20th International Conference on Information Fusion
Xi'an, China - July 10-13, 2017

A Medical Image Fusion Method Based on Convolutional Neural Networks

Yu Liu, Xun Chen, Juan Cheng, Hu Peng
Department of Biomedical Engineering, Hefei University of Technology, Hefei 230009, China
Email: {yuliu, xun.chen, chengjuan, hpeng}@hfut.edu.cn

***Abstract*—Medical image fusion technique plays an an increasingly critical role in many clinical applications by deriving the complementary information from medical images with different modalities. In this paper, a medical image fusion method based on convolutional neural networks (CNNs) is proposed. In our method, a siamese convolutional network is adopted to generate a weight map which integrates the pixel activity information from two source images. The fusion process is conducted in a multi-scale manner via image pyramids to be more consistent with human visual perception. In addition, a local similarity based strategy is applied to adaptively adjust the fusion mode for the decomposed coefficients. Experimental results demonstrate that the proposed method can achieve promising results in terms of both visual quality and objective assessment.**

I. Introduction

With the rapid development of sensor and computer technology, medical imaging has emerged as an irreplaceable component in various clinical applications including diagnosis, treatment planning and surgical navigation. To provide medical practitioners sufficient information for clinical purposes, medical images obtained with multiple modalities are usually required, such as X-ray, computed tomography (CT), magnetic resonance (MR), positron emission tomography (PET), single photon emission computed tomography (SPECT), etc. Due to the difference in imaging mechanism, medical images with different modalities focus on different categories of organ/tissue information. For instance, the CT images are commonly used for the precise localization of dense structures like bones and implants, the MR images can provide excellent soft-tissue details with high-resolution anatomical information, while the functional information on blood flow and metabolic changes can be offered by PET and SPECT images but with low spatial resolution. Multi-modal medical image fusion aims at combining the complementary information contained in different source images by generating a composite image for visualization, which can help physicians make easier and better decisions for various purposes [1].

In recent years, a variety of medical image fusion methods have been proposed [2]–[17]. Due to the difference in imaging mechanism, the intensities of different source images at the same location often vary significantly. For this reason, most of these fusion algorithms are introduced in a multi-scale manner to pursue perceptually good results. In general, these multi-scale transform (MST)-based fusion methods consist of three steps, namely, decomposition, fusion and reconstruction. Multi-scale transforms which are frequently studied in image fusion include pyramids [17]–[19], wavelets [9], [20], [21], multi-scale geometrical transforms like contourlet and shearlet [2], [6], [10], [16]. In image fusion research, sparse representation is another popular image modelling approach, which has also been successfully applied to fuse multi-modal medical images [4], [5], [15], [22].

One of the most crucial issues in image fusion is calculating a weight map which integrates the pixel activity information from different sources. In most existing fusion methods, this target is achieved by two steps known as *activity level measurement* and *weight assignment*. In conventional transform domain fusion methods, the absolute value of a decomposed coefficient (or the sum of those values within a small window) is employed to measure its activity, and then a "choose-max" or "weighted-average" fusion rule is applied to assign weights to different sources based on the obtained measurement. Clearly, this kind of activity measurement and weight assignment are usually not very robust resulting from many factors like noise, mis-registration and the difference between source pixel intensities. To improve the fusion performance, many complex decomposition approaches and elaborate weight assignment strategies have been recently proposed in the literature [6], [8]–[13], [15], [16]. However, it is actually not an easy task to design a ideal activity level measurement or weight assignment strategy which can comprehensively take all the key issues of fusion into account. Moreover, these two steps are designed individually without a strong association by many fusion methods, which may greatly limit the algorithm performance.

In this paper, this issue is addressed from another viewpoint to overcome the difficulty in designing robust activity level measurements and weight assignment strategies. Specifically, a convolutional neural network (CNN) [23] is trained to encode a direct mapping from source images to the weight map. In this way, the activity level measurement and weight assignment can be jointly achieved in an "optimal" manner via learning network parameters. Considering the different imaging modalities of multi-modal medical images, we adopt a multi-scale approach via image pyramids to make fusion process more consistent with human visual perception. In addition, a local similarity based strategy is applied to adaptively adjust the fusion mode for the decomposed coefficients of source images.

Corresponding author: Xun Chen

978-0-9964-5270-0©2017 ISIF

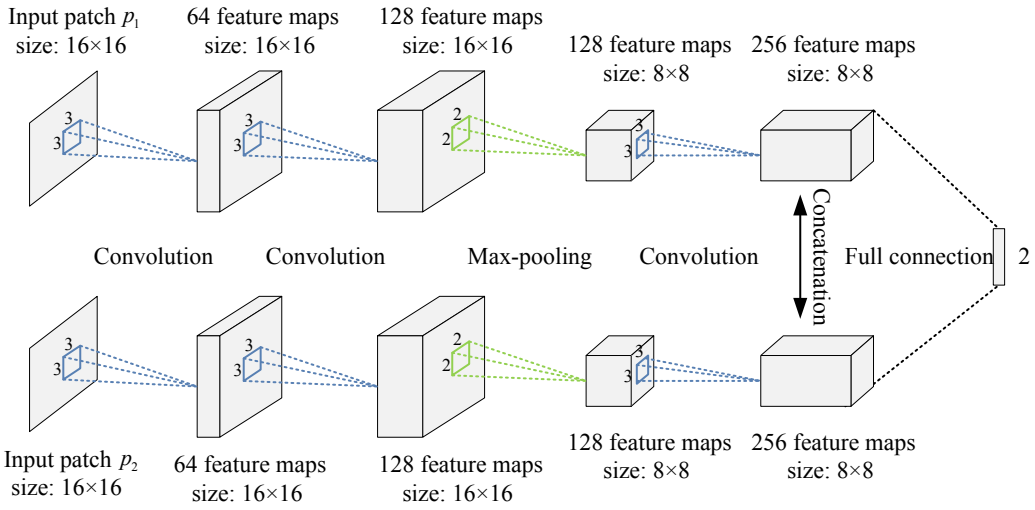

Fig. 1. The architecture of the siamese network for training.

The rest of this paper is organized as follows. In Section 2, some related work and the basic idea of the proposed fusion method are introduced. Section 3 presents the fusion method in detail. Experimental results and discussions are provided in Section 4. Finally, Section 5 concludes the paper.

II. Related work and motivation

In our recent work [24], a CNN-based multi-focus image fusion method which can obtain state-of-the-art results was proposed. In the method, two source images are fed to the two branches of a siamese convolutional network in which the two branches share the same architecture and weights [25], respectively. Each branch contains three convolutional layers and the obtained feature maps essentially act as the role of activity level measures. The feature maps of two branches are concatenated and then pass through two fully-connected layers (they are converted into equivalent convolutional layers in the fusion process to allow arbitrary input size [26]), which can be viewed as the weight assignment part of a fusion method. As a result, the value of each coefficient in the network output map indicates the focus property of a pair of source image patches at a corresponding location. By assigning the value as the weights of all the pixels within the patch location and then averaging the overlapped pixels, a focus map with the same size of source images is generated. The final fused image is obtained based on the focus map using the weighted-average rule along with two consistency verification techniques [20]. In [24], the feasibility and superiority of CNNs used for image fusion have been explicitly presented. Please refer to [24] for more details.

The target of this paper is to extend the CNN model to medical image fusion. However, the method proposed in [24] cannot be directly used to fuse medical images primarily due to the following two reasons.

1) The fusion method [24] is performed in spatial domain. As medical images are obtained with different imaging modalities, transform domain based fusion methods are more likely to produce results with less undesirable artifacts for their good consistency with human visual perception.
2) The two inputs of the CNN model in [24] are assumed to have similar local structures. As mentioned above, the intensities of multi-modal medical images at the same location often vary significantly, so the assumption of local similarity between two source images is not always valid.

To address the first problem, we apply a pyramid-based multi-scale approach [27] to pursue perceptually better results. Specifically, each source image is decomposed into a Laplacian pyramid while the weight map obtained from the network is decomposed into a Gaussian pyramid. The fusion procedure is conducted at every decomposition level.

For the second issue, we adopt a local similarity-based fusion strategy to determine the fusion mode for the decomposed coefficients [18]. When the contents of source images have high similarity, the "weighted-average" fusion mode is applied to avoid losing useful information. In this situation, the weights obtained by the CNN is more reliable than the coefficient-based measure, so they are employed as the merging weights. When the similarity of image contents is low, the "choose-max" or "selection" fusion mode is preferred to mostly preserve the salient details from source images. In this situation, the CNN output is not reliable and the pixel activity is directly measured by the absolute values of the decomposed coefficients.

Based on the above ideas, the CNN model presented in [24] can be applied to the fusion of medical images. It is worthwhile to note that both the pyramid-based decomposition and the similarity-based fusion mode determination are just "naive" techniques which are commonly-used in the field of image fusion. Nevertheless, it will be demonstrated that a reasonable usage of these techniques incorporated with the CNN model can result in state-of-the-art fusion performance.

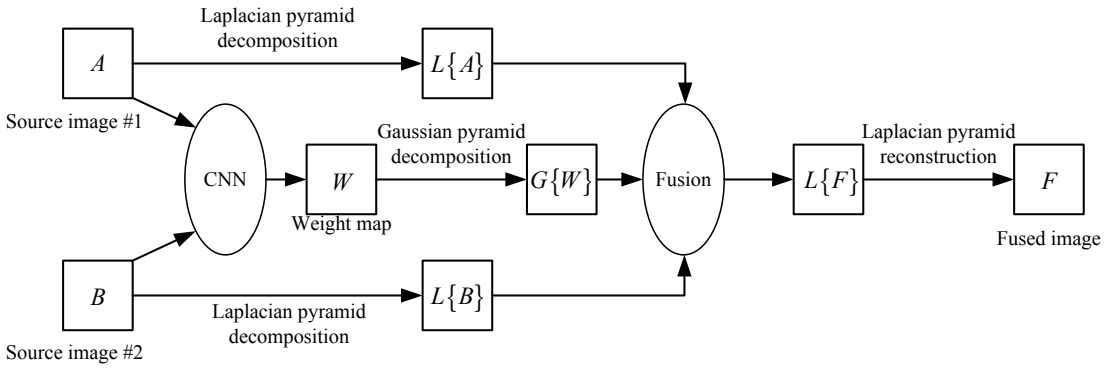

Fig. 2. Schematic diagram of the proposed medical image fusion algorithm.

III. The proposed method

A. The CNN model for medical image fusion

Fig. 1 shows the convolutional network used in the proposed fusion algorithm. It is a siamese network in which the weights of the two branches are constrained to the same. Each branch consists of three convolutional layers and one max-pooling layer, which is the same as the network used in [24]. To reduce the memory consumption as well as increase the computational efficiency, we adopt a much slighter model in this work by removing a fully-connected layer from the network in [24]. The 512 feature maps after concatenation are directly connected to a 2-dimensional vector. It can be calculated that the slight mode only takes up about 1.66 MB of physical memory in single precision, which is significantly less than the 33.6 MB model employed in [24]. Finally, this 2-dimensional vector is fed to a 2-way softmax layer (not shown in Fig. 1), which produces a probability distribution over two classes. The two classes correspond two kinds of normalized weight assignment results, namely, "first patch 1 and second patch 0" and "first patch 0 and second patch 1", respectively. The probability of each class indicates the possibility of each weight assignment. In this situation, also considering that the sum of two output probabilities is 1, the probability of each class just indicates the weight assigned to its corresponding input patch.

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

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

B. Detailed fusion scheme

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

Step 1: CNN-based weight map generation.

Feed the two source images A and B to the two branches of the convolutional network, respectively. The weight map W is generated using the approach described in Section 3.1.

Step 2: Pyramid decomposition.

Decompose each source image into a Laplacian pyramid. Let $L\{A\}^l$ and $L\{B\}^l$ respectively denote the pyramids of A and B, where l indicates the l-th decomposition level. Decompose the weight map W into a Gaussian pyramid $G\{W\}^l$. The total decomposition level of each pyramid is set to the highest possible value $\lfloor \log_2 \min(H, W) \rfloor$, where $H \times W$ is the spatial size of source images and $\lfloor \cdot \rfloor$ denotes the flooring operation.

Step 3: Coefficient fusion.

For each decomposition level l, calculate the local energy map (sum of the squares of the coefficients within a small window) [18] of $L\{A\}^l$ and $L\{B\}^l$, respectively.

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

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

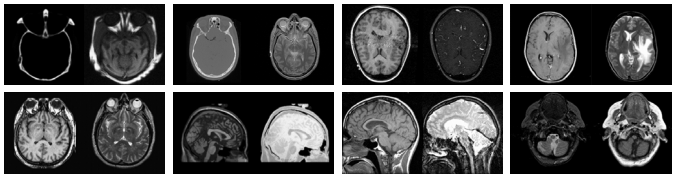

Fig. 3. Source images used in the experiments.

The similarity measure used for fusion mode determination is calculated as

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

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

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

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

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

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

Step 4: Laplacian pyramid reconstruction.

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

IV. Experiments

A. Experimental settings

To verify the effectiveness of the proposed fusion method, eight pairs of medical images shown in Fig. 3 are mainly used in our experiments.

In the experiments, five objective fusion metrics which are commonly-used in medical image fusion are adopted to make quantitative evaluations. They are the information entropy (EN) of the fused image, the feature mutual information (FMI) [29], Xydeas-Petrovic's gradient based metric Q_G [30], Piella-Heijmans's similarity based metric Q_E [31], and the visual information fidelity fusion ($VIFF$) metric [32]. A higher score indicates a better fusion performance for all of these metrics.

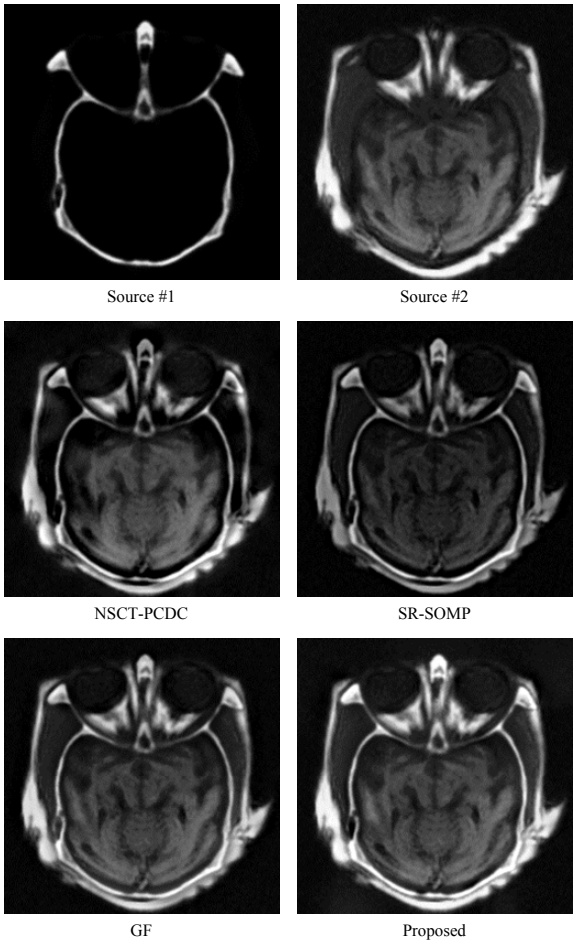

Fig. 4. Fusion results of different methods on a pair of CT and MR images.

The proposed fusion method is compared with other three recently proposed medical image fusion methods, which are the non-subsampled contourlet transform based method with phase congruency and directive contrast (NSCT-PCDC) [6], the sparse representation based method with simultaneous orthogonal matching pursuit (SR-SOMP) [4], and the guided filtering (GF)-based method [7]. The parameters of these three methods are all set to the default values reported in the related publications. In the proposed fusion method, the parameter of threshold t is experimentally set to 0.6 by comparing the visual experience and objective performance using different settings.

B. Experimental results and discussions

Figures 4-6 show the fusion results of different methods on three pairs of multi-modal medical images (CT and MR images in Fig. 4, MR-Gad and MR-T2 images in Fig. 5,

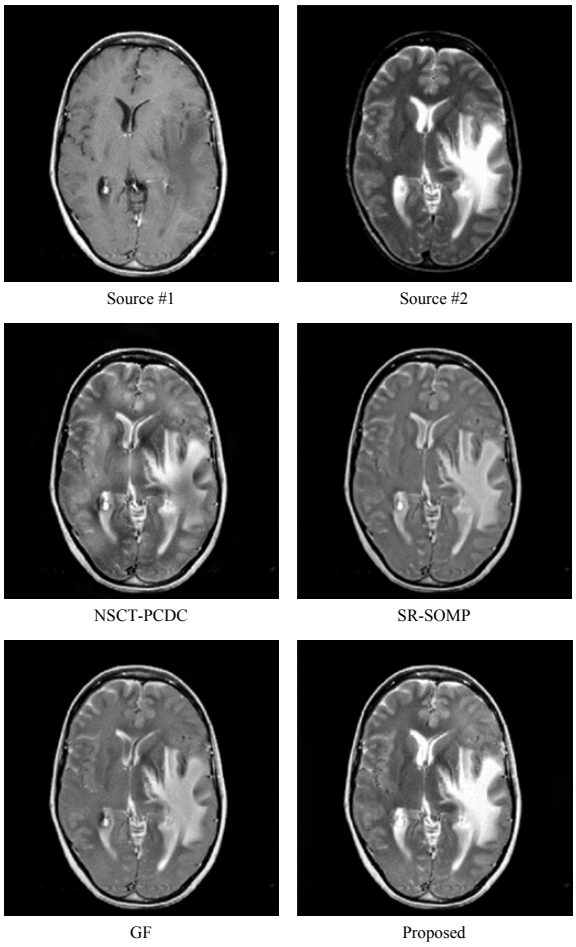

Fig. 5. Fusion results of different methods on a pair of MR-Gad and MR-T2 images.

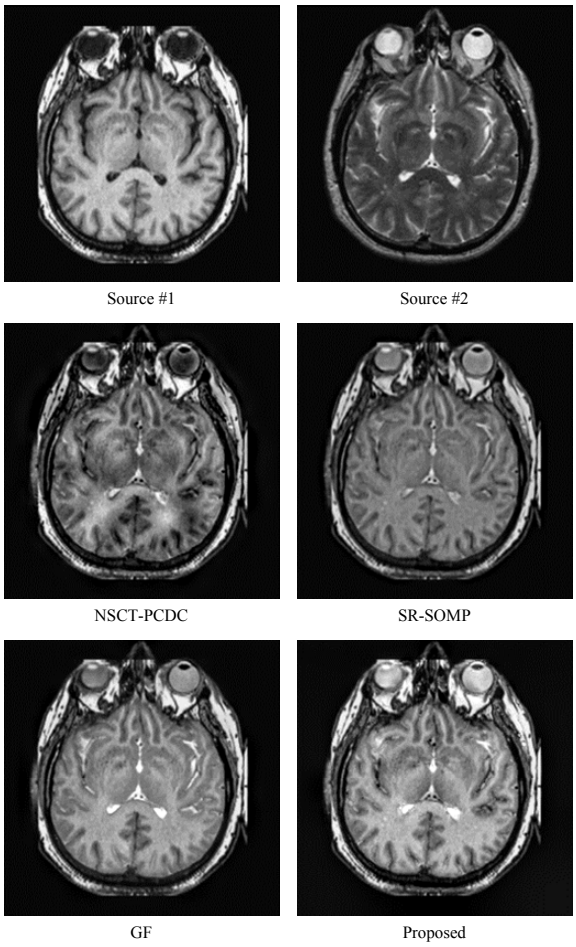

Fig. 6. Fusion results of different methods on a pair of MR-T1 and MR-T2 images.

MR-T1 and MR-T2 images in Fig. 6), respectively. It can be seen that the NSCT-PCDC method could extract sufficient spatial details from source images, but the fused images suffer from some undesirable artifacts which degrades the visual perception to some extent. The SR-SOMP method can well prevent visual artifacts, but it tend to lose the energy contained in source images, leading to decreasing brightness and contrast of some regions in the fused image. The main defects of the GF method is its limited ability in detail preservation, and it can be observed that many small details of the source images are blurred in the fused images. The proposed fusion method generally do well in both detail and energy preservation without introducing undesirable visual artifacts.

TABLE I
OBJECTIVE ASSESSMENT OF DIFFERENT FUSION METHODS.

Method	NSCT-PCDC	SR-SOMP	GF	Proposed
EN	5.9811	5.4970	5.8209	**6.1741**
FMI	0.8768	0.8809	0.8864	**0.8872**
Q_G	0.5531	0.6023	**0.6423**	0.6309
Q_E	0.5780	0.5955	0.6191	**0.6457**
$VIFF$	0.4806	0.5106	0.5465	**0.6412**

Table 1 lists the objective assessment of different fusion methods. For each metric, the average score of each method over eight source image pairs is provided and the value shown as bold denotes the best score among all the methods. The proposed method outperforms other three methods on all the metrics except for Q_G on which our method gets the second place just inferior to the GF method. Moreover, the advantages are generally very clear.

At last, we compare the computational efficiency of different fusion methods. The proposed method is implemented by two individual means, namely, C++ and MATLAB. In the C++ implementation, some interfaces of the Caffe library are called to make source images go forward through the network using GPU parallel computation. The other part of the algorithm is serially implemented. The MATLAB version is implemented in a purely serial way without utilizing GPU and C/C++. The other three methods are primarily implemented with MATLAB, while some functions in the NSCT-PCDC and SR-SOMP methods are complied from C/C++ programming files for acceleration. All the experiments are conducted on a computer which is equipped with an Inter Core i7-4790k CPU, a NIVIDA GTX TITAN Black GPU, and 16 GB RAM.

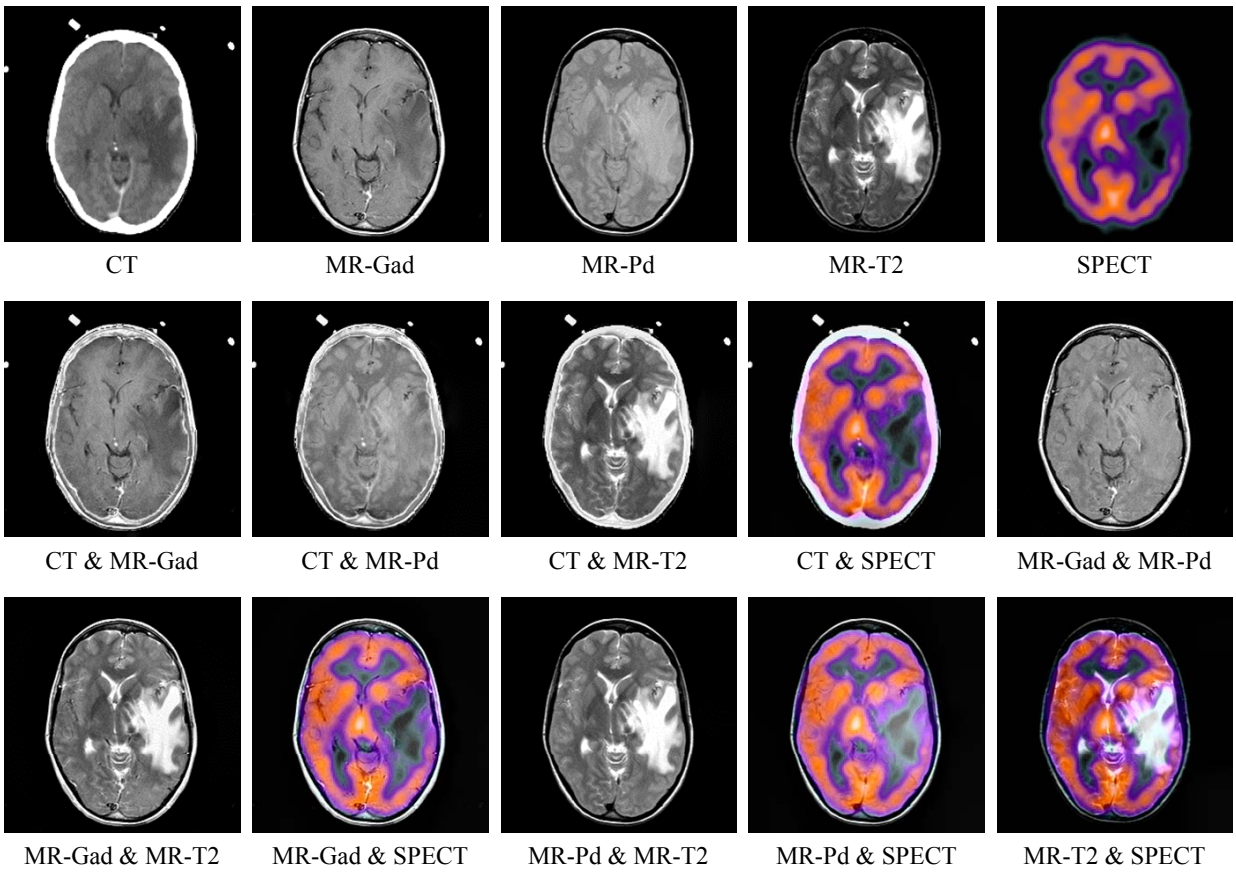

Fig. 7. A pairwise fusion of five medical images with different modalities.

TABLE II
THE AVERAGE RUNNING TIME OF DIFFERENT METHODS FOR FUSING TWO SOURCE IMAGES OF SIZE 256×256 PIXELS (UNIT: SECONDS).

Method	NSCT-PCDC	SR-SOMP	GF	Proposed (Matlab)	Proposed (C++)
Time	15.8	46.3	0.06	12.1	0.08

Table 2 lists the average running time of different methods for fusing two source images of size 256×256 pixels. It can be seen that the proposed method is relatively time consuming when implemented in MATLAB, but still more efficient than the NSCT-PCDC and SR-SOMP methods. The C++ version accelerated by GPU computing takes less than 0.1s to accomplish the fusion, which exhibits the potential of the proposed method for practical usage.

C. More fusion results

To further demonstrate the effectiveness of the proposed method on fusing medical images with various modalities, as shown in Fig. 7, we conduct a pairwise fusion among five source images which obtained with CT, MR-Gad, MR-Pd, MR-T2 and SPECT imaging technologies, respectively. These five source images are downloaded from Harvard Medical School's project of the whole brain atlas [33] and the related clinical case is metastatic bronchogenic carcinoma. The images were obtained from a 42 year old woman with a long history of tobacco use began having headaches. To fuse color images such as SPECT, we extend our fusion method by applying the YUV color space. Specifically, the color source image is firstly converted into YUV space from the original RGB space. Then, the Y component is merged with the other grayscale image. The final fused image is obtained by performing YUV to RGB conversion using the intermediate fusion result along with the original U and V components. It can be seen from Fig. 7 that the proposed method can achieve promising results on every pairwise fusion.

V. CONCLUSION

In this paper, a medical image fusion method based on convolutional neural networks is proposed. We employ a siamese network to generate a direct mapping from source images to a weight map which contains the integrated pixel activity information. The main novelty of this approach is it can jointly implement activity level measurement and weight assignment via network learning, which can overcome the difficulty of artificial design. To achieve perceptually good results, some popular techniques in image fusion such as multi-scale processing and adaptive fusion mode selection are appropriately adopted. Experimental results demonstrate that

the proposed method can obtain high-quality results in terms of visual quality and objective metrics. In addition to the proposed algorithm itself, another contribution of this work is that it exhibits the great potential of some deep learning techniques for image fusion, which will be further studied in the future.

ACKNOWLEDGMENT

The authors would like to thank Gaurav Bhatnagar, Bin Yang and Xudong Kang for sharing the codes of their methods for comparison in this work. This work was supported by the Fundamental Research Funds for the Central Universities (Grants JZ2017HGTA0176, JZ2016HGBZ1025 and JZ2016HGPA0731), the National Natural Science Foundation of China (Grants 81571760, 61501164 and 61401138), and the Research and Development Foundation of Hefei Research Institutes (Grant IMICZ2015111).

REFERENCES

[1] A. James and B. Dasarathy, "Medical image fusion: a survey of the state of the art," *Information Fusion*, vol. 19, pp. 4–19, 2014.

[2] L. Yang, B. Guo, and W. Ni, "Multimodality medical image fusion based on multiscale geometric analysis of contourlet transform," *Neurocomputing*, vol. 72, pp. 203–211, 2008.

[3] Z. Wang and Y. Ma, "Medical image fusion using m-pcnn," *Information Fusion*, vol. 9, pp. 176–185, 2008.

[4] B. Yang and S. Li, "Pixel-level image fusion with simultaneous orthogonal matching pursuit," *Information Fusion*, vol. 13, pp. 10–19, 2012.

[5] S. Li, H. Yin, and L. Fang, "Group-sparse representation with dictionary learning for medical image denoising and fusion," *IEEE Transactions on Biomedical Engineering*, vol. 59, pp. 3450–3459, 2012.

[6] Z. L. G. Bhatnagar, Q. Wu, "Directive contrast based multimodal medical image fusion in nsct domain," *IEEE Transactions on Multimedia*, vol. 15, pp. 1014–1024, 2013.

[7] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2864–2875, 2013.

[8] R. Shen, I. Cheng, and A. Basu, "Cross-scale coefficient selection for volumetric medical image fusion," *IEEE Transactions on Biomedical Engineering*, vol. 60, pp. 1069–1079, 2013.

[9] R. Singh and A. Khare, "Fusion of multimodal medical images using daubechies complex wavelet transform c a multiresolution approach," *Information Fusion*, vol. 19, pp. 49–60, 2014.

[10] L. Wang, B. Li, and L. Tan, "Multimodal medical volumetric data fusion using 3-d discrete shearlet transform and global-to-local rule," *IEEE Transactions on Biomedical Engineering*, vol. 61, pp. 197–206, 2014.

[11] Z. Liu, H. Yin, Y. Chai, and S. Yang, "A novel approach for multimodal medical image fusion," *Expert Systems with Applications*, vol. 41, pp. 7425–7435, 2014.

[12] G. Bhatnagar, Q. Wu, and Z. Liu, "A new contrast based multimodal medical image fusion framework," *Neurocomputing*, vol. 157, pp. 143–152, 2015.

[13] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," *Information Fusion*, vol. 24, no. 1, pp. 147–164, 2015.

[14] Q. Wang, S. Li, H. Qin, and A. Hao, "Robust multi-modal medical image fusion via anisotropic heat diffusion guided low-rank structural analysis," *Information Fusion*, vol. 26, pp. 103–121, 2015.

[15] Y. Liu and Z. Wang, "Simultaneous image fusion and denosing with adaptive sparse representation," *IET Image Process.*, vol. 9, no. 5, pp. 347–357, 2015.

[16] Y. Yang, Y. Que, S. Huang, and P. Lin, "Multimodal sensor medical image fusion based on type-2 fuzzy logic in nsct domain," *IEEE Sensors Journal*, vol. 16, pp. 3735–3745, 2016.

[17] J. Du, W. Li, B. Xiao, and Q. Nawaz, "Union laplacian pyramid with multiple features for medical image fusion," *Neurocomputing*, vol. 194, pp. 326–339, 2016.

[18] P. Burt and R. Kolczynski, "Enhanced image capture through fusion," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 1993, pp. 173–182.

[19] A. Toet, "A morphological pyramidal image decomposition," *Pattern Recognition Letters*, vol. 9, no. 4, pp. 255–261, 1989.

[20] H. Li, B. Manjunath, and S. Mitra, "Multisensor image fusion using the wavelet transform," *Graphical Models and Image Processing*, vol. 57, no. 3, pp. 235–245, 1995.

[21] J. Lewis, R. OCallaghan, S. Nikolov, D. Bull, and N. Canagarajah, "Pixel- and region-based image fusion with complex wavelets," *Information Fusion*, vol. 8, no. 2, pp. 119–130, 2007.

[22] Y. Liu, X. Chen, R. Ward, and Z. Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Processing Letters*, vol. 23, no. 12, pp. 1882–1886, 2016.

[23] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based leaning applied to document recognition," *Proceedings of The IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[24] Y. Liu, X. Chen, J. Cheng, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Information Fusion*, vol. 36, pp. 191–207, 2017.

[25] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4353–4361.

[26] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. L, "Overfeat: Integrated recognition, localizaton and detection using convolutional networks," *arXiv*, vol. 1312.6299v4, pp. 1–16, 2014.

[27] T. Mertens, J. Kautz, and F. V. Reeth, "Exposure fusion," in *Proc. Pacific Graphics*, 2007, pp. 382–390.

[28] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the ACM International Conference on Multimedia*, 2014, pp. 675–678.

[29] M. Haghighat, A. Aghagolzadeh, and H.Seyedarabi, "A non-reference image fusion metric based on mutual information of image features," *Computers and Electrical Engineering*, vol. 37, pp. 744–756, 2011.

[30] C. S. Xydeas and V. S. Petrovic, "Objective image fusion performance measure," *Electronics Letters*, vol. 36, no. 4, pp. 308–309, 2000.

[31] G. Piella and H. Heijmans, "A new quality metric for image fusion," in *Proceedings of 10th International Conference on Image Processing*, 2003, pp. 173–176.

[32] Y. Han, Y. Cai, Y. Cao, and X. Xu, "A new image fusion performance metric based on visual information fidelity," *Information Fusion*, vol. 14, pp. 127–135, 2013.

[33] http://www.med.harvard.edu/AANLIB/.