

# Efficient Fusion of Medical Images Based on CNN

Randa Ali

Dept. of Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt  
randaaly60@gmail.com

Taha E. Taha

Dept. of Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt  
taha117@hotmail.com

Fathi E. Abd El-Samie

Dept. of Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt  
fathi\_sayed@yahoo.com

Walid El-Shafai

Dept. of Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf, Egypt  
walid.elshafai@el-eng.menofia.edu.eg

**Abstract**— Fusion of images has an important role in medical image diagnosis. Medical imaging systems have some limitations due to degradations and noise. Hence, it is necessary to apply some techniques such as image fusion to integrate information from different image modalities into the fusion result. With this strategy, the image features will be more defined, and this in turn, helps for better diagnosis. This paper presents an image fusion technique for different kinds of images to enhance the quality of brain tumor images. This technique depends on convolutional neural network (CNN) with more layers to obtain high-quality images for better diagnosis. The learning process of the CNN helps to measure the activity level and fusion ratio. The obtained fusion results are compared with those of traditional fusion methods. Different quality evaluation metrics are used in the assessment of the proposed technique.

**Keywords**—Image fusion, PCA, CNN, DWT, DT-CWT.

## 1. Introduction

Medical imaging is a new trend in the last decades that can help specialist for better diagnosis. Both Magnetic Resonance (MR) and Computed Tomography (CT) types of images have different characteristics and limitations. For example, MR images contribute much more details about soft tissues. On the other hand, CT images give information about bone structures [1]. The main limitations here are the absence of information about both soft tissues and dense layers. That is why the fusion of registered MR and CT images is necessary for information integration.

We use image fusion as a tool to get useful information from medical images. Generally, MR images are rich in information about soft tissues, while CT images are rich in information about dense tissues. Hence, the fusion of both types of images is necessary to obtain complementary information from both of them in a single image [2]. Table (1) illustrates a comparison between MR and CT images.

Different techniques have been presented for MR and CT image fusion. Wavelet transform has been used for multi-modality medical image fusion. It depends on the decomposition of both images with different scales. Fusion is performed using a certain fusion rule such as maximum and average fusion rules. The wavelet fusion with maximum fusion rule reduces any possible amount of blurring in each observation.

Unfortunately, wavelet fusion cannot deal with curved lines, correctly. Hence, other transforms such as curvelet transform can be used for medical image fusion.

The algorithm of curvelet fusion of MR and CT images [3] divides the images into tiles and applies Additive Wavelet Transform (AWT) for decomposition. Maximum fusion rule has been adopted in this work. Interpolation is utilized after the fusion process to obtain High Resolution (HR) images. Different interpolation techniques including LMMSE, maximum entropy and regularized interpolation have been used for this purpose.

Wavelet and curvelet fusion are traditional fusion techniques, which do not consider the local image activity levels in the fusion process. Hence, there is a need to use sophisticated image fusion tools such as deep neural networks to accommodate for the image local activity levels.

**Table 1** Comparison between CT and MR images.

	MR images	CT images
Soft tissue	Maximum	Minimum
Calcination	Insensitive	Sensitive
Bone	Distinct	Unclear
Gristle	Distinct	Unclear
Tendons	Contrast agent not needed	Needed contrast agent
Bleeding	Displayable	Distinct

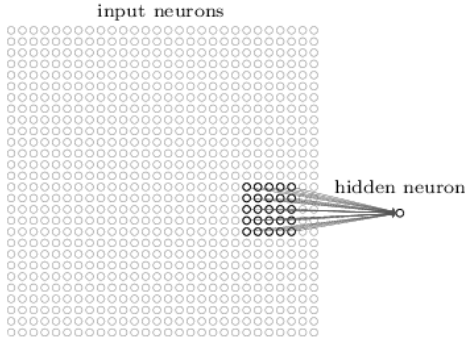
Fusion of medical images helps for disease diagnosis. So, the fusion should be performed, carefully. The resolution and details of medical images should be high [4]. The most important characteristics of the organ of interest should be represented in the fused images, and our rule is to select the best fusion technique.

## 2. Preliminaries

### 2.1 Convolutional Neural Network

A CNN is a trainable multi-stage neural network, which is fed-forward. Neuron is a coefficient or unit in the feature map. Convolution, activation, and spatial-domain pooling are used to create the feature maps. Shared weights, sub-sampling and local receptive fields are the basic building blocks of CNNs [5]. The input neurons are basically the pixel intensities of the input image. Each neuron is connected to the local receptive field (region in the input image) [6]. Spatially-invariant feature maps are generated from the convolutional kernels [7]. Mathematically,  $x^i$  is

the  $i$ -th feature map input and  $y^j$  is the  $j$ -th feature map output of the convolutional layer. The ReLU activation function used in the CNN is a non-linear function explained as follows [8]:



**Fig. 1.** Local receptive field

$$y^j = \max(0, b^j \sum_i k^{ij} \otimes x^i)(1)$$

where  $b^j$  is the bias and  $k^{ij}$  is the convolutional kernel between  $x^i$  and  $y^j$ . A convolution matrix is a small matrix used for edge detection, sharpening, blurring and embossing. This is done by a convolution between an image and a kernel. The symbol  $*$  indicates convolution operation. Another type of feature extraction is the pooling, which is used to decrease the size of the image by combining neighboring pixels into a single pixel. Average and max-pooling are operations performed in the CNN.

### 2.2 CNNs for Image Fusion

In the fusion of images, the focus map generation can be viewed as a classification task [9]. Fusion with a CNN is similar to classification. Spatial-domain fusion methods are performed by the measurement of the activity levels to extract the high-frequency details by designing the local filters. In a CNN model, the basic operation is the convolution [10]. Fusion based on CNN surpasses existing fusion methods, as it overcomes the difficulties in designing filters to extract effective features. Learning of a CNN is performed to allow activity level measurement and fusion. So, the fusion method based on the CNN gives higher quality results than those of traditional fusion methods.

### 3- Proposed CNN-Based Medical Image Fusion Method

The CNN-based medical image fusion method is illustrated in Fig. 2. We used two input images ( $A$  and  $B$ ). The proposed technique goes through four stages: focusing detection, segmentation, consistency verification and fusion.

In the first step, we input the two images ( $A$  and  $B$ ) to a CNN trained before to get a score map. The score map includes the focus information of the two input images. The average overlapping in the score map gives the focus map with the same size as that of the two input images.

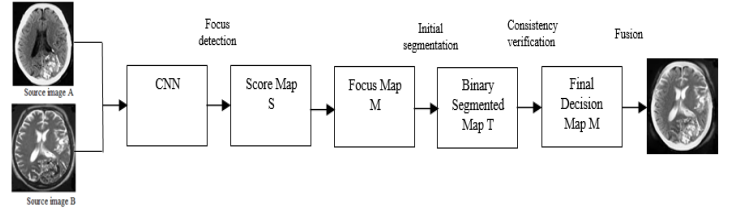
In the second step, we transform the focus map into a binary map with a threshold level of 0.5.

In the third step, we perform removal of the small region through a filtering strategy to refine the binary segmented map.

In the fourth step, consistency verification is implemented through pixel-wise weighted averaging to get the fused image with the final decision map.

### 3.1 Design of the Network

The input images have a variable spatial size. Images are divided into overlapping patches, and then we apply a sliding window technique.



**Fig. 2.** Proposed CNN image fusion.

As the input and output data of the fully-connected layers have fixed dimensions, to solve this problem, we perform reshaping of parameters after the conversion to feed the fully-connected layer from the convolutional layer. The network contains max-pooling and convolutional layers and can treat input images with different sizes to get the dense predictions [11]. The score map output is obtained, and it shows the focus property for the patch pairs of the input images to the network. In [12], to compare the similarity of the patches, we use a Siamese model for its advantages over other models, including 2-channel and pseudo-Siamese models. Siamese model is easier to train in image fusion, leading to an easy convergence.

The batch size selection is a vital problem in any classification problem. Large batch sizes of size  $32 \times 32$  are appropriate for getting high classification performance. Unfortunately, the process may include focused and unfocused regions of images. Hence, there is a difficulty in the determination of the number of max-pooling layers.

On the other hand, small patch sizes are not also appropriate as much details are missed. The solution that we will adopt in this paper is a patch size of  $16 \times 16$ . Max pooling is implemented to reduce the amount of features used in the classification process. Fig. 3 shows the proposed CNN model. The network consists of a single max-pooling layer, and three convolutional layers. The convolutional layer kernel size and stride are set to  $3 \times 3$  and 1, respectively. The max-pooling layer kernel size and the stride are set to  $2 \times 2$  and 2, respectively. After converting the fully-connected layers into convolutional ones in the test/fusion process, the network can be fed with two input images with arbitrary sizes to generate a dense score map. When the input images are of size  $H \times W$ , the map size of the output score is  $(\lceil H/2 \rceil - 8 + 1) \times (\lceil W/2 \rceil - 8 + 1)$ , where  $\lceil \cdot \rceil$  is the ceiling operation. Fig. 3 shows the obtained score map. The output score in the score map keeps the patch size of the input images to be  $16 \times 16$  to go through the different layers of the network.

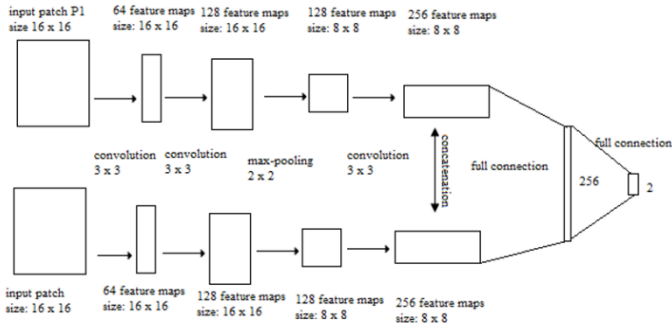


Fig.3. The CNN diagram used in the proposed fusion technique

### 3.2 Fusion Model

We have 2 input images  $A$  and  $B$  in gray-scale format. When feeding the input images to the trained CNN model, a score map  $S$  is obtained. The values of the score map range from 0 to 1. When the value is close to 1 or 0, the patch is more focused to the input image  $A$  or  $B$ . The focus map  $M$  has the same size as the input image, and it is generated by assigning the score map  $S$  coefficient values within the corresponding patch in  $M$  and averaging the overlapping pixels. Regions with more details are close to 1 (white) or 0 (black), while the regions are close to 0.5 (gray).

“Choose-max” strategy is used to process the focus map  $M$  to keep useful information according to a fixed threshold level of 0.5, to get a binary map  $T$ . Good performance of the CNN is obtained, when it is applied to classify gray levels in the focus map.

$$T(x, y) = \begin{cases} 1 & M(x, y) > 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

In this paper, the threshold of the input image is set to  $0.01 \times W \times H$ , where  $W$  and  $H$  are the width and height of every input image, respectively.

### 4 Simulation Results and Comparisons

Several simulation experiments are presented on three different cases comprising MR and CT images for fusion. The proposed fusion technique is implemented to fuse a pair of MR and CT images in each experiment. Different evaluation metrics including average gradient ( $avg. G$ ), local contrast ( $local C$ ), standard deviation (STD), and entropy ( $E$ ) have been considered in the assessment of results. In addition, the visual images obtained are also included. Simulation results of these experiments are presented in Tables (2, 3, 4).

It is clear that the processing time is acceptable, and entropy values are high and close to 8. This means a large amount of information in the fusion results. In addition, much details are found in the fusion results, and this is proved through STD and variance values.

## 6 Comparison with Other Techniques

### 6.1 Principal Component Analysis (PCA)

It is a spatial fusion method used to improve the resolution of the image and reduce the redundancy [13]. The spatial domain is the image plane, where some values are specified to different positions  $(x, y)$  [14].

### 6.2 Discrete Wavelet Transform (DWT)

The wavelet transform of discrete type (DWT) is frequently used in medical image fusion. It gives information about the image from all perspectives.

### 6.3 DT-CWT

Unlike the DWT which combines positive and negative frequencies, the DT-CWT treats positive and negative frequencies separately at each level.

### 6.4 Proposed Technique

The CNN fusion technique is compared with the PCA [13], and two other traditional fusion techniques.

Table 2 PCA method

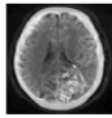
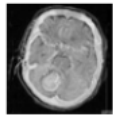
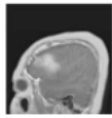
PCA	Case 1	Case 2	Case 3
Proc. time (Sec.)	0.542	0.523	0.498
Avg. G	0.038	0.033	0.028
XYdeas	0.241	0.185	0.277
E	7.564	7.059	6.073
STD	0.257	0.289	0.253
Local C	0.665	0.607	0.530
SSIM	0.993	0.995	0.998
MI	0.635	0.841	0.745
Visual results			

Table 3 DWT method.

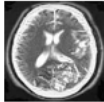
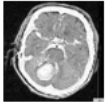
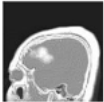
DWT	Case 1	Case 2	Case 3
Proc. time (Sec.)	0.372	0.172	0.166
Avg. G	0.064	0.057	0.044
<u>XYdeas</u>	0.474	0.307	0.315
E	7.737	7.313	6.043
STD	0.294	0.336	0.320
Local C	0.744	0.706	0.609
SSIM	0.993	0.992	0.997
MI	0.808	0.852	0.913
Visual results			

Table 5 proposal Method

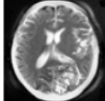
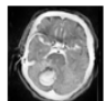
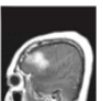
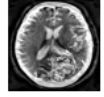
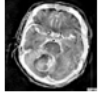
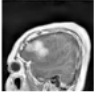
Proposed	Case 1	Case 2	Case 3
Processing time (Sec.)	10.152	9.767	9.608
Avg. G	0.047	0.042	0.046
<u>XYdeas</u>	0.497	0.480	0.760
E	7.620	7.433	6.542
STD	0.301	0.318	0.296
Local C	0.686	0.669	0.690
SSIM	0.526	0.470	0.462
MI	0.917	0.967	0.904
Visual results			

Table 4 DT-CWT Method

DT-CWT	Case 1	Case 2	Case 3
Processing time (Sec.)	0.786	0.779	0.780
Avg. G	0.068	0.0637	0.050
<u>XYdeas</u>	0.226	0.237	0.394
E	7.420	6.891	7.422
STD	0.317	0.347	0.312
Local C	1.037	1.114	0.789
SSIM	0.993	0.994	0.997
MI	0.535	0.565	0.653
Visual results			

### 7 Quantitative Evaluations

A comparison is presented between the proposed technique and the PCA, DWT and DT-CWT techniques, and the results are given in Tables (6, 7, 8).

Table 6 Evaluation of different methods incase 1.

Case 1	Proc. time	E	Avg	XY	STD	Local c	ssim	MI
<b>Proposed</b>	10.152	7.620	0.047	<b>0.497</b>	0.301	0.686	0.526	0.917
<b>PCA</b>	0.542	7.564	0.038	0.241	0.257	0.665	0.993	0.635
<b>DWT</b>	0.372	7.737	0.064	0.474	0.294	0.744	0.993	0.808
<b>DT-CWT</b>	0.786	7.420	0.068	0.226	0.317	1.037	0.993	0.535

Table 7 Evaluation of different methods incase 2

Case 2	Proc. time	E	Avg	XY	STD	Local c	ssim	MI
<b>Proposed</b>	9.767	<b>7.433</b>	0.042	<b>0.480</b>	0.318	0.669	0.470	<b>0.967</b>
<b>PCA</b>	0.523	7.059	0.033	0.185	0.289	0.607	0.995	0.841
<b>DWT</b>	0.172	7.313	0.057	0.307	0.336	0.706	0.992	0.852
<b>DT-CWT</b>	0.779	6.891	0.064	0.237	0.347	1.114	0.994	0.565

Table 8 Evaluation of different methods in case 3

Case 3	Proc. time	E	Avg	XY	STD	Local c	ssim	MI
<b>Proposed</b>	9.608	6.542	0.046	<b>0.760</b>	0.296	0.690	0.462	0.904
<b>PCA</b>	0.498	6.073	0.274	0.277	0.253	0.530	0.998	0.745
<b>DWT</b>	0.166	6.043	0.044	0.315	0.320	0.609	0.996	0.913
<b>DT-CWT</b>	0.780	7.422	0.050	0.394	0.312	0.789	0.997	0.653

## 8 Conclusion

This paper introduced a complete optimized fusion system for achieving image fusion by learning a CNN. The obtained results are good, when compared with those of other techniques. In addition, the processing time is low. So, the proposed technique is a recommended solution to image fusion for images of different kinds like medical images of different modalities.

## References

1- H. M. El-Hoseny, S. M. EL-Rabaie, W.A. El-rahman, F.E. Abd-El-Samie, "Medical image fusion techniques based on combined discrete transform domains", National Radio Science Conference (NRSC). IEEE, pp. 471–480, 2017.

2- K. Parmar, R.Kher, "A Comparative Analysis of Multimodality Medical Image Fusion Methods", Sixth Asia Modelling Symposium IEEE, pp. 93-97, 2012.

3- F. E. Ali, I. M. El-Dokany, A. A. Saad, and F. E. Abd El-Samie, "Curvelet fusion of MR and CT images", Progress In Electromagnetic Research C, vol. 3, pp. 215–224, 2008.

4- Y. L. Ping, L. B. Sheng, Z. D. Hua (2007) Novel image fusion algorithm with novel performance evaluation method. Systems Engineering and Electronics, pp.509–513,2007.

5- Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition". Proceedings of the IEEE, pp. 2278-2324, 1998.

6- <https://www.quora.com/What-is-a-receptive-field-in-a-convolutional-neural-network> 20-12-2019.

7- Y. Liu, X. Chen, H. Peng, Z. Wang, "Multi-focus image fusion with a deep convolutional neural network", Information Fusion, pp. 191-207,36,2017.

8- V. Nair, G. Hinton, "Rectified linear units improve restricted boltzmann machines", 27th International Conference on Machine Learning, pp.807–814, 2010.

9- S. Li, J. Kwok, Y. Wang, "Multi focus image fusion using artificial neural networks" Pattern Recognition Letters, pp.985–997, 2002, 23, 8.

10- J. Long, E. Shelhamer, T. Darrell, "Fully convolutional networks for semantic segmentation", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.3431–3440, (2015).

11- P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Y. L. "OverFeat: integrated recognition localization and detection using convolutional networks", arXiv preprint, vol. 4, pp.1–16, (2014).

12- S. Zagoruyko, N. Komodakis, "Learning to compare image patches via convolutional neural networks", IEEE Conference on Computer Vision and Pattern Recognition, pp. 4353–4361, (2015).

13- S. S. Bedi, "Contrast enhancement for PCA fusion of medical images", Journal of global research in computer science, pp. 25-29, (2013).

14- <https://www.quora.com/What-is-spatial-domain-in-image-processing> 20-1-2020.