Medical Image Fusion through Siamese Network-Based Spatial Domain Fusion with Gaussian Pyramid Decomposition

Ramakrishna B[1], Danthala Ramya[2], Gajibinkar Divya[3]

[1,2,3]Assistant Professor, Department of ECE, Malla Reddy Engineering College and Management Sciences, Hyderabad, Telangana.

Abstract

Spatial domain-based medical image fusion methods were among the earliest research focal points. However, spatial domain techniques tend to introduce spectral and spatial distortions into fused images. This research introduces the Siamese network, one of the three models for comparing patch similarity in convolutional neural networks (CNN). Its two identical weight branches ensure uniform feature extraction and activity level measurements, offering advantages over pseudo-Siamese and 2-channel models. The ease of training makes the Siamese model a preferred choice in fusion applications. The proposed method utilizes weight maps, Gaussian pyramid decomposition, and pyramid transform for multiscale decomposition, aligning the fusion process with human visual perception. Additionally, a localized similarity-based fusion strategy is employed to adaptively adjust decomposed coefficients. This algorithm combines pyramid-based and similarity-based fusion techniques with CNN models, resulting in an advanced fusion approach.

Keywords: Computed Tomography, Magnetic Resonance Imaging, Positron Emission Tomography, Single Photon Emission Computed Tomography, Convolutional Neural Network, Medical Image Fusion, Gaussian Pyramid Decomposition.

1. Introduction

With the rapid development of sensor and computer technology, medical imaging has emerged as an irreplaceable component in various clinical applications including diagnosis, treatment planning and surgical navigation. To provide medical practitioners sufficient information for clinical purposes, medical images obtained with multiple modalities are usually required, such as X-ray, computed tomography (CT), magnetic resonance (MR), positron emission tomography (PET), single photon emission computed tomography (SPECT), etc [1]. Due to the difference in imaging mechanism, medical images with different modalities focus on different categories of organ/tissue information. For instance, the CT images are commonly used for the precise localization of dense structures like bones and implants, the MR images can provide excellent soft-tissue details with high-resolution anatomical information, while the functional information on blood flow and metabolic changes can be offered by PET and SPECT images but with low spatial resolution. Multi-modal medical image fusion aims at combining the complementary information contained in different source images by generating a composite image for visualization, which can help physicians make easier and better decisions for various purposes [2]. In recent years, a variety of medical image fusion methods have been proposed [3]. Due to the difference in imaging mechanism, the intensities of different source images at the same location often vary significantly. For this reason, most of these fusion algorithms are introduced in a multi-scale manner to pursue perceptually good results. In general, these multi-scale transform (MST)-based fusion methods consist of three steps, namely, decomposition, fusion, and reconstruction. Multi-scale transforms which are frequently studied in image fusion include pyramids [4], wavelets multi-scale geometrical transforms like contourlet and shearlet [5]. In image fusion research, sparse representation is another popular image modelling approach, which has also been successfully applied to fuse multi-modal medical images [6]. One of the most crucial issues in image fusion is calculating a weight map which Integrates the pixel activity information from different

sources. In most existing fusion methods, this target is achieved by two steps known as activity level measurement and weight assignment. In conventional transform domain fusion methods, the absolute value of a decomposed coefficient (or the sum of those values within a small window) is employed to measure its activity, and then a "choose-max" or "weighted-average" fusion rule is applied to assign weights to different sources based on the obtained measurement [7]. Clearly, this kind of activity measurement and weight assignment are usually not very robust resulting from many factors like noise, mis-registration, and the difference between source pixel intensities. To improve the fusion performance, many complex decomposition approaches and elaborate weight assignment strategies have been recently proposed in the literature [8].

However, it is actually not an easy task to design a ideal activity level measurement or weight assignment strategy which can comprehensively take all the key issues of fusion into account. Moreover, these two steps are designed individually without a strong association by many fusion methods, which may greatly limit the algorithm performance [9]. In this paper, this issue is addressed from another viewpoint to overcome the difficulty in designing robust activity level measurements and weight assignment strategies. Specifically, a convolutional neural network (CNN) [10] is trained to encode a direct mapping from source images to the weight map. In this way, the activity level measurement and weight assignment can be jointly achieved in an "optimal" manner via learning network parameters. Considering the different imaging modalities of multi-modal medical images, we adopt a multi-scale approach via image pyramids to make fusion process more consistent with human visual perception. In addition, a local similarity-based strategy is applied to adaptively adjust the fusion mode for the decomposed coefficients of source images.

Rest of the paper is organized as follows: Section 2 details about literature survey, section 3 details about the proposed methodology, section 4 details about the results with discussion, and section 5 concludes article with references.

2. Literature Survey

Li, Xiaosong et al. (2021) [11] introduced a two-layer decomposition scheme by the joint bilateral filter, the energy layer containing rich intensity information, and the structure layer capturing ample details. Then a novel local gradient energy operator based on the structure tensor and neighbour energy is proposed to fuse the structure layer and the l1-max rule is introduced to fuse the energy layer. A total of 118 co-registered pairs of medical images covering five different categories of medical image fusion problems were tested in experiments. Zhang Hao et al. (2021) [12] introduced the concept of image fusion and classify the methods from the perspectives of the deep architectures adopted and fusion scenarios. Then the state-of-the-art on the use of deep learning in various types of image fusion scenarios, including digital photography image fusion, multi-modal image fusion and the sharpening fusion is reviewed. Subsequently, the evaluation for some representative methods in specific fusion tasks are performed qualitatively and quantitatively. Subbiah Parvathy, et al. (2020) [13] proposed a novel fusion model based on optimal thresholding with deep learning concepts. An enhanced monarch butterfly optimization (EMBO) is utilized to decide the optimal threshold of fusion rules in shearlet transform. Then, low and high-frequency sub-bands were fused based on feature maps and were given by the extraction part of the deep learning method. Deng, Xin, et al. (2020) [14] proposed a novel deep convolutional neural network to solve the general multi-modal image restoration (MIR) and multi-modal image fusion (MIF) problems. The key feature of the proposed network is that it can automatically split the common information shared among different modalities, from the unique information that belongs to each single modality, and is therefore denoted with CU-Net, i.e., common, and unique information splitting network. Specifically, the CU-Net is composed of three modules, i.e., the unique feature extraction module (UFEM), common feature preservation

module (CFPM), and image reconstruction module (IRM). The architecture of each module is derived from the corresponding part in the MCSC model, which consists of several learned convolutional sparse coding (LCSC) blocks. Dinh, Phu-Hu et al. (2021) [15] proposed two novel algorithms to tackle the current image fusion approaches. The first algorithm is based on the Equilibrium optimizer algorithm (EOA) to find optimal parameters to fuse low-frequency components. This allows the fused image to have good contrast. The second algorithm is based on the sum of local energy functions using the Prewitt compass operator to create an efficient rule for the fusion of high-frequency components. This allows the fused image to significantly preserve details transferred from input images.

Ding, Zhaisheng et al. (2020) [16] introduced a new framework for medical image fusion is proposed which combines convolutional neural networks (CNNs) and non-subsampled shearlet transform (NSST) to simultaneously cover the advantages of them both. This method effectively retains the functional information of the CT image and reduces the loss of brain structure information and spatial distortion of the MRI image. The initial weights integrate the pixel activity information from two source images that are generated by a dual-branch convolutional network and are decomposed by NSST. Muzammil, Shah Rukh et al. (2020) [17] proposed a novel algorithm, namely Convolutional Sparse Image Decomposition (CSID), that fuses CT and MR images. CSID uses contrast stretching and the spatial gradient method to identify edges in source images and employs cartoon-texture decomposition, which creates an overcomplete dictionary. Yu, Hang et al. (2021) [18] provided a survey on convolutional neural networks in medical image analysis which involves the commonly used CNNs in medical image processing, including AlexNet, GoogleNet, ResNet, R-CNN, and FCNN and an overview of the use of CNNs, for image classification, segmentation, detection, and other tasks such as registration, content-based image retrieval, image generation and enhancement, in some typical medical diagnosis areas such as brain, breast, and abdominal. Goyal, Bhawna et al. (2021) [19] introduced a multimodal medical image fusion method that integrates multimodal medical images having low resolution with reduced computational complexity to improve the accuracy of target recognition and for providing a basis for clinical diagnosis. Initially salient structure extraction (SSE) approach, which employ a rolling guidance filter (RGF) over the source images for removing small scale structures while preserving the image textures and thereby recovering the salient edges has been implemented. Subsequently, an image gradient operator is employed to restore large-scale structures from the filtered images. A DTF (Domain Transfer Filtering) is further used to recover the small-scale details in the neighborhood of large-scale structures of the images. Boveiri, Hamid Reza et al. (2020) [20] provided a comprehensive review on the state-of-the-art literature known as medical image registration using deep neural networks is presented. The review is systematic and encompasses all the related works previously published in the field. Key concepts, statistical analysis from different points of view, confining challenges, novelties and main contributions, key-enabling techniques, future directions, and prospective trends all are discussed and surveyed in detail in this comprehensive review. It requires a deep understanding and insight for the readers active in the field who are investigating state-of-the-art and seeking to contribute the future literature.

3. Proposed Methodology

3.1 The CNN model for medical image fusion

Fig. 1 shows the convolutional network used in the proposed fusion algorithm. It is a Siamese network in which the weights of the two branches are constrained to the same. Each branch consists of three convolutional layers and one max-pooling layer which is the same as the network used in [24]. To reduce the memory consumption as well as increase the computational efficiency, we adopt a much slighter model in this work by removing a fully connected layer from the network used in [24].

The 512 feature maps after concatenation are directly connected to a 2-dimensional vector. It can be calculated that the slight mode only takes up about 1.66 MB of physical memory in single precision, which is significantly less than the 33.6 MB model employed in [24]. Finally, this 2dimensional vector is fed to a 2-way SoftMax layer (not shown in Fig. 1), which produces a probability distribution over two classes. The two classes correspond two kinds of normalized weight assignment results, namely, "first patch 1 and second patch 0" and "first patch 0 and second patch 1", respectively. The probability of each class indicates the possibility of each weight assignment. In this situation, also considering that the sum of two output probabilities is 1, the probability of each class just indicates the weight assigned to its corresponding input patch.

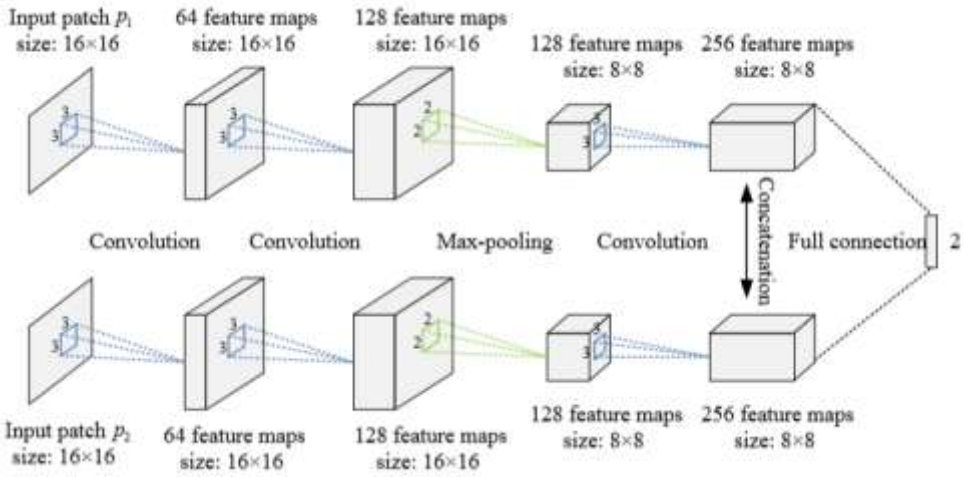


Figure 1. Architecture for CNN training

The network is trained by high-quality image patches and their blurred versions using the approach in [24]. In the training process, the spatial size of the input patch is set to 16 × 16 according to the analysis in [24]. The creation of training examples is based on multi-scale Gaussian filtering and random sampling. The SoftMax loss function is employed as the optimization objective, and we adopt the stochastic gradient descent (SGD) algorithm to minimize it. The training process is operated on the popular deep learning framework Caffe [28]. Please refer to [24] for the details of example generation and network training.

Since the network has a fully connected layer that have fixed dimensions (pre-defined) on input and output data, the input of the network must have a fixed size to ensure that the input data of a fully connected layer is fixed. In image fusion, to handle source images of arbitrary size, one can divide the images into overlapping patches and input each patch pair into the network, but it will introduce many repeated calculations. To solve this problem, we first convert the fully connected layer into a equivalent convolutional layer containing two kernels of size 8 × 8 × 512 [26]. After the conversion, the network can process source images of arbitrary size to generate a dense prediction map, in which each prediction (a 2-dimensional vector) contains the relative clarity information of a source patch pair at the corresponding location. As there are only two dimensions in each prediction and their sum is normalized to 1, the output can be simplified as the weight of the first (or second) source. Finally, to obtain a weight map with the same size of source images, we assign the value as the weights of all the pixels within the patch location and average the overlapped pixels.

3.2 Detailed fusion scheme

The schematic diagram of the proposed medical image fusion algorithm is shown in Fig. 2. The algorithm can be summarized as the following four steps.

Step 1: CNN-based weight map generation. Feed the two source images A and B to the two branches of the convolutional network, respectively. The weight map W is generated using the approach described above.

Step 2: Pyramid decomposition. Decompose each source image into a Laplacian pyramid. Let $L\{A\}^l$ and $L\{B\}^l$ respectively denote the pyramids of A and B, where l indicates the l-th decomposition level. Decompose the weight map W into a Gaussian pyramid $G\{W\}^l$. The total decomposition level of each pyramid is set to the highest possible value$\lfloor \log 2 \min(H, W) \rfloor$, where $H \times W$ is the spatial size of source images and$\lfloor \cdot \rfloor$denotes the flooring operation.

Step 3: Coefficient fusion. For each decomposition level l, calculate the local energy map (sum of the squares of the coefficients within a small window) of $L\{A\}^l$ and $L\{B\}^l$, respectively.

$$E_A^l(x,y) = \sum_m \sum_n L\{A\}^l(x+m, y+n)^2,$$
$$E_B^l(x,y) = \sum_m \sum_n L\{B\}^l(x+m, y+n)^2. \tag{1}$$

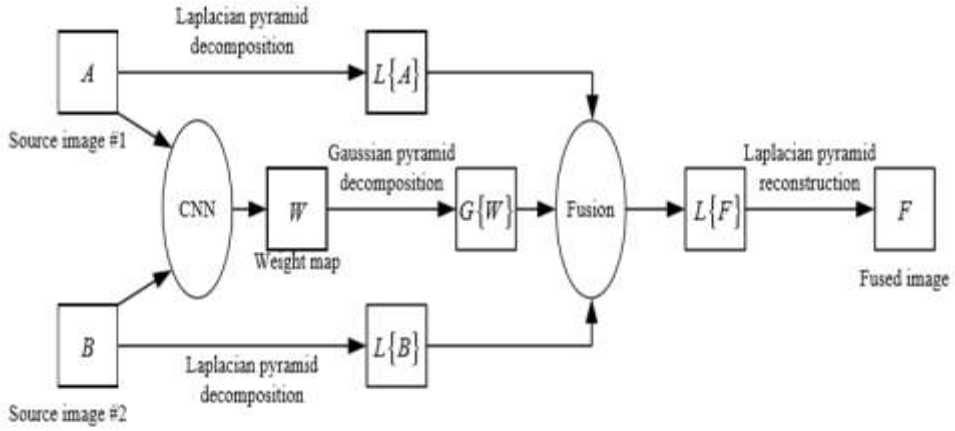


Figure 2. Proposed medical image fusion.

The similarity measure used for fusion mode determination is calculated as

$$M^l(x,y) = \frac{2\sum_m \sum_n L\{A\}^l(x+m, y+n) L\{B\}^l(x+m, y+n)}{E_A^l(x,y) + E_B^l(x,y)}. \tag{2}$$

The range of this measure is [-1 1] and a value closer to 1 indicates a higher similarity. A threshold t is set to determine the fusion mode to be used. If $M^l(x, y) \geq t$, the "weighted average" fusion mode based on the weight map W is adopted as

$$L\{F\}^l(x,y) = G\{W\}^l(x,y) \cdot L\{A\}^l(x,y) + \left(1 - G\{W\}^l(x,y)\right) \cdot L\{B\}^l(x,y). \tag{3}$$

If $M^l(x, y) < t$, the "selection" fusion mode via comparing the local energy in Eq. (1) is applied as

$$L\{F\}^l(x,y) = \begin{cases} L\{A\}^l(x,y), \text{ if } E_A^l(x,y) \geq E_B^l(x,y) \\ L\{B\}^l(x,y), \text{ if } E_A^l(x,y) < E_B^l(x,y) \end{cases} \quad (4)$$

The fusion strategy can be summarized shown in Eq. (5).

Step 4: Laplacian pyramid reconstruction. Reconstruct the fused image F from the Laplacian pyramid $L\{F\}^l$.

$$L\{F\}^l(x,y) = \begin{cases} G\{W\}^l(x,y) \cdot L\{A\}^l(x,y) + \left(1 - G\{W\}^l(x,y)\right) \cdot L\{B\}^l(x,y), \text{ if } M^l(x,y) \geq t \\ L\{A\}^l(x,y), \text{ if } M^l(x,y) < t \ \& \ E_A^l(x,y) \geq E_B^l(x,y) \\ L\{B\}^l(x,y), \text{ if } M^l(x,y) < t \ \& \ E_A^l(x,y) < E_B^l(x,y) \end{cases} \quad (5)$$

3.3 Pyramid Decomposition

Pyramid, or pyramid representation, is a type of multi-scale signal representation developed by the computer vision, image processing and signal processing communities, in which a signal or an image is subject to repeated smoothing and subsampling. Pyramid representation is a predecessor to scale-space representation and multiresolution analysis.

3.3.1 Pyramid Generation

There are two main types of pyramids: lowpass and bandpass.

A lowpass pyramid is made by smoothing the image with an appropriate smoothing filter and then subsampling the smoothed image, usually by a factor of 2 along each coordinate direction. The resulting image is then subjected to the same procedure, and the cycle is repeated multiple times. Each cycle of this process results in a smaller image with increased smoothing, but with decreased spatial sampling density (that is, decreased image resolution). If illustrated graphically, the entire multi-scale representation will look like a pyramid, with the original image on the bottom and each cycle's resulting smaller image stacked one atop the other.

A bandpass pyramid is made by forming the difference between images at adjacent levels in the pyramid and performing image interpolation between adjacent levels of resolution, to enable computation of pixelwise differences.

3.3.2 Pyramid Generation Kernels

A variety of different smoothing kernels have been proposed for generating pyramids. Among the suggestions that have been given, the binomial kernels arising from the binomial coefficients stand out as a particularly useful and theoretically well-founded class.Thus, given a two-dimensional image, we may apply the (normalized) binomial filter (1/4, 1/2, 1/4) typically twice or more along each spatial dimension and then subsample the image by a factor of two. This operation may then proceed as many times as desired, leading to a compact and efficient multi-scale representation. If motivated by specific requirements, intermediate scale levels may also be generated where the subsampling stage is sometimes left out, leading to an oversampled or hybrid pyramid. With the increasing computational efficiency of CPUs available today, it is in some situations also feasible to use wider support Gaussian filters as smoothing kernels in the pyramid generation steps.

Gaussian Pyramid

In a Gaussian pyramid, subsequent images are weighted down using a Gaussian average (Gaussian Blur) and scaled down. Each pixel containing a local average corresponds to a neighbourhood pixel on a lower level of the pyramid. This technique is used especially in texture synthesis.

The Gaussian pyramid is computed as follows. The original image is convolved with a Gaussian kernel. As described above the resulting image is a low pass filtered version of the original image. The cut-off frequency can be controlled using the parameter σ. The Laplacian is then computed as the difference between the original image and the low pass filtered image. This process is continued to obtain a set of band-pass filtered images (since each is the difference between two levels of the Gaussian pyramid). Thus, the Laplacian pyramid is a set of band pass filters.

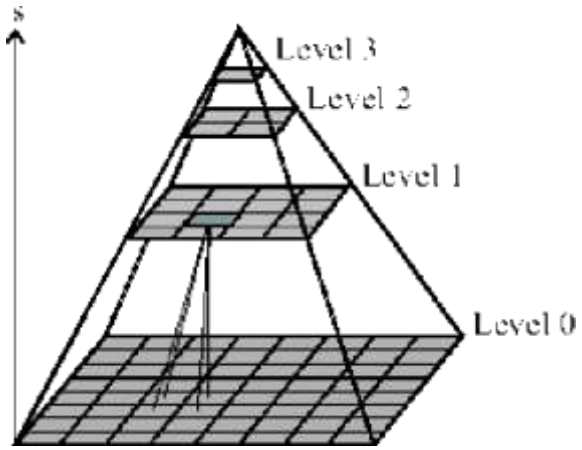


Figure 3. The filtered images stacked one on top of the other form a tapering pyramid structure, hence the name

Implementation Let $I(x, y)$ be the original image. The Gaussian pyramid on image I is defined as:

$$G_0(x, y) = I \tag{6}$$

$$G_{i+1}(x, y) = REDUCE\big(G_i(x, y)\big) \tag{7}$$

The REDUCE operation is carried out by convolving the image with a Gaussian low pass filter. The filter mask is designed such that the center pixel gets more weight than the neighbouring ones and the remaining terms are chosen so that their sum is 1. The Gaussian kernel is given by:

$$w(r, c) = w(r)w(c) \tag{8}$$

where, $w(r) = \left(\frac{1}{4} - \frac{a}{2}\ \frac{1}{4}\ \ a\ \ \frac{1}{4}\ \frac{1}{4} - \frac{a}{2}\right)$, a is chosen in the range 0.3 to 0.6. The prediction error $L_0(x, y)$is then given by.

$$L_i(x, y) = G_i(x, y) - EXPAND\big(G_{i+1}(x, y)\big) \tag{9}$$

The EXPAND operation is defined as follows:

$$G_{i+1}(x, y) = 4\sum_{m=-2}^{2}\sum_{n=-2}^{2} w(m, n)G_i\left(\frac{x-m}{2}, \frac{y-n}{2}\right) \tag{10}$$

Only terms for which (x-m)/2 and (y-n)/2 are integers are included in the sum. Rather than encode G_0,L_0 and G_1 is encoded. This results in a net data compression because:

- $L_0$is largely uncorrelated, and so may be represented pixel by pixel with many fewer bitsthanG_0
- G_1islow pass filtered, and so may be encoded at a reduced sample rate. Further data compression is achieved by iterating this process.

By repeating these steps several times, a sequence of images$L_0, L_1, L_2, \dots, L_n$ are obtained. If we now imagine these images stacked one above another, the result is a tapering pyramid data structure -

hence the name. The Laplacian pyramid can thus be used to represent images as a series of band-pass filtered images, each sampled at successively sparser densities.

Laplacian Pyramid

The Laplacian Pyramid (LP) was first proposed by Burt et al. for compact image representation. A Laplacian pyramid is very similar to a Gaussian pyramid but saves the difference image of the blurred versions between each level. Only the smallest level is not a difference image to enable reconstruction of the high-resolution image using the difference images on higher levels. This technique can be used in image compression.

The basic steps of the LP are as follows:

1. Convolve the original image g_0 with a lowpass filter w (e.g., the Gaussian filter) and subsample it by two to create a reduced lowpass version of the image $-g_1$.

2. This image is then up sampled by inserting zeros in between each row and column and interpolating the missing values by convolving it with the same filter w to create the expanded lowpass image g'_1 which is subtracted pixel by pixel from the original to give the detail image $- L_0$ given by

$$L_0=g_0-g'_1 \tag{11}$$

In order to achieve compression, rather than encoding g_0, the images L_0 and g_1 are encoded. Since g_1 is the lowpass version of the image it can be encoded at a reduced sampling rate and since L0 is largely decorrelated it can be represented by far fewer bits than required to encode g_0. The above steps can be performed recursively on the lowpass and subsampled image g_1 a maximum of N number of times if the image size is $2^N \times 2^N$ to achieve further compression. Thus, the result is a number of detail images $L_0, L_1, \ldots, L_N$ and the lowpass image g_N. Each recursively obtained image in the series is smaller in size by a factor of four compared to the previous image and its centre frequency reduced by an octave.

The inverse transform to obtain the original image g_0 from the N detail images $L_0, L_1, \ldots, L_N$ and the lowpass image g_N is as follows:

1. g_N is up sampled by inserting zeros between the sample values and interpolating the missing values by convolving it with the filter w to obtain the image g'_N.

2. The image g'_N is added to the lowest level detail image L_N to obtain the approximation image at the next upper level:

$$g_{N-1} = L_N + g'_N \tag{12}$$

Steps 1 and 2 are repeated on the detail images $L_0, L_1, \ldots, L_{N-1}$ to obtain the original image.

4. Results and discussions

All the experiments have been done in MATLAB 2016b version under the high-speed CPU conditions for faster running time with test images shown in figure 4 Aim of any fusion algorithm is to integrate required information from both source images in the output image. Fused image cannot be judged exclusively by seeing the output image or by measuring fusion metrics. It should be judged qualitatively using visual display and quantitatively using fusion metrics. In this section, we are presenting both visual quality and quantitative analysis of proposed and existing algorithms such as, Wavelet based methods discrete wavelet transform (DWT), stationary wavelet transform (SWT). Analysis of fusion metrics along with image quality assessment (IQA) metrics such as peak signal-to-

noise ratio (PSNR), structural similarity index (SSIM), correlation coefficient (CC), root mean square error (RMSE) and entropy (E) are considered to verify the effectiveness of the proposed algorithm. The objective of any fusion algorithm is to generate a qualitative fused image. For better quality, fused image should have optimal values for all these metrics. The fusion metric with best value is highlighted in bold letter. Visual quality of fused images obtained using state-of-art algorithms such as DWT, SWT and proposed method has demonstrated in figure 5, figure 6 and figure 7 with data set 1, data set 2 and dataset 3.

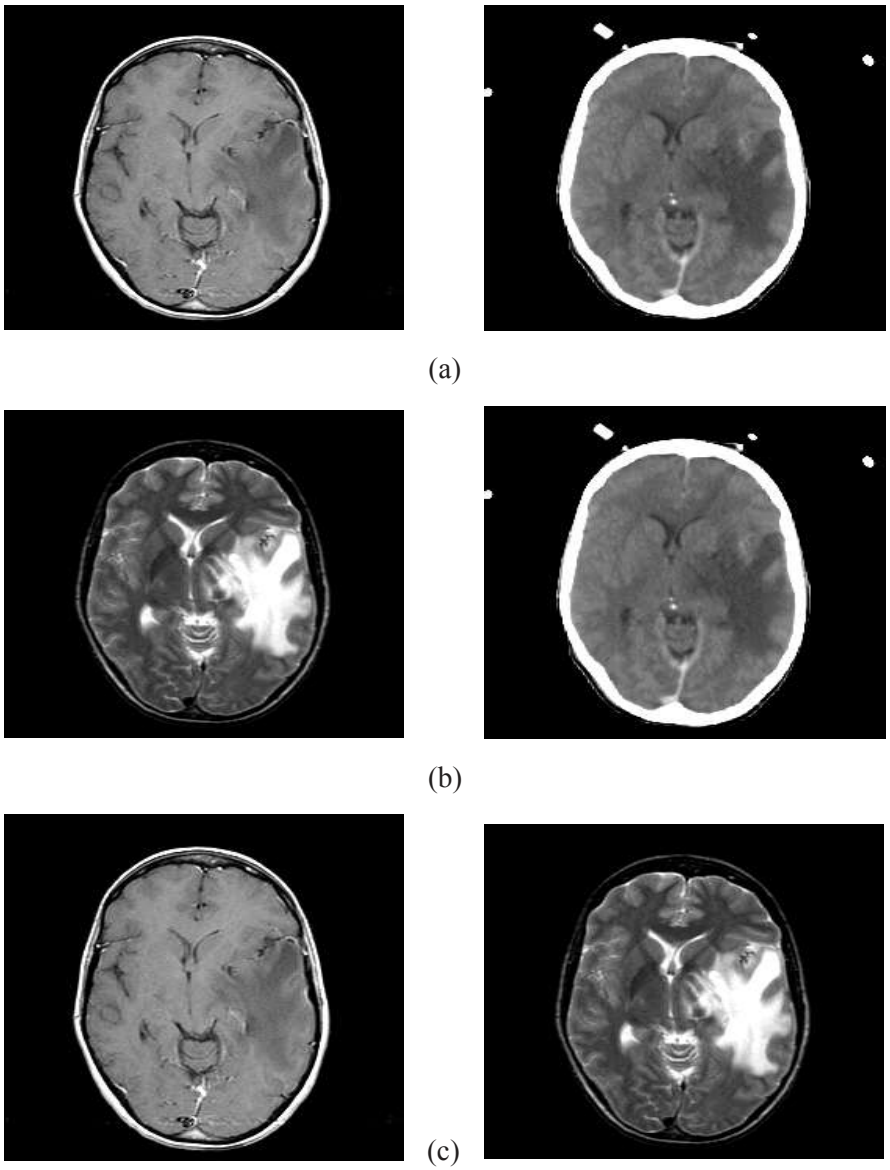


Figure 4. Test images (a) dataset 1 (MR-Gad & CT) (b) dataset 2 (MR-T2 & CT) (c) dataset 3 (MR-Gad and MR-T2)

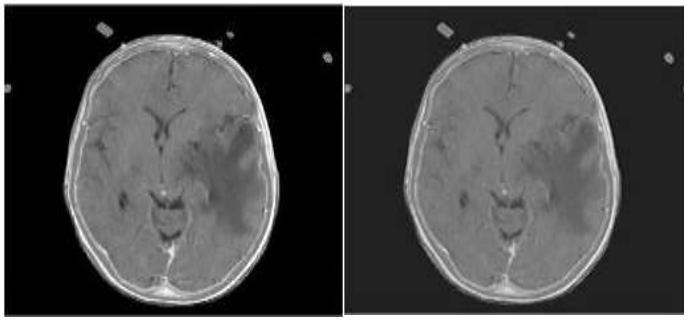
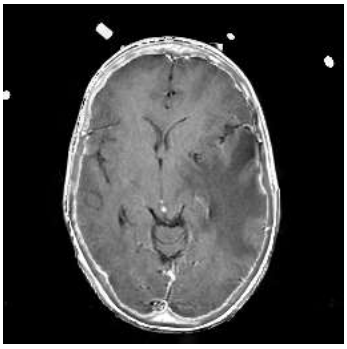


Figure 5. Obtained fused images for dataset 1using (a) existing (b) proposed (c) extension method.



(a) (b) (c)

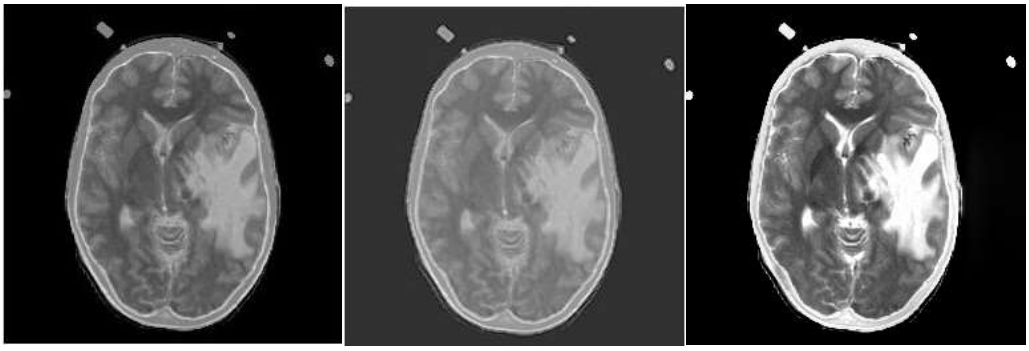


Figure 6. Obtained fused images for dataset 2 using (a) DWT (b) SWT (c) Proposed method

(a) (b) (c)

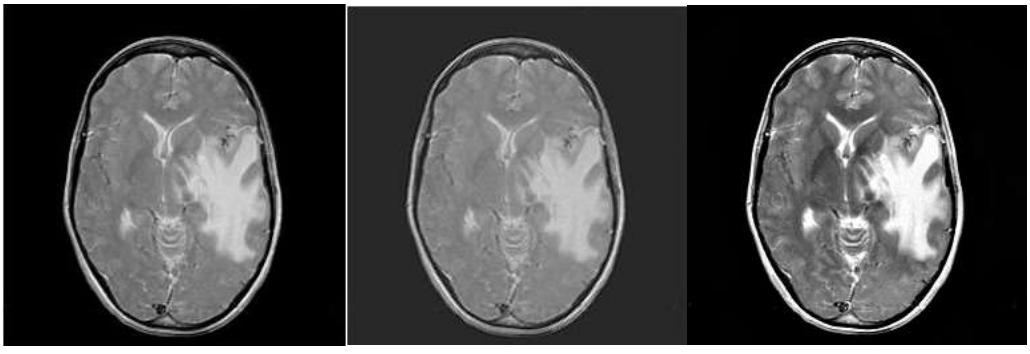


Figure 7. Obtained fused images for dataset 3 using (a) DWT (b) SWT (c) Proposed method.

(a) (b) (c)

However, all the existing fusion methods outputs not good at visual perception, lack of contrast with edge information and texture preservation. Our proposed method with 3 different

datasets which are presented in figure 5(c), figure 6(c) and figure 7(c) looks more quality in visualization, good contrast with proper edge information and excellent texture preservation as the value of entropy is much higher.

TABLE 1: QUANTITATIVE ANALYSIS OF FUSION METHODS FOR DATASET 1

Methodology	PSNR (in dB)	RMSE	CC	SSIM	Entropy
SWT	68.95	0.0909	0.94	0.988	1.12
DWT	68.69	0.093	0.944	0.98	1.11
Proposed method	**90.51**	**0.00047**	**1**	**1**	**4.37**

TABLE 2: QUANTITATIVE ANALYSIS OF FUSION METHODS FOR DATASET 2

Methodology	PSNR (in dB)	RMSE	CC	SSIM	Entropy
SWT	63.56	0.169	0.867	0.975	1.03
DWT	63.60	0.16	0.871	0.975	1.02
Proposed method	**86.60**	**0.0007**	**1**	**1**	**4.76**

TABLE 3: QUANTITATIVE ANALYSIS OF FUSION METHODS FOR DATASET 3

Methodology	PSNR (in dB)	RMSE	CC	SSIM	Entropy
SWT	65.05	0.142	0.885	0.979	1.009
DWT	65.02	0.14	0.89	0.979	0.99
Proposed method	**88.19**	**0**	**1**	**1**	**4.90**

Quantitative analysis with IQA sown in table 1 for the test results presented in figure 6.2, which gives the analysis of dataset 1. Table 1 consists of various fusion metric parameters such as PSNR, RMSE, CC, SSIM and entropy. The best values are highlighted in bold letters. Our proposed method obtained far better values over all the existing fusion methods discussed in the literature. We also tested the qualitative analysis of dataset 2 with the similar fusion metric parameters considered for dataset 1.

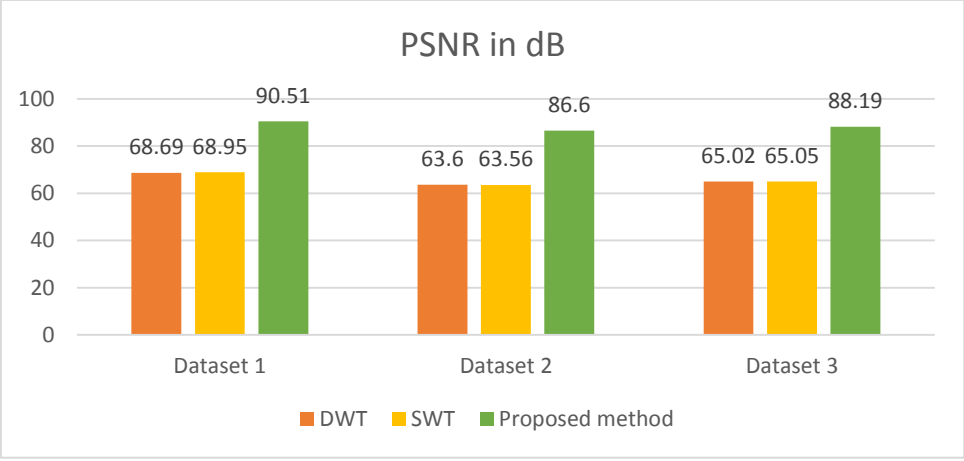


Figure 8. Performance analysis of PSNR for dataset 1, dataset 2 and dataset 3 with existing and proposed fusion methodologies.

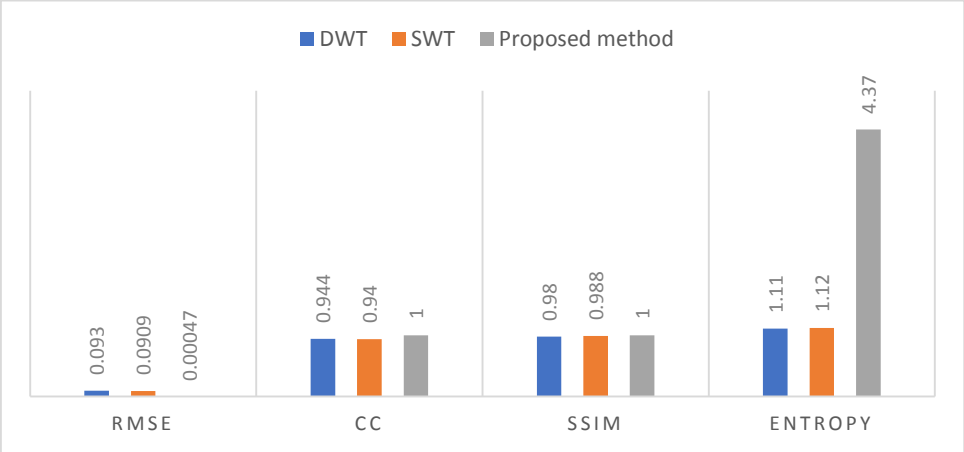


Figure 9. Graphical representation of other metrics.

5. Conclusion

In this paper, a medical image fusion method based on convolutional neural networks is proposed. We employ a Siamese network to generate a direct mapping from source images to a weight map which contains the integrated pixel activity information. The main novelty of this approach is it can jointly implement activity level measurement and weight assignment via network learning, which can overcome the difficulty of artificial design. To achieve perceptually good results, some popular techniques in image fusion such as multi-scale processing and adaptive fusion mode selection are appropriately adopted. Experimental results demonstrate that the proposed method can obtain high-quality results in terms of visual quality and objective metrics. In addition to the proposed algorithm itself, another contribution of this work is that it exhibits the great potential of some deep learning techniques for image fusion, which will be further studied in the future.

References

[1]. Wang, Zeyu, et al. "Medical image fusion based on convolutional neural networks and non-subsampled contourlet transform." *Expert Systems with Applications* 171 (2021): 114574.

[2]. Zhang, Yu, et al. "IFCNN: A general image fusion framework based on convolutional neural network." *Information Fusion* 54 (2020): 99-118.

[3]. Tan, Wei, et al. "Multimodal medical image fusion algorithm in the era of big data." *Neural Computing and Applications* (2020): 1-21.

[4]. Hermessi, Haithem, Olfa Mourali, and Ezzeddine Zagrouba. "Multimodal medical image fusion review: Theoretical background and recent advances." *Signal Processing* 183 (2021): 108036.

[5]. Tirupal, T., B. Chandra Mohan, and S. Srinivas Kumar. "Multimodal medical image fusion techniques–a review." *Current Signal Transduction Therapy* 16.2 (2021): 142-163.

[6]. Kaur, Manjit, and Dilbag Singh. "Multi-modality medical image fusion technique using multi-objective differential evolution based deep neural networks." *Journal of Ambient Intelligence and Humanized Computing* 12 (2021): 2483-2493.

[7]. Alseelawi, Nawar, Hussein Tuama Hazim, and Haider TH Salim ALRikabi. "A Novel Method of Multimodal Medical Image Fusion Based on Hybrid Approach of NSCT and DTCWT." *International Journal of Online & Biomedical Engineering* 18.3 (2022).

[8]. Xia, Jingming, Yi Lu, and Ling Tan. "Research of multimodal medical image fusion based on parameter-adaptive pulse-coupled neural network and convolutional sparse representation." *Computational and Mathematical Methods in Medicine* 2020 (2020).

[9]. Jose, Jais, et al. "An image quality enhancement scheme employing adolescent identity search algorithm in the NSST domain for multimodal medical image fusion." *Biomedical Signal Processing and Control* 66 (2021): 102480.

[10]. Li, Xinhua, and Jing Zhao. "A novel multi-modal medical image fusion algorithm." *Journal of Ambient Intelligence and Humanized Computing* 12 (2021): 1995-2002.

[11]. Li, Xiaosong, et al. "Multimodal medical image fusion based on joint bilateral filter and local gradient energy." *Information Sciences* 569 (2021): 302-325.

[12]. Zhang, Hao, et al. "Image fusion meets deep learning: A survey and perspective." *Information Fusion* 76 (2021): 323-336.

[13]. Subbiah Parvathy, Velmurugan, Sivakumar Pothiraj, and Jenyfal Sampson. "A novel approach in multimodality medical image fusion using optimal shearlet and deep learning." *International Journal of Imaging Systems and Technology* 30.4 (2020): 847-859.

[14]. Deng, Xin, and Pier Luigi Dragotti. "Deep convolutional neural network for multi-modal image restoration and fusion." *IEEE transactions on pattern analysis and machine intelligence* 43.10 (2020): 3333-3348.

[15]. Dinh, Phu-Hung. "Multi-modal medical image fusion based on equilibrium optimizer algorithm and local energy functions." *Applied Intelligence* 51.11 (2021): 8416-8431.

[16]. Ding, Zhaisheng, et al. "Brain medical image fusion based on dual-branch CNNs in NSST domain." *BioMed research international* 2020 (2020).

[17]. Muzammil, Shah Rukh, et al. "CSID: A novel multimodal image fusion algorithm for enhanced clinical diagnosis." *Diagnostics* 10.11 (2020): 904.

[18]. Yu, Hang, et al. "Convolutional neural networks for medical image analysis: state-of-the-art, comparisons, improvement and perspectives." *Neurocomputing* 444 (2021): 92-110.

[19]. Goyal, Bhawna, et al. "Measurement and analysis of multi-modal image fusion metrics based on structure awareness using domain transform filtering." *Measurement* 182 (2021): 109663.

[20]. Boveiri, Hamid Reza, et al. "Medical image registration using deep neural networks: a comprehensive review." *Computers & Electrical Engineering* 87 (2020): 106767.