

Medical image fusion method by deep learning

Yi Li^{a,b,*}, Junli Zhao^a, Zhihan Lv^a, Jinhua Li^a

^a College of Data Science Software Engineering, Qingdao University, Qingdao, Shandong 266071, PR China

^b Business School, Qingdao University, Qingdao, Shandong 266071, PR China

ARTICLE INFO

Keywords:

Deep learning
Multi-modal medical image
Image fusion

ABSTRACT

Deep learning technology has been extensively explored in pattern recognition and image processing areas. A multi-mode medical image fusion with deep learning will be proposed, according to the characters of multi-modal medical image, medical diagnostic technology and practical implementation, according to the practical needs for medical diagnosis. It cannot be only made up for the deficiencies of MRI, CT and SPECT image fusion, but also can be implemented in different types of multi-modal medical image fusion problems in batch processing mode, and can be effectively overcome the limitation of only one-page processing. The proposed method can greatly improve the fusion effect, image detail clarity and time efficiency. The experiments on multi-modal medical images are implemented to analyze performance, algorithm stability and so on. The experimental results prove the superiority of our proposed method in terms of visual quality and a variety of quantitative evaluation criteria.

1. Introduction

Most images have unilateral or limited information content. For instance, focus positions are different, while objects closer or further than that appear as blurred in one image. Different types of multi-modal images are classified as different types of images if they are strictly distinguished. Different types of images reflect different types of information. It makes the information too scattered and hampers the doctor's judgment. This problem has attracted the attentions of plenty of researchers in the field of medical diagnose (Liu et al., 2018; Peter et al., 2019; Sandhya et al., 2019). With the continuous recommendation of medical image processing research in recent years, image fusion is an effective solution which can automatically detect the information in different images and integrates them to produce one composite image in which all objects in interest are clear. Image Fusion (Farid M et al., 2019; Liu et al., 2019; Ma et al., 2019; Ouerghi et al., 2017; Pan & Shen, 2019; Yang et al., 2019) is a specific algorithm to combine two or more images into a new image. Potentially, multi-modal medical image fusion is an important branch in the field of image fusion. Due to the wide use of multi-modal medical images, this problem has become a hot topic in recent years. Based on their domain, these mainly approaches can be categorized into three classes: deep learning algorithm (Ahmad et al., 2017; Asif et al., 2018; Chen & Konukoglu, 2018; Chen et al., 2019; Cheng et al., 2018; Hou et al., 2018; Ijjina E., 2016; Jiao et al., 2018; Liu et al., 2017; Liu et al., 2016; Liu et al., 2019; Ma et al., 2018; Saadat et al., 2017; Salvado et al., 2018; Schlemper et al., 2018; Schramm et al., 2018; Shao & Cai, 2018; Sun et al., 2019;

Thirukovalluru et al., 2016; Wohlberg, 2016; Yang et al., 2018; Ye et al., 2018), transform domain algorithm (Argal et al., 2018; Geng et al., 2018; Li et al., 2018) and spatial domain algorithms (Wang & Bovik, 2002). At present, in the field of image fusion, deep learning method is a representative method, and it is also one of the research focuses in recent years. Many scholars at home and abroad have conducted on deep learning research, and widely apply the research results in image processing and other fields. The recent research in image fusion based on deep learning is listed as follows: pixel-level image fusion (Liu et al., 2018), convolutional neural networks (CNN) (Cheng et al., 2018; Hou et al., 2018; Liu et al., 2017; Schlemper et al., 2018; Shao & Cai, 2018; Sun et al., 2019), convolutional sparse representation (CSR) (Liu et al., 2016; Liu et al., 2019; Wohlberg, 2016), stacked auto-encoders (SAEs) (Ahmad et al., 2017; Chen et al., 2019; Ijjina E., 2016; Jiao et al., 2018; Ma et al., 2018; Thirukovalluru et al., 2016) and deep boltzmann machine (DBM) (Asif et al., 2018; Saadat et al., 2017; Ye et al., 2018). Most algorithms based pixel-level work in four steps: training, classification, weight and fusion. Liu (Liu et al., 2018) presents a systematic review of the DL-based pixel-level image fusion literature. In this paper, application of conventional image fusion techniques according to specific fusion issues is enlarged. As an example in CNN research, Shao (Shao & Cai, 2018) proposes a remote sensing image fusion method based on CNN. Cheng (Cheng et al., 2018) focuses on challenges within-class diversity and between-class similarity. The proposed D-CNN models are trained by optimizing a new discriminative objective function. A metric learning regularization term on the CNN features is imposed in this research in order to outperform the existing baseline methods and achieve

* Corresponding author at: College of Data Science Software Engineering, Qingdao University, Qingdao, Shandong 266071, PR China.

E-mail address: lyqgx@126.com (Y. Li).

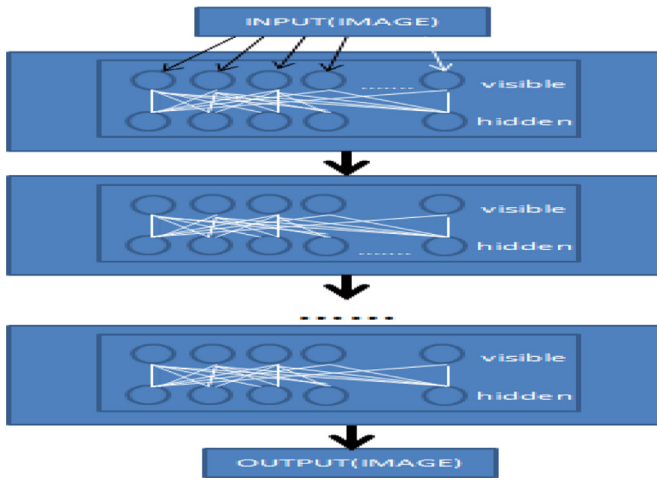


Fig. 1. Training process of DBM.

state-of-the-art results. **Level measurement and fusion rule are still important problem in CNN fusion modal.** Emphasis in work (Liu et al., 2017) is laid on these two aspects of research. It proposes a new multi-focus image fusion method to overcome the difficulty faced by the existing fusion methods. In CSR (Wohlberg, 2016) research, framework in CSR-based image fusion is listed as follows: first, one type of image transform is performed on the source images. Then, the convolutional sparse coding is performed on learned dictionary filters. Next, fusion strategy and fusion rules are designed to obtain the fused bands, using sparse coefficient maps or dictionary filters. Finally, the fused image is reconstructed by performing the inverse transform. Liu (Liu et al., 2016) points out two drawbacks in the traditional sparse representation model: one is limited ability in detail preservation. Two is high sensitivity to registration. A convolutional sparse representation (CSR) into image fusion to address this problem is introduced. However, it's regrettable in the problem about multi-component and global sparse representations of source images at the same time. Since then, this problem is carried out in the work (Liu et al., 2019). A new medical image fusion method based on the CS-MCA model is proposed in this letter. The application of SAEs in image fusion is still rare. SAEs is one important research branch in deep learning. Research on structural framework and its improved version is in the work (Ahmad et al., 2017; Chen et al., 2019; Ijjina E., 2016; Jiao et al., 2018; Ma et al., 2018; Thirukovalluru et al., 2016). Research on its application in image detection and other fields is quite common. For instance, In this paper (Ma et al., 2018), a fast unsupervised deep fusion framework for change detection is presented. Compared with shallow network, the proposed framework can extract more available features and get better results in detection. DBM model has been successfully applied in the field of image interpretation and good results have been achieved in work (Ye et al., 2018). A multi-modal

feature fusion based framework to improve the geographic image annotation is presented. Deep correlations between high-level features from both shallow and deep modalities are used to achieve a final representation, which is its biggest bright spot.

The above research mainly focuses on the application of deep learning model in single image processing, and seldom involves the research of multi-image batch processing. But medical images have specific practical requirements, information richness and high clarity. In these documents, the requirements of multi-modal images for information and clarity have been repeatedly emphasized. Image fusion can increase the amount of information in single image. To solve this practical medical problem, we propose the method of completion in fusion. In order to effectively meet the needs of the above-mentioned medical images and make tentative research on the development of automatic diagnostic technology, supervised deep learning methods are used to achieve image fusion. In this paper, deep learning model is intended to be introduced into the field of image fusion. It is intended to develop a new idea of image fusion based on supervised deep learning. We can obtain a new model through the establishment of an image training data bases in successful fusion results, then can complete the fusion of batch multi-medical images using the training model. It is suitable for multi-modal image fusion to improve the efficiency and accuracy of image process.

2. Multi-modal medical image

MRI Resolution Imaging (Asif et al., 2018; Saadat et al., 2017) is magnetic resonance imaging. Magnetic resonance imaging is a type of tomography that uses magnetic resonance phenomena to obtain electromagnetic signals from the human body and can reconstruct human information. In 1946, Professor Felix Bloch of stanford university and professor Edward Purcell of harvard university independently discovered the phenomenon of nuclear magnetic resonance. Magnetic resonance imaging technology is based on this physical phenomenon. In 1972, Professor Paul Lauterbur developed a method of spatial coding of nuclear magnetic resonance signals, which can be used to reconstruct human images. The specific imaging principle is that the nucleus has positive electricity and the atomic nuclear energy of many elements spins. Usually the arrangement of the spin axis of the nucleus is irregular, however when it is placed in an external magnetic field, the orientation of the nuclear spin space transitions will be from disorder to order. In this way, the spin core is also rotated around the applied magnetic field vector in the angle of the spin axis and the vector direction of the applied magnetic field. When the magnetization vector of the spin system gradually increases from zero, the magnetization intensity reaches a stable value when the system reaches equilibrium. If the nuclear spin system is affected by the outside world at this time, the spin core will also rotate in the direction of the radio frequency. After the radio frequency pulse stops, the nuclear nucleus that has intensified the spin system can't maintain this state, and it will return to the original arrangement state in the magnetic field to release weak energy, which will become a radio signal and detect the signal. And so that it can be spatially resolved, the

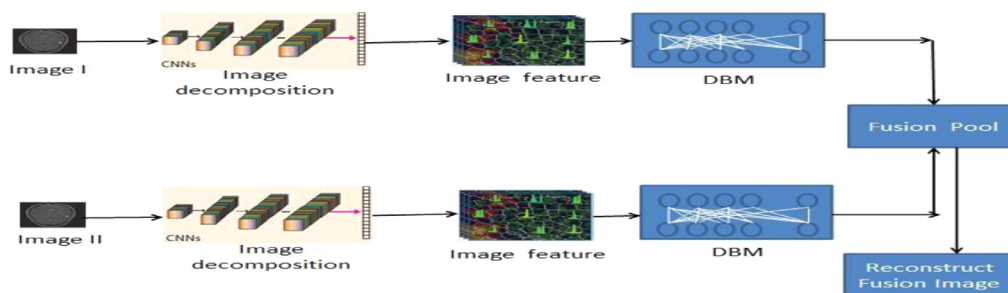


Fig. 2. Model of image fusion based on deep learning.

Table 1
Databases of learning images.

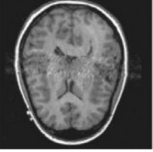
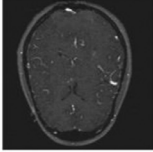
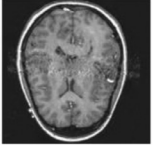
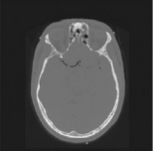
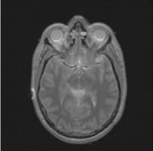
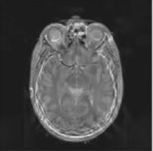
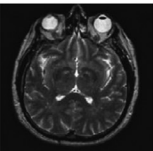
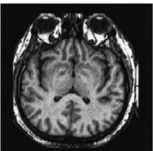
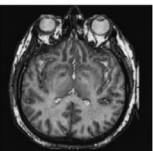
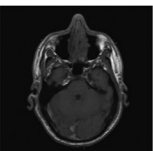
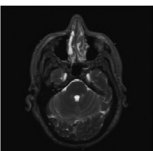
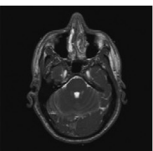
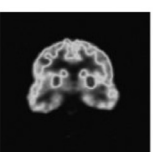
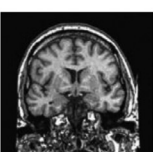
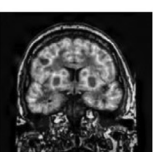
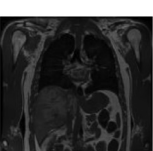
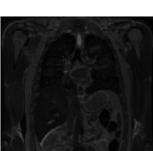
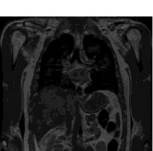



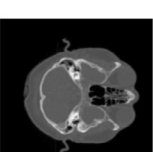
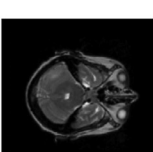
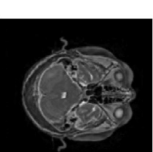
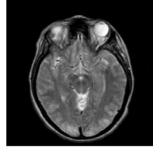
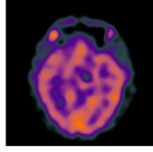
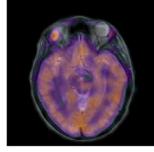
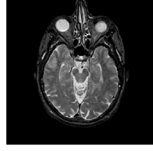

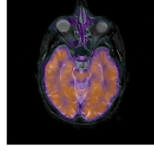
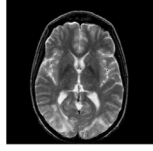
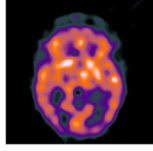
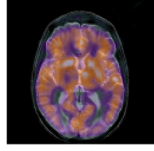
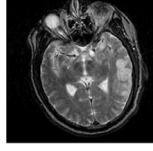
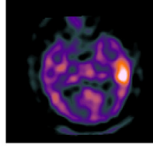
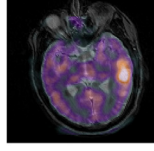
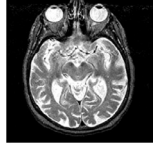

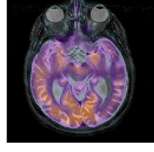
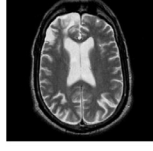
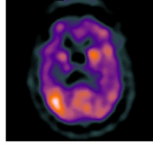
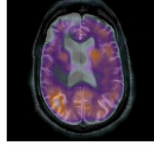
| input image 1 | input image 2 | target image |
|---|---|---|
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

image of the distribution of nucleus in motion can be obtained. We call the process of returning the nucleus from an intense state to a balanced arrangement as "Relaxation Process." MRI completes image information through this process.

2.1. Comparison of MRI images with CT images

Compared to imaging technology, CT (Yang et al., 2018) can all show the distribution of a certain physical quantity in space, but

Table 2
Databases of learning images.

| input image 1 | input image 2 | target image |
|--|---|---|
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

MRI magnetic resonance imaging can obtain a three-dimensional image of a fault in any direction or a four-dimensional image of the spectral distribution of space. CT can't achieve this. In addition, MRI imaging methods are more diverse and imaging principles are more complex, so the resulting image information will be more abundant. This is the biggest difference between CT images and MRI images.

Another large class of multi-mode image imaging techniques is Single-Photon Emission Computer Tomography(SPECT) and Positron Emission Tomography(PET). Both of these technologies are among the CT imaging techniques. Since they all complete the imaging of gamma rays emitted from the patient's body, they are collectively referred to as "Emission Computed Tomography"(ECT). The imaging principle of SPECT (Salvado et al., 2018) is that each sensitive point of the gamma camera probe detects gamma photons coming along a projection line, and its measurement value represents the sum of the radioactivity of the human body on the projection line. The sensitive points on the same line can detect the radioactive drugs of the human body on a fault. Their output is called the one-dimensional projection of the fault. Each projection line is perpendicular to the detector and parallel to each other. The intersection angle θ of the normal line of the detector with the X axis is called the observation angle. If you want to know the structure of the

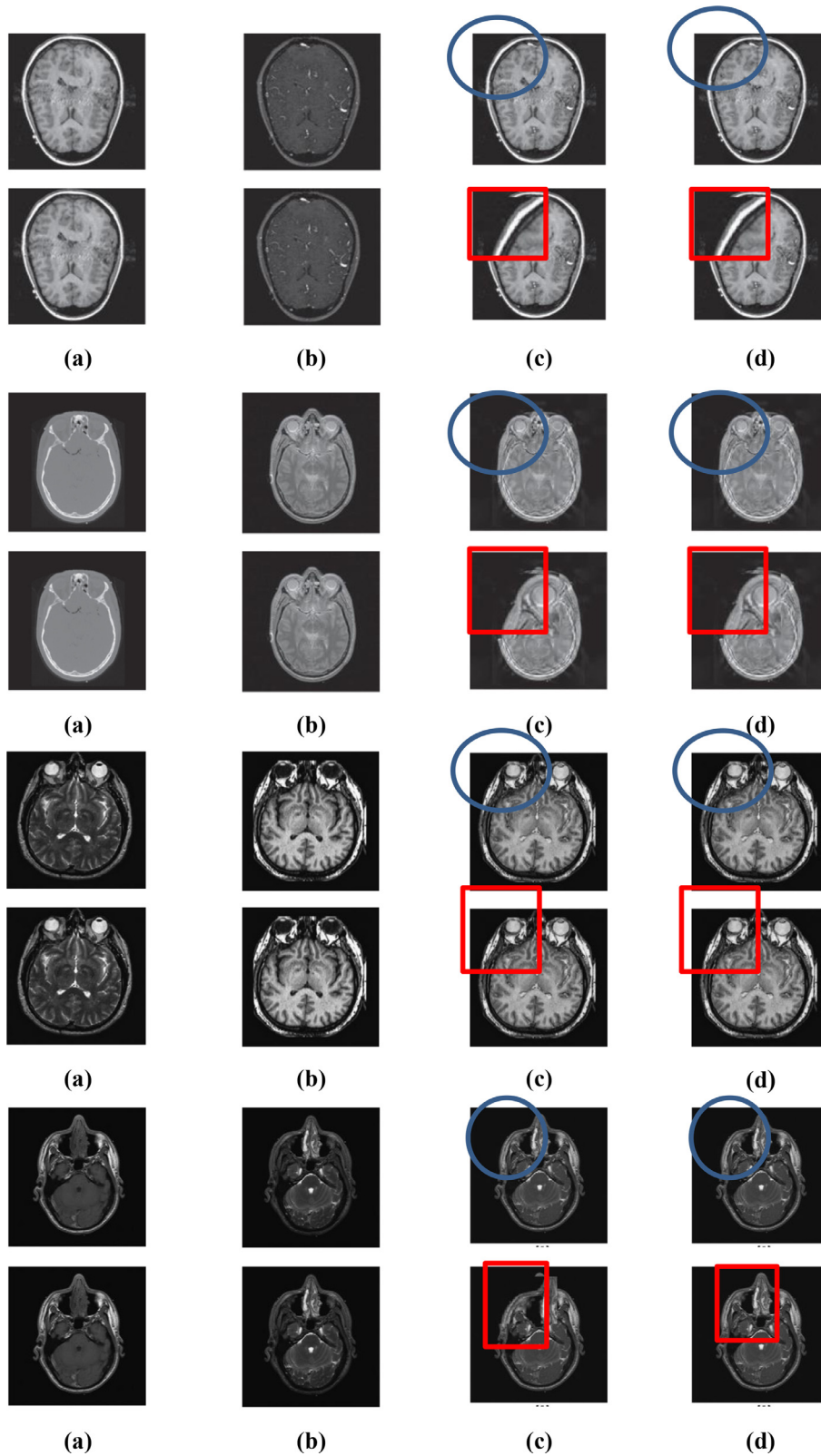


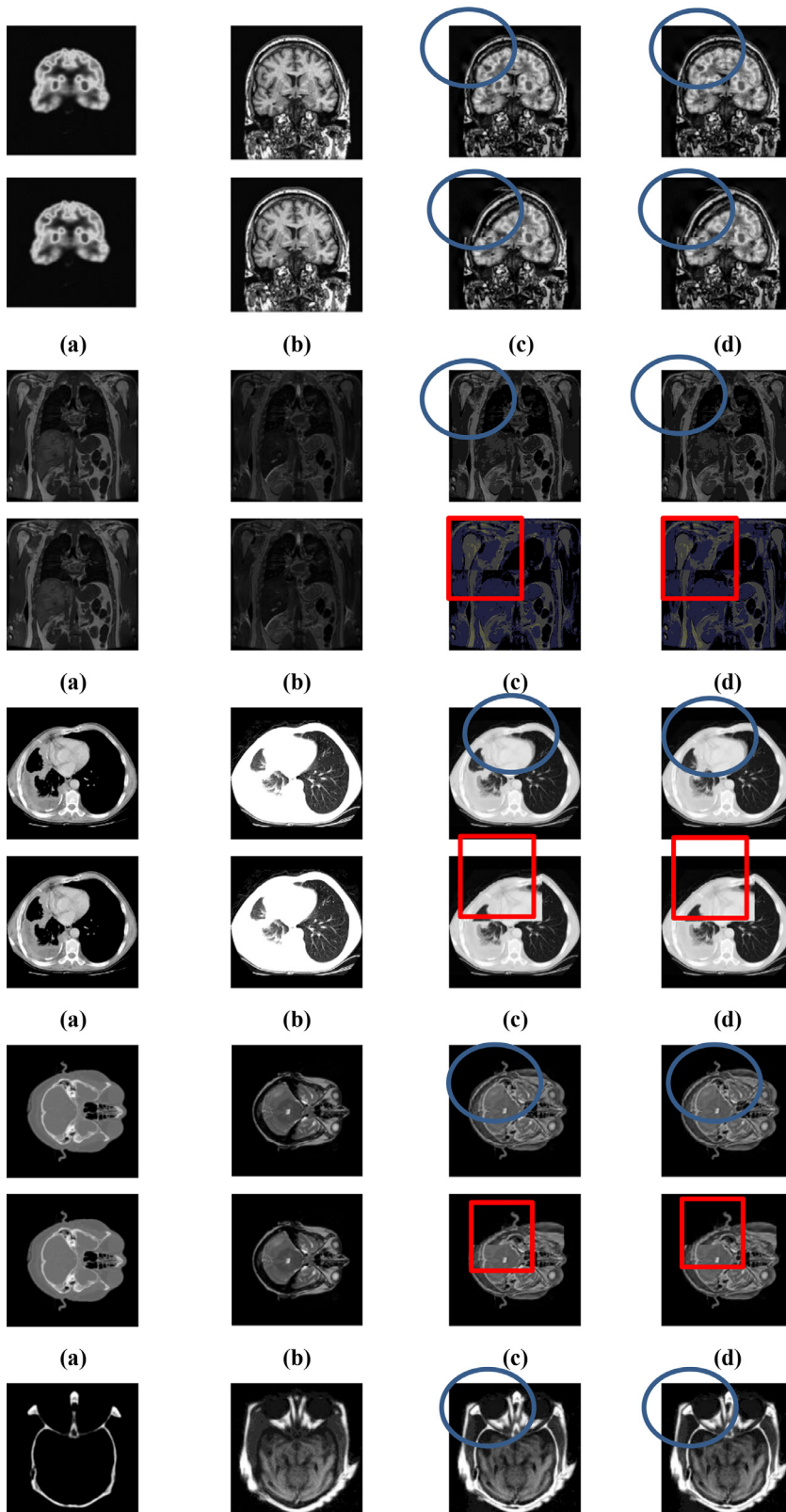
Fig. 3. Experiment results on images:(a) image-A;(b) image-B;(c) experiment results of image fusion;(d) SR-image fusion methods.

human body in the depth direction, you need to repeatedly observe from different angles. The current theory can prove that a one-dimensional projection of a fault at all observation angles is known and the image of the fault can be deduced. We refer to the process of reverse-solving fault images from the projection as "Reconstruction Process", also known as "Computed Tomography(CT)". Its main function is to obtain projection data and reconstruct fault images.

2.2. Comparison of SPECT images with CT images

SPECT images are obtained using radionuclides that emit single photons (Gamma Rays), while CT images are obtained using X-rays; SPECT images are emitted from the human body, while CT is transmitted from the human body. SPECT images belong to functional imaging, while CT images are anatomical imaging.

Fig. 3. Continued



2.3. Comparison of PET images with CT images

- (1) CT images provide a rich anatomical structure of the human body. Clinically, images of various anatomical structures can be observed through bone Windows, soft tissue Windows, and lung Windows.

However, PET only provides a single radioactive drug distribution image. PET images are relatively monotonous, simple and easy to analyze, but they do not have clinical diagnosis and positioning. CT images have become an effective tool for clinical diagnosis of diseases.

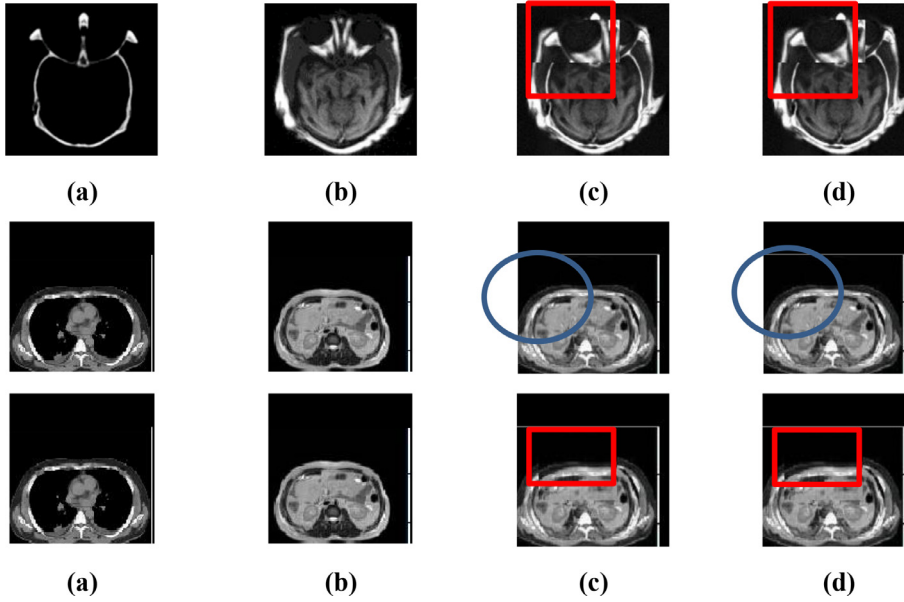


Fig. 3. Continued

- (2) CT image acquisition is relatively easy, the use of contrast agents is not limited by time. However, the use of positron radioactive drugs used by PET (Schramm et al., 2018) is limited by time in clinical use.
- (3) PET and CT combined equipment” PET-CT “is the direction of image equipment development. In particular, the observation of cell metabolism, enzyme receptors, and gene expression imaging in the case of high-resolution anatomical results can overcome the lack of simple CT or PET images.

3. Deep learning

Deep learning is a hot topic in recent years. Many scholars have focused on their research in several common models such as AE, RBM, and DBM. Next, this article summarizes these three models as follows.

The auto-encoders(AE) (Chen & Konukoglu, 2018; Li et al., 2018) model is a kind of feedforward neural network model. Similar to other feedforward neural network models, it is divided into two stages: encoding and decoding. The encoding stage process converts the input signal X to H through a non-linear map.

$$H = \Psi(Wx + b) \quad (1)$$

Eq. (1) represents a nonlinear function, which can usually be taken as softmax,relu,Tanh, sigmoid, etc.. The decoding process is exactly the opposite of the encoding process. The process of introducing input X is known in H , and the equation is shown in (2).

$$Z = \Psi(W'x + b') \quad (2)$$

The model involves a total of four parameters. The parameter matrix is expressed in terms of θ . It can be expressed as $\theta = [W, b, W', b']$ in specific. The average reconstruction error usually used based on the data set N is the square error method. The optimization problem can be attributed to the following l2 normal form:

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^N ||x_i - f_{\theta}(x_i)||_2^2 \quad (3)$$

$$z = f_{\theta}(x) \quad (4)$$

Eq. (3) here represents the error between the decoding reconstruction signal Z and the original input signal X , and the value θ of the minimum parameter matrix of the error is found by the l2 normal form.

Relative to AE, the more complex network model is Restricted Boltzmann Machines (RBM) (Argal et al., 2018; Geng et al., 2018), which is

a two-layer neural network form, data units are divides into two types: visual units and hidden units. There is a symmetric association between these two types of nodes. In addition, there is no connection between the same type of node. The energy function of the model can be expressed as Eqs. (5)–(8).

$$E(v, h; \theta) = - \sum_{i=1}^I \sum_{j=1}^J w_{ij} v_i h_j - \sum_{i=1}^I b_i v_i - \sum_{j=1}^J a_j h_j \quad (5)$$

$$p(v, h; \theta) = \frac{\exp(-E(v, h; \theta))}{Z} \quad (6)$$

$$p(h_j = 1 | v; \theta) = \delta \left(\sum_{i=1}^I w_{ij} v_i + a_j \right) \quad (7)$$

$$p(v_j = 1 | v; \theta) = \delta \left(\sum_{j=1}^J w_{ij} v_j + b_i \right) \quad (8)$$

Here, the link weight w_{ij} is a visual unit with a total of I , which linked to a hidden unit with a total of J . And the a_j and b_i represent the offset parameters in visual unit and the hidden unit, respectively. The energy calculation of each unit link is completed by Eq. (6). Here $Z = \sum_{h,v} \exp(-E(v, h; \theta))$ is the normalized factor function. The equation for calculating the probability of hidden units and visual units is Eqs. (7) and (8). This δ is defined as a logical function, usually in exponential expression. The learning of the parameter matrix W can be achieved by contrastive divergence(CD).

Based on the further development of the RBM neural network, the Deep Boltzmann machine (DBM) network was formed. DBM can be implemented by superimposing a multi-layer RBM network. The output of the L -level hidden unit in the network is the input of the $L + 1$ layer visual unit. It can be studied with supervision in the greedy layer. As shown in Fig. 1, after the pre-training, the parameters of this in-depth learning architecture can be further fine-tuned for logarithmic representation in DBM or for the calibration of training data labels by adding a "softmax" layer in the top layer.

The traditional DBM consists of multiple RBM layers. A typical neural network type is shown in Fig. 1. These networks are "limited" to a visual layer and a hidden layer, with connections between layers, but no connections between units within the layer. Hidden layer units are trained to capture the correlation of higher-order data expressed in the visual layer.

Table 3
Comparison of results in CT and MRI images.

| Indices Test Images No. | Indices EOG | RMSE | PSNR | Entropy | SF | REL | SSIM | H | MI | MEAN | STD | GRAD | TIME (s) |
|-------------------------------|----------------|---------|---------|---------|---------|--------|--------|--------|--------|----------|---------|--------|----------|
| Test-1 | 0.0382 | 41.0644 | 15.8615 | 1.9226 | 15.2158 | 0.7618 | 0.9950 | 6.7339 | 1.4421 | 96.6570 | 40.9483 | 5.1724 | 1.9843 |
| Test-2 | 0.0392 | 23.6111 | 20.6685 | 1.4407 | 10.4070 | 0.9126 | 0.9870 | 5.9261 | 1.6457 | 87.5130 | 51.2122 | 3.3713 | 1.9081 |
| Test-3 | 0.0229 | 44.9188 | 15.0822 | 1.1018 | 14.4439 | 0.7732 | 0.9906 | 7.1931 | 1.8145 | 84.1888 | 53.5341 | 6.4110 | 1.8060 |
| Test-4 | 0.0420 | 21.7137 | 21.3961 | 0.8414 | 11.8741 | 0.8639 | 0.8510 | 5.7112 | 1.4747 | 40.1885 | 36.4815 | 3.6244 | 1.9074 |
| Test-5 | 0.0509 | 42.0040 | 15.6650 | 1.1411 | 17.6166 | 0.8149 | 0.9062 | 7.4402 | 1.6701 | 81.4598 | 58.0940 | 0.6123 | 1.9013 |
| Test-6 | 0.0190 | 17.4056 | 17.0547 | 0.5633 | 10.5152 | 0.8978 | 0.9725 | 6.1563 | 1.3734 | 39.9128 | 18.9756 | 5.1018 | 1.7789 |
| Test-7 | 0.0452 | 38.2454 | 16.4792 | 1.9966 | 21.2161 | 0.9749 | 0.6002 | 7.4146 | 1.9814 | 103.7848 | 84.9175 | 7.2501 | 1.9624 |
| Test-8 | 0.0389 | 35.3230 | 17.1697 | 1.3424 | 14.0835 | 0.8400 | 0.9003 | 6.2220 | 1.4598 | 46.9382 | 52.6670 | 3.8931 | 1.8790 |
| Test-9 | 0.0195 | 36.6529 | 16.8486 | 2.3213 | 10.0064 | 0.7714 | 0.9987 | 6.8299 | 1.6847 | 50.5727 | 43.4916 | 3.9461 | 1.7992 |
| Test-10 | 0.0483 | 39.9292 | 16.1050 | 0.5429 | 15.3944 | 0.8343 | 0.9020 | 6.2898 | 1.7874 | 58.7925 | 62.1224 | 4.9946 | 1.9038 |

4. Frames and algorithms of the proposed method

The proposed framework is shown in Fig. 2. It consists of two major steps: model learning, fusion test. In first step, the parameters in the DBM model are learned by training the multiple groups of images in the train datasets. Registration processing and pixel alignment in these train images have achieved in advance. In second fusion test process the multiple groups of test images are entered into the model that learn and train have achieved, then the fusion process is completed. Next, the final synthesis obtains the fused image.

The new method proposed in this paper can realize the fusion task of multi-modal medical images. The specific implementation process is listed as follows:

The proposed method:

In this paper, noise removal, registration, standardization and other pre-processing work will be carried out for a large number of images in datasets. These images will be divided into training data sets and testing data sets in the same level according to the standards of depth learning model.

The image block size is determined on the training data set and the test data set image respectively. On the two data sets, we use the same block size to complete the model calculation. The size is determined by the standard and the result of the main reference image fusion. If high precision is needed, it must be selected smaller. If higher calculation speed is needed, it can be selected larger. In order to better meet the needs of medical diagnosis, different sizes of calculation model can be used alternately.

After the training data collection is completed, a fusion calculation model is generated and related parameters are further optimized. Parameter optimization can be achieved by using automatic optimization.

The final step in two separate cases is also a key step in the process of model testing. First, the two sets of multi-modal images were entered into the fusion model to complete the fusion results. Next, batch processing in multiple group images are achieved. The fusion results were obtained by refactoring and then these results were output.

The experimental data analysis was conducted and the fusion time was counted.

5. Experimental and analysis

Extensive experiments are conducted on different multi-medical image pairs in this part, e.g., CT, MRI and SPECT. The evaluation metrics used in this paper are EOG, RMSE, PSNR, ENT, SF, REL, SSIM, H, MI, MEAN, STD, GRAD, Q_0 , Q_E , Q_W , $Q^{AB/F}$ and Time(s) (Liang et al., 2016; Ma et al., 2015; Piella & Heijmans, 2003; Wang & Bovik, 2002). The ranges of Q_0 , Q_E , Q_W and $Q^{AB/F}$ are in [0,1], and the larger the value, the better the fused result. To reduce variation, each experiment is repeated 50 times and their mean values are recorded.

5.1. Databases of learning

Among the extensive multi-modal medical images, the classic images can be divided into two categories: MRI images and CT images. MRI images are more accurate, and its information is more abundant and accurate, especially for human tissue structure and details. CT images provide rich anatomical structure images of the human body. Clinically,

images of various anatomical structures can be observed through bone Windows, soft tissue Windows, and lung Windows, and the details of organs can be reflected in detail from a certain angle. SPECT images are one of the typical image types in CT imaging technology. Therefore, the construction of the data sets in the experiment we composite with more classical MRI, CT and SPECT image in the multi-mode image. The images were all derived from the standard medical image database and were registered before the experiment. Typical part sets of images in image databases are selected to display in this paper, as shown in Tables 1 and 2. The first image in Tables 1 and 2 is an MRI image, the second image is SPECT image, and the third image is a target image.

- Typical part sets of images of CT and MRI images databases
- Typical part sets of images of MRI and SPECT images databases

6. Experimental results on CT and MRI images

We show the typical two sets of experimental results in the paper, as shown in Fig. 3 and Table 3. In the experiment, we also carried out this method and compared with typical image super-resolution ‘Yin’ method (SR) in literature (Yin et al., 2013). It shows that as long as we ensure the image quality in the training data set, the clarity of our method to a certain extent can be guaranteed. It shows that there are many similar elements in the image to be fused. How to make full use of them and avoid duplication of information fusion and lack of key information will greatly improve the fusion effect.

6.1. Analysis of experimental results

6.1.1. Visual quality of the fused images

First of all, the visual quality of the fused images obtained by the proposed method is better than the other methods. The fused images obtained by our method look more natural. They produce sharper edges and higher resolution. In addition, the detailed information and interested features are better preserved to some extent.

In particular, from the area circled by the red calibration frame, it can be seen that the fusion results are clear, the edges are clear, the information is clearly contrasted, the contrast is obvious, and the key information in the image can be reflected, and the virtual shadow is effectively removed. Moreover, it is clear that the information contained in the fusion image already covers most of the information in the two multi-modal images of CT and MRI, MRI and SPECT, it can be effectively supplementing the deficiencies of the single MRI /CT/SPECT image information. The increase in the amount of information brings changes to the improvement of medical imaging diagnosis undoubtedly. More valuable information can be used to support effective diagnosis. It also brings possibilities for the study of “automatic diagnosis” technology and conducts tentative research. The result of test 1–10 shown in Fig. 3 is the key main similarity between the fusion result and the target image. It can be reflected that the result of the method fusion in this paper has

covered the main key information of the target image in a more comprehensive way. This can explain the validity of this method furtherly.

6.1.2. Analysis of evaluation data

For the first experiment, the proposed method achieves excellent results when using evaluation metrics EOG, RMSE, PSNR, ENT, SF, REL, SSIM, H, MI, MEAN, STD, GRAD, Q_0 , Q_E , Q_W and $Q^{AB/F}$. The results show that our methods are effective in image fusion.

Specifically, the result in test 1–8 of CT and MRI is somewhat better than SR with respect to Q_W .

For the test 2–10 experiment result of MRI and SPECT, the proposed method yields outstanding results in terms of Q_W , EOG, RESE, PSNR, REL.

Hence, our method can realize the image fusion task and capture the details of the image compared to other fusion methods.

The results also verify that the process of learning, testing, fusion in this method not only can introduce fusion effectively, but also can make the important information and details in the images integrated by the depth learning model prominently. Fusion result can be covered in key information in two types of medical images, and it can obtain more informative and complete medical image information. However, it is slightly inferior in terms of REL and STD, indicating that the fusion process will cause a loss of information, how to learn model parameters, increase the size of the training data and improve the degree of training accuracy are our issues for further study.

7. Conclusions

The application of Deep Learning techniques to Multi-modal medical image has proposed in this paper. This paper reviews the recent advances achieved in DL image fusion and puts forward some prospects for future study in the field. The primary contributions of this work can be summarized as the following three points.

- 1 Deep learning models can extract the most effective features automatically from data to overcome the difficulty of manual design.
- 2 The method in this paper can achieve the multi-modal medical image fusion. This method can effectively achieve the batch operation of image fusion, meet the actual needs of medical diagnosis, and greatly improve the efficiency of medical image fusion. It has a good meaning in improving the accuracy of medical diagnosis.
- 3 Experimental results indicate that the proposed method achieves state-in-art performance in terms of both visual quality and quantitative evaluation metrics.
- 4 In conclusion, the recent research achieved in DL image fusion exhibits a promising trend in the field of image fusion with a huge potential for future improvement. It is highly expected that more related researches would continue in the coming years to promote the development of image fusion.

Declaration of Competing Interest

None.

Acknowledgments

The authors first sincerely thank the editors and anonymous reviewers for their constructive comments and suggestions, which are of great value to us. The authors would also like to thank Prof. Xiaojun Wu from Jiangnan University.

This work is supported by the National Natural Science Foundation of China (Grants 61702293, No. 61902203) and the Shandong Provincial Natural Science Foundation of China (Grants ZR2017QF015) and Key Research and Development Plan-Major Science and Technological Innovation Projects of Shandong Province (2019JZZY020101).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijcce.2020.12.004.

References

- Ahmad, M., Yang, J., Ai, D., et al. (2017). Deep-stacked auto encoder for liver segmentation. In *Proceedings of the Chinese conference on image and graphics technologies* (pp. 243–251). Singapore: Springer.
- Argal, A., Gupta, S., Modi, A., et al. (2018). Intelligent travel chatbot for predictive recommendation in echo platform. In *Proceedings of the IEEE 8th annual computing and communication workshop and conference (CCWC)* (pp. 176–183). IEEE.
- Asif, U., Bennamoun, M., & Soheli, F. A. (2018). A multi-modal, discriminative and spatially invariant CNN for RGB-D object labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(9), 2051–2065.
- Chen, H., Jiao, L., Liang, M., et al. (2019). Fast unsupervised deep fusion network for change detection of multitemporal SAR images. *Neurocomputing*, 332, 56–70.
- Chen, X., & Konukoglu, E. (2018). Unsupervised detection of lesions in brain MRI using constrained adversarial auto-encoders. arXiv:1806.04972.
- Cheng, G., Yang, C., Yao, X., et al. (2018). When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs. *IEEE Transactions on Geoscience and Remote Sensing*, 56(5), 2811–2821.
- Farid M, S., Mahmood, A., & Al-Maadeed S, A. (2019). Multi-focus image fusion using content adaptive blurring. *Information Fusion*, 45, 96–112.
- Geng, Z., Li, Z., & Han, Y. (2018). A new deep belief network based on RBM with glial chains. *Information Sciences*, 463, 294–306.
- Hou, Y., Li, Z., Wang, P., et al. (2018). Skeleton optical spectra-based action recognition using convolutional neural networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(3), 807–811.
- Ijjina, E. P. (2016). Classification of human actions using pose-based features and stacked auto encoder. *Pattern Recognition Letters*, 83, 268–277.
- Jiao, R., Huang, X., Ma, X., et al. (2018). A model combining stacked auto encoder and back propagation algorithm for short-term wind power forecasting. *IEEE Access: Practical Innovations, Open Solutions*, 6, 17851–17858.
- Li, F., Qiao, H., & Zhang, B. (2018). Discriminatively boosted image clustering with fully convolutional auto-encoders. *Pattern Recognition*, 83, 161–173.
- Liang, R. Z., Shi, L., Wang, H., et al. (2016). Optimizing top precision performance measure of content-based image retrieval by learning similarity function. In *Proceedings of the 23rd international conference on pattern recognition (ICPR)* (pp. 2954–2958). IEEE.
- Liu, Y., Chen, X., Peng, H., et al. (2017). Multi-focus image fusion with a deep convolutional neural network. *Information Fusion*, 36, 191–207.
- Liu, Y., Chen, X., Wang, Z., et al. (2018). Deep learning for pixel-level image fusion: Recent advances and future prospects. *Information Fusion*, 42, 158–173.
- Liu, Y., Chen, X., Ward, R. K., et al. (2016). Image fusion with convolutional sparse representation. *IEEE Signal Processing Letters*, 23(12), 1882–1886.
- Liu, Y., Chen, X., Ward, R. K., et al. (2019). Medical image fusion via convolutional sparsity based morphological component analysis. *IEEE Signal Processing Letters*.
- Liu, Y., Chen, X., Ward, R. K., et al. (2019). Medical image fusion via convolutional sparsity based morphological component analysis. *IEEE Signal Processing Letters*.
- Ma, J., Yu, W., Liang, P., et al. (2019). FusionGAN: A generative adversarial network for infrared and visible image fusion. *Information Fusion*, 48, 11–26.
- Ma, K., Zeng, K., & Wang, Z. (2015). Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 24(11), 3345–3356.
- Ma, S., Chen, M., Wu, J., et al. (2018). High-voltage circuit breaker fault diagnosis using a hybrid feature transformation approach based on random forest and stacked auto-encoder. *IEEE Transactions on Industrial Electronics*.
- Ouerghi, H., Mourali, O., & Zagrouba, E. (2017). Multimodal medical image fusion using modified PCNN based on linking strength estimation by MSVD transform.
- Pan, Z. W., & Shen, H. L. (2019). Multispectral Image super-resolution via RGB image fusion and radiometric calibration. *IEEE Transactions on Image Processing*, 28(4), 1783–1797.
- Peter, J. D., Fernandes, S. L., Thomaz, C. E., et al. (2019). *Computer aided intervention and diagnostics in clinical and medical images*. Springer.
- Piella, G., & Heijmans, H. (2003). A new quality metric for image fusion. In *Proceedings of the IEEE international conference on image processing*: 3 (pp. 173–176).
- Saadat, S., Pickering, M. R., Perriman, D., et al. (2017). Fast and robust multi-modal image registration for 3d knee kinematics. In *Proceedings of the international conference on digital image computing: techniques and applications (DICTA)* (pp. 1–5). IEEE.
- Salvado, D., Erlandsson, K., Occhipinti, M., et al. (2018). Development of a practical calibration procedure for a clinical SPECT/MRI system using a single INSERT prototype detector and multi-mini slit-slat collimator. *IEEE Transactions on Radiation and Plasma Medical Sciences*.
- Sandhya, S., Kumar, M. S., & Karthikeyan, L. (2019). A hybrid fusion of multimodal medical images for the enhancement of visual quality in medical diagnosis. In *Proceedings of the computer aided intervention and diagnostics in clinical and medical images* (pp. 61–70). Cham: Springer.
- Schlemper, J., Caballero, J., Hajnal, J. V., et al. (2018). A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE transactions on Medical Imaging*, 37(2), 491–503.
- Schramm, G., Holler, M., Rezaei, A., et al. (2018). Evaluation of parallel level sets and Bowsher's method as segmentation-free anatomical priors for time-of-flight PET reconstruction. *IEEE Transactions on Medical Imaging*, 37(2), 590–603.

- Shao, Z., & Cai, J. (2018). Remote sensing image fusion with deep convolutional neural network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(5), 1656–1669.
- Sun, G., Huang, H., Zhang, A., et al. (2019). Fusion of multiscale convolutional neural networks for building extraction in very high-resolution images. *Remote Sensing*, 11(3), 227.
- Thirukovalluru, R., Dixit, S., Sevakula, R. K., et al. (2016). Generating feature sets for fault diagnosis using denoising stacked auto-encoder. In *Proceedings of the IEEE international conference on prognostics and health management (ICPHM)* (pp. 1–7). IEEE.
- Wang, Z., & Bovik, A. C. (2002). A universal image quality index. *IEEE Signal Processing Letters*, 9(3), 81–84.
- Wohlberg, B. (2016). Efficient algorithms for convolutional sparse representations. *IEEE Transactions on Image Processing*, 25(1), 301–315.
- Yang, Q., Yan, P., Zhang, Y., et al. (2018). Low dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Transactions on Medical Imaging*.
- Yang, Y., Nie, Z., Huang, S., et al. (2019). Multi-level features convolutional neural network for multi-focus image fusion. *IEEE Transactions on Computational Imaging*.
- Ye, D., Fuh, J. Y. H., Zhang, Y., et al. (2018). In situ monitoring of selective laser melting using plume and spatter signatures by deep belief networks. *ISA Transactions*, 81, 96–104.
- Yin, H., Li, S., & Fang, L. (2013). Simultaneous image fusion and super-resolution using sparse representation. *Information Fusion*, 14(3), 229–240.