Biomedical Signal Processing and Control 72 (2022) 103357

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

An improved approach for medical image fusion using sparse representation and Siamese convolutional neural network

Ahmed Sabeeh Yousif *, Zaid Omar, Usman Ullah Sheikh

School of Electrical Engineering, Computer Vision, Video & Image Processing (CvviP) Lab, Universiti Teknologi Malaysia, Skudai, Johor Bahru 81310, Malaysia

ARTICLE INFO

Keywords:
Siamese convolutional neural network
Sparse representation
Medical imaging
Image fusion applications

ABSTRACT

Multimodal image fusion is a contemporary branch of medical imaging that aims to increase the accuracy of clinical diagnosis of the disease stage development. The fusion of different image modalities can be a viable medical imaging approach. It combines the best features to produce a composite image with higher quality than its predecessors and can significantly improve medical diagnosis. Recently, sparse representation (SR) and Siamese Convolutional Neural Network (SCNN) methods have been introduced independently for image fusion. However, some of the results from these approaches have recorded defects, such as edge blur, less visibility, and blocking artifacts. To remedy these deficiencies, in this paper, a smart blending approach based on a combination of SR and SCNN is introduced for image fusion, which comprises three steps as follows. Firstly, entire source images are fed into the classical orthogonal matching pursuit (OMP), where the SR-fused image is obtained using the max-rule that aims to improve pixel localization. Secondly, a novel scheme of SCNN-based K-SVD dictionary learning is re-employed for each source image. The method has shown good non-linearity behavior, contributing to increasing the fused output's sparsity characteristics and demonstrating better extraction and transfer of image details to the output fused image. Lastly, the fusion rule step employs a linear combination between steps 1 and 2 to obtain the final fused image. The results depict that the proposed method is advantageous, compared to other previous methods, notably by suppressing the artifacts produced by the traditional SR and SCNN model.

1. Introduction

It is known that doctors and physicians have mostly been required to sequentially analyze medical imaging techniques such as computed tomography (CT), Magnetic resonance imaging (MRI), and positron emission tomography (PET) captured using various equipment. However, these ways may still introduce inconvenience in many complex brain diseases cases. A practical method to avoid this problem is to apply the medical image fusion technique [1], which aims to integrate the complementary information from multiple medical images with different modalities for visualization. This may assist the physicians in making easier decisions and improvements for various purposes. For example, the CT images can express the precise localization of dense structures such as implants and bones. On the other hand, MR images can show an excellent soft-tissue of anatomical information but are less sensitive to bones' diagnosis than CT.

Currently, it is necessary to improve the visibility of the medical imaging that can be used to visualize the disease. Abnormal versus normal tissues in different clinical analysis applications is essential for doctors in complex diseases involving diagnosis, planning-based surgery, treatment, and surgical navigation. Recently, multimodality fusion images are considered the main tools for obtaining rich information for clinical analysis purposes. Image fusion can be defined as a process of extracting/detecting an essential feature from source images and fusing/transferring these details into a synthetic image [1,2]. There are different fusion applications, such as multimodal medical images, infrared–visible, and multi-focus fusion schemes.

Many imaging techniques are available, such as CT, MRI, PET fMRI (functional magnetic resonance imaging), and SPECT (Single-photon emission computed tomography), and many others [2–4]. For instance, CT and MRI scanned images are mainly used for more detailed stroke disease and the localization of core brain structures like white and gray matters with high-resolution anatomical details. The diversified image fusion methods have been proposed in past years and are generally classified as spatial and transform domains. The spatial domain was further divided into pixel-based, region-based, and blocks-based [5,6]. These methods conduct fusion images in the intensity space of source images with transformation. Based on the space theory, these methods

* Corresponding author.
E-mail address: Syahmed@graduate.utm.my (A. Sabeeh Yousif).

https://doi.org/10.1016/j.bspc.2021.103357
Received 11 June 2021; Received in revised form 9 September 2021; Accepted 5 November 2021
Available online 17 November 2021
1746-8094/

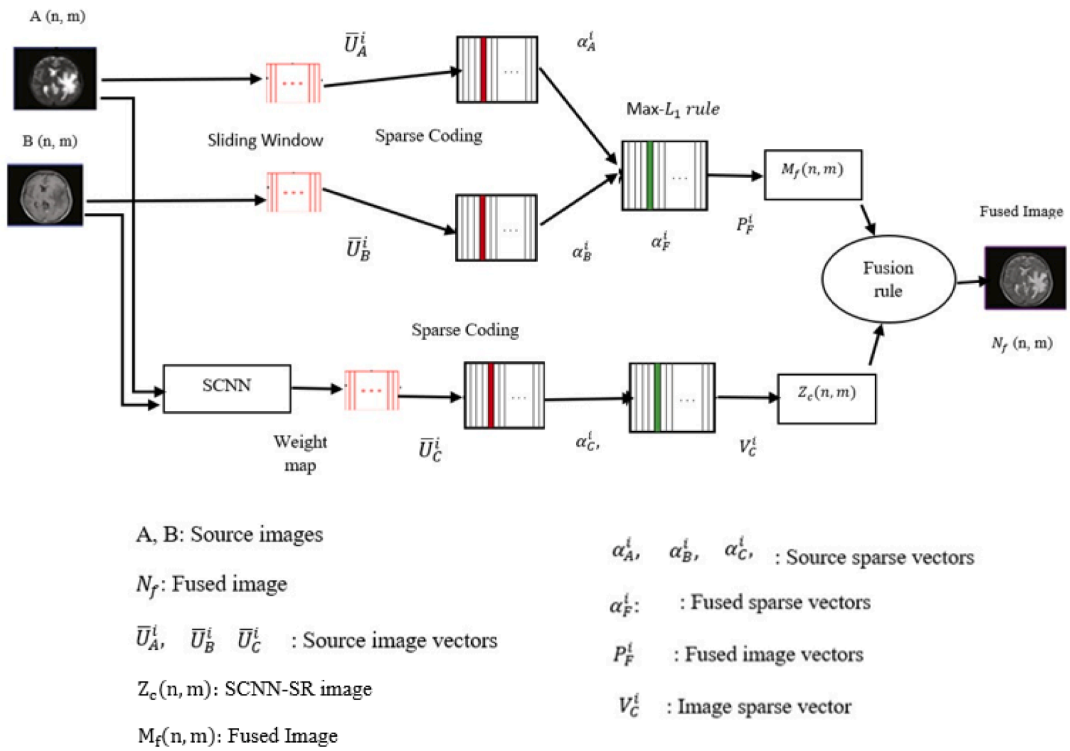

Fig. 1. Overview schematic diagram of the proposed SCNN based sparse coding and OMP method.

were rapidly developed to more complex domains such as multiscale transforms, sparse representation, intensity-hue-saturation color space, principal component space, and learning-based methods.

The most classic transform domain methods in the fusion area are defined as the fused signal produced via three steps: 1) transform decomposition, 2) coefficients, 3) inverse transformation. These methods can avoid the spatial artifacts resulting from spatial methods effectively. The multiscale transform (MST) fusion technique emerged recently and is employed for computer vision and image processing areas. Moreover, there are various transforms based fusion images, for instance a discrete wavelet (DWT)-based [7], complex wavelet transform (CWT)-based [8], dual-tree complex wavelet transform (DTCWT)-based [9], the non-subsampled contourlet transform (NSCT)-based method [10], curvelet transform (CVT)-based [11], contourlet transform based [12], dense SIFT-based [13], shear-let transform-based [14], singular value decomposition (SVD) [15], and sparse representation based techniques [16]. However, the shortcomings of these methods are always observed in the design of the transform basis, namely activity level measurement and fusion rule. These have shown low performance in the final fused images, which led to lower visibility of the important cells and wrong clinical decisions, for example, blocking artifact and edge blur in the final fused images [7,8].

Owing to the development in image representation methods, sparse representation (SR) techniques have drawn attention. For example, Olshausen and David exhibited that images can be represented by sparse coding [17]. SR model an image patch as a linear combination of few atoms with various constructed in an over-complete dictionary. There are two categories in designing an over-complete dictionary, namely, mathematical and model-based learning. The dictionary's limitation designed by the mathematical model can lose some of the natural images' characteristics. Therefore, model-based learning can effectively remedy this deficiency, such as K-SVD [18] and MOD [19].

Recently, methods of (SR) [20–22] proposed a novel method to overcome the two keys issue: over-complete construction and sparse coefficient solution. Moreover, Yang and Liu [18] employed an orthogonal matching pursuit (OMP) to obtain a multi-focus application coarse coefficient. However, the characteristic of an over-complete dictionary could affect the quality of the fused image. In [23], the SR has been developed using simultaneous OMP (SOMP) to ensure that various images are sparsely decomposed into the same sub-set of dictionary bases. Furthermore, Yin et al. [24] introduced an image fusion method based on joint sparsity, where the effect of smooth regions caused inaccurate segmentation and lower visual. In addition, Lie et al. [25] proposed different multiscale transform as a relative application on NSCT-based fusion, which achieved superior over the state of arts. Apart from that, Liu [26] presented the novel framework, which investigated the combination of sparse representation and multiscale transform to avoid drawbacks, such as low contrast, blocking artifact, and the defects mentioned earlier in recent works.

Convolutional Neural networks (CNNs) have led to the state of arts resulting in image processing topics, as it has greater capability to feature representation. Successful use of CNNs over traditional image fusion research can advance visual recognition issues since CNN can lean significant features from massive data. Furthermore, CNNs could characterize complicated relationships between input and fused images due to the strong representation of image details. Liu et al. [27] also applied the CNN network for multi-focus image fusion. However, it has been viewed as a classification problem. Besides, Yu et al. [28] proposed a method-based Siamese convolutional network by applying a multiscale manner with image pyramids to achieve more human visual perception consistency. The method recorded limit in preserved images details like edges and overlap occurring between spatial pixels.

However, SR and SCNN-based fusion have achieved good performance but have some drawbacks in the image fusion applications. Firstly, the spatial difference in the multimodal medical fused images is inherited by the max- L_1 fusion rule in the SR methods when different imaging modalities take the source images. Secondly, the standard SR-based image fusion is directly implemented into several steps that cannot incorporate detailed information. For SCNN based fusion, the main drawbacks are that undesirable artifacts for acceptable consistency in the human visual perception have been recorded [27,28]. Moreover, the irregularity of the edge is presented in the fused images. These issues

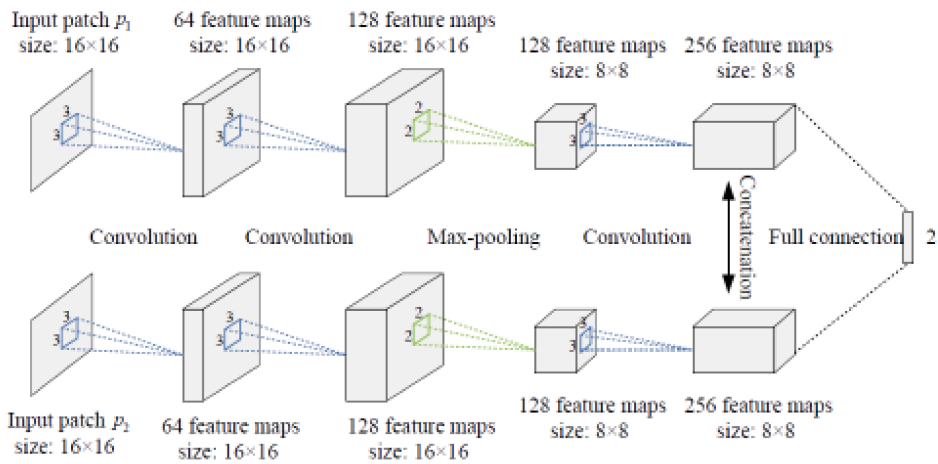

Fig. 2. Highlights the Siamese network type of the same network in [36].

are introduced because the intensities of multimodal medical images at the exact location vary significantly [28]. This paper addresses these problems from another viewpoint to determine the optimal solution that overcomes the difficulty in designing activity level measurements and fusion rule strategies. Comprehensively, the main contributions of the proposed scheme focus on the following four aspects:

1) A novel framework of multimodality medical image fusion-based SR and SCNN is presented. The proposed fusion model has the capability to integrate geometric features from the source image sensors.
2) We introduce the OMP method to improve the image natural sparsity, which is in accordance with the physiological characteristics of the human visual system. OMP separates source images, where the final image is obtained by max L_1 -rule.
3) For the SCNN based K-SVD, we present a novel algorithm for adapting dictionaries to represent medical images sparsely. It uses a dictionary learning method based on the K-SVD method and SCNN for image representation. First, the medical images are transformed using SCNN. Then, a new weighted map is estimated by K-SVD for increased strength of the sparsity representation (i.e., sparse vector), which involves rich information from the source images. A novel SCNN-based dictionary learning method has shown more non-linearity, which leads to a good sparsity on the entire image. Final images have to fit the requirements of sparsity, which are localization, geometric invariance, and over-completeness.
4) Finally, applying the fusion rule, the fused image is formed. The proposed method produces better fusion results for multimodal image datasets than existing techniques. Fig. 1 highlights the overall block diagram of the proposed fusion method.

The rest of the paper is structured as follows. Section 2 briefly explains the research mythology and motivation for the SR and SCNN-based medical image fusion. Meanwhile, Section 3 illustrates the proposed method in detail. The experimental setting and results are explored in Section 4. Finally, the conclusion is given in Section 5.

2. Methodology

2.1. Introduction to medical image fusion

Multimodal medical image fusion aims to combine complementary information from different modalities in a single image [7–12], helping the physicians make better decisions and various clinical purposes. Furthermore, it is also used to reduce storage costs by reducing storage to a single fused image instead of multiple-source images.

2.2. Sparse representation based-image fusion

SR has drawn the attention of both theoretical signal/image processing and practical application [32]. Its source was used for the theory of compressed sensing [30]. It should include dictionary atoms as soon as possible to provide a more powerful signal composition to assure the vector of coefficients is sparse. The theory expressed a signal x as a linear combination of a few atoms from a dictionary [31].

The main objective of the proposed sparse representation-based image fusion is to extract more comprehensive information from the signal. To obtain the optimal representation, this question could be converted to the following optimization problem:

$$\min s.t, \|B\|_0, \|X - DB\|_2^2 \le \mathscr{E}. \quad (1)$$

In the above equation, B refers to the sparse coefficients of signal X, $\|B\|_0$ represents the L_0 -norm of B, the size of dictionary D is $z < k$ so that it can refer to that dictionary as overcomplete (size of column vectors of the dictionary), while z is the patch dimension, and ε refers to error tolerance with $\varepsilon > 0$.

There are many approaches to design proper dictionary-based examples, such as MOD and K-SVD [18,19]. The sparse coding is used to solve the NP-hard problem obtained from the sparse coefficients.

Given the dictionary D, the purpose of sparse coding is to obtain the sparse coefficient B for signal X. There are different algorithms of sparse-coding like matching pursuit (MP) [29], orthogonal matching pursuit (OMP), least angle regression (LAR) [33], and basis pursuit (BP) [34]. This paper employs the OMP algorithm to calculate the sparse coefficients of each source image.

In our paper, the setting of K-SVD based dictionary learning model is used for both OMP and SCNN as follows:

1) Load the original images CT and MRI with size $n \times m$, where the source image size is 256×256.
2) Divide each image to obtain the overlapping patches of size $z \times z$, where the total number of patches is $(n + z - 1) \times (m + z - 1)$. Patches are sequentially extracted from the image, vectorized in lexicographic order, and used as columns, one after the other, to define the new matrix.
3) K-SVD is mobilized to train an overcomplete dictionary. The training data is constructed as a set of 71.289 patches of size 12×12, sampled randomly from various locations of the 40 CT and MRI 2D- images obtained from the Medical Harvard school database. As we know, the representation ability of the over-completed dictionary relies much on the number of atoms in it, but a dictionary with a large size will directly increase the computational cost. Also, the image representation ability is not sufficient for improve the textures and edges of

fine details in the reconstruction result. Thus, a compromise on dictionary size is required. Therefore, in our paper, the dictionary size is set to 256 when the input 144 dimensional (12 ×12 patch).

4) Finally, running the K-SVD to reshape the resulting atoms of the final dictionary learning. The main contribution of K-SVD with achieved column by column update based the OMP and SCNN output images by updating the atom and its associated sparse coefficients simultaneously, it can be shown the better fit data, and efficient algorithm for the final fused image.

2.3. The SCNN Model-based image fusion

CNN [35] is a trainable multilayer scheme aiming to learn a multistage feature representation of the input signal, where each stage is composed of several feature maps. The coefficient in a feature map is referred to as a neuron. Feature maps at different stages are connected by achieving several kinds of estimation like convolution, non-linear activation, and spatial pooling. A fully connected layer requires predefined dimensions of both input and output.

Fig. 2 shows the proposed Siamese convolutional network used in the proposed fusion algorithm using the same network in [36]. The Siamese network is much preferred to be engaged in solving the image fusion issues. This is because it uses the same weights to see which one is more powerful in demonstrating the feature extraction approach. In addition, the activity level measure is the same for two source images, and (2) Siamese network is often easier to be trained than the other types of networks.

The SCNN has two similar branches in terms of architecture, which are one max-pooling and three convolutional layers. To maximize computational output and improve the memory occupancy, 512 map features have been used to provide the highly flexible SCNN network, obtained from the concatenation process and connected to provide the 2D vectors. These new 2D feature maps enter a new layer of the two-way SoftMax (this layer is not highlighted in Fig. 2) process, aiming to produce probability values of over two final classes. Each probability distribution value of every class was assigned the likelihood of each weight assignment Formula. Here, the first patch equals 1, while the second patch equals 0, as well as the first patch equals 0, while the second patch equals 1.

The 104 pairs of each CT and MRI scan 2-D images are collected from the brain dataset of Medical Harvard School. Our proposed method artificially created training and testing set to be applied for the medical image fusion issues. Therefore, they tend to have a stronger ability to improve the fusion outcome. The proposed network design of SCNN employs the original CT and MRI scan image patches and their blurred versions based on multiscale Gaussian filtering and random sampling. Here, the suggested Gaussian filter has a standard deviation of two, and the cutoff is equal to 7×7. Then, 40 pairs of spatial patches of size 16×16 are randomly selected. The reason for using size 16 is to enhance the fusion quality of boundary regions compared to 32, as mentioned in [37].

In this paper, a total of 8320 pairs of the patches from the CT and MRI scan stroke datasets were obtained and fed to the proposed network design.

The training process of 6320 pairs of patches is fed to the proposed popular deep-learning namely; Caffe method [36]. Meanwhile, in the test fusion process, the same procedure was used for dividing into pre-specific patches. Thus, the number of test images equals 2000 by using the unknown CT and MRI scan for the images for the fusion process.

The optimization method for the Soft-Max-based loss function is used for the stochastic gradient descent (SGD) scheme. However, due to the repeated large number of the calculation process handled by only an arbitrary size for the original input image, the only solution is to convert the full connection to an equal convolutional layer and utilize the two kernels of dimension size equivalent to $8 \times 8 \times 512$ [38].

Thus, the output dense prediction map, meaning each prediction of a 2D vector, is obtained. The final prediction map contains more image details and clear information on the source patch in the spatial location, which corresponds to an interest in the specialized multimodality of the stroke image fusion. Here, the two final dimensions are extracted from each prediction, and their result is summed and normalized to 1. The output can be determined easily as the weight value of the first and second input images. Lastly, it has been assigned the value of the weights of all the pixels within the patch location and the average of the overlapped pixels. This is to reconstruct the weight map to correspond similarly to the size of the input images.

To have more insights, one may refer to Fig. 2. The examples of medical images captured by CT and MRI imaging techniques are used as inputs for each convolutional layer of both channels, and the same goes for Siamese CNN. The new corresponding feature maps are obtained, with each map normalized to [1,0]. The first convolutional layer produced feature maps containing high-frequency information, while some produced the same input images. The second convolutional layer provided feature maps with more extracted spatial details covering the different gradient orientations. Here, the second convolutional layer has shown promising results of the features maps, which can be characterized as the spatial details than the first convolutional layer. On the other hand, the third convolutional layer integrates the gradient information as its output feature maps characterizes the complete information of each CT and MRI image successfully. Finally, with the following two fully connected layers, an outcome of the score map could obtain an abundance of spatial details.

In the proposed algorithm, the SCNN model produced the focus map from the learned SCNN network. Thus, it contains complementary information about spatial details, which is essential for medical image application. Although the experimental setting for the training and testing process is different from the state of the art for applying the divided patches for both stages, the final fused images illustrate an efficient medical image fusion when a small number of testing samples are used.

3. Proposed fusion scheme

3.1. Overview

The nature of medical images from various sensors and mechanisms may have different intensities at the same location in the source image, which often changes significantly. It also introduced the blocking artifacts, low contrast, and irregularity of the edges in spatial details. In addition, the lack of image representation, decomposition, and the activity level measurement have caused several limitations. For instance, the max- L_1 rule in the SR method limits the size of the patch, constructing an overcomplete dictionary.

Our proposed method, which fulfills the objective of the paper, used the best characteristics of the dictionary in SR and the efficiency of learning features via the SCNN model to extract more of the image representation, fewer blur edges, and also avoid the overlapped pixels to maximize the human visual quality. Moreover, this observed more features representation in the output fusion images. Fig. 1 illustrates the schematic diagram of the detailed proposed method.

3.2. Fusion details

Let S_i be the original image having dimensions $n \times m$, $n = 1,2,3,4,\cdots, n$, $m = 1,2,3,4,\cdots,n$. Here, $i \in [A, B]$ refers to the multimodality CT/MRI images. The following steps exploited the details of the proposed image fusion method:

A. Orthogonal matching pursuit (OMP)

The following steps illustrate the proposed OMP procedures:

1- Divide the source images S_A, S_B using the sliding window to obtain the patches of size $z \times z$, which is the same size as an atom in a dictionary. Thus, it has W patches described as $\{S_A^i\}_{i=1}^W$ and $\{S_B^i\}_{i=1}^W$ in S_A and S_B, respectively.
2- At each position in i, S_A^i and S_B^i patches are re-arranged into column vectors to obtain $\left\{P_A^i, P_B^i\right\}$.
3- The normalization for each vector's mean value to zero is applied, yielding $\left\{\overline{U}_A^i, \overline{U}_B^i\right\}$ such that

$$\overline{U}_A^i = P_A^i - \mu_A^i \tag{2}$$

$$\overline{U}_B^i = P_B^i - \mu_B^i.1, \tag{3}$$

where μ_A^i and μ_B^i are the mean values of all elements, and **1**:refers to all one valued $n \times 1$ vectors.

4- The sparse coefficients vector $\left\{\alpha_A^i, \alpha_B^i\right\}$ of $\left\{\overline{U}_A^i, \overline{U}_B^i\right\}$ is calculated using the orthogonal matching pursuit (OMP) in [18,19], such that

$$\alpha_{A=}^i argmin\|\propto\|_0 \quad s.t \ \|U_A^i\ , \ - DH\|_2 \quad < \mathscr{E} \tag{4}$$

$$\alpha_{B=}^i argmin\|\propto\|_0 s.t\|\overline{U}_B^i - DH\|_2 < \mathscr{E}, \tag{5}$$

where D is the learned dictionary and α represents the sparse coefficients of image U.

5- The sparse vector is fused by combining α_A^i and α_B^i using max- L_1 rule as follows:

$$\alpha_F^i = \left\{ \begin{array}{c} \alpha_A^i if \|\alpha_A^i\| > \|\alpha_B^i\| \\ \alpha_B^i otherwise \end{array} \right\} \tag{6}$$

The fused image of P_A^i and P_B^i is computed via

$$P_F^i = D\propto_F^i + U_F^i \tag{7}$$

where the mean value of U_F^i is computed using

$$\overline{U}_F^i = \left\{ \begin{array}{c} U_A^i if \propto_F^i = \propto_A^i \\ U_B^i otherwise \end{array} \right\} \tag{8}$$

6- Repeat all above processes for source patch images S_A and S_B to obtain all fused vectors $\left\{\overline{U}_F^i\right\}_{i=1}^W$.
7- Let M_F be the final fused image by OMP, each P_F^i is reshaped into S_F^i patches and plug S_F^i, which are then inserted into their original position in M_F.

B. SCNN model-based learning dictionary

The following procedures express the SCNN based dictionary as follows:

1- First, feed the source images to the SCNN network, explained in detail in Section 1.2.3. Then, the focus map of the CNN-designed network contains more image details with the informative region.
2- Find the sparse coding of particle image S_C (i.e., focus map) that contains rich information. Here, A_f is the output image such that

$$argmin\|A\|_0 with = \|\overline{X} - DA\|_2 < \mathscr{E}. \tag{9}$$

Here, D refers to the same K-SVD dictionary earlier, $\overline{X}$ is the sparse image recovery, the size of the dictionary D is required to be $n > k$, implying the dictionary is over-complete. Moreover, A is the focus map representing the sparse coefficients of the output SCNN image $\overline{X}$, while $\|A\|_0$ refers to L_0 −norm of A.

3- Divide the source images S_C using the sliding window to obtain the image patches of size $z \times z$ with the step length of r pixels. Thus, it has W patches described as $\{S_C^i\}_{i=1}^W$.
4- At each position i, the $\{S_C^i\}$ is re-arranged into a column vector P_C^i, where the vector's mean value is normalized to zero to obtain the $\overline{U}_C^i$ given by

$$\overline{U}_C^i = P_C^i - \mu_C^i \tag{10}$$

where μ_C^i is the mean value of all elements in P_C^i and **1** refers to all one valued $n \times 1$ vectors.

5- The sparse coefficients α_C^i is determined, and the final sparse vector is calculated using

$$V_C^i = DC + \mu_C^i \tag{11}$$

Again, all the above processes are iterated for all patches in $\{S_C^i\}_{i=1}^W$ to obtain the final image vector $\{V_C^i\}_{i=1}^W$. Here, each V_C^i is reshaped into S_C^i and placed into the final image of $Z_c(n,m)$.

C. Fusion rule

For the reconstruction process, the fused image $N_f(n,m)$ was obtained via a linear combination of each pixel's value in $M_f(x,y)$ and sparse coding of CNN model $Z_c(x,y)$. It is then averaged over its accumulation times of all patches in the source images $S_A(n,m)$, $S_B(n,m)$, and $S_C(n,m)$ in the $S_f(n,m)$. The fused image showed a natural look, less sharp, and more acceptable for the human visual system. The fusion rule formula is formed such that

$$N_f(n,m) = \frac{M_f(n,m) + Z_c(n,m)}{S_f(n,m)} \tag{12}$$

3.3. Objective evaluation metrics

The experiments were conducted through some quantitative measurements to verify the proposed method [39]. Five metrics were applied for objective evaluation assessment, which comprises Spatial Structural Similarity (Q_F^{AB}) [42], Feature Mutual Information (FMI) [45], Visual Information Fidelity (VIF) [36], Edge-Strength Similarity-Based Image Quality Metric (ESSIM) [42] and a metric Peilla (Q_E) [41]. The larger values of metrics usually highlight a better result.

Q_F^{AB} measures the amount of the edge information that transfers from the input images to the fused images through the Sobel edge detection operator. Assuming the input images A, B of $n \times m$ and $N_f(n,m)$ represents the fused image. The larger value refers to the better edge preserved and information retrieved, resulting in more detailed information converted from the source image. Generally, a stronger edge has impacts on the Q_F^{AB} than the lower strength edge. Thus, the Q_F^{AB} can be expressed as follows:

$$Q_F^{AB} = \frac{\sum_{n=1}^N \sum_{m=1}^M (Q^A(n,m)W^A(n,m) + Q^B(n,m)W^B(N,M))}{\sum_{n=1}^N \sum_{m=1}^M (W^A(i,j) + W^B(i,j))},$$

where $Q^A(n,m), Q^B(n,m)$ are edge information value storages, while $W^A(n,m), W^B(n,m)$ refer to the weighting map.

Feature mutual information (FMI) is used to estimate the degree of dependency between the input images and the fused images. A larger value of FMI refers to a better quality of the fused image. Here, we have

$$FMI_F^{AB} = FMI_{FA}(f:a) + FMI_{FB}(f:b)$$

where the $FMI_{FA}(f:a)$ and $FMI_{FB}(f:b)$ are the amount of information about A and B.

The principle of visual information fidelity (VIF) is considered a full-reference image quality metric based on the Natural Scene Statistics or (NSS) theory. The main steps to illustrate the VIF in terms of the measure of good quality fusion can be summarized as follow: 1) VIF decomposes the source image into various sub-bands and divide each sub-band into blocks, 2) The VIF measure the visual information by estimating the mutual information in various models in each block and each sub-band, 3) The outcome image quality is computed by integrating the visual information for all the blocks and sub-bands, respectively. VIF presented three models to measure the visual information: 1) the Gaussian scale mixture model GSM model, the human visual system HVS model, and the distortion model.

The *GSM* is an NSS model in the wavelet domain. AGSM is a random field (*RFs*) and has two independents *RFs*: Gaussian and scale, which can be defined as

$$C_i = s_i U_i \tag{12}$$

where C_i denotes the $i_{th}RFs$ of the reference signal, s_i is the i_{th} random positive scalar and the U_i refers to Gaussian vector RFs and their variances.

A distortion model is used to demonstrate the extent to which distortion operators can disturb an image, given by

$$D_i = g_i C_i + V_i \tag{13}$$

where C_i denotes the i_{th} RFs of the reference signal, D_i is the *RFs* of the sub-band of the tested image, g_i is the scalar value and V_i denotes a stationary additive zero-mean Gaussian noise field with variance $c_{vi} = \sigma^2_{vi}$.

For the HVS, it can show the impact of the VIF, which quantifies the signal through the HVS model. Thus, HVS is modelled to be an additive component in the distortion channel.

Through the edge-strength exploited, visual fidelity between the fused image f and the source image g can be estimated by the similarity between their edge strength- maps. Thus, the ESSIM can be defined as:

$$ESSIM(g,f) = \frac{1}{N}\sum_{i=1}^{N}\frac{2E(f,i)E(g,i)+C}{(E(f,i))^2+(E(g,i))^2+C}. \tag{14}$$

The different magnitude in C parameter leads to the difference in the ESSIM score due to two senses: 1) it is presented to avoid zero denominator, 2) it can be shown as a scaling parameter.

$$C = (BL)^2 \tag{15}$$

where B is the predefined constant, while L is the dynamic range of the edge-strength. The edge-strength $E(f,i)$ satisfies $0 < E(f,i) <= 255^p$, namely $L = 255^p$. Hence, it can be estimated as:

$$C = (BL)^2 = \left(B^{\frac{1}{p}}255\right)^{2p} = (B_1 255)^{2p}, \text{with} B_1 = B^{\frac{1}{p}}.$$

The ESSIM index I is close to the gradient methods. Hence, the gradient itself can be considered as an edge-strength measure.

The novel non-reference quality assessment, Piella's metric-based image fusion, was proposed in [40]. The fused images used local measures to estimate how the salient details information is better than the source image introduced in the output fused images.

Previously, both authors have introduced a new scheme of fusion quality index, which adapted the scheme proposed in [40]. This study was based on the universal image quality index (UIQI). UIQI-based metric has investigated the evidence that the human visual system is highly adapted to structural information. Therefore, the proper measurement of structural information loss can provide a good approximation of the perceived image distortion. UIQI can be expressed as

$$Q = \frac{4\sigma_{xy}\overline{xy}}{\left(\sigma_x^2+\sigma_y^2\right)\left(\overline{x}^2+\overline{y}^2\right)} = \frac{\sigma_{xy}}{\sigma_x\sigma_y}\cdot\frac{2\overline{xy}}{\overline{x}^2+\overline{y}^2}\cdot\frac{2\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}$$

where $\overline{x,y}$ are the average values, while σ_x, σ_y and σ_{xy} refer to the covariance values, respectively. There are three items used to estimate the linear correlation between the source images (i.e., x and y), which means to the how close the luminance and contrasts of the images are. The aforementioned equation has been improved to incorporate the structural similarity index measure-based metric SSIM, which can be a better result when used in general conditions. The quality measurement is determined by applying to the interested local regions using a sliding window process. The size 8×8 is used for the top left to the bottom right of the source image.

Hence, the measurement of the average quality index is obtained and written by the following Eq:

$$Q_o = \frac{1}{|W|}\sum_{w\in W} Q_o(a,b|w) \tag{18}$$

Where w refers to the estimation by sliding window process, W is the family of all windows and $|W|$ represents the number of elements in the group (i.e., the cardinality of W).

The Piella matric is derived from the three main fusion indexes with UIQI theory as follows:

$$Q(a,b,f) = \frac{1}{|W|}\sum_{w\in W}[\lambda(w)Q_o(a,f|w)+(1-\lambda(w)Q_0(b,f|w)], \tag{19}$$

$$Q_w(a,b,f) = \sum_{w\in W} c(w)[\lambda(w)Q_o(a,f|w)+(1-\lambda(w)Q_0(b,f|w)], \tag{20}$$

$$Q_E(a,b,f) = Q_w(a,b,f)Q_w(a^{'},b^{'},f^{'})^{\alpha}, \tag{21}$$

where f refers to the output fused image of source images a and b, while $a^{'}, b^{'}, f^{'}$ are the result of edge map image a, b, f, respectively. The variance value can be estimated (i.e., $\lambda(w)$) as:

$$\lambda(w) = \frac{s(a|w)}{s(a|w)+s(b|w)} \tag{22}$$

where $s(a|w)$ and $s(b|w)$ represent the local-based salience information of the image a and b within the window $w = 8\times8$ as the window size. Here, $C(w)$ is the max of the $s(a|w)$, $s(b|w)$, and $c(w) = C(w)/(\sum_{w^{'}\in W}C(W^{'}))$.

4. Experimental results and analysis

This section demonstrates that SCNN and sparse representation (SCNN-SR) have better capability in medical image fusion applications. Several experiments were conducted to verify the feasibility of the suggested multimodality image fusion-based SCNN-SR. The experiments were used on different 104-pairs of medical images from different slices of the CT/MRI images. The availability of these datasets has existed online in [42]. The size of the images is 2-D with 256×256 and a depth equal to 8-bit. Fig. 2 displays test sample examples of different medical image pairs (CT/MRI) from different types of datasets. The performance evaluation tests for the proposed method under lab specifications are given as follows. Assume the sources images A, B of size $n\times m$, the size of sliding window $z\times z$ is set to a size of 8×8, the total number of obtained patches is calculated by $(n+z+1)(m+z-1)/s$, where s refers to the step length and set to 1 to avoid the blocking effects of the spatial domain in the final fused image N_f. The training data set are randomly selected as 40 CT and MRI images from Harvard medical school dataset. The dictionary size is set to 256 when the input image is 64 dimensional (8×8) patch, the number of iterations of K-SVD is fixed to 180, and error tolerance is fixed at $\varepsilon = 0.1$.

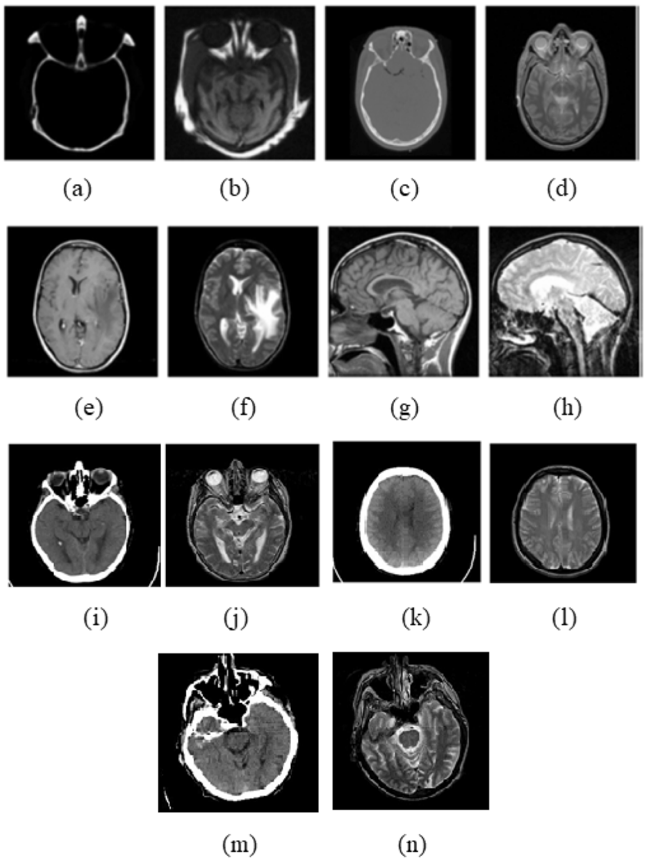

Fig. 3. Testing medical image samples of the CT and MRI: (a, b) pair-1, (c, d) pair-2, (e, f) pair-3, (g, h) pair −4, (i, j) pair 5, (k, l) pair −6, and (m, n) pair-7.

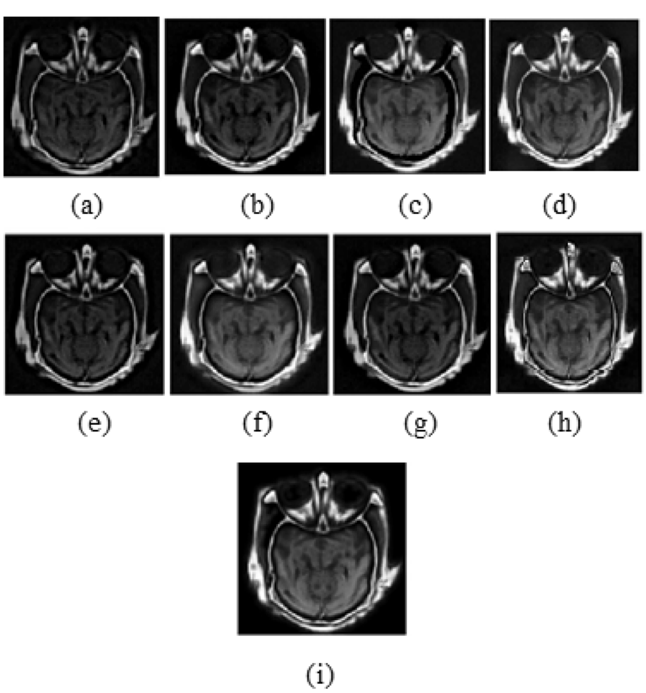

Fig. 4. Fusion results for Fig. 3(a) and (b). (a) DCTWT; (b) LP; (c) MST-SR1; (d) CNN + LP; (e) NSCT; (f) NSST-PAPCNN; (f) CSMCA; (g) MST-SR; (i) proposed method.

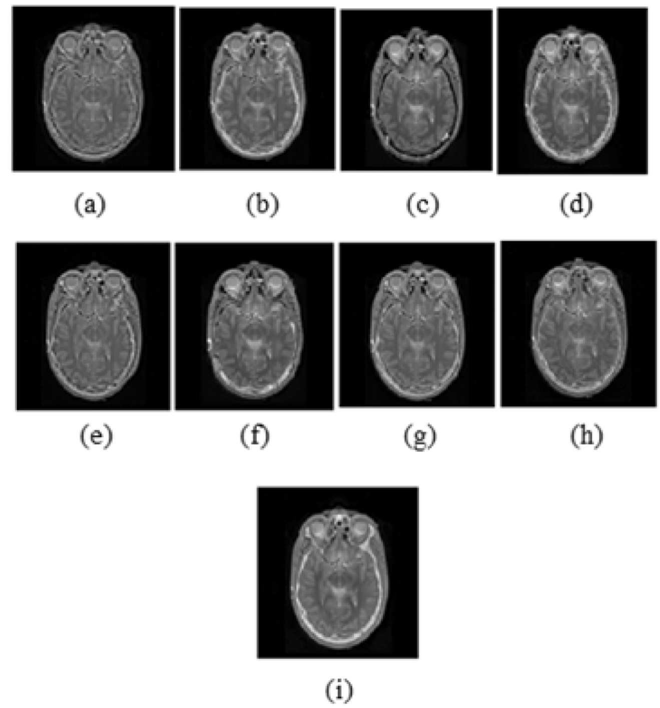

Fig. 5. Fusion results for Fig. 3(a) and (b). (a) DCTWT; (b) LP; (c) MST-SR1; (d) CNN + LP; (e) NSCT; (f) NSST-PAPCNN; (f) CSMCA; (g) MST-SR; (i) proposed method.

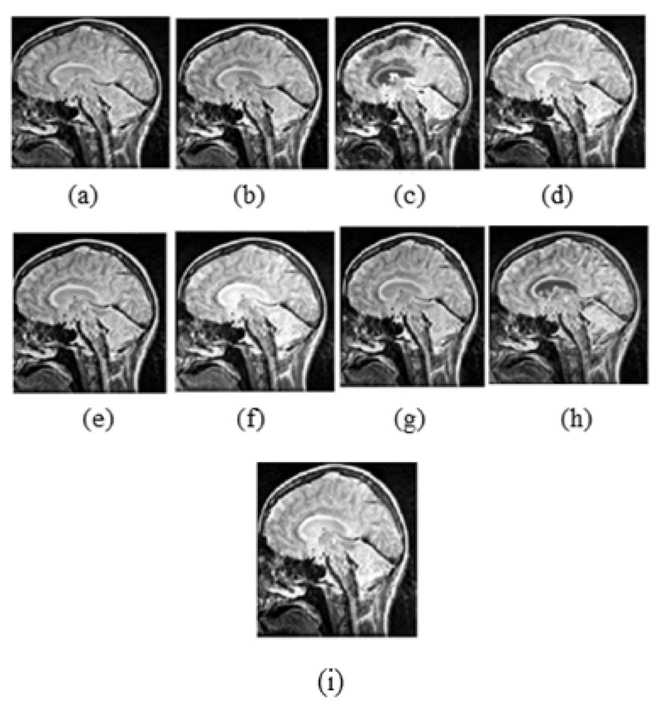

Fig. 6. Fusion results for Fig. 3(g) and (h). (a) DCTWT; (b) LP; (c) MST-SR1; (d) CNN + LP; (e) NSCT; (f) NSST-PAPCNN; (f) CSMCA; (g) MST-SR; (i) proposed method.

To study the efficiency of the proposed method, an extensive comparison is achieved with the existing image fusion methods. This includes dual-tree complex wavelet transform (DTCWT) [43], Laplacian pyramid (LP) [44], non-subsampled contourlet transform (NSCT) [28], LP-SR [45], NSST-PAPCNN [10], convolutional sparsity-based morphological component analysis (CSMCA) [46], Convolutional

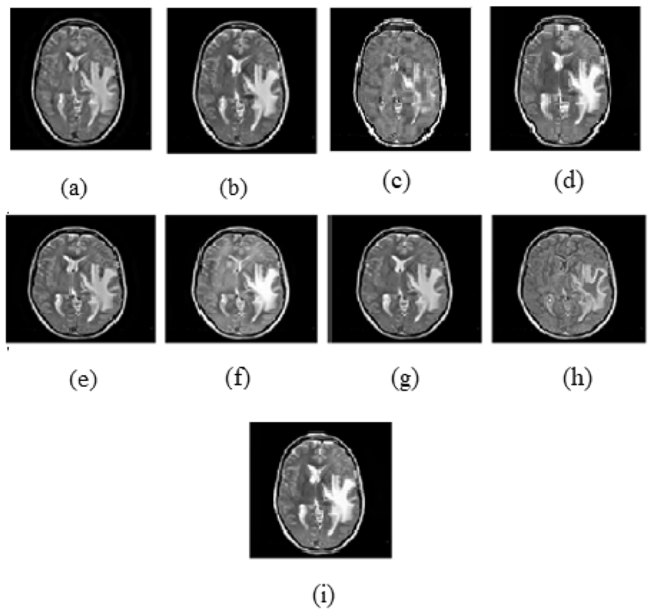

Fig. 7. Fusion results for Fig. 3(e) and (f). (a) DCTWT; (b) LP; (c) LP-SR; (d) CNN + LP; (e) NSCT; (f) NSST-PAPCNN; (f) CSMCA; (g) MST-SR; (i) proposed method.

Neural Networks (CNNs) based image pyramids [36], as well as image fusion based on multi-scale transform and sparse representation [42]. Finally, the results for the proposed method over the state of arts are tested by visual observation and objective evaluation assessments.

4.1. Visual observation assessment

To verify the superiority of our proposed fusion algorithm, this paper highlights seven pairs of source medical images displayed in Fig. 3, which are mainly used in our experiments, including one CT scan and the MRI scan images. While the other two pairs of multi-focus, as well as Infrared and visible images shown in, are used to verify the effectiveness of the proposed fusion framework. For each pair, the two source images are assumed to be pre-registered in our experiments.

Some examples are demonstrated to show the outcome fused results in Figs. 4–7 and Figs. 9–11 for medical image fused application. The lack of fewer feature details representations has been recorded for the DTCWT (Figs. 4–7(a)) and LP (Figs. 4–7(b). It leads to minimizing the visual quality and the information transferred to the fused result. Due to the Gibbs phenomenon, the details and essential information like edge, texture, and lines would be lost in the fused image obtained in NSCT Fig. 12. Thus, the details in Figs. 4–7(e) are not illustrated clearly. For instance, brain tissues' gray and white matter are mostly overlapped in Fig. 7(e). CNN-LP (Figs. 4–7(d)) has shown an excellent image representation over DTCWT, LP, and LP-SR (Figs. 4–7(c)) but fails to preserve the visual information and visual quality is not up to the required mark. Moreover, NSST-PAPCNN (Figs. 4–7(f)) and CSMAC (Figs. 4–7(g)) are better in terms of visual quality and contrast than the remaining methods, except that these methods are not achieved through a local regularity/smoothness for the edge's details at the fused image. Due to the L_1 −norm that is used as a fusion strategy, it can cause spatial inconsistency for the fused image in MST-SR (Figs. 4–7(h)). This side effect leads to loss of characteristics of the edge [40], for instance, more local regularity/smoothness/continuity in a certain direction when different sensors were employed. The proposed method (Figs. 4–7(i)) provided fusion results with more image information, natural looks, higher visibility, and pixel energy preserve than other fusion methods. To have more insights into the stroke disease analysis, for instance, in Fig. 8, the arrow highlighted the infarct on the splenium of the corpus callosum in the fused image. Here, CT and MRI images integrated the important information of splenium tissues in the final fused image. Therefore, the method showed a good visual quality to the radiologist to assess various stroke diseases. Fig. 8 demonstrates the final fused image of stroke disease with elaborated splenium corpus callosum in the brain. On the other hand, Fig. 9 shows the subjective evaluation of the proposed method y_9 over the state of arts.

4.2. Objective performance via fusion metrics

All fusion methods are implemented in MATLAB on a computer equipped with an Intel (R) Core (TM) i7-6500U.

The step lengths of window size have been studied previously, where the best performances have been recorded for three methods: NSCT-SR, DTCWT-SR, and LP-SR for various applications set to 1, 2, 4, and 8 pixels [41]. This paper is proved experimentally by setting the step length to 1, 2, and 8 pixels, respectively. When the step length is set with no more than 1, the running time is even shorter, while the quality of the fused results is still promising for the proposed method. For instance, in our experiment based medical imaging, we compared the step length that equals to 1 pixel over 2 or 8 pixels. When using step length that equals 2 pixels, the fusion metrics of the proposed method such as FMI, VIF, Q_F^{AB}, Piella (Q_E) and ESSIM are recorded, yielding 0.9605, 0.584, 0.8541, 0.6567, and 0.7462, respectively. For the step length that equals 8 pixels, the fusion metrics are 0.8564, 0.4721, 0.6317, 0.4142, and 0.7149, respectively. In summary, the proposed method shows higher performance with a step length equal to 1 than the state of arts.

The investigation of five metrics can evaluate the information transferred from source images of pairs 1–4 in Figs. 4–7. Therefore, the performance of the proposed method is assessed by the five measurement indexes mentioned.

Due to this paper's length limitation, only the first 1, 2, 3, 4, 5, 6, and 7 pairs of Figs. 4–7 and Figs. 9–11 of medical images are recorded and analyzed in Table 1. Pair-1 of the normal brain with CT/MRI modalities

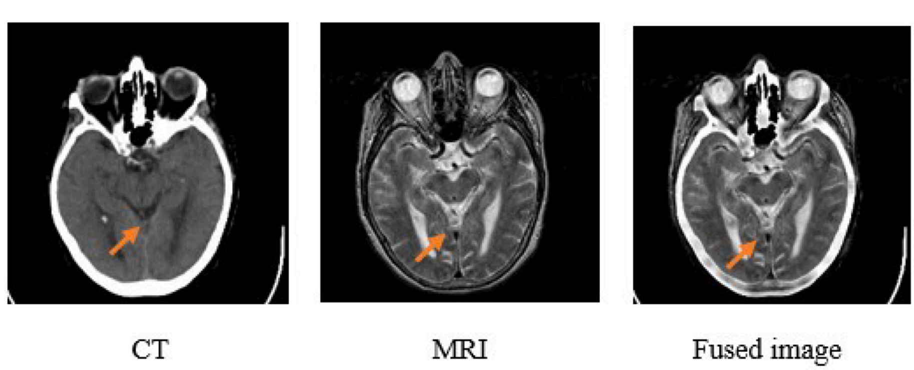

Fig. 8. Fused image N_f of a stroke disease case, it shows the infarct involving the splenium of the corpus callosum.

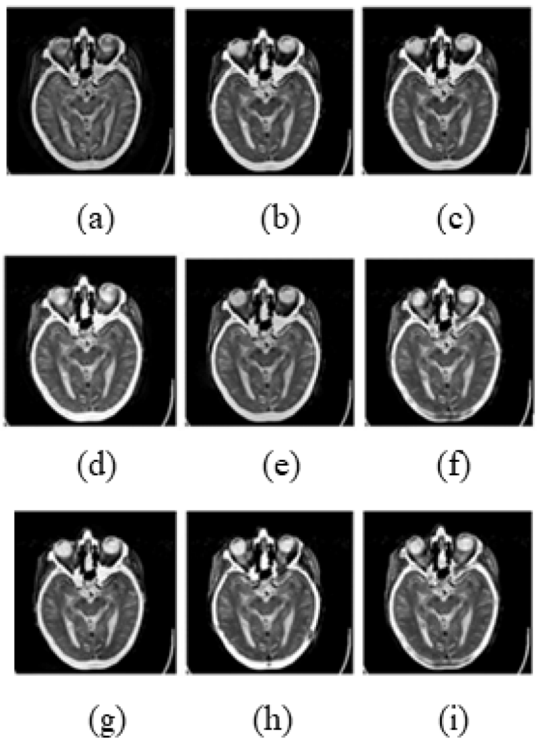

Fig. 9. Fusion results for (a) DCTWT; (b) LP; (c) LP-SR; (d) CNN + LP; (e) NSCT; (f) NSST-PAPCNN; (g) CSMCA; (h) MST-SR; (i) proposed method.

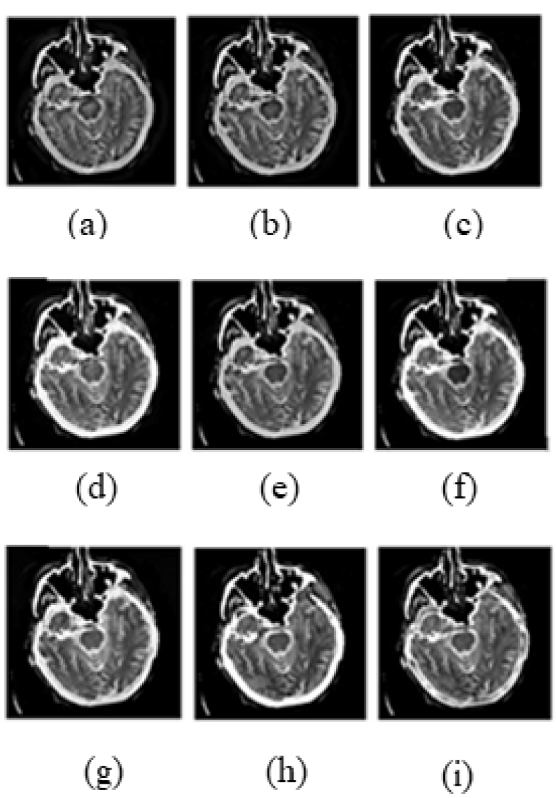

Fig. 10. Fusion results for (a) DCTWT; (b) LP; (c) LP-SR; (d) CNN + LP; (e) NSCT; (f) NSST-PAPCNN; (g) CSMCA; (h) MST-SR; and (i) for the proposed method.

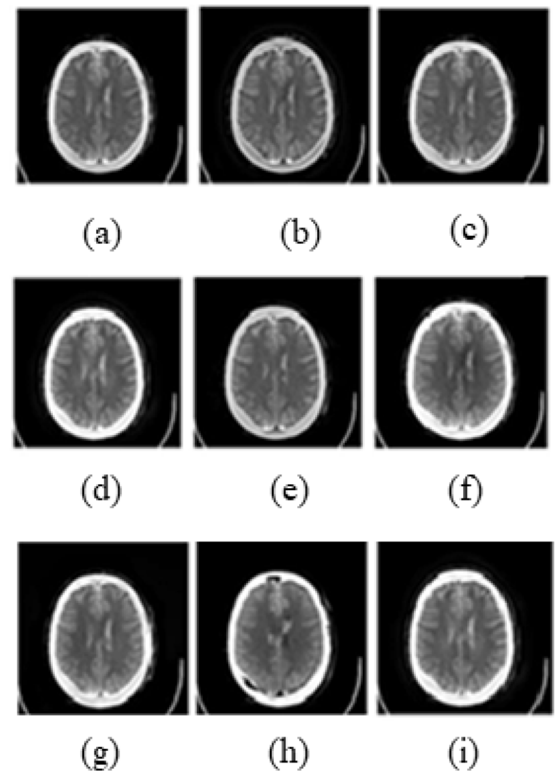

Fig. 11. Fusion results for (a) DCTWT; (b) LP; (c) LP-SR; (d) CNN + LP; (e) NSCT; (f) NSST-PAPCNN; (g) CSMCA; (h) MST-SR; and (i) for the proposed method.

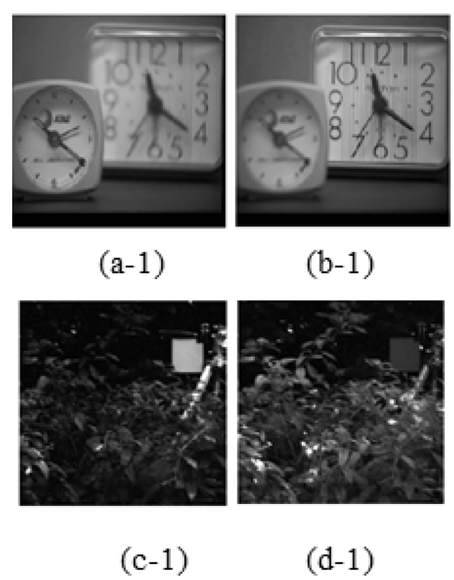

Fig. 12. Testing samples from multi-focus, Infrared and visible image datasets. Pair-1 (a-1) and (b-1) represents the multi-focus clock, while pair-2 (c-1) and (d-1) refers to visible and infrared.

are used. From Table 1, FMI, VIF, Q^{AB}_{F}, Piella (Q_E) and ESSIM recorded a higher score for the proposed method (Fig. 4 (i)) over the state of the arts. Meanwhile, MST-SR (Fig. 4(h)) obtained higher scores over CNN-LP (Fig. 4(d)), MST-SR1 (Fig. 4(c)), CSMCA (Fig. 4(g)), DTWCT (Fig. 4 (a)) and NSCT (Fig. 4(e)) in terms of FMI, VIF and Piella (Q_E) except for a lower ESSIM. Note that lower ESSIM introduced less informative edge details and blocking artifacts for the fused image. Moreover, the LP

Table 1
The details of quantitative assessment results for various fusion methods.

Data-sets	Metrics	DTCWT [43]	NSCT [28]	LP [44]	NSST-PAPCN [10]	CSMCA [46]	LP-SR [45]	CNN-LP [36]	MST-SR [42].	Proposed
Pair-1	**1-FMI**	0.8341	0.7612	0.7412	0.4559	0.4751	0.9705	0.8963	0.9116	**0.9715**
	2-VIF	0.3976	0.3864	0.4141	0.9015	0.9088	0.595	0.69500.8842	0.9645	**0.9858**
		0.6454	0.6872	0.6321	0.6968	0.7373	0.8703	0.9932	0.7997	**0.8945**
	3- $Q^{AB/F}$	0.5342	0.6543	0.5534	0.7345	0.7763	0.746	0.8327	0.8531	**0.9994**
	4-Piella	0.8334	0.8893	0.6543	0.6623	0.8652	0.8327		0.7767	**1.000**
	5-ESSIM									
Pair-2	**1-FMI**	0.9361	0.9197	0.8914	0.5536	0.4679	0.9548	0.9573	0.9744	**0.9756**
	2-VIF	0.5348	0.5132	0.4832	0.8825	0.9048	0.5052	0.9825	0.9891	**0.9908**
		0.7414	0.7101	0.7499	0.5956	0.6397	0.7845	0.7989	0.7842	**0.8168**
	3- $Q^{AB/F}$	0.9943	0.7896	0.9886	0.8912	0.9992	0.8914	0.9993	0.8845	**0.9994**
	4-Piella	0.8012	0.7986	0.7754	0.8100	0.8100	0.6272	0.7989	0.7662	**0.8110**
	5-ESSIM									
Pair-3	**1-FMI**	0.8479	0.8498	0.8568	0.5597	0.4728	0.9145	0.9454	0.8756	**0.9555**
	2-VIF	0.5521	0.5435	0.4352	0.8393	0.8615	0.532	0.9794	0.8951	**0.9813**
		0.6756	0.6837	0.6571	0.5136	0.5772	0.7518	0.7701	0.7169	**0.8651**
	3- $Q^{AB/F}$	0.7986	0,0.9887	0.8876	0.6782	0.9943	0.735	0.9962	0.6347	**0.9975**
	4-Piella	0.5462	0.3876	0.4823	0.7043	0.7110	0.3998	0.6969	0.4376	**0.7142**
	5-ESSIM									
Pair-4	**1-FMI**	0.8339	0.8395	0.7690	0.5401	0.4939	0.9697	0.9698	0.9721	**0.9730**
	2-VIF	0.6679	0.7769	0.9255	0.8960	0.9027	0.5768	0.9904	0.9737	**0.9966**
		0.6921	0.7093	0.6391	0.6076	0.6601	0.7701	0.8322	0.7654	**0.8913**
	3- $Q^{AB/F}$	0.8876	0.4567	0.8435	0.9854	0.9943	0.8142	0.9983	0.9970	**0.9989**
	4-Piella	0.7765	0.7534	0.7823	0.7888	0.6354	0.7447	0.7171	0.7546	**0.7903**
	5-ESSIM									
Pair-5	**1-FMI**	0.8045	0.8641	0.8665	0.8681	0.8703	0.8712	0.8660	0.8564	**0.8691**
	2-VIF	0.41020.	0.4931	0.6142	0.6848	0.4768	0.5131	0.66280.6198	0.5084	**0.9846**
	3- $Q^{AB/F}$	0.5299	0.5953	0.6182	0.4855	0.5893	0.5944	0.9963	0.5776	**0.8786**
	4-Piella	0.6317	0.6893	0.6439	0.7739	0.8163	0.8515	**1.0000**	0.8225	**0.9974**
	5-ESSIM	0.9981	0.9980	0.9980	**1.0000**	0.9985	0.9982		0.9979	**1.0000**
Pair-6	**1-FMI**	0.8604	0.8567	0.8628	0.8504	0.8601	0.8629	0.8615	0.8619	**0.8657**
	2-VIF	0.4271	0.3583	0.5889	0.7008	0.4693	0.6373	0.6542	0.5566	**0.9213**
	3- $Q^{AB/F}$	0.5221	0.5737	0.5930	0.5638	0.5737	0.5936	**0.5919**	0.5845	0.5789
	4-Piella	0.5560	0.6084	0.5282	0.9949	0.7769	0.6782	0.9952	0.7673	**0.9962**
	5-ESSIM	0.9986	0.9983	0.9983	**1.0000**	0.9988	0.9984	**1.0000**	0.9988	**1.0000**
Pair-7	**1-FMI**	0.8976	0.8102	0.9107	0.8210	0.9104	0.9012	0.9102	0.9090	**0.9110**
	2-VIF	0.3421	0.5213	0.4213	0.4404	0.3624	0.4952	0.5235	0.3907	**0.9307**
	3- $Q^{AB/F}$	0.4678	0.5793	0.5870	0.5298	0.5187	0.5922	0.7872	0.5157	**0.5634**
	4-Piella	0.6899	0.6591	0.6292	0.7719	0.8075	0.7527	0.9950	0.7851	**0.9918**
	5-ESSIM	0.9983	0.9987	0.9988	0.9991	0.9994	0.9989	1.0000	0.9992	**1.0000**
	6-GLCM									

method (Fig. 4(d)) increased the visibility of tissue in the fused image by registering a higher score of FMI metric over CSMCA (Fig. 4(g)).

Pair-2 and Fig. 5(a-i) demonstrated the comparison between the proposed method in Fig. 5(i) over the state of arts, which is shown in Fig. 5(a-h). The experiment used the CT and MRI modalities-based brain infected by Stroke disease. The previous methods have integrated the important information and have shown good visual quality. From Table 1, the proposed method depicted higher scores in all five metrics, implying well preserved fused image visibility like edges, textures, lines, and boundaries. It also improved the HVS for fused images and reduced the blur artifact.

For pair-3 of brain disease, which was infected by fatal Stroke, CT/MRI were also used. The aforementioned method is shown in (Fig. 6(a)-(h)), while the proposed method is illustrated in (Fig. 6(i)). The superiority of the proposed method is shown in Table 1, where ESSIM recorded a higher score over CSMCA (Fig. 6(g)), CNN-LP (Fig. 6 (d)). This implies that more local regularity is preserved for the edge overlapped along the direction. Note that CNN-LP achieved a higher score than MST-SR, CSMCA, NSST-PAPCN (Fig. 6(f)), and NSCT (Fig. 6(f)) as it has higher pixels energy preserving. On the other hand, DTCWT and LP have shown less integration of essential information for the fused image like a white, gray matter for the hard and soft tissues scene, as depicted in (Fig. 6(a, c)).

For the brain head-neck, pair-4 with CT/MRI is also processed as shown in (Fig. 7(a)-(i)). We found that CSMCA (Fig. 7(g)) recorded a higher score for the VIF metric than MST-SR1(Fig. 7(c)), DTCWT (Fig. 7 (a)), NSCT (Fig. 7(e)), LP (Fig. 7(b)) and NSST-PAPCN (Fig. 7(f)). It achieved less image sharpness and more information transferred to the fused image. The CNN-LP method showed better performance than MST-SR in terms of FMI, VIF, Piella (Q_E) and ESSIM. It illustrated the advantages of CNN-based fusion in maximizing the visual quality and contrast in extracting information from sources images. Moreover, pairs 5, 6, and 7 are also processed for the stroke disease dataset, highlighted in Figs. 9-11. It indicates that the fused image has the highest quality by SCNN-SR over the previous works, except for the methods CNN-LP [36] and MST-SR [42] in the pair-6. These methods recorded higher scores of $Q^{AB/F}$ based-fusion metric (i.e., 0.5919, and 0.5845, respectively). Overall, the proposed multimodality medical image fusion provided a good visual, natural contrast, superior over the state of arts of fusion schemes. The highest values of objective evaluation metrics imply that

Table 2
The details of quantitative assessment results for various fusion methods for multi-focus images.

Datasets	Metrics	LP-SR [45]	CSMCA [46]	MST-SR [42].	Proposed
Multi-focus	**1-FMI**	0.8895	0.8909	**0.8902**	0.8873
	2-VIF	0.9700	0.9276	0.9380	**0.9992**
	3- Q_F^{AB}	0.7204	0.7085	0.7115	**0.8723**
	4-Piella	0.9501	0.9443	0.9474	**1.0000**
	5-ESSIM	0.9991	0.9991	0.9991	**1.0000**

Table 3
The details of quantitative assessment results for various fusion methods for Infrared and visible images.

Datasets	Metrics	LP-SR [45]	CSMCA [46]	MST-SR [42].	Proposed
Infrared and visible image fusion	**1-FMI**	0.8872	0.8871	0.8837	**0.9980**
	2-VIF	0.9398	0.7953	0.8841	**0.9992**
	3- Q_F^{AB}	0.7102	0.6977	0.6824	**0.8979**
	4-Piella	0.8874	0.8567	0.8777	**0.9995**
	5-ESSIM	0.9987	0.9987	0.9986	**1.0000**

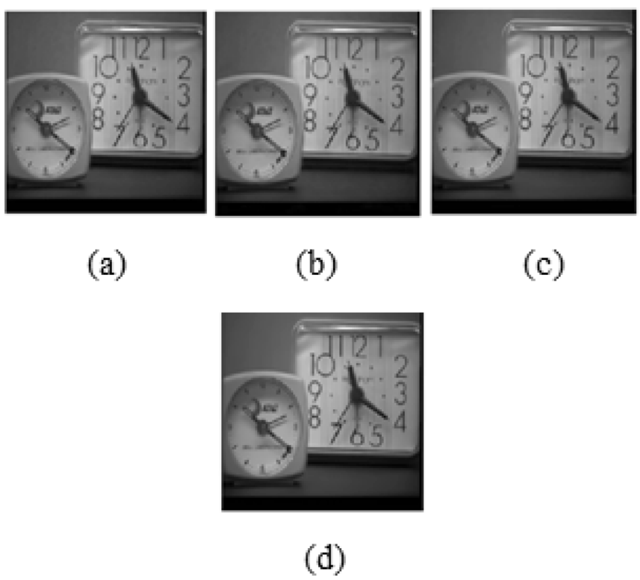

Fig. 13. Fusion results for Fig. 9(a-1), and (b-1). (a) LP-SR [45]; (b) CSMCA [46]; (c) MST-SR [42]; (d) proposed method.

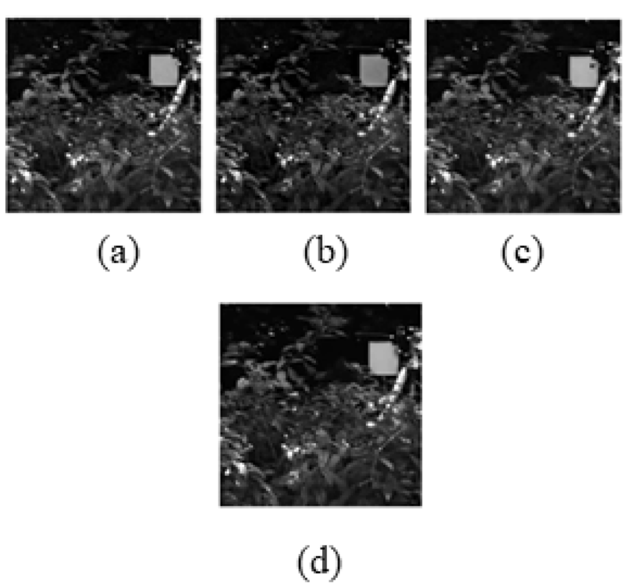

Fig. 14. Fusion results for Fig. 9(c-1), and (d-1). (a) LP-SR [45]; (b) CSMCA [46]; (c) MST-SR [42]; (d) proposed method.

the proposed scheme of SCNN-SR produces fused images with fine edges, sharp features and less blur. This implies that the fused image is considered a higher-performing imaging tool containing the largest amount of information used for clinical diagnosis. Table 1 demonstrated that the proposed method obtains a high-quality fusion result and eliminates the fusion artifact (Table 2).

4.3. Extension to another fusion applications

To demonstrate the generalization capability of the proposed SCNN-OMP fusion framework, we expand its usage to the multi-focus, infrared and visible image fusion applications. The image fusion issues have been obtained from the different mechanisms that contain their own characteristics and differences, for example, studying the scheme and basic architectures. The aim is to study the proposed design-related SCNN and OMP for other types of image fusion issues. OMP and SCNN are strong tools consistent with human visual perception for multi-model image fusion. This paper employs the weighted map obtained from the SCNN model-based design of the K-SVD dictionary and OMP method as they improved the weight maps that indicate pixels' activity level and fusion rule. During the fusion process, the new activity level measurement is procced for each OMP and then continued with the SCNN based-K-SVD. The new local linear of fusion rule method can combine the information from two source images. The fused image is finally obtained by fusion rule reconstruction. Two testing images examples are shown in Fig. 12, where (a-1, b-1) are the multi-focus clock input images, while (c1, d-1) are the visible and infrared input images. Figs. 13 and 14 are the fused image examples for the proposed method over the state of arts (Table 3).

It can be highlighted from Figs. 13 and 14 that the output of the fused image well preserves important local information in the input images. The fusion quality of multi-exposure and visible-infrared images are corresponding to be relatively high. Thus, the fused images extract an important spatial detail without presenting undesirable spatial detail artifacts. For example, in Figs. 13 and 14, the proposed method has recorded promising results for the VIF metric, which is interpreted as favorable results compared to the state of arts (i.e., VIF = 0.9992 and 0.9992). Those two values can mean a lot when transferring the important spatial details to the fused image. Therefore, it indicates that the proposed technique applied by different image fusion issues are not the same.

Nevertheless, they share the excellent output results by the proposed SCNN and OMP based method in the mapping process starting from source images decomposition to the reconstruction of the final fused image. In this section, we aim to apply our proposed technique for multi-focus image fusion or visible and infrared fusion issues to highlight the superiority of our proposed fusion algorithm in various applications. Moreover, the proposed method has shown promising subjective and objective evaluation results than the state of arts.

5. Conclusion

In this paper, the core advantages of the novel proposed hybrid SCNN based K-SVD and OMP increases the medical images fusion's visibility. This avoids the spatial inconsistency for the detailed fused images, such as edge blur or edge overlap. As a result, the proposed method can help the physicians diagnose complex patient's diseases correctly, for instance, distinguishing the stages of stroke disease development (acute stages to fatal stages). Furthermore, SR-CNN-based multimodality fusion has been proposed to achieve a good representation capability for the fusion result. We jointly constructed a novel activity level measurement called CNN-based overcomplete-dictionary with another activity level measurement, namely OMP, which further contributed to the transfer of more edge, texture, and details into the fused image. Simulation results through various experiments demonstrated that the proposed method is more efficient, has better performance, and has good potential in medical image applications.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Wardlaw JM, del Zoppo G, Yamaguchi T. Thrombolytic therapy in acute ischaemic stroke. Part 1. Thrombolysis versus control. In: Warlow C, Van Gijn J, Sandercock P, eds. Stroke module of the Cochrane Database of Systematic Reviews, [updated 10 October 1999]. Available in The Cochrane Library [database on disk and CD-ROM]. The Cochrane Collaboration; Issue 4. Oxford: Update Software; 1999. Updated quarterly—third substantive revision.

[2] Sekhon, M. S., Heran, M., Sekhon, M., & Griesdale, M. (n.d.). Neuroimaging. Neurocritical Care Essentials, 38–46. DOI:10.1017/cbo9781316146187.006.

[3] J. Jan, Medical Image Processing, Reconstruction, and Restoration, CRC Press, 2005.

[4] J.V. Hajnal, D.L.G. Hill, D.J. Hawkes, Medical Image Registration, CRC Press, 2001.

[5] Y. Liu, X. Chen, Z. Wang, Z.J. Wang, R.K. Ward, X. Wang, Deep learning for pixel-level image fusion: Recent advances and prospects, Inform Fusion 42 (2018) 158–173, https://doi.org/10.1016/j.inffus.2017.10.007.

[6] Y. Li, Q. Miao, R. Liu, J. Song, Y. Quan, Y. Huang, A multiscale fusion scheme based on haze-relevant features for single image de-hazing, Neurocomputing 283 (2018) 73–86.

[7] R. Singh, A. Khare, Fusion of multimodal medical images using Daubechies complex wavelet transform – a multi-resolution approach, Inform. Fusion 19 (3) (2014) 49–60.

[8] Siddalingesh G., Mallikarjun A., Sanjeevkumar H., Kotresh S. (2014) Feature-Level Image Fusion Using DWT, SWT, and DT-CWT. In: Sridhar V., Sheshadri H., Padma M. (eds) Emerging Research in Electronics, Computer Science and Technology. Lecture Notes in Electrical Engineering, vol 248. Springer, New Delhi.

[9] B. Yu, B.o. Jia, L.u. Ding, Z. Cai, Q.i. Wu, R. Law, J. Huang, L. Song, S. Fu, Hybrid dual-tree complex wavelet transform and support vector machine for digital multi-focus image fusion, Neurocomputing 182 (2016) 1–9.

[10] J. Adu, S. Xie, J. Gan, Image fusion based on salient visual features and the cross-contrast, J. Vis. Commun. Image Repres. 40 (2016) 218–224.

[11] M. Choi, R.Y. Kim, M.G. Kim, et al., The Curvelet Transform for Image Fusion (2004).

[12] Khare A., Srivastava R., Singh R. (2012) Edge Preserving Image Fusion Based on Contourlet Transform. In: Elmoataz A., Mammass D., Lezoray O., Nouboud F., Aboutajdine D. (eds) Image and Signal Processing. ICISP 2012. Lecture Notes in Computer Science, vol 7340. Springer, Berlin, Heidelberg.

[13] Y.u. Liu, Z. Wang, Dense SIFT for ghost-free multi-exposure fusion, J. Vis. Commun. Image Repres. 31 (2015) 208–224.

[14] G. Guorong, X.u. Luping, F. Dongzhu, Multi-focus image fusion based on non-subsampled shearlet transform, IET Image Process 7 (6) (2013) 633–639.

[15] X. Luo, Z. Zhang, C. Zhang, X. Wu, Multi-focus image fusion using HOSVD and edge intensity, J. Vis. Commun. Image Repres. 45 (2017) 46–61.

[16] B. Yang, S. Li, Multifocus image fusion and restoration with sparse representation, IEEE Trans. Instrum. Meas. 59 (4) (2010) 884–892.

[17] B.A. Olshausen, D.J. Field, Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature 381 (6583) (1996) 607–609.

[18] R. Rubinstein, A.M. Bruckstein, M. Elad, Dictionaries for sparse representation modeling, Proc. IEEE 98 (6) (2010) 1045–1057.

[19] K. Engan, S. Aase, J. Husøy, Multi-frame compression: theory and design, Signal Process. 80 (10) (2000) 2121–2140.

[20] M. Nejati, S. Samavi, S. Shirani, Multi-focus image fusion using dictionary-based sparse representation, Inf. Fusion 25 (2015) 72–84.

[21] Y. Liu, X. Chen, R.K. Ward, Z.J. Wang, Image fusion with sparse convolutional representation, IEEE Signal Process Lett. 23 (12) (2016) 1882–1886.

[22] B. Yang, S. Li, Multi-focus image fusion and restoration with sparse representation, IEEE Trans. Instrum. Meas. 59 (4) (2010) 884–892.

[23] B. Yang, S. Li, Pixel-level image fusion with simultaneous orthogonal matching pursuit, Inf. Fusion 13 (1) (2012) 10–19.

[24] H. Yin, S. Li, Multimodal image fusion with joint sparsity model, Opt. Eng. 50 (6) (2011), 067007.

[25] S. Li, B. Yang, J. Hu, Performance comparison of different multi-resolution transform for image fusion, Information Fusion 12 (2) (2011) 74–84.

[26] Y.u. Liu, S. Liu, Z. Wang, A general framework for image fusion based on multiscale transform and sparse representation, Inf. Fusion 24 (2015) 147–164.

[27] S. Li, J.T. Kwok, Y. Wang, Multifocus image fusion using artificial neural networks, Pattern Recognit. Lett. 23 (8) (2002) 985–997.

[28] Y. Liu, X. Chen, J. Cheng, H. Peng, Z. Wang, Infrared and visible image fusion with convolutional neural networks, Int. J. Wavelets Multiresolut. Inf. Process. 16 (03) (2018) 1850018, https://doi.org/10.1142/s0219691318500182.

[29] K. Wang, Rock Particle Image Fusion Based on Sparse Representation and Non-subsampled Contourlet Transform, Optik 178 (2019) 513–523, https://doi.org/10.1016/j.ijleo.2018.09.121.

[30] B. Yang, L.i. Shutao, Multifocus Image Fusion and Restoration With Sparse Representation, IEEE Trans. Instrum. Meas. 59 (4) (2010) 884–892, https://doi.org/10.1109/Tim.2009.2026612.

[31] B.A. Olshausen, D.J. Field, Emergence of simple-cell receptive field properties by learning a sparse coding for natural images, Nature 381 (6583) (Jun. 1996) 607–609.

[32] I. Daubechies, M. Defrise, C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Commun. Pure Appl. Math. 57 (11) (2004) 1413–1457.

[33] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, Ann. Stat. 32 (2) (2004) 407–499.

[34] M. Nejati, S. Samavi, S.M. Reza Soroushmehr, Najarian Kayvan, Coherence Regularized Dictionary Learning, IEEE International Conference on Acoustics, Speech and Signal Processing, 2016, pp. 4717–4721.

[35] Y. Le Cun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.

[36] Y. Liu, X. Chen, J. Cheng, H. Peng, A medical image fusion method based on convolutional neural networks. 2017 20th International Conference on Information Fusion (Fusion).doi:10.23919/icif.2017.8009769, 2017.

[37] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Y. L, Over feat: Integrated recognition, localization, and detection using convolutional networks, arXiv vol. 1312.6299v4 (2014) 1–16.

[38] Vladimir S. Petrović, Costas S. Xydeas, Sensor noise effects on signal-level image fusion performance, Inform. Fusion 4 (3) (2003) 167–183.

[39] Mohammad Bagher Akbari Haghighat, Ali Aghagolzadeh, Hadi Seyedarabi, A non-reference image fusion metric based on mutual information of image features, Comput. Electr. Eng. 37 (5) (2011) 744–756.

[40] Yu Han, Yunze Cai, Yin Cao, Xiaoming Xu, A new image fusion performance metric based on visual information fidelity, Inform. Fusion 14 (2) (2013) 127–135.

[41] X. Zhang, X. Feng, W. Wang, W. Xue, Edge Strength Similarity for Image Quality Assessment, IEEE Signal Process Lett. 20 (4) (2013) 319–322, https://doi.org/10.1109/lsp.2013.2244081.

[42] Sarmad Maqsood, Umer Javed, Multimodal Medical Image Fusion based on Two-scale Image Decomposition and Sparse Representation, Biomed. Signal Process. Control 57 (2020) 101810, https://doi.org/10.1016/j.bspc.2019.101810.

[43] Jiao Du, Weisheng Li, Bin Xiao, Qamar Nawaz, Union laplacian pyramid with multiple features for medical image fusion, Neurocomputing 194 (2016) 326–339.

[44] Huafeng Li, Hongmei Qiu, Zhengtao Yu, Yafei Zhang, Infrared and visible image fusion scheme based on NSCT and low-level visual features, Infrared Phys. Technol. 76 (2016) 174–184.

[45] Ming Yin, Xiaoning Liu, Yu Liu, Xun Chen, Medical Image Fusion With Parameter-Adaptive Pulse Coupled-Neural Network in Nonsubsampled Shearlet Transform Domain, IEEE Trans. Instrum. Meas. 68 (1) (2019) 49–64.

[46] Yu Liu, Xun Chen, Rabab K. Ward, Z. Jane Wang, Medical Image Fusion via Convolutional Sparsity Based Morphological Component Analysis, IEEE Signal Process Lett. 26 (3) (2019) 485–489.