelet transform with pca," in *2016 International Conference on Circuits, Controls, Communications and Computing (I4C)*. IEEE, 2016, pp. 1–5.

[41] S. Patel, "Diabetic retinopathy detection and classification using pre-trained convolutional neural networks," *International Journal on Emerging Technologies*, vol. 11, no. 3, pp. 1082–1087, 2020.

[42] S. Huang, G. Chu, Y. Fei, X. Zhang, and H. Wang, "Medical image fusion based on sparse representation and guided filtering," in *Journal of Physics: Conference Series*, vol. 1302, no. 2. IOP Publishing, 2019, p. 022045.

[43] J. Piao, Y. Chen, and H. Shin, "A new deep learning based multi-spectral image fusion method," *Entropy*, vol. 21, no. 6, p. 570, 2019.

[44] J. Li, G. Yuan, and H. Fan, "Multifocus image fusion using wavelet-domain-based deep cnn," *Computational intelligence and neuroscience*, vol. 2019, 2019.

[45] X. Song, X.-J. Wu, and H. Li, "Msdnet for medical image fusion," *Lecture Notes in Computer Science*, p. 278–288, 2019.

Nalini Santosh Jagtap

Nalini Santosh Jagtap working as Assistant Professor in Computer Engineering at Dr D Y Patil Institute of Engineering Management at Research, SPPU, Pune, India and pursuing her doctorate at Pimpri Chinchwad College of Engineering, SPPU, Pune, India. Her current research interest includes image fusion and content based video retrieval.

Sudeep D Thepade

Sudeep D Thepade is working as a Professor in Computer Engineering at Pimpri Chinchwad College of Engineering, Savitribai Phule Pune University, Pune, India. His current research interests include image classification, CBIR, image enhancements and biometric identification. He has more than 350 research papers to his credit. He has guided two Ph.D. Scholars, guiding six Ph.D. Scholars and several Masters Dissertation Projects.