Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2021) 29: 2780 – 2794

doi:10.3906/elk-2105-170

Medical image fusion with convolutional neural network in multiscale transform domain

Asan Ihsan ABAS[1,*], **Hasan Erdinç KOÇER**[2], **Nurdan AKHAN BAYKAN**[1]
[1]Department of Computer Engineering, Engineering Faculty, Konya Technical University, Konya, Turkey
[2]Department of Electrical and Electronics Engineering, Technology Faculty, Selçuk University, Konya, Turkey

Received: 18.05.2021 • **Accepted/Published Online:** 09.08.2021 • **Final Version:** 04.10.2021

Abstract: Multimodal medical image fusion approaches have been commonly used to diagnose diseases and involve merging multiple images of different modes to achieve superior image quality and to reduce uncertainty and redundancy in order to increase the clinical applicability. In this paper, we proposed a new medical image fusion algorithm based on a convolutional neural network (CNN) to obtain a weight map for multiscale transform (curvelet/ non-subsampled shearlet transform) domains that enhance the textual and edge property. The aim of the method is achieving the best visualization and highest details in a single fused image without losing spectral and anatomical details. In the proposed method, firstly, non-subsampled shearlet transform (NSST) and curvelet transform (CvT) were used to decompose the source image into low-frequency and high-frequency coefficients. Secondly, the low-frequency and high-frequency coefficients were fused by the weight map generated by Siamese Convolutional Neural Network (SCNN), where the weight map get by a series of feature maps and fuses the pixel activity information from different sources. Finally, the fused image was reconstructed by inverse multi-scale transform (MST). For testing of proposed method, standard gray-scaled magnetic resonance (MR) images and colored positron emission tomography (PET) images taken from Brain Atlas Datasets were used. The proposed method can effectively preserve the detailed structure information and performs well in terms of both visual quality and objective assessment. The fusion experimental results were evaluated (according to quality metrics) with quantitative and qualitative criteria.

Key words: Medical image fusion, convolutional neural networks, multiscale transform

1. Introduction

Medical image fusion is an important way to get both high spatial (MR) and high spectral (PET) information with as many details as possible for correct medical diagnosis and therapy. A PET scan is mainly used to show how blood circulates through the brain's arteries and veins or to get functional information. MRI scan is used to show the physical anatomy or structural information of the brain. Medical image fusion is used to extract the high special resolution part of the MR and added to the high spectral part of the PET image to get a more informative fused image than any of the input images. In multiscale transforms (MST) fusion methods, the most significant information of input multiscale coefficients is transferred to the fused coefficients. In general, the methods most commonly used as MST in image fusion include discrete wavelet transform (DWT) [1, 2], Laplacian pyramid (LP) [3, 4], curvelet transform (CVT) [5, 6], non-subsampled shearlet transform (NSST) [7] and non-subsampled contourlet transform (NSCT) [8]. The general image fusion based MST is performed as

*Correspondence: e138129002004@ktun.edu.tr

follows: image decomposition into low and high-frequency sub-bands at different resolutions, applying image fusion rules for combining the coefficients of different sub-bands, image reconstruction to achieve a merged image by taking the inverse MST of composite coefficients. In general fusion methods, a simple fusion rule such as choosing max or weighted average is applied to obtain the fused coefficients. These fusion strategies may cause loss of part of spectrum information, details, and the contrast in the fused image. To address this problem and enhance the fusion performance, many fusion methods have been proposed in the literature. Shen et al. [9] presented a cross-scale fusion rule for volumetric image fusion T1-weighted and T2-weighted MR images. That is used to select an optimal set of coefficients by effective exploitation of neighborhood information for each decomposition level, and it guarantees interscale consistencies. Du et al. [10] introduced the methods combine Laplacian pyramid with multiple features to completely transfer salient features from the source medical images into a fused image. Singh et al. [11] proposed MR-CT medical image fusion based on a shearlet transform and spiking neural network. In this work, high-frequency sub-band coefficients of shearlet transform are fused with a biologically motivated pulse coupled neural network as a bio-inspired neural network. Metaheuristic algorithms were also used in fusion studies, and successful results were obtained [12–14].

Recently, image fusion based deep learning is gaining popularity because it is able automatically to extract elective features from data without any intervention. Deep learning models can provide potential image representation approaches, which could be useful to the study of image fusion. This models can be listed as convolutional neural network (CNN) [15], convolutional sparse representation (CSR) and stacked auto encoder (SAE) [16]. Liu et al. [17] introduced CNNs to the field of image fusion for the first time and proposed a multifocus image fusion method based on CNN mode to achieve the state-of-the-art results in terms of visual quality and objective evaluation. In the same year, Liu also proposed a CNN based medical image fusion algorithm, which applies image pyramids to fuse multi-modal medical image in a multi-scale manner [18]. Liu et al. [19] presented an infrared (IR) and visible (VIS) image fusion method by using Siamese CNN (SCNN). In this method, SCNN-based image feature coding and feature classification approach were applied to create a weight map, and the resulting weight map was combined with activity measurement for image fusion. The fusion method aimed to retain thermal radiation information and texture detail information in IR images. After then, Li et al. [20] proposed a multimode medical image fusion by CNN and supervised learning, in order to solve the problem of practical medical diagnosis, to overcome the difficulty of manual design that the deep learning models using to extract the most effective features automatically from data.

In this paper, to get the best fusion result, we proposed medical image fusion based CNN with multiscale transforms (Curvelet (CvT) and non-subsampled shearlet transform (NSST)). A siamese convolutional neural network is applied to obtain a weight map from the focus map, then these weights are used in fusion process to fuse the pixel activity information of the source medical images. In fusion process, the medical source images and weight maps are decomposed using curvelet transform (CvT) and non-subsampled shearlet transform (NSST). Then, the weighted average fusion rule is used to merge the information from two source images. Finally, the fused image based on the focus map using the weighted-average fusion rule of high frequency and using activity level measurement of low-frequency coefficient is obtained. In the proposed method, it is aimed to improve the quality of the fusion image by using the weight map created by CNN and MST. By using MST (CvT and NSST) together with CNN, edges and details of source images were better preserved in the fusion result image at the combination of high and low frequency coefficients.

2. Material and methods

2.1. Convolutional neural networks architecture

Siamese convolutional neural network (SCNN) model is a pre-training model proposed by [17–19]. SCNN has two branches and these branches share the same set of weights. The uniform architecture, as well as each branch, has three convolutional layers with a non-linear layer like ReLU. The ReLU layers always follow a convolutional layer, max-pooling layers, and fully connected layers. The first convolutional layer is used for feature extraction and it is used to extract simple features of the image. In the second layer, the number of the feature maps are increasing with the depth of the network. The feature maps of two branches are combined and then pass through one fully-connected layer [17]. The size of each convolutional layer is set as 3×3 and max-pooling layer is set as 2×2. The convolution layers are used to extract the features from an input image, and it has a set of filters. Max-pooling layer is used to calculate the maximum value for each patch of the feature map of the image and reduces the dimension of the feature maps. The output network having fully-connected layer with the 256-dimensional vector is fed a (2-way) softmax layer. The full- connection has a fixed size on input and output data of images. The 256 feature maps obtained by each branch are concatenated and then fully-connected to a 256-dimensional feature vector. To avoid repeated calculations and to allow arbitrary input size, the fully-connected layer is converted into an equivalent convolutional layer containing kernels of size $8 \times 8 \times 512$. The output of the network is a 2-dimensional vector fed into a 2-way softmax layer, which calculates a probability distribution of different characteristics over two classes. The two classes correspond to two kinds of normalized weight assignment results for a pair of image patches *(p1, p2)*. The method proposed in this study aimed to obtain the output (weight map) of the CNN, which contains the value of the integrated pixel and is in the range of [0,1]. A weight map is calculated using the pre-trained CNN model. The network aims to measure the activity level and generate weight map automatically. The network structure, which is deeper than the "shallow" network as proposed in [17–19] is given in Figure 1.

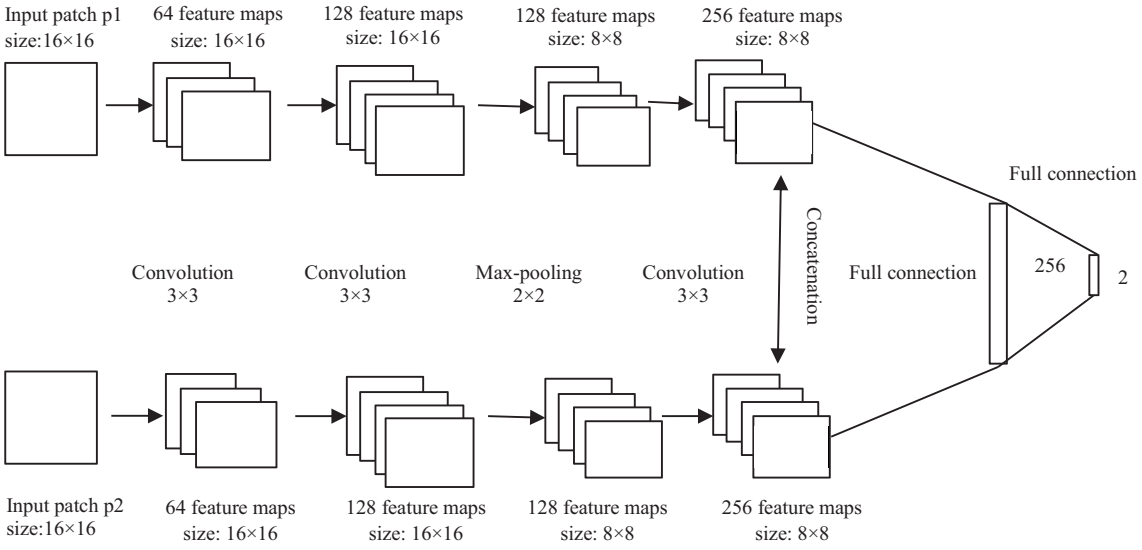


Figure 1. The architecture of the siamese neural network.

In this study, SCNN parameters given in Table 1 were used as suggested in [17, 18].

Table 1. Parameter settings of SCNN.

Layers	Patch size	Kernel size	Stride	Feature dimension
Conv 1	16×16	3×3	1	64
Conv 2	16×16	3×3	1	128
Max-Pooling	8×8	2×2	2	128
Conv 3	8×8	3×3	1	256
Concatenation	8×8	-	-	512
FC1	8×8	8×8	-	256

2.2. Training of CNN

In this study, training examples are generated using high-quality images with blurred version patches by the network [21]. Initially, each image is converted into grayscale image. Then, a 7 × 7 Gaussian filter with a standard deviation of 2 is applied to the input image to obtain five blurred images with different blur levels [21]. Next, for each original and blurred image, the 20 pair's patches of 16 ×16 image blocks are randomly sampled. A training example is labeled as Equation (1).

$$\begin{array}{l} \text{If } (p_1 = p_c) \text{ and } (p_2 = p_b) \text{ then Label} = 1 \text{ (positive)} \\ \text{If } (p_1 = p_b) \text{ and } (p_2 = p_c) \text{ then Label} = 0 \text{ (negative)} \end{array} \tag{1}$$

Here, p_1 and p_2 are the input patch images; p_c is the clear patch of image and p_b is the blurred patch of the image. In this work, the softmax loss function is used as the optimization objective of the network. The stochastic gradient descent (SGD) is applied to minimize the loss function. The batch size is set to 16 ×16. The momentum is set to 0.9, and the weight decay is 0.0005. The weights are updated with the Equations (2) and (3):

$$v_{i+1} = 0.9 * v_i - 0.0005 * a * w_i - a * \frac{\partial L}{\partial w_i} \tag{2}$$

$$w_{i+1} = w_i + vi + 1 \tag{3}$$

Where v_i represents the momentum variable at ith iteration, w_i is the weight at ith iteration, α denotes the learning rate and L denotes the loss function. CNN model is trained using the popular deep learning framework Caffe [22]. Weight initialization of each convolutional layer is achieved with the Xavier algorithm [23]. The biases in each layer are initialized with the value of 0. The weight map is the output of the training phase. The output of the training network is a weight map for the final decision.

2.3. The weighted map

The input source images are subject to a 2-dimensional vector that fed into a 2-way softmax layer, which produces a probability distribution between two classes in a range of [0, 1]. In the output network, the value of each coefficient map contains the relative clarity information of a pair source image patches. Finally, to get a

network output with the same size of input images, a weight map is achieved by assigning the value of all the pixels and averaging the overlapping patches (image pixel).

2.4. Image fusion based on MST decomposition

MultiScale transforms (MST) algorithms are used in frequency domain to represent images in different scales and levels to extract features and information such as edges and corners; these algorithms are also used to attenuate noise. The MST methods decompose an image to the frequency sub-bands where the high-pass sub-band gives details about edges and the low-pass bands contain outlines about texture and energy.

General image fusion methods of MST can be summarized as follows: First, the source images are decomposed into high and low frequency sub bands using MST transform. Second, the coefficients of sub band images are fused using the average (mean) for low-frequency sub bands, and max absolute value for high-frequency sub bands [24]. Finally, the fused image is reconstructed by the inverse MST.

The curvelet transform (CvT) and non-subsampled shearlet transform (NSST) have a good performance in representing edges and curves; therefore, in this work, we have used NSST and CvT transform based on CNN to obtain a fused image, and we have compared the results of performance methods with each other.

2.4.1. Curvelet transform (CVT)

Curvelet Transform (CvT) is a multiscale directional transform and it is used to represent the image at different angles and scales. It is developed by Candès and Donoho to solve the problem limitation of Wavelet and Gabor transform [6]. The proposed CvT, based on an anisotropic geometric wavelet called Ridgelet transform, decomposes the image into a set of wavelet bands and to analyze each band by a local ridgelet transform. In this work, we used a 2D discrete CvT version implemented via the "wrapping" transform (FDCT-Wrap) that utilizes fast Fourier transform and proposed by Candès et al. in 2005 [25]. This method is faster and more robust than curvelet based ridgelet. In the frequency domain decomposition, the 2D curvelet transform adopts local Fourier transform and is applied at four steps (FDCT-Wrap):

1. The Fourier coefficients are obtained by applying 2D Fast Fourier transform to the source image and obtain Fourier coefficient $\hat{f}[i,j]$ (Equation (4)).

$$\hat{f}[i,j], -\frac{n}{2} \leq i,j < \frac{n}{2} \tag{4}$$

   Where i,j are the index of the pixel; $\hat{f}[i,j]$ is the Fourier coefficient and n is the Fourier sample.

2. For each scale j and angle i, multiply the interpolated object with the parabolic window (Equation (5)).

$$\hat{f}[i,j] = \hat{f}[i, j - i\tan\theta_i] x \hat{U}[i,j] \tag{5}$$

   Where $\hat{U}[i,j]$ is the parabolic window like "Cartesian", θ is the orientation in the range $\left(-\frac{\pi}{4}, \frac{\pi}{4}\right)$.

3. Wrap this data for the origin and obtain re-index data (Equation (6)).

$$\hat{f}[i,j] = W[\hat{U}\hat{f}]x[i,j] \tag{6}$$

4. In the last, the inverse 2D FFT is applying to each $\hat{f}$ to collect the discrete coefficient to obtain the 2D curvelet.

2.4.2. Non-subsampled shearlet transform (NSST)

Wavelet transform (WT) is not very efficient in capturing edges and other anisotropic features in the image. Direction information in an image is not determined with WT since it depends on scale and transformation parameters. Shearlet transform was proposed to overcome these limitations of WT. The shearlet transform is based on an affine system with composite dilations, and it has become the most successful and effective method in recent years to represent multidimensional data efficiently. Non-subsampled shearlet transform (NSST) is proposed by Easley [26] based on non-subsampled Laplacian pyramid (NSLP) and shift-invariant shearlet filter banks (SFB) to provide multiscale decomposition. It is used to represent an image in multidimensional space. NSST combines the NSLP filter and several shearing filters in shearlet transform [27, 28]. The NSST performs sub-band decomposition similar to the shearlet transform, but downsampling and upsampling that are used in shearlet transform are not used here because it causes a lack of shift-invariance, which is a very critical feature to prevent the undesirable Gibbs phenomenon of image fusion. In high-frequency components of NSST, the geometric information like edges, curves and textures of images is preserved while a salient feature and energy are preserved in the low-frequency.

At each NSLP decomposition level, one high frequency and one low-frequency sub-images are produced, and further, the low-frequency sub-band is decomposed iteratively. At the decomposition level m = 3, an image is decomposed into m + 1 = 4 sub-bands with the same size of the source image in which one sub-band image is the low-frequency component and other m images are the high-frequency sub-band images. Shearing filter is also used in higher frequency sub-images decomposition without sub-sampling, which satisfies the shift-invariance property. The steps of obtaining standard fusion image with NSST are given below:

1. Images A and B are decomposed using NSST into one low frequency and a series of high-frequency sub-band images.

2. After the NSST is analyzed, fusing the high-frequency coefficients based on the maximum selection (MAX(A, B)), and fusing the low-frequency coefficients based on the average ((A+B)/2) fusion rules.

3. Inverse NSST is performed on the fused low and high-frequency coefficient to get the fused image.

3. Proposed fusion methods

Fusion rule plays an important role in image fusion algorithms. The low-frequency band represents the approximate component, which contains the base information of the source image and which can control in contrast to the fused image. High-frequency sub-images represent detailed components and contain much salient information of edge details information of the input image at the different directions and scales. For fusion, the fusion rules like average or weighted average are commonly used in the low-frequency domain. This method might not be suitable for all types of images and usually cause loss of contrast preserving the overall brightness. Max absolute fusion rule is often used in the high-frequency domain and may transfer less fine-scale details between source images. To solve this problem in the fused image, the method based on MST and SCNN is proposed. In this study, images with the size of M = 256 and N = 256 are used. The schematic diagram of the proposed fusion method is shown in Figure 2. The proposed image fusion method is described as follows:

1. Apply siamese CNN network model to calculate a weight map (W) of the original images (MR/PET).
2. Convert original PET image from RGB (Red-Green-Blue) to IHS (Intensity-Hue-Saturation) color space.
3. Decompose the both source images (MRI, PET) into sub-bands using NSST and CVT to obtain low and high-frequency components.
4. Transform a weight map W into the frequency domain using NSST and CVT to obtain weights in frequency domains.
5. Multiply the values of the activity level measurement (WLE) (Equation (7)) and the weight map extracted by CNN from the source images to fuse the coefficients of the low frequency band. For this, apply the fusion rule given in Equation (8).

$$WLE_S(i,j) = \sum_{m=-r}^{r} \sum_{n=-r}^{r} W \times (m+r+1, n+r+1) L_S(i+m, j+n)^2 \tag{7}$$

$$LF_F^{MST} = \begin{cases} LF_A(i,j), & WLE_A(i,j).W_{cnn}(i,j) \geq WLE_B(i,j).W_{cnn}(i,j) \\ LF_B(i,j), & otherwise \end{cases} \tag{8}$$

Where WLE denotes activity level measurement based on local energy proposed in [29]. The $S \in \{A, B\}$ and W is a $(2r+1)$ x $(2r+1)$ weighting matrix with radius r. $W_{CNN}(i,j)$ is a weight map generated by CNN.

6. Apply the the "weighted average" fusion mode [37] based on the weight map W to fuse the coefficients of the high frequency band. For this, apply the fusion rule given in Equation (9).

$$HF_F^{MST} = W_{cnn}(i,j) x HF_A(i,j) + (1 - W_{cnn}(i,j)) x HF_B(i,j) \tag{9}$$

7. To obtain the fused image in greyscale, perform the inverse MST on the fused low-frequency coefficients and high-frequency coefficients.

$$Fused = MST^{-1}(LF_F^{MST}, HF_F^{MST}) \tag{10}$$

8. Finally, apply inverse IHS transform to get the fused RGB color image.

At Algorithm 1 , the pseudo-code of fusion process with CNN is given.

In this study, image fusion was performed using MRI and PET images. Within the scope of the study, image fusions were performed by using both MST and CNN-MST together, and the results were given comparatively. In order to examine the effects of high and low frequency coefficients in the image fusion with CNN, the results were obtained by using the coefficients separately in the fusion process. Different fusion studies carried out within the scope of the study are given below.

Algorithm 1 Steps of the proposed medical image fusion algorithm using CNN with MST

Input: the source images: *MR* and *PET*.

Part CNN feature extraction to get weight map

1:Inputs two same size source images MR and PET (PET resized to 256 $\times$256 and converted to greyscale image) to the trained siamese network;

2:Obtained feature map from each branch contains convolutional layers;

3:Generates a dense prediction map S, where each prediction has two dimensions;

4:**for** any prediction Si **do**

5:each prediction normalization processing to obtain a corresponding image weight with a dimension value of 1;

6:end for

7:for an overlapping region of two adjacent predictions, Sj and Sj+1 **do**

8:Averaging process to obtain the mean value of the overlapping image patches(pixels) to obtain the output weight map (W) of the same size of the source images;

9:same size of source images is obtained from the score map by averaging the overlapping patches.

10:end for

Parameters: the number of MST decomposition levels: L, the number of directions at each decomposition level

Part 2: MST decomposition

11:Convert image (PET) from RGB (Red-Green-Blue) space to IHS (Intensity-Hue-Saturation) space to get Grayscale (intensity) image;

12:For each level $l = 1 : L$

13:For each direction $k = 1 : K(l)$

For Weighted map (W) and each source image *MR and PET;*

14:Decomposed source images (MR and PET) to obtain $\{High, Low\}$ coefficients;

15:**end for**

16:**end for**

Part 3: Fusion of high-frequency bands

17:For each source image (MR and PET) with weight map (W);

18:For each level $l = 1 : L$

For each direction $k = 1 : K(l)$

19:Weighted-average rule is used to merge ($\mathrm{HF_A(i,j)}$,$\mathrm{HF_B(i,j)}$) according to Equation (9) to obtain $\mathrm{HF_F^{MST}}$;

20:**end for**

21:end for

Part 4: Fusion of low-frequency bands

For each source image (MR and PET) with weight map (W);

22:Calculate the WLE with weight map (W) for $(LF_A(i,j), LF_B(i,j))$ using Equation (7) and Equation (8); to obtain LF_F^{MST};

23:Perform the inverse MST on $\left(LF_F^{MST}, HF_F^{MST}\right)$ according to Equation (10) ;

24:The inverse IHS transform is used to get the fused RGB color space image.

Output: The *Fused* image.

1. Only using MST (CvT / NSST with basic fusion rules for low / high bands): While applying the "average" rule for low frequency components; the "max-absolute" rule was chosen for high-frequency components [38].

2. Using CNN for only high frequency (CvT / NSST + basic fusion rule for low band + (CNN_{HF})) : The "average" rule was applied for low frequency components and the proposed method given in Equation (9) was applied for high-frequency components.

3. Using CNN for both low and high frequency (CvT / NSST + CNN_{HF} + CNN_{LF}), Equation (8) was applied to low-frequency and Equation (9) was applied to high-frequency. The flowchart of the proposed medical image fusion is given in Figure 2.

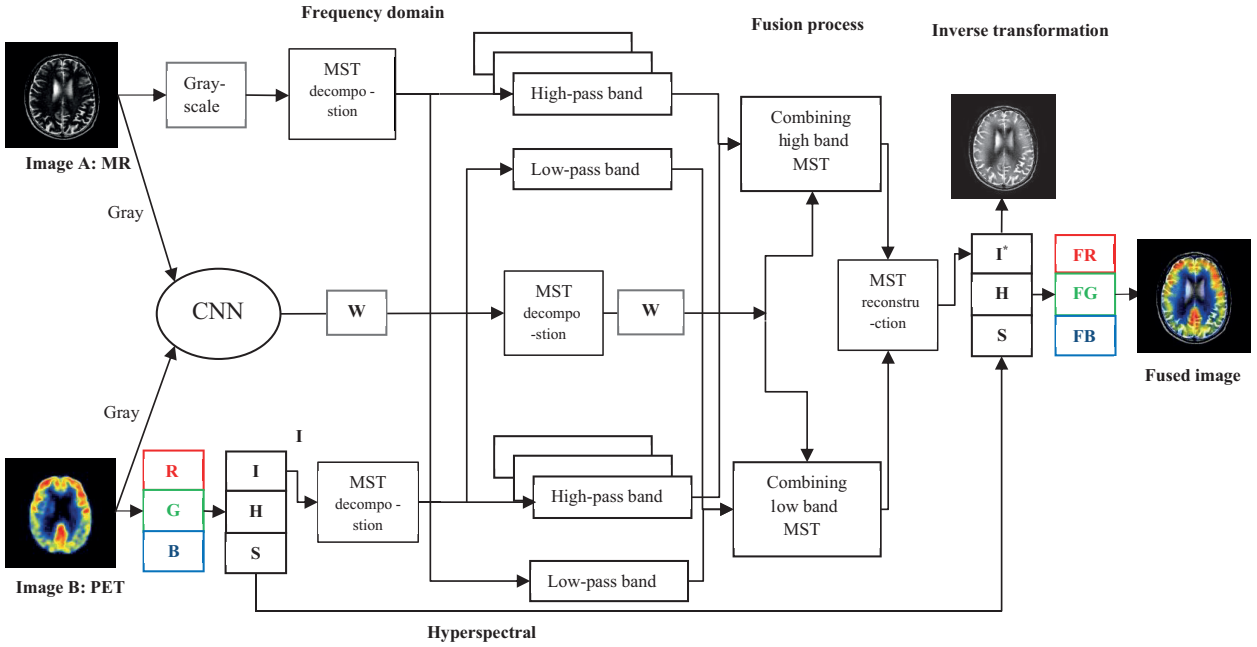


Figure 2. Flowchart of the proposed medical image fusion.

4. Experimental results

The source and reference medical images ("Alzheimer" image set) were taken from the Whole Brain Atlas database [30] that is available publicly, which is created by Harvard Medical School. Alzheimer images are of a 70 years old man who began experiencing difficulty with memory. While PET images are 128x128 pixels, MRI images are 256x256 pixels. Therefore, PET is resized to 256 $\times$256. The reference images were used to help evaluate the quality of the test image, and MATLAB MathWorks, Inc., Natick, MA, USA) was used for the experiments. The sample MRI and PET images from "Alzheimer" image data set are shown in Figure 3.

Comparisons show that the results are obtained by the proposed method, which has a better performance than basic MST methods. This means that the proposed method retains more information than other MST algorithms from source images. Evaluations of fusion performance were calculated by some fusion metrics as given below.

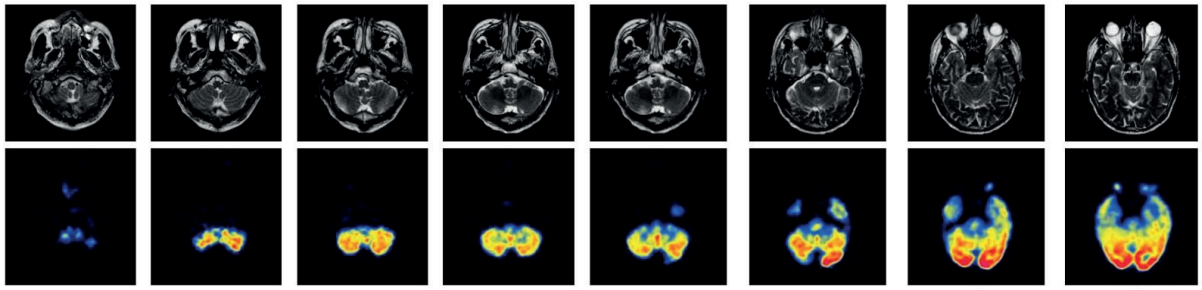


Figure 3. Sample MR and SPECT images of Alzheimer image data set [30].

1. Entropy: It's an important measurement used to evaluate the richness and information detail of pixel value in the image given in Equation (11).

$$E = -\sum_{i=0}^{L-1} p_{i\log_2(p_i)} \tag{11}$$

Where i is the probability density of the grayscale of a pixel value (Pi) and L is the number of intensity levels in the image.

2. Fusion factor(FF): This is used to measure the dependency and amount of mutual information (MI) between the source and fused image where MI measures the amount of information transferred from source to fused image as given in [31] (Equation (12)).

$$FF = MI_{AF} + MI_{BF} \tag{12}$$

Where MI_{AF} and MI_{BF} are mutual information is given in Equation (13).

$$MI_{A,BF} = \sum_{I,J} P_{A,BF}(I,J)(\log \frac{P_{AF,BF}(I.J)}{P_{A,B}(I)P_F(J)}) \tag{13}$$

Where $P_{A,B}$, and $P_{AF,BF}$ are the normalized joint grayscale of reference and fused images. The larger value of FF indicates a good amount of information comes from two source images.

3. Objective edge-based measure($Q^{AB/F}$): This is used to evaluate the amount of edge information transferred from the input images into the fused image at the pixel level by measuring the relative amount of edge or gradient information and employing a Sobel edge detector [32] (Equation (14)).

$$Q^{AB/F} = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M} Q^{AF}(n,m)w^A(n,m) + Q^{BF}(n,m)w^B(n,m)}{\sum_{i=1}^{N}\sum_{j=1}^{M}(w^A(i,j) + w^B(i,j))} \tag{14}$$

where Q^{AF} and Q^{BF} denoted to the edge preservation values, respectively. n, m are image pixel location, w^A and w^B are the weighting factors.

4. Standart deviation (STD): This is the most widely used method to measure the contrast and distribution between the pixel values and the mean values of the fused image [33] (Equation (15)).

$$STD = \sqrt{\sum_{x=1}^{M}\sum_{y=1}^{N}(F(x,y) - \overline{F(x,y)})^2} \tag{15}$$

STD represents the standard deviation of the fused image. $\overline{F(x,y)}$ is the mean of the fused image.

5. Piella's metric (QW): This is used to measure the structural similarity between the fused image and source images by addressing contrast, coefficient correlation, and illumination all at once [34] (Equation (16)).

$$Q_W(a,b,f) = \sum_{w \in W} c(w)(\lambda(w)Q_0(a,f|w) + (1 - \lambda(w)Q_0(b,f|w))) \tag{16}$$

Where $c(w)$ is overall saliency, a and b represent input images, f represents the fused image, $\lambda(w)$ represents a local weight between 0 and 1. The value Q_0 represents the similarity of the fused and source images and takes values between –1 and 1. w represents sliding window of pixels.

The sample results of fusion images getting from the study are given in Figure 4. All resulted images of fusion process are also given in Appendix. The MR source image (given in Figure 4 (a) as the gray-scaled image) and PET source image (given in Figure 4 (b) as the colored image) were fused both by only MST and proposed fusion methods (MST with CNN).

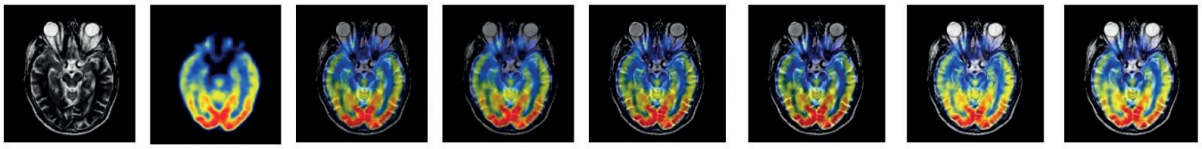


Figure 4. The sample input MR (a) and SPECT (b) images and the resulting images of fusion algorithms: CVT (c), NSST (d), CVT+High-Pass+CNN (e), NSST+High-Pass+CNN (f), CVT+Low-Pass&High-Pass+CNN (g) and NSST+Low-Pass&High-Pass+CNN (h).

Experimental results (average of objective evaluations of twenty three MR- PET image fusions) are given in Table 2. As seen from the results, the fused image obtained by our proposed methods has less color distortion and richer anatomical structural information. The box-plot graphics of performance metrics are shown in Figure 5.

The results we obtained by the methods used in the study were compared with the ones in other studies in the literature. According to the literature, it has been seen that different types of medical images (MR, CT, PET and PET) are used for image fusion [9, 11, 20, 29, 35, 36]. However, in terms of the reliability of the comparison, the results were compared to the ones obtained by Yin et al. [29], Wang et al. [35] and Asha et al. [36] and the comparison results are given in Table 3. The comparisons have been done according to literature studies on the image fusion methods where only three images of the dataset were used.

Table 2. Experimental results of CVT, NSST and proposed methods.

Fusion methods	MI/Fusion Factor	QAB/F	Entropy	Piella (QW)	STD	Running Time [s]
CVT	1.085207	0.495873	4.195350	0.67998	58.23761	2.23338
NSST	1.192435	0.398118	3.973277	0.926117	57.34500	5.27105
CVT+High-Pass +CNN	1.058632	0.551544	4.078167	0.894842	59.4448	56.17472
NSST+High-Pass+CNN	1.057745	0.566145	3.776068	0.886778	61.71	67.37383
CVT+Low-Pass& High-Pass+CNN	**1.297150**	0.590004	**4.304213**	0.920476	73.91037	62.90322
NSST+Low-Pass& High-Pass+CNN	1.2849701	**0.605715**	4.197245	**0.93356**	**74.22369**	64.31998

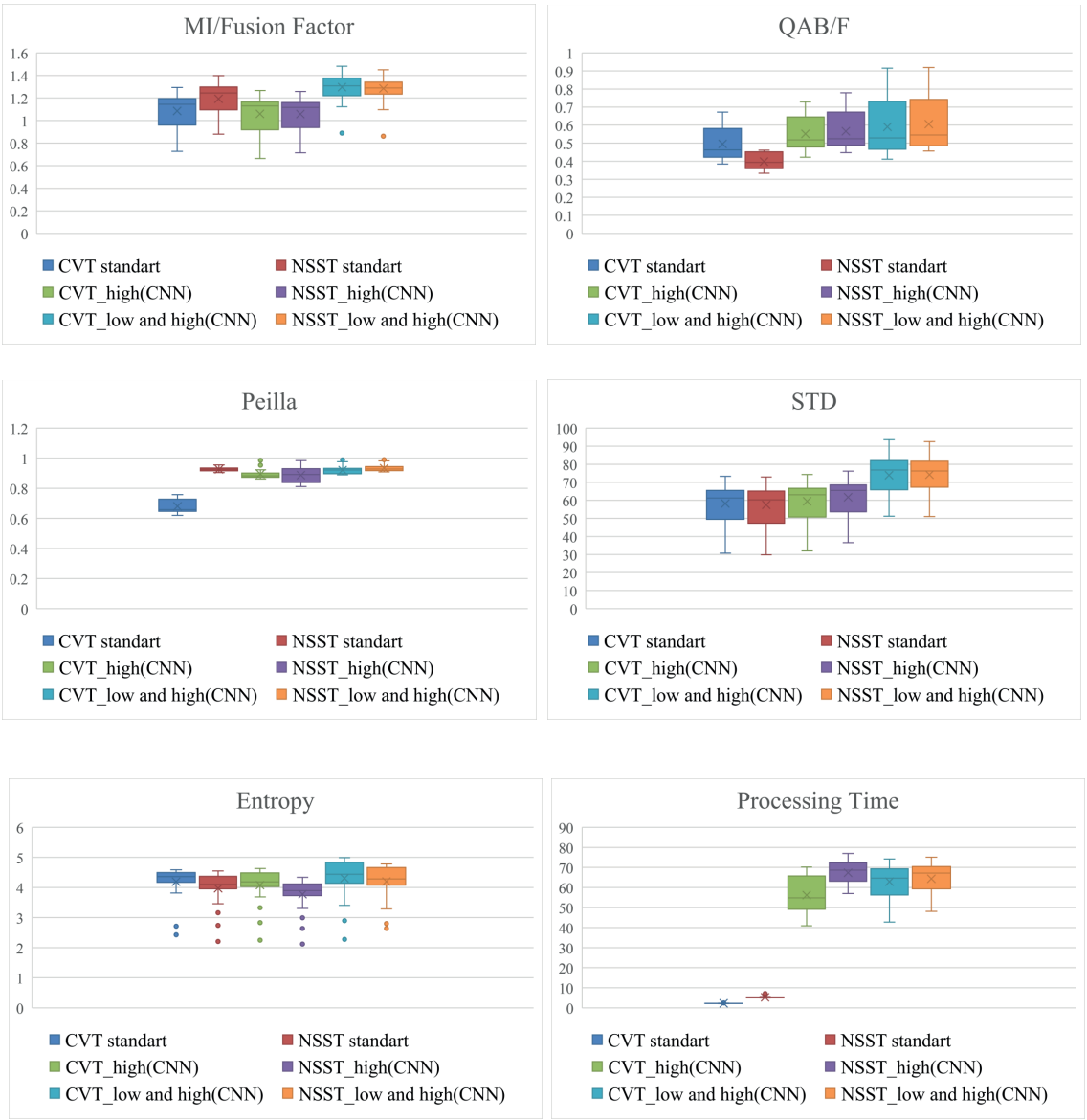


Figure 5. The graphic representation of performance metrics.

Table 3. Comparison of image fusion results with the literature.

Fusion methods	**MI/Fusion Factor**	**QAB/F**	**Entropy**	**Piella (QW)**	**STD**
CNN + Contrast Pyramid [35]	1.0925	0.4449	-	-	-
NSST + Chaotic Grey Wolf Optimization [36]	-	-	4.8686	0.8097	77.61502
NSST + Parameter-adaptive Pulse-coupled NN [29]	-	-	**4.9461**	0.8769	62.4796
CvT + CNN_{LF} + CNN_{HF} (Proposed)	**1.306892**	0.507295	4.920098	0.893203	**82.0025**
NSST + CNN_{LF} + CNN_{HF} (Proposed)	1.285977	**0.551556**	4.829476	**0.914767**	81.7243

As can be seen from the Table 3, the proposed approach gave better results in terms of Piella and standard deviation metrics. In terms of entropy metrics, the result achieved by Yin et al. [29] is high, but it is close to the result of our proposed method. These experimental results show that the methods we propose have yielded very successful and satisfying results.

5. Conclusion

In this paper, a new image fusion technique based on deep learning is proposed. In the proposed method, MST decomposes the source images into sub-band images and low and high sub-band coefficients of the source images are fused based on the weight map obtained by a siamese convolutional neural network. This method improves the quality of the final fused image. The fused images were evaluated visually and quantitatively by employing base fusion metrics. The simulation results show that the proposed techniques provide better fusion performance than the other methods; therefore, they can be utilized for better medical diagnosis.

Acknowledgment

This study was carried out as a part of the PhD thesis of Asan Ihsan Abas.

References

[1] Pajares G, De La Cruz JM. A wavelet-based image fusion tutorial. Pattern recognition 2004; 37 (9): 1855-72. doi: 10.1016/j.patcog.2004.03.010

[2] Vijayarajan R, Muttan S. Discrete wavelet transform based principal component averaging fusion for medical images. AEU-International Journal of Electronics and Communications 2015; 69 (6): 896-902. doi: 10.1016/j.aeue.2015.02.007

[3] Rockinger (1999). Image fusion toolbox. Website http://www.metapix.de/toolbox.htm [accessed 01 2021].

[4] Burt PJ, Adelson EH. The Laplacian pyramid as a compact image code. InReadings in computer vision, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc, 1987, pp. 671-679

[5] Yang Y, Tong S, Huang S, Lin P, Fang Y. A hybrid method for multi-focus image fusion based on fast discrete curvelet transform. IEEE Access 2017; 5: 14898-913. doi: 10.1109/ACCESS.2017.2698217

[6] Candes EJ, Donoho DL. Curvelets: A Surprisingly Effective Nonadaptive Representation for Objects With Edges. Nashville, Tenn, USA: Stanford Univ Ca Dept of Statistics, 2000.

[7] Guorong G, Luping X, Dongzhu F. Multi-focus image fusion based on non-subsampled shearlet transform. IET Image Process 2013; 7 (6): 633-639. doi: 10.1049/iet-ipr.2012.0558

[8] Zhang Q, Guo B. Multifocus image fusion using the nonsubsampled contourlet transform. Signal Process 2009; 89 (7): 1334-1346. doi: 10.1016/j.sigpro.2009.01.012

[9] Shen R, Cheng I, Basu A. Cross-scale coefficient selection for volumetric medical image fusion. IEEE T Bio-Med Eng 2012; 60 (4): 1069-1079.

[10] Du J, Li W, Xiao B, Nawaz Q. Union Laplacian pyramid with multiple features for medical image fusion. Neurocomputing 2016; 194: 326-339.

[11] Singh S, Gupta D, Anand R, Kumar V. Nonsubsampled shearlet based CT and MR medical image fusion using biologically inspired spiking neural network. Biomed Signal Proces 2015; 18: 91-101. doi: 10.1016/j.bspc.2014.11.009

[12] An H, Qi Y, Cheng Z. A novel image fusion method based on particle swarm optimization. Heidelberg, Berlin, Germany: Springer-In Advances in Wireless Networks and Information Systems, 2010, pp. 527-535.

[13] Madheswari K, Venkateswaran N. An optimal weighted averaging fusion strategy for thermal and visible images using dual tree discrete wavelet transform and self tunning particle swarm optimization. Multimedia Tools and Applications 2017; 76 (20): 20989-21010. doi: 10.1007/s11042-016-4030-x

[14] Abas AI, Baykan NA. Multi-Focus Image Fusion with Multi-Scale Transform Optimized by Metaheuristic Algorithms. Traitement du Signal 2021; 38 (2). doi: 10.18280/ts.380201

[15] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE 1998; 86 (11): 2278-324.

[16] Kaur H, Koundal D, Kadyan V. Image fusion techniques: a survey. Archives of Computational Methods in Engineering 2021: 1-23. doi: 10.1007/s11831-021-09540-7

[17] Liu Y, Chen X, Peng H, Wang Z. Multi-focus image fusion with a deep convolutional neural network. Inform Fusion 2017; 36: 191-207.

[18] Liu Y, Chen X, Cheng J, Peng H. A medical image fusion method based on convolutional neural networks. In: IEEE 2017 20th international conference on information fusion (Fusion); Xi'an, China; 2017. pp. 1-7.

[19] Liu Y, Chen X, Cheng J, Peng H, Wang Z. Infrared and visible image fusion with convolutional neural networks. International Journal of Wavelets, Multiresolution and Information Processing 2018; 16 (03): 1850018.

[20] Li Y, Zhao J, Lv Z, Pan Z. Multimodal Medical Supervised Image Fusion Method by CNN. Frontiers in Neuroscience 2021; 15: 303. doi: 10.3389/fnins.2021.638976

[21] Liu Y, Chen X, Peng H, Wang Z. Multi-focus image fusion with a deep convolutional neural network. Information Fusion 2017; 36: 191-207. doi: 10.1016/j.inffus.2016.12.001

[22] Jia Y, Shelhamer E, Donahue J et al. Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM international conference on Multimedia; Florida, USA; 2014. pp. 675-678.

[23] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. Proceedings of the thirteenth international conference on artificial intelligence and statistics: JMLR Workshop and Conference Proceedings; Sardinia, Italy; 2010. pp. 249-256.

[24] Zhang Z, Blum RS. A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application. Proceedings of the IEEE 1999; 87 (8): 1315-1326.

[25] Candes E, Demanet L, Donoho D, Ying L. Fast discrete curvelet transforms. Multiscale Model Sim 2006; 5 (3): 861-899.

[26] Easley G, Labate D, Lim W-Q. Sparse directional image representations using the discrete shearlet transform. Applied and Computational Harmonic Analysis 2008; 25 (1): 25-46.

[27] Cao Y, Li S, Hu J. Multi-focus image fusion by nonsubsampled shearlet transform. In: IEEE 2011 Sixth International Conference on Image and Graphics; Anhui, China; 2011. pp. 17-21.

[28] Kong W, Lei Y. Technique for image fusion between gray-scale visual light and infrared images based on NSST and improved RF. Optik 2013; 124 (23): 6423-6431.

[29] Yin M, Liu X, Liu Y, Chen X. Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain. IEEE Transactions on Instrumentation and Measurement 2018; 68 (1): 49-64. doi: 10.1109/TIM.2018.2838778

[30] Summers D. Harvard Whole Brain Atlas: www. med. harvard. edu/AANLIB/home.html. Journal of Neurology, Neurosurgery and Psychiatry 2003; 74 (3): 288-288.

[31] Chaudhuri S, Kotwal K. Hyperspectral image fusion. Berlin, Germany: Springer, 2013.

[32] Xydeas Ca, Petrovic V. Objective image fusion performance measure. Electronics letters 2000; 36 (4): 308-309.

[33] Liu Y, Liu S, Wang Z. A general framework for image fusion based on multi-scale transform and sparse representation. Information fusion 2015; 24: 147-64. doi: 10.1016/j.inffus.2014.09.004

[34] Piella G, Heijmans H. A new quality metric for image fusion. Proceedings of the IEEE 2003 International Conference on Image Processing (Cat No 03CH37429); Barcelona, Spain; 2003. pp. III-173.

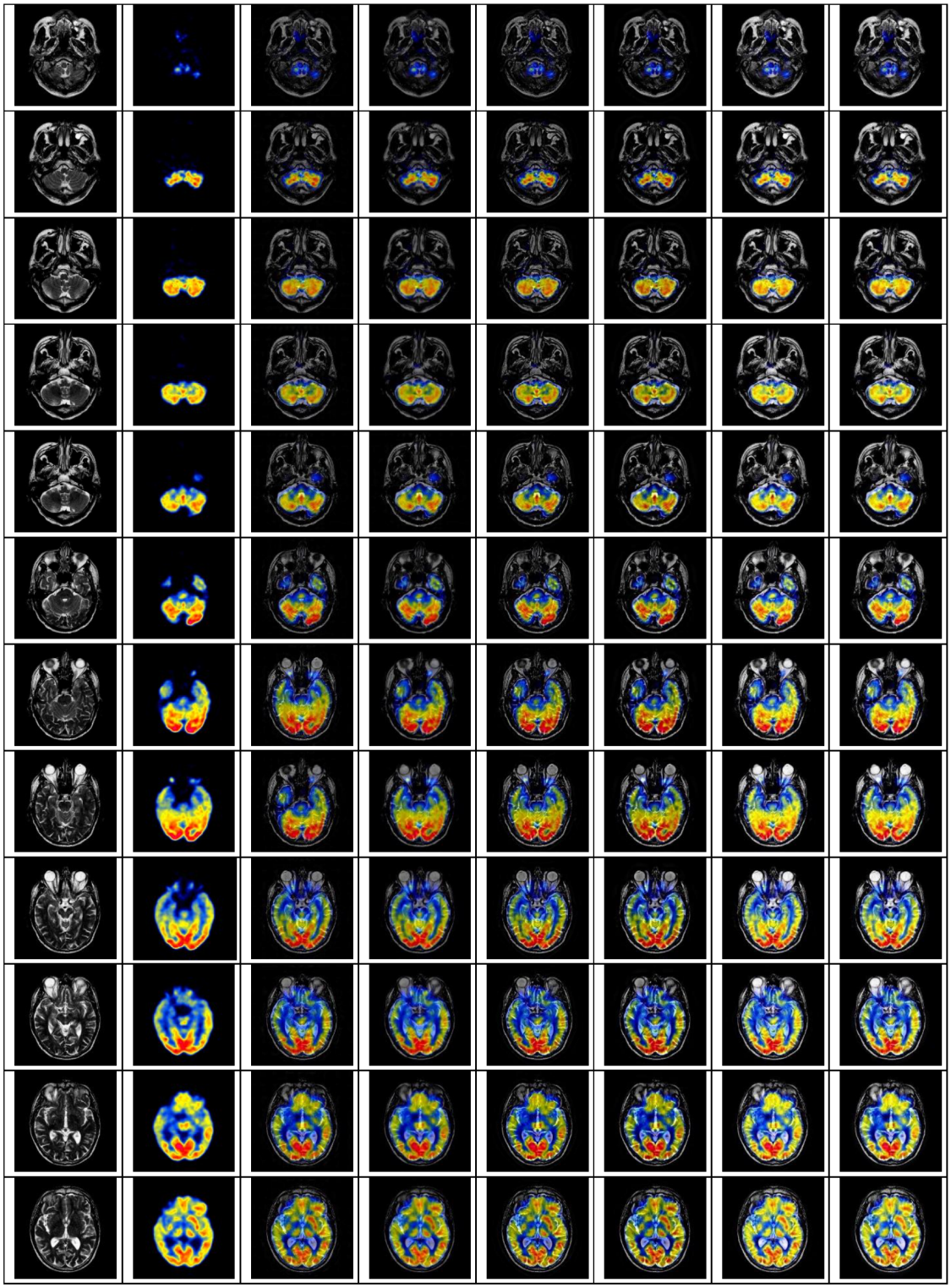
[35] Wang K, Zheng M, Wei H, Qi G, Li Y. Multi-modality medical image fusion using convolutional neural network and contrast pyramid. Sensors-Basel 2020; 20 (8): 2169.

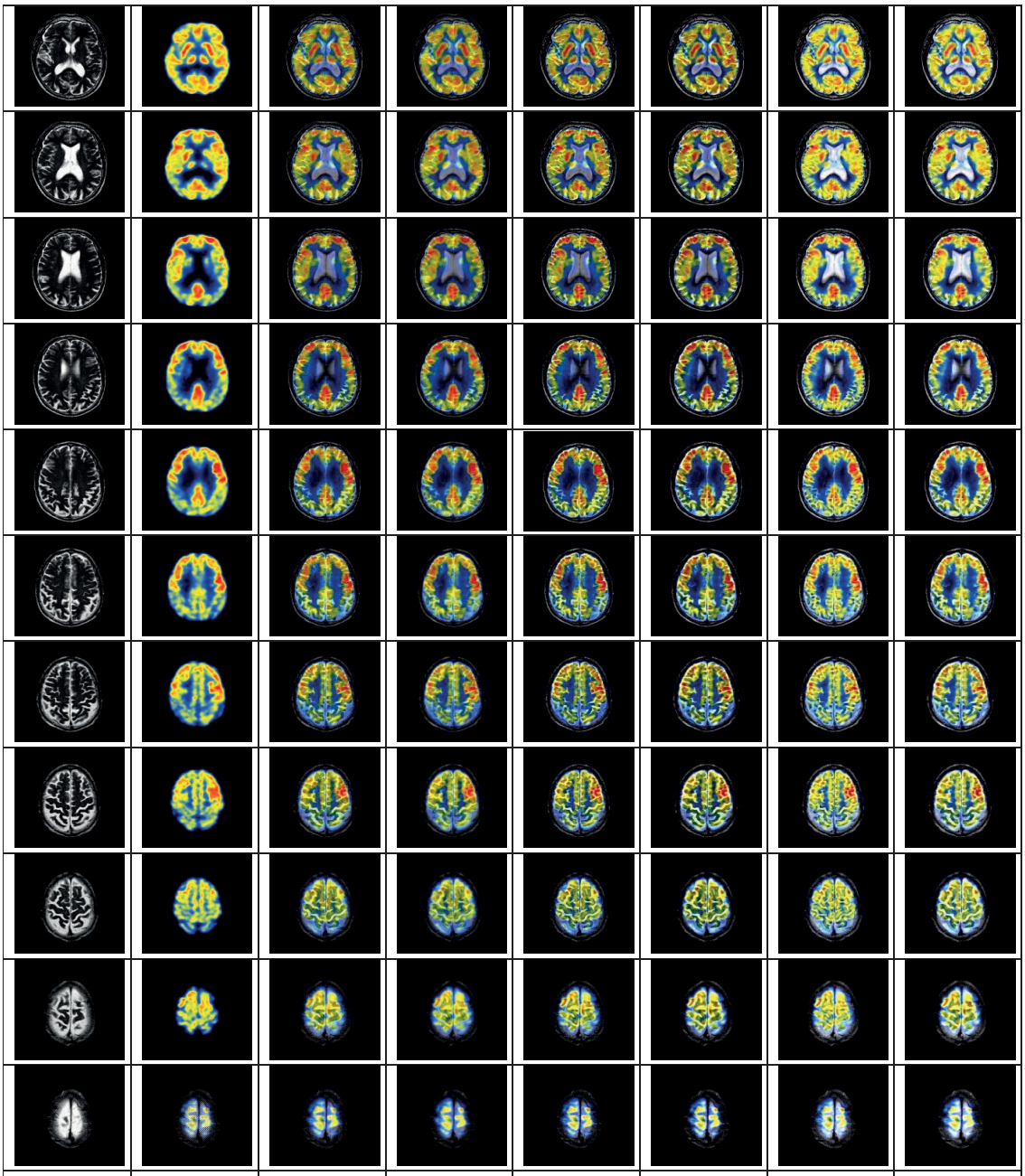
[36] Asha C, Lal S, Gurupur VP, Saxena PP. Multi-modal medical image fusion with adaptive weighted combination of NSST bands using chaotic grey wolf optimization. IEEE Access 2019; 7: 40782-40796.

[37] Lu B, Wang H, Miao C. Medical image fusion with adaptive local geometrical structure and wavelet transform. Procedia Environmental Sciences 2011; 8: 262-9. doi: 10.1016/j.proenv.2011.10.042

[38] Li S, Yang B, Hu J. Performance comparison of different multi-resolution transforms for image fusion. Inform Fusion 2011; 12 (2): 74-84.

Appendix. Result images of fusion process





Appendix: (a) MR source image; (b) PET source image; (c) Result of CVT fusion; (d) Result of NSST fusion; (e) Result of CVT+High+CNN fusion; (f) Result of NSST+High+CNN fusion; (g) Result of CVT+Low& High+CNN fusion; (h) Result of NSST+Low&High+CNN fusion.