Multimodal Brain Image Fusion Algorithm Based on Multiscale Contextual Inference

Jiancong Fan , Jianjun Liu , Miaoxin Guo , Yang Li *

Posted Date: 1 January 2025

doi: 10.20944/preprints202412.2638.v1

Keywords: medical image fusion; multi-scale contextual reasoning; feature expression; loss function

Article

Multimodal Brain Image Fusion Algorithm Based on Multiscale Contextual Inference

Jiancong Fan [1,2], Jianjun Liu [1,2], Miaoxin Guo [1,2] and Yang Li [1,2,*]

[1] School of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong 266590, China
[2] Provincial Key Laboratory for Information Technology of Wisdom Mining of Shandong Province, Shandong University of Science and Technology, Qingdao, China
[*] Correspondence: deryang@163.com

Abstract: Data from magnetic resonance imaging (MRI) and positron emission tomography (PET) scans can effectively assist physicians in diagnosing and treating brain tumors. However, images from different modalities have their advantages and limitations. Multimodal medical image fusion is the process of extracting and merging the information of every single modality medical image and retaining the characteristic information of each modality to the maximum extent. Therefore, this paper proposes a medical image fusion method based on multiscale contextual reasoning to address the problems of the scattered size distribution of pathological regions, inconspicuous detail features, and extensive visual differences between similar tissue images. Firstly, the original image is decomposed by the method to get the global part and the local part. Secondly, the multi-scale feature extraction network (MSFE-Net) mines the different regions between multi-level features and improves the network's ability to extract pathological features at different scales. Meanwhile, the attention module is introduced to perform channel-weighted summation of the network feature maps to improve the feature expression ability of key channels so that the network can accurately capture the pathological feature regions. Thirdly, in the loss function design, multiple losses are used further to optimize the distribution of the sample feature space. This paper conducted experiments using clinical images from computed tomography/magnetic resonance/ of the brain. The experimental results show that the medical image fusion method based on multi-scale contextual inference works better than other advanced fusion methods.

Keywords: medical image fusion; multi-scale contextual reasoning; feature expression; loss function

1. Introduction

Medical imaging has become an integral part of modern medicine. It is used throughout clinical work, not only for diagnosing disease but also for planning, implementing, and evaluating the efficacy of surgical procedures and radiotherapy. In recent years, medical technology has advanced with the development of computer technology, and medical imaging has become increasingly important in various clinical applications. Because of the different imaging mechanisms of various medical imaging devices, different modalities of medical images reflect detailed information about other tissues and organs in the human body [1]. The massive amount of data and the variety of categorization criteria pose a tremendous challenge for the effective organization and management of medical images. With this comes an extreme shortage of specialized imaging physicians.

As these individual imaging modalities provide distinct information about the human body, analyzing only a single imaging modality has incomplete information about the state of the patient, which affects the judgment of the clinicians. In the absence of medical resources and highly skilled personnel support, making a diagnosis only on the information reflected by single-modality images is not only prone to lack of security but also can lead to medical errors. Therefore, medical image fusion technology is widely used to combine multiple pre-registered medical images obtained from several different modalities into a single image. Through the application of medical image fusion, the complementarity between different image data is fully utilized to integrate the anatomical structure information and physiological and biochemical information in medical images to provide clinicians with more comprehensive and accurate medical information [2,3].

In this article, an example is used to describe it in detail. As shown in Figure 1, the brain image shows a large mass with edema. The mass and edema pressure force the midline to move, and the adjacent midbrain structures are compressed. On proton density (PD) and magnetic resonance (MR) T2-weighted (T2) images, a large area of the left temporal region shows high signal intensity. On enhanced images, the lesion contained a cystic component. The narrow cerebral sulcus in the left cerebral hemisphere suggested severe swelling of the left cerebral hemisphere. In addition, the blood volume in the lesion area was shallow, according to positron emission tomography (PET) imaging. It can be seen that the fused image can narrow the diagnosis and eliminate the interference information better than the unfused source image.

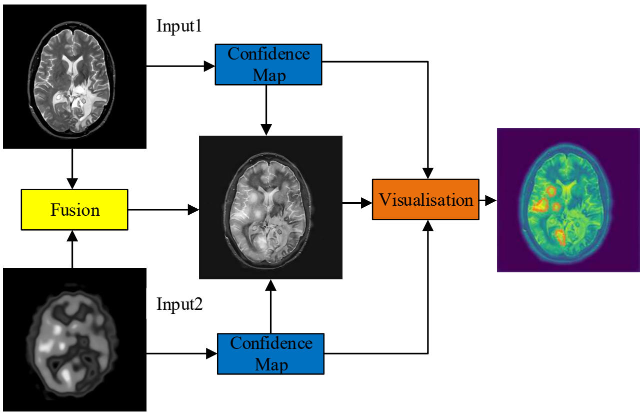


Figure 1. Multimodal fusion and visualization process of brain tumors.

Image fusion is broadly classified into two categories: traditional image fusion methods [4,5] and deep learning-based image fusion methods [6,7]. However, existing fusion technologies have many shortcomings. One of the main drawbacks of pixel-level medical image fusion is that clinicians are uncertain about the pixel-level contribution of the input image sources to the fused image. This can cause misinterpretation of the fused image provided to the clinicians. In addition, traditional feature extraction methods rely mainly on manual extraction. This requires specialized domain knowledge and a complex parameter tuning process. Moreover, each method is specific to a particular application scenario, and its generalization ability and robustness are poor.

This paper proposes a medical image fusion method based on multi-scale contextual inference to solve the above problems. The method designs a novel feature extraction network to extract the input image's local and global features. Then an attention mechanism is introduced to improve the attention of the network to critical regions and suppress the interference of outside areas to the fusion results. Finally, the design of the loss function combines the advantages of cross-entropy loss and central loss to effectively mitigate the errors generated during the fusion process. The contributions of our study are as follows.

a. A medical image fusion network model based on multiscale contextual inference is proposed to preserve the structural information of the images. The model uses large-sized extraction frames to extract global structural information for filtered images and small-sized extraction frames to extract local structural information for detail layer images.

b. For feature maps of different scales, this paper proposes a self-attentive module to further filter different channel features in the feature maps to improve the feature representation of critical channels and further guide the network to focus its attention on the regions containing essential information.

c. Because most medical images have high background similarity, the data of the same class will present significant visual differences due to different acquisition objects, which leads to the mixing of sample features between other classes. Based on the fact that the distance in the feature space is enlarged due to the large visual differences between data of the same class, a new loss function is designed by combining the advantages of cross-entropy loss and central loss to deal with this problem.

2. Related Work

With the development of machine learning technology, image processing techniques have been widely used, among which image fusion techniques are now commonly used in remote sensing, computer vision, and medical fields. Medical image fusion is to integrate the useful human body information contained in multiple source images obtained from different imaging modalities into a single image by some technical means. Image fusion technology can realize the organic combination of multimodal medical image information, enrich helpful details to assist disease diagnosis, reduce the randomness and redundancy of information, and at the same time improve the efficiency of diagnosis of specific complex diseases. The existing fusion methods include traditional fusion methods and deep learning-based fusion methods.

2.1. Space domain-based medical image fusion method

Space domain-based methods usually take image pixels, blocks, or image regions as units and design algorithms to process and compute them directly. They keep the units with high weights in the fused image. Among them, the Intensity-Hue-Saturation (IHS) method is commonly used to decompose functional images to achieve the fusion of functional and anatomical images [8,9]. Daneshvar et al. combined the IHS method with a retinal model to improve the content of spatial and functional information of the fused images [10]. Chen combined the IHS method with the logarithmic Gabor wavelet transform to obtain the intensity components of PET images with appropriate scale decomposition to achieve the fusion of PET images with MRI images [11]. Similarly, Gillespie et al. found that color images with high contrast could also be generated by the Brovey transform (BT) [12]. In addition to the spatial transformation approach, some dimensionality reduction techniques are widely used for image feature processing. For example, Nandi et al. discussed the application of the commonly used Principal Component Analysis (PCA) in the field of medical image processing [13], which can maximize the spatial resolution of fused images [14] and reduce redundant information [15,16]. He et al. experimentally found that combining IHS and PCA could preserve more spatial features and desired functional information of the image without distortion [17]. Benjamin et al. designed a cascaded PCA framework to enrich the information of the fused image, keep the edges of the image, and improve the quality of the fused image. Unlike the principle of the PCA method to make the data uncorrelated, Independent Component Analysis (ICA) adopted higher-order statistics to discover features in the observed data [18]. Cui et al. used the ICA technique to feature the image components after wavelet decomposition to improve the effectiveness of the fusion algorithm [19].

In addition to the traditional methods of image feature representation, image decomposition methods can effectively distinguish different classes of image features to enhance the quality of fused images. The proposal and use of filters in the null domain have proven as a new research direction in medical image processing. The Bilateral Filter (BF) proposed by Tomasi et al. is a nonlinear method that performs a smoothing operation on the image while preserving the image edge information [20]. Studies have shown that various improved BF methods can fuse multimodal medical images [21,22] effectively. Li et al. proposed an efficient and fast Guided Filter (GF)-based weighted averaging method that fully uses spatial coherence to fuse the base and detail layers of the decomposed image separately and achieved excellent fusion results [23]. On this basis, Li et al. used a spectral residual algorithm and a graph-based visual saliency model to extract saliency features in the smoothed and detailed layers at different scales of the source image after GF decomposition and construct fusion rules [24]. Jian et al. proposed a GF-based rolling-guided filtering method to achieve multiscale decomposition of the source image and use it in conjunction with bilateral filtering, which preserves the detailed information of the source image and can effectively suppress artifacts [25]. Zhao et al. used a multi-scale alternating order filter to extract useful features from the input medical images and constructed a recursive filter-based weight decision map to guide the fusion of salient information from the source images [26]. Biswas et al. applied the Wiener filter to the transform domain to achieve spine medical image fusion [27]. Liu et al. proposed a filtering method based on gradient minimum smoothing, which enabled the fused MRI-CT images to retain more detailed information about the source images [28]. Jiang et al. also achieved a similar result using a weighted least squares filtering method [29]. The above methods have better feature representation and texture information recognition capability, and the obtained fused images have precise details.

Algorithms designed based on image null domain features can work as fusion rules in the fusion framework, such as the commonly used weighted average method [30] and the maximum/minimum selection method [31]. Obviously, the null domain-based algorithm [32] is relatively simple to implement and has high computational efficiency. However, it does not perform well enough in preserving the luminance and contrast information of the source image. The fused image may suffer from a lack of helpful information loss or color distortion.

2.2. Medical image fusion method based on change domain

Transform domain-based image fusion methods can effectively compensate for the possible shortcomings of the null domain methods. Such methods usually consist of three steps: first, the source image is decomposed into a series of high-frequency subbands and a low-frequency subband using one or more transform tools; then, different fusion schemes are designed according to the characteristics of varying frequency subbands; finally, the fused image is obtained by performing the inverse transform operations corresponding to the transform tools in step 1. Among them, Pajares et al. pointed out that the Discrete Wavelet Transform (DWT) is a multi-scale (multi-resolution) method that enables the decomposition of images under different coefficients and is suitable for multi-class image fusion tasks [33]. On this basis, Wang et al. investigated the application of the wavelet transform (WT) fusion algorithm in CT and MRI image fusion and the selection of wavelet function. In this paper, authors demonstrated the effectiveness of the wavelet transform-based fusion scheme through experiments [34]. Cheng et al. obtained fused PET-CT images by wavelet transform method, which was proven as an effective way to quickly detect and accurately localize the diseased tissue in the fused image. [35]. Yang et al. combined the characteristics of the human eye visual system and the physical significance of wavelet coefficients to propose a new wavelet-based medical image fusion algorithm that reduces the image's noise while ensuring the fused image's homogeneity [36]. Vijayarajan et al. used the DWT method to decompose the source image into multi-scale inputs. In this paper, authors performed the multi-scale evaluation of their principal components and used their the average of which was used as the weight of the fusion rule to achieve the fusion of CT-MRI and MRI images [37]. To further enhance the subjective perception of the fused images, Bhavana et al. added an image preprocessing step to the conventional image fusion framework to improve the quality of the input images and then fused the enhanced images using the DWT method. Experimental results showed that this method substantially improved the gradient information of the fused images and reduced the spectral differences without reducing color distortion and losing any anatomical information [38]. Although the DWT method can realize the multi-scale representation and better extract the feature information of the image, this method has fewer decomposition base directions. The selection of the number of directions is not flexible, so there are many improved methods based on DWT gradually proposed, such as Discrete Fractional Order Wavelet Transform (DFWT), Dual Tree-Complex Wavelet Transform (DT-CWT), and so on. Although the above methods compensate for the limitations of DWT methods in feature representation, they do not change the nature of wavelet functions that cannot capture the curves and edges of images well.

2.3. Deep learning based medical image fusion method

In addition to traditional fusion methods, deep learning methods based on convolutional neural networks (CNN), which have emerged in recent years and have been widely used in various image fusion fields [39], perform better in extracting image features and are able to train model parameters. Generally, such methods can from a large amount of labeled data without human intervention and adaptively represent features at different scales of images. However, due to the particular characteristics of medical images, manual labeling of training data requires specialized background knowledge, and the size of publicly available aligned medical image datasets is small, so it is difficult to train a CNN framework from scratch to achieve medical image fusion. Singh et al. used a pretrained CNN model to extract salient features from the decomposed base layer [41] and obtained good fusion results. It can be seen that CNN can be used either alone to process the source image directly or applied to the transform domain [42]. Fu et al. proposed a multiscale residual pyramidal attention network to improve the performance of a single kind of network-based framework [43] approach for end-to-end fusion. However, this framework is complex and time-consuming. To further enhance the stability and efficiency of network training, Wang et al. proposed a Generative Adversarial Network (GAN)-based medical image fusion method that can effectively suppress artifacts and distortions in fused images [44]. Le et al. designed a conditional with multiple generators and multiple discriminators generating adversarial networks for different types of medical image fusion tasks [45]. Overall, though deep learning-based methods are effective in extracting image features, their interpretation is still one of the hot spots of research at this stage. Moreover, the learning method by migration or natural images supplementing the dataset may lead to unsatisfactory fusion results. Still, its idea of acquiring parameters by self-learning is very promising, which inspires us that traditional fusion algorithms can also process images adaptively.

Considering the advantages and limitations of methods based on a single type of image features, more and more fusion methods based on image hybrid features have been proposed. El-Hoseny et al. discussed the characteristics of hybrid techniques and proposed a hybrid method with additive and dual-tree complex wavelet transform to provide rich details for fused images [46]. Wang et al. proposed a method based on Laplace pyramid decomposition and adaptive sparse representation for multimodal image fusion, which can reduce the noise in the high-frequency information of the fused image [47]. Similarly, Zhang and Yan combined spatial domain and fractional-order derivatives to achieve noise suppression [48]. Li et al. used a joint filter to decompose the source image into an energy layer and a structure layer. Although this two-layer decomposition framework has high computational efficiency [48], it leads to low contrast and discontinuous illumination information in the fused image. Tan et al. proposed a multi-stage edge-preserving filter to distinguish better the fine and coarse structures of the input image, but the fused image is prone to color distortion [49]. Dinh improved the above defects by balancing the optimization algorithm and the selection of color space [50], ensuring that the perceptual quality of the fused images is consistent with that of the human visual system.

3. Methodology

Due to the limitations of the methods based on a single type of image feature, this paper proposes a medical image fusion method based on multi-scale contextual inference. The method first decomposes the original image to obtain the global part and local part of the original image. Then, the feature representation of each layer is enhanced by imposing saliency region constraints on multi-level features through the MSFE-Net network to extract the detailed characteristics of the image entirely. Meanwhile, an attention module is introduced to capture task-relevant fine features. Finally, the sample features are optimized by the design of loss functions. The MSFE-Net network consists of a feature encoder, a multi-scale feature extraction, a feature enhancement, a self-attention module, and decoding. The network structure diagram is shown in Figure 2.

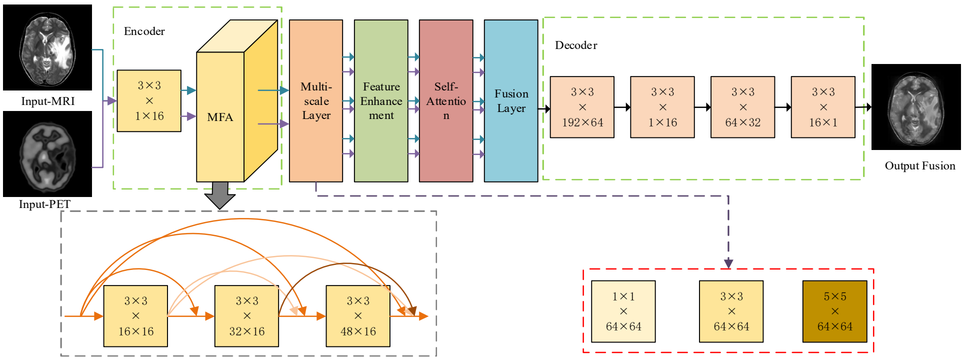


Figure 2. The architecture of the MSFE-Net.

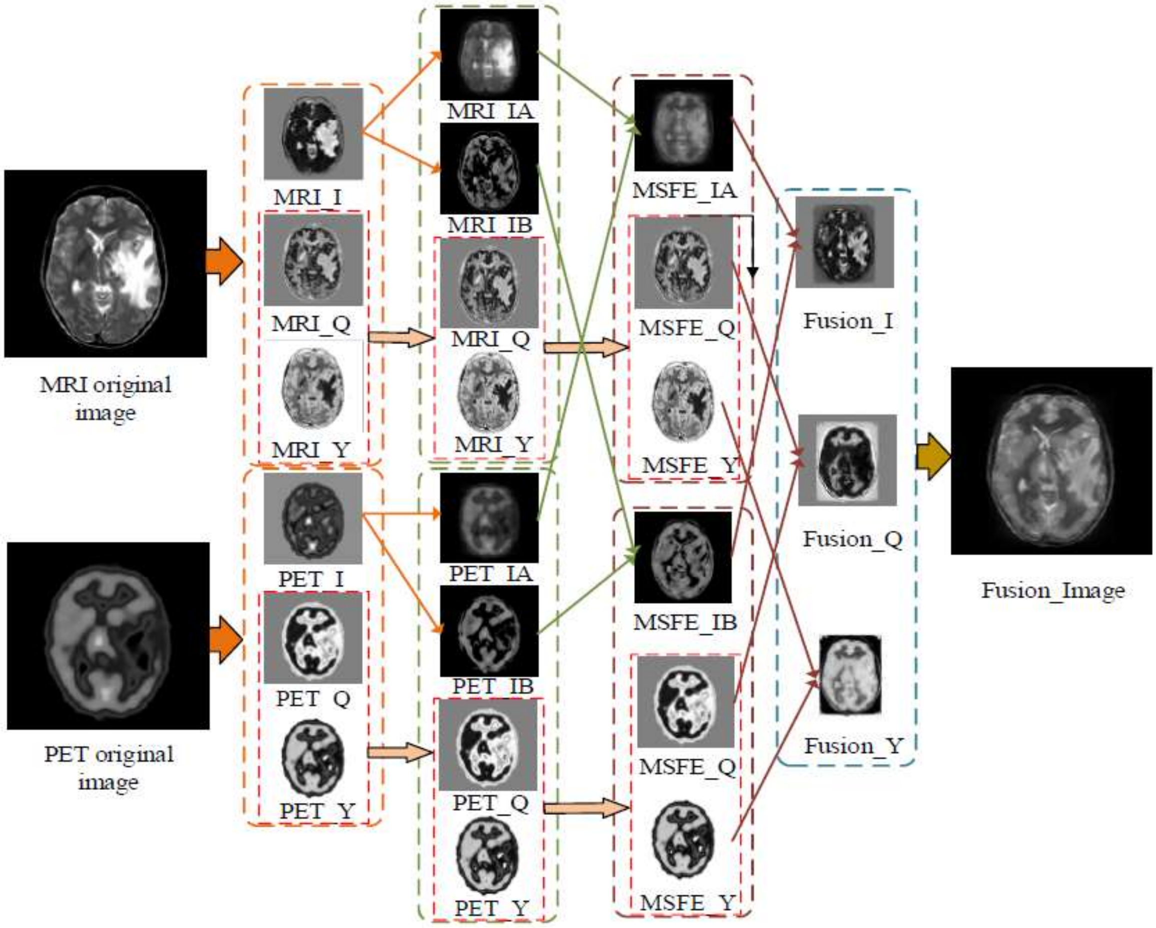


Figure 3. The flowchart of the proposed method.

In this paper, MRI and PET images of the brain are fused by the proposed MSFE-Net fusion network, and the detailed process is shown in Figure 3.

As shown in Figure 3, the first step of this paper is to input MRI and PET images, convert the input images to NTSC color space, and then extract the I, Q, and Y channels of NTSC space, respectively. In the second step, this paper extracts the global and local parts from the I channel of the obtained MRI and PET images, respectively. In the third step, the international and local parts corresponding to the second step are fused by the MSFE-Net method to obtain the I, Q, and Y channels of the reconstructed MRI and PET images. In the fourth step, the MSFE-Net method is used to fuse the I, Q, and Y component maps of the reconstructed MRI and PET images to obtain the I, Q, and Y component maps of the fused image. The fifth step is to convert the I, Q, and Y channels of the fused image into a fused grayscale image.

3.1. Feature Encoder

Marginal Fisher Analysis (MFA): The MFA diagram, as shown in Figure 4, is a basic diagram of separation between similar samples and Fisher's separation of heterogeneous samples. The eigenmaps describe the intra-class neighbor relationship, and each sample in the graph is connected to its nearest K_1 similar samples. The penalty diagram describes the relationship of inter-class boundary points, and each sample in the diagram is connected to its nearest K_2 different samples. Given the training data $\mathrm{X}_\mathrm{n} = [\mathrm{x}_1, \dots, \mathrm{x}_\mathrm{n}]$, labeled as $\mathrm{Y}_\mathrm{n} = [\mathrm{y}_1, \dots, \mathrm{y}_\mathrm{n}]$. With X_n as input, the features obtained through the network are $\mathrm{F}_\mathrm{n} = [\mathrm{f}_1, \dots, \mathrm{f}_\mathrm{n}]$.

The intra-class compactness and inter-class separability are denoted by S_c and S_p, respectively.

$$\begin{aligned} S_c = &\sum_{i\ i \in N_{K_1}^{+}(j)} \sum_{j\ j \in N_{K_1}^{+}(i)} ||\omega^T x_i - \omega^T x_j||_2^2 \\ &= 2\omega^T X (D^c - W^c) X^T \omega \\ &= \sum_i^n \sum_j^n (D_{ij}^c - W_{ij}^c) ||f_i - f_j||_2^2 \end{aligned} \tag{1}$$

$$=1_n^T(L^c \cdot Q)1_n$$

$$W_{ij}^c = \begin{cases} 1, & if\ i\epsilon N_{K_1}^+(j) or\ j\epsilon N_{K_1}^+(i) \\ 0, & else \end{cases} \tag{2}$$

$$D_{ii}^c = \sum_{j\neq i} W_{ij}^c,\ \ \forall i \tag{3}$$

$$L_{ij}^c = D_{ij}^c - W_{ij}^c \tag{4}$$

$$Q_{ij} = ||f_i - f_j||_2^2 \tag{5}$$

Where $N_{K_1}^+(j)$ denotes the index set of K_1 nearest neighbors in the same kind of sample x_j, and 1_n represents an n-dimensional vector with all elements being 1.

$$S_p = \sum_{i\ (i,j)\in P_{K_2}(c_i) or (i,j)\in P_{K_2}(c_j)} \sum_{j\ (i,j)\in P_{K_2}(c_i) or (i,j)\in P_{K_2}(c_j)} |||\omega^T x_i - \omega^T x_j||_2^2$$

$$=2\omega^T X(D^P - W^P)X^T\omega \tag{6}$$

$$=\sum_i^n \sum_j^n (D_{ij}^P - W_{ij}^P)||f_i - f_j||_2^2$$

$$=1_n^T(L^P \cdot Q)1_n$$

$$W_{ij}^P = \begin{cases} 1\ , & if(i,j) \in P_{K_2}(c_i) or (i,j) \in P_{K_2}(c_j) \\ 0\ , & else \end{cases} \tag{7}$$

$$D_{ii}^P = \sum_{j\neq i} W_{ij}^P\ \ ,\ \ \forall i \tag{8}$$

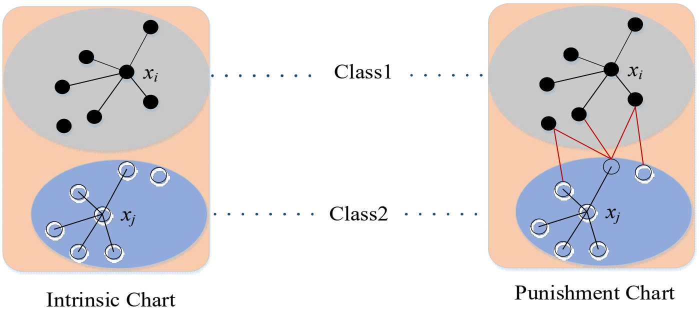
$$L_{ij}^P = D_{ij}^P - W_{ij}^P \tag{9}$$

Where $P_{K_2}(c)$ is the K_2 nearest neighbor satisfying the conditions $\{(i.j), i \in \pi_c, j \notin \pi_c\}$, and π_c refers to the index set of samples belonging to class c.

From S_c, S_p, we can obtain the marginal Fisher criterion $JMF = \frac{S_c}{S_p}$, the optimal weight matrix can be obtained when JMF takes the minimum value.

$$\omega^* = argmin_\omega JMF = argmin_\omega \frac{\omega^T(X)(D^C - W^C)\omega^T(X)}{\omega^T(X)(D^P - W^P)\omega^T(X)} \tag{10}$$

Where $\omega^T(\cdot)$ denotes the mapping function from the sample space to the feature space. Using the optimal weight matrix ω^* obtained by the minimum JMF constraint in the above equation, the sample data can be mapped from the original sample space to the feature space with better separability.



Intrinsic Chart

Punishment Chart

Figure 4. Eigen and penalty diagrams in marginal Fisher analysis.

The main idea of the feature encoder is to incorporate MFA into the supervised fine-tuning training of the stack denoising autoencoder and propose the feature encoder method. This encoder uses MFA to constrain the representation of various types of samples in the feature space so that the features learned by the network are more discriminative. Its components are shown in Figure 1(A). $\{F_i|i = 1,2,3,4,5\}$ denotes the features extracted by the network at different levels. $\{F_i \epsilon R^{\frac{w}{2^{i-1}} \times \frac{h}{2^{i-1}} \times \frac{c}{2^{1-i}}} | i = 1,2,3,4,5\}$, where *w*, *h*, and *c* represent the width, height, and number of channels of the feature map.

In this paper, JMF is added to the objective function of the auto-encoder so that the obtained optimal parameters ensure that the JMF is as small as possible. In JMF, the numerator S_c and the denominator S_p describe the inter-like and inter-species distances, respectively. Therefore, in the adjusted new feature space, the inter-species distance will be pulled in, and the inter-species distance will be sparse. This way, overfitting can be better prevented during the training process, and the classification effect can be improved.

Let ω_{enc} be the weight matrix of the encoding part of the autoencoder, and the mapping from the sample space to the feature space is $\emptyset(\cdot\,|\omega_{\mathrm{enc}})$. By moving the MFA to the autoencoder, we can get the corresponding JMF.

$$JMF = \frac{\emptyset_{enc}^{T}(X)(D^{C} - W^{C})\emptyset_{enc}^{T}(X)}{\emptyset_{enc}^{T}(X)(D^{P} - W^{P})\emptyset_{enc}^{T}(X)} \tag{11}$$

The new objective function is obtained by adding JMF to the equation $\mathrm{J} = \frac{1}{2}\mathrm{R} + \lambda\mathrm{D}$.

$$E = JMF + J = JMF + \frac{\alpha}{2}R + \lambda D \tag{12}$$

Where α and λ are the regularization coefficients. Optimal weight matrix.

$$W^{*} = argmin_{\omega}(JMF + \frac{\alpha}{2}R + \lambda D) \tag{13}$$

To calculate the gradient $\frac{\partial \mathrm{J}}{\partial \mathrm{W}}$, we need to calculate $\frac{\partial \mathrm{JMF}}{\partial \mathrm{F_n}}$ first, which can be obtained by solving it computationally.

$$\frac{\partial S_c}{\partial f_i} = \sum\nolimits_{j} (L_{ij}^{c} + L_{ji}^{c}) \frac{\partial ||f_i - f_j||_2^2}{\partial f_i} = f_i \cdot 1_n^T (L^c + (L^c)^T)_i - F_n (L^c + (L^c)^T)_i \tag{14}$$

$$\frac{\partial S_c}{\partial F_n} = F_n[Diag\big(1_n^T (L^c + (L^c)^T)\big) - (L^c + (L^c)^T)] \tag{15}$$

Same reason,

$$\frac{\partial S_p}{\partial F_n} = F_n[Diag\big(1_n^T (L^P + (L^P)^T)\big) - (L^P + (L^P)^T)] \tag{16}$$

$$\frac{\partial JMF}{\partial F_n} = \frac{\partial (\frac{S_c}{S_p})}{\partial F_n} = \frac{S_c \cdot \frac{\partial S_p}{\partial F_n} - S_p \cdot \frac{\partial S_c}{\partial F_n}}{(S_p)^2} \tag{17}$$

$$\frac{\partial JMF}{\partial W} = \frac{\partial JMF}{\partial F_n}\frac{\partial F_n}{\partial W} \tag{18}$$

$\frac{\partial \mathrm{F_n}}{\partial \mathrm{W}}$ can be found based on the gradient iteration. From $\mathrm{R} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{L}(\mathrm{h}(\mathrm{x}^{(\mathrm{i})}), \mathrm{label}^{(\mathrm{i})})$, it can be obtained.

$$\frac{\partial R}{\partial W} = \frac{2}{n}\sum\nolimits_{i=1}^{n} (h(f^{(i)}) - label^{(i)}) \frac{\partial h(f^{(i)})}{\partial W} \tag{19}$$

It can be obtained by $\mathrm{D} = \frac{1}{2}||\mathrm{W}||_2^2$.

$$\frac{\partial D}{\partial W} = W \tag{20}$$

From equations (18), (19), and (20), the derivation of E to the weight parameter W in the objective function equation (12) can be calculated:

$$\frac{\partial E}{\partial W} = \frac{\partial JMF}{\partial W} + \frac{\alpha}{2}\frac{\partial R}{\partial W} + \lambda \frac{\partial D}{\partial W} \tag{21}$$

3.2. Multi-scale feature extraction network

In this paper, we design a multi-scale feature extraction network. We hope to obtain the deep semantic features in medical images through a deep network. However, as the network layers become deeper, it also brings the problem of gradient explosion or gradient dispersion. The gradient gradually decreases in the process of passing from deep to shallow layers, making the external network unable to be trained effectively. Due to the instability of the gradient and the inefficiency of the backpropagation, the network is difficult to converge. We construct a multi-scale feature extraction network to address these problems, as shown in Figure 1.

The following points mainly cause the instability of the gradient in the propagation process: first, the weights are given larger values in the process of random initialization of the weights, resulting in the back-propagation of the gradient multiplied by the weights being more extensive than and amplified layer by layer in the subsequent propagation process leading to the gradient explosion, the multi-scale feature extraction network can better avoid the problem of gradient explosion by Gaussian initialization of the weights; second, the sigmoid activation function Secondly, the characteristics of the sigmoid activation function determine that it exhibits low sensitivity to gradient for larger or smaller input values, resulting in the gradient not being effectively back-propagated through the sigmoid activation function. Based on this, the multi-scale feature extraction network is designed to limit the input of the activation function to the gradient-sensitive interval of the activation function by

BatchNorm operation and introduce a simpler and more efficient relu activation function for gradient back-propagation to alleviate the problem of high gradient loss through the activation function.

Although the number of relu activation functions is chosen to alleviate the loss of gradient when passing through the activation function, it still does not entirely solve the problem of insufficient learning in external networks brought about by the deepening of the network. For this purpose, the prediction map is generated by convolution operation on the feature map. Then the up-sampling of the generated prediction map is kept at the same ratio as the input image by interpolation sampling. During training, a more discriminative feature map is generated by training the difference between it and the actual value.

3.3. Feature Enhancement

Considering that the complex morphology of the brain can make the fusion of solid brain images difficult, several enhancement modules are used to enhance the global information in the encoder, and the global feature enhancement module is proposed, as shown in Figure 1(B). The feature enhancement module uses global averaging pooling without dimensionality reduction to aggregate the local features of the network and form the final global features. The global information from different layers is then fully fused. Such a strategy has been widely used in similar layer attention modules, but these layer attention modules usually use a fully connected layer to accomplish the fusion between layers. Since, the fully-connected layer requires vast computational cost, it is not an efficient approach for layer interaction. In this paper, a convolutional layer with a convolutional kernel size of M is used instead of a fully connected layer, and M can be a fixed value or an adaptive value that is positively related to the number of layers. The interacted feature map is then passed through the Sigmoid activation function to obtain a two-dimensional weight map, multiplied with the feature map transmitted from the feature encoder to the feature enhancement module to get the final feature output.

Every medical image consists of an object area and a background area, but many images suffer from class imbalance. In some images, the background region is relatively large while the object occupies a small proportion, or the object occupies a large proportion while the background occupies a small proportion. This class imbalance phenomenon can be very disruptive to the fusion task, a suitable loss function is designed in this paper to alleviate this phenomenon. Compared with the method of adjusting the network structure, improving the loss function can be implemented more easily while avoiding the higher computational cost effectively. The designed loss function contains a Focal term, and the Focal term is defined as follows.

$$L_{Focal} = -(1-y_i)^{\gamma} log y_i \tag{22}$$

The Focal term consists of the log-term and the preceding coefficients, and y_i denotes the predicted probability of the pixel. The starting point of the Focal term is used to alleviate the class imbalance and make the training process more stable. With the addition of the coefficients, Focal can improve the fairness of the fusion process so that the network can treat the majority and minority classes equally and put the network in a balanced state.

Although the Focal term can mitigate class imbalance, it does not sufficiently consider the spatial relationship between pixels. Especially for medical images, when the color or texture information of the target in the image is not uniform, it is usually not possible to fuse the whole object using a loss similar to cross-entropy. At the same time, it is crucial to combine the spatial information of the shape of the brain because of its complex morphology and differentiated shapes. Therefore, a variance term is added to constrain the shape of the brain, and the variance term is defined as

$$L_{var}(y,t) = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{|S_n|}\sum_{i=1}^{|S_n|}(\mu_n - y_i)^2 \tag{23}$$

In equation (23), N represents the number of brain images in the data, and S_n denotes the number of pixels in the nth brain image. y_i indicates the predicted probability of the ith pixel in the nth brain instance, and n represents the mean value of the predicted probability of all pixels in the nth brain image, i.e., the mean value of y_i in the nth brain image. The shape of a single object is constrained during the training process of the network, i.e., the concept of instance is added to the process of semantic fusion. The shape constraints of the objects guide the fusion process of individual brain images.

Therefore, the final form of the FV loss function is:

$$L = -\alpha(1-y_i)^\gamma log y_i + \frac{\beta}{N}\sum_{n=1}^{N}\frac{1}{|S_n|}\sum_{i=1}^{|S_n|}(\mu_n - y_i)^2 \tag{24}$$

The parameters α and β are used to adjust the weights of the two terms, and the variance term makes the pixels in a single brain image converge to the same category. The continuity of elements in a single brain image is enhanced, which helps to alleviate the phenomenon that a single brain image is divided into multiple brain images. The Focal term, on the other hand, alleviates the class imbalance, allowing the majority and minority classes to converge during the fusion process. After the loss map is calculated, the loss function is weighted using the pre-generated weight map, which enables the network to pay more attention to the boundary information, thus alleviating the problem of brain tissue adhesion. The later experimental results show that the combination of Focal term, variance term, and weight map can effectively improve the fusion results.

3.4. Self-Attention Module

For different scales of feature maps output by the network, this paper improves the feature representation of key channels by designing a self-attentive module to filter further the features of different channels in the feature maps and further guides the network to focus its attention on the regions containing key information. More global auxiliary information is obtained to compensate for the lack of information acquisition of small convolutional kernels by calculating the interaction between any two feature channels to capture the remote dependencies directly. Then all feature channels are assigned with more reasonable weights. The structure of the self-attentive module is shown in Figure 5.

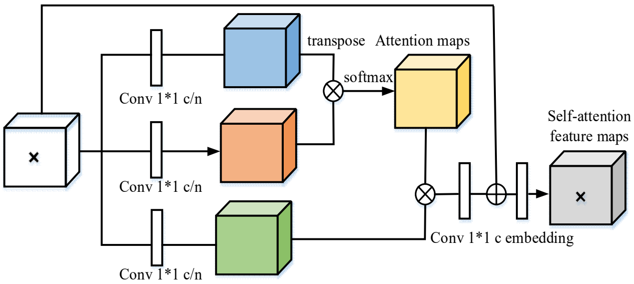


Figure 5. Structure diagram of self-attention-module.

First, the input feature map is compressed by three different branches f(x), g(x)and h(x). The channel is compressed by three sets of $1*1$ convolution with the same number. The channel dimension is retained to flatten the width and height into one dimension, mainly to reduce the information redundancy of the input feature map and to reduce the complexity of the similarity calculation later. Secondly, the matrix of branch f(x) is transposed and then multiplied with that of branch g(x), after which, the result is normalized by softmax. We are calculating the similarity of the feature maps between different channels in the NLM algorithm. Finally, the attention matrix after normalization is multiplied with the feature map obtained by branching h(x), which is the redistribution of the weights of different channels according to the similarity. Also the number of channels is expanded to the input feature map by softmax and $1*1$ convolution. The attention reallocation is achieved.

3.5. Decoders

The decoder is a symmetric structure of the encoder. During the construction of the encoder, the spatial size information passed to the following layer decreases as the convolution layers are added. The decoder should adjust the image size and the number of convolution layers to reconstruct the original image. Transpose Convolution can increase the spatial size and convolution, but transpose convolution layers cause artifacts in the final output image. In order to keep more critical information about the original image in the reconstructed image, upsampling techniques are needed in the decoder. The most common upsampling methods are UnPooling and UnSampling.

The process of UnPooling is shown in Figure 6, where the maximum position information is retained during Max pooling. Then the feature map is expanded with this information during the

UnPooling phase, and the rest is complemented by 0. This is the inverse of pooling, and restoring all the original data through the pooling result is impossible. If you want to recover all the information from the pooled primary information, the maximum information integrity can only be achieved through complementary positions.

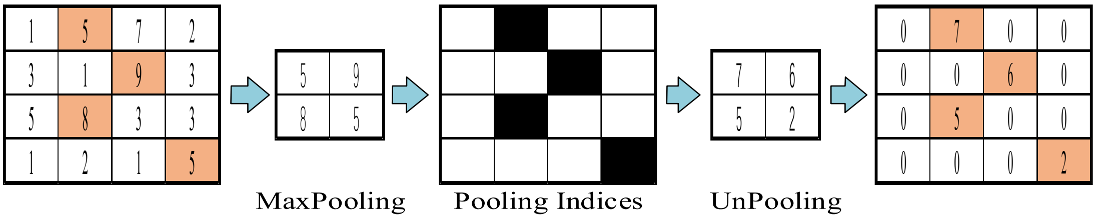


Figure 6. UnPooling operation procedure.

The process of UnSampling is shown in Figure 7. This operation does not use the location information of Max Pooling and directly copies the maximum expansion feature map.

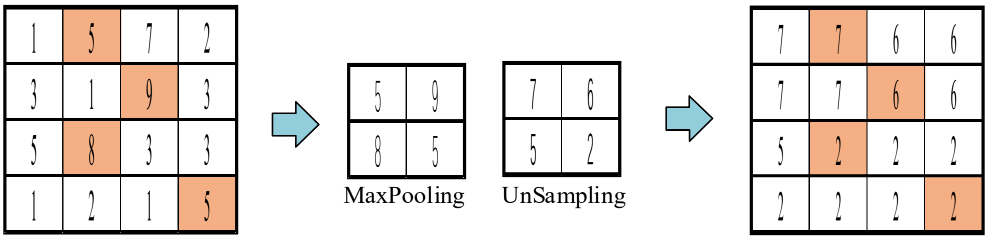


Figure 7. UnSampling operation procedure.

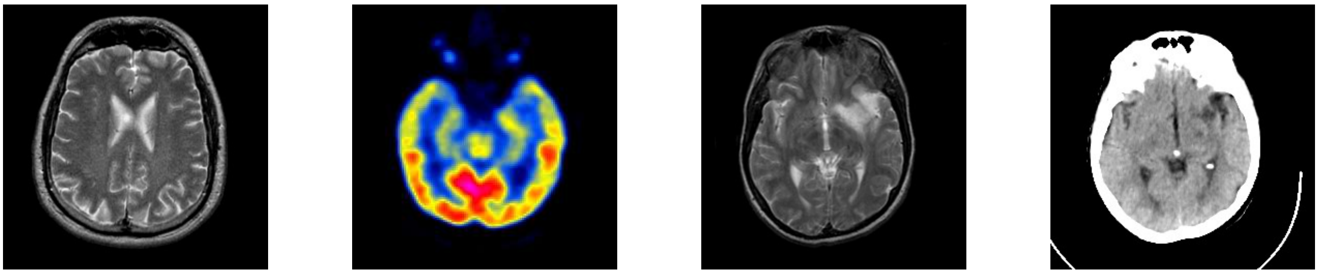
In this paper, different upsampling operations are designed for decoders of different layers according to the characteristics of shallow and deep layer features in the network. Considering the fact that the information extracted from the shallow network structure is coarser and contains more useless information, while the feature maps extracted from the deep network structure are smaller, rich in semantic information, and contain relationships between global contexts, the up-sampling in the deep $D4(\cdot)$ and $D3(\cdot)$ layers in the process of designing the decoder uses the UnSampling operation, which directly copies the input feature map content to expand the upsampling layer feature map; and in the shallow $D2(\cdot)$ and $D1(\cdot)$ upsampling using the UnPooling operation, which records the location information of the maximum value at the time of maximum pooling, and then uses this information in the upsampling phase to expand The feature map of the upsampling layer is complemented by 0 except for the position of the maximum value. UnSampling preserves the overall data features, and UnPooling preserves features with vital semantic information, typically texture features. These two upsampling methods combine to compensate for the lost spatial feature information while preserving important information in the original feature space.

4. Experiments

In this section, we conduct extensive experiments to evaluate medical image fusion methods based on multiscale contextual reasoning.

4.1. Experiment Settings

Datasets: To validate the validity of the methods in this paper, we acquired 1000 MRI-PET and MRI-CT image pairs that are publicly published in the Alzheimer's Disease Neuroimaging Initiative (ADNI) with subjects aged 55-90 years of either sex. The spatial resolution of all source images was set to 256 × 256, and the pairs of source images were aligned. The three medical image categories are shown in Figure 8. In this paper, 60% of the acquired data are used as the training set, 10% as the test set, and 30% of the data is used as the test set. The CT-MRI and MRI-PET image pairs used for testing are cerebrovascular disease (stroke) and oncological disease (brain tumor).



(a) MRI-PET

(b) MRI-CT

Figure 8. Multimodal images of the brain. (a) is the MRI and PET image pair. (b) is the MRI and CT image pair.

Environment Configuration: All the algorithms are coded using python in PyCharm Community Edition 2020. For each algorithm configuration and each instance, we carry out five independent replications on the same AMD Ryzen 5 3500X 6-Core Processor CPU @ 3.60 GHz with 16.00-GB RAM and NVIDIA GeForce GTX 1660 SUPER GPU in the 64-bit Windows 10 professional Operation System.

Model-related Parameters: The kernel filter of the fusion network in this paper is initialized to a truncated normal distribution with a standard deviation of 0.01 and a deviation of zero. Each layer has a step size of 1 and no padding during convolution because each downsampling layer removes detailed information from the input image, which is crucial for medical image fusion. We use batch normalization with a slope of 0.2 and ReLU activation to avoid the problem of gradient disappearance. The network is trained for 200 epochs on a single GPU with a batch size of 1 and different $\lambda \in$ [0,1]. The Adam optimizer is used as the optimization function in the backpropagation step with a learning rate of 0.002.

4.2. Evaluation Criteria

This paper compares the developed medical image fusion method based on multiscale contextual inference with LatLrr, IFCNN, NestFuse, atsIF, FusionDN, FusionGAN, and FunFuseAn algorithms. The evaluation indices are EN, SD, MI, rSFe, SM, and VIF. EN and SD represent the information in the input image, and the larger value represents the better fusion effect. the larger value of MI represents the better preservation of the original information and features of the source image. rSFe is the spatial frequency measurement, and when $rSFe > 0$, it will cause the image. When $rSFe > 0$, it will cause distortion or add some noise to the image, resulting in excessive image fusion. When $rSFe < 0$, it will cause the loss of image information, resulting in insufficient image fusion. Because all the results in our experiments are in the state of under-fusion, the smaller the absolute value, the better the fusion effect. The larger the SM and VIF values, the more information about the source image structure is retained by the fusion algorithm, and more natural features can be generated. Since the values of different metrics vary widely, I have adjusted some values for comparative analysis. The values of each metric are adjusted by linear transformation: $y = mx + n$. x is the original metric value, y is the transformed metric, and m and n are two coefficients. The coefficients are shown in Table 1.

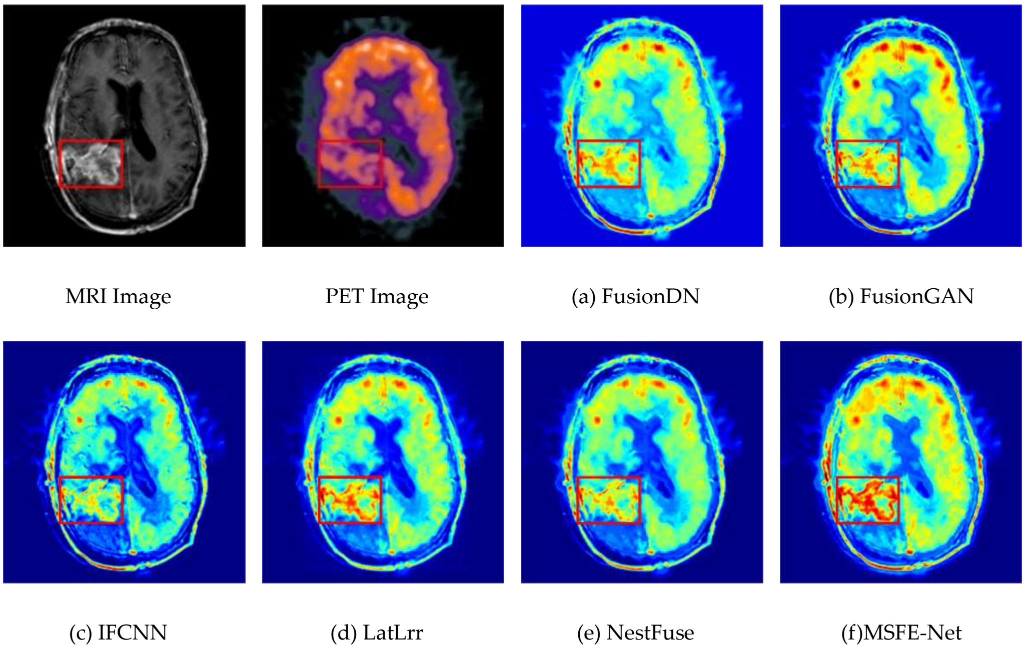
Table 1. The coefficients of linear transformation.

Coefficients	EN	SD	MI	rSFe	SM	VIF
m	1	0.1	1	1000	1000	1000
n	0	0	-6	995	-2	1

4.3. Experimental results between the proposed method and existing methods

To demonstrate the effectiveness of the medical image fusion method based on multiscale contextual inference, this paper compares this method with five other advanced and representative methods. The comparison methods are FusionDN [42], FusionGAN [20], IFCNN [21], LatLrr [10], and NestFuse [22]. the results of the MSFE-Net fusion method and the fusion methods of the other five methods are shown in Figure 6. The results of the MSFE-Net fusion method with the other five methods are shown in Figure 6 and Figure 7, and their evaluation metrics are shown in Tables 2 and 3.

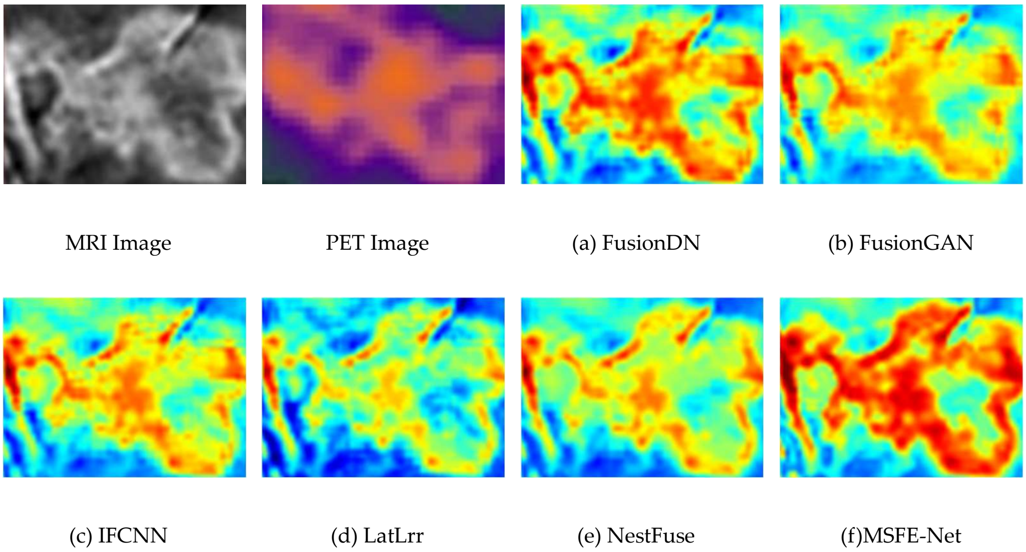
Experiments on MRI-PET medical image fusion: MRI is used to obtain electromagnetic signals based on the different attenuation of energy materials in different structural environments of the body. MRI images have a high resolution and clear soft tissue information to locate lesions better. PET images are obtained by injecting a radioisotope drug into the body, which is metabolized by human tissues to cause the drug to decay and produce γ photons, which are then converted to light/electricity and processed by a computer into PET images. PET images have a lower resolution and poorer localization ability. However, the fused MRI and PET raw images provide structural and activity information with high spatial resolution. Figure 9 shows the subjective evaluation of MRI-PET fusion. Figure 10 shows the fusion results of MRI-PET brain tumor locations. Figure 9 and Figure 10 show an example, including the input and fused images and their corresponding scaled regions marked by red borders.



MRI Image PET Image (a) FusionDN (b) FusionGAN

(c) IFCNN (d) LatLrr (e) NestFuse (f)MSFE-Net

Figure 9. Fused results of MRI-PET medical image fusion.

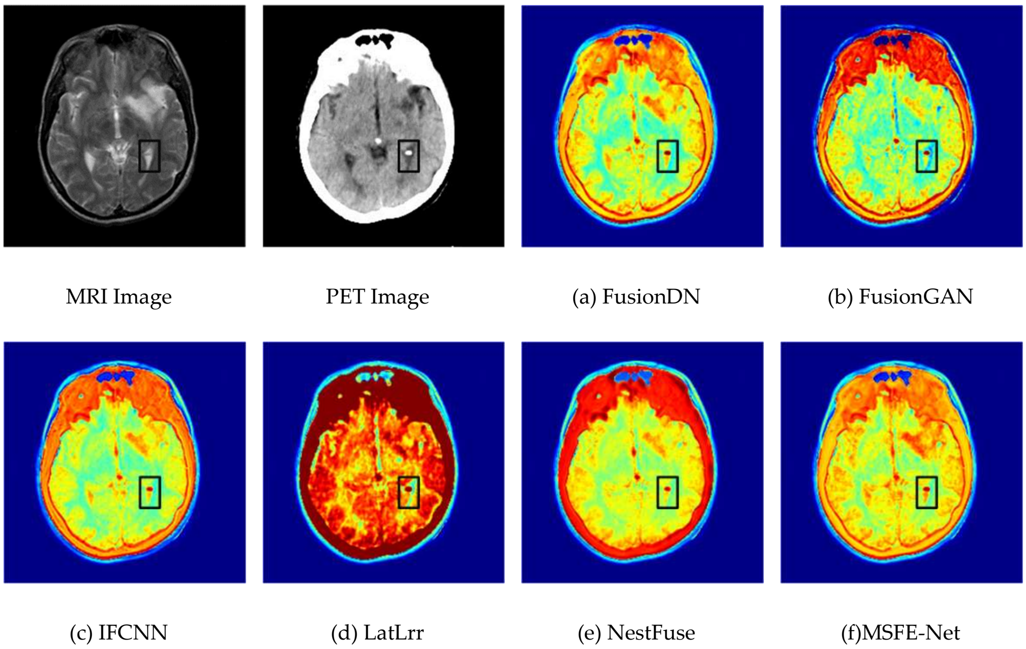


MRI Image PET Image (a) FusionDN (b) FusionGAN

(c) IFCNN (d) LatLrr (e) NestFuse (f)MSFE-Net

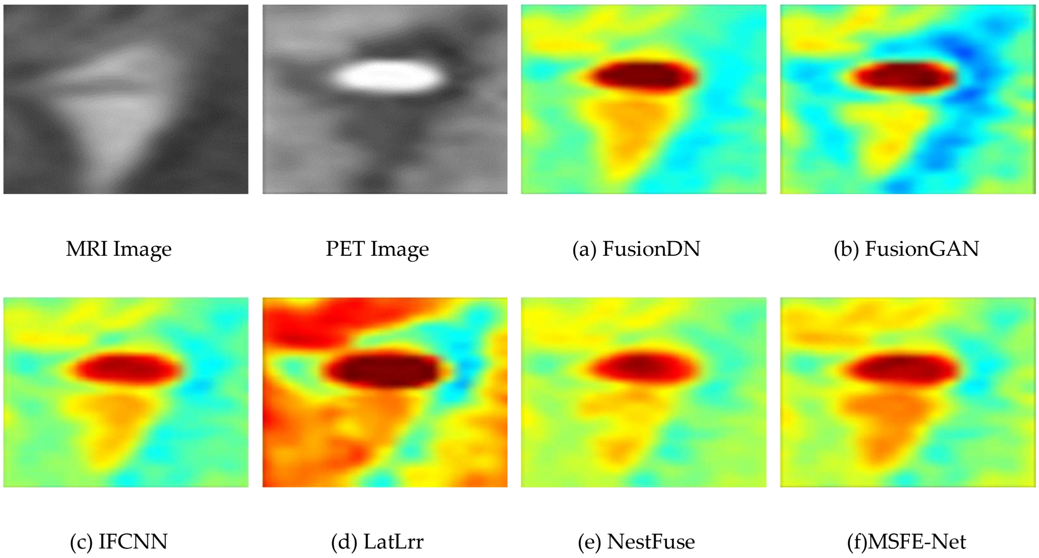
Figure 10. Fused results of MRI-PET brain tumor locations.

In general, the quality of fused images is evaluated in terms of three main aspects: image visualization, image edge contour, and detailed information. Figure 6 shows the fusion results of MRI-PET images of brain tumor patients, including the input and fused images and the corresponding scaled areas marked by red bounding boxes. Based on this set of results, it can be seen that the contours of the fused images based on the NestFuse method are clear, but some artifacts make the resulting images blurred. The FusionGAN-based fusion image has moderate brightness, and the contour information is clear, but the specific texture details are not obvious. The texture details of the IFCNN method are obvious. But more PET information is discarded. The LatLrr-based method has the blurred interior of the fused image with clear contours but contains insufficient information on PET. The fused image based on the NestFuse method is not rich in detailed information, and the internal structure is incomplete. The fused image based on the MSFE-Net method can retain more features, and the visual effect, edge contour, and detail information can be greatly improved.


MRI Image PET Image (a) FusionDN (b) FusionGAN

(c) IFCNN (d) LatLrr (e) NestFuse (f)MSFE-Net

Figure 11. Fused results of MRI-CT medical image fusion.


MRI Image PET Image (a) FusionDN (b) FusionGAN

(c) IFCNN (d) LatLrr (e) NestFuse (f)MSFE-Net

Figure 12. Fused results of MRI-CT brain tumor locations.

Experiments on MRI-CT medical image fusion: CT images are cross-sectional images of human tissues, which can well display the results of organs in various parts of the human body and examine secret changes in tissues in addition to discovering morphological changes, expanding the scope of imaging examinations. However, CT images make detecting functional changes in organ tissue structure more challenging. However, MRI-CT fusion systems produce final output images that contain structural and blood flow information with high spatial resolution. The subjective results of MRI-CT medical image fusion are shown in Figure 11. Figure 12 shows the fusion results of MRI-CT brain tumor locations. Figure 11 shows an example of MRI-CT fusion, including the input and fused images and their corresponding scaled regions marked by black borders.

Figure 11 shows the MRI-CT fusion results. The left column shows the original MRI and CT images, and the other columns show the fusion result maps obtained by different fusion methods. The brain tumor region is presented in Figure 12 in this paper. The enlarged image of the area also illustrates the comparison. Figure 11(a) and Figure 12(a) show the fusion results obtained by FusionDN based on a dense connectivity network. This fusion result shows that the FusionDN algorithm can improve the brightness of the fusion result, but there is some lack of detail in its fusion. Figure 11 (b) and Figure 12(b) show the fusion results obtained by the FusionGAN method, a fusion method for generating response networks with a loss function that includes a discriminant function. Still, the discriminant function is not good enough, so the fusion results do not highlight important information. Figure 11 (c) and Figure 12(c) show the fusion results obtained by IFCNN. The fusion strategy is divided into three cases: maximum value, average value, and the sum of two feature values. The IFCNN algorithm uses maximum value when processing medical images. Therefore, I only list the fusion results for the maximum value in the experiments. From the visual effect, the edges and internal regions are clear. Figure 11 (d) and Figure 12(d) show the fusion results obtained by LatLrr. The significant and low-rank parts of the image are extracted using LRR, and then the two elements are combined to reconstruct the fused image. In Figure 10 (d), it can be seen that some details are lost, the reconstructed image is incomplete, and the brightness of the fused image is lower than that of the original image. Figure 11 (e) and Figure 12(e) show the fusion results obtained by NestFuse using the maximum fusion strategy. The idea of the NestFuse algorithm is encoding, decoding, and fusion. The encoding includes double convolution and pooling operations, and the decoding includes dual convolution and sampling operations. The network structure is slightly more complex, but deeper features can be obtained. From the results, it looks similar to LatLrr, without clear texture details. The final result image is the application of the MSFE-Net method in this paper. Compared with other algorithms, the fusion results in this paper have more suitable brightness, clearer contours, and finer textures. In addition, our results maintain and enhance crucial medical information very well.

4.4. Comparison of the proposed method with advanced method statistics

The above subjective evaluation has proved that the algorithm of this paper has certain advantages, but it is relatively one-sided to judge only from the subjective assessment. Therefore, to further verify the superiority of the algorithm, this paper selects six indicators as the objective evaluation. The objective evaluation indicators of MR-PET image fusion are shown in Table 2. The objective evaluation indexes of MRI-CT image fusion are shown in Table 3.

Table 2. Objective evaluation results of image MRI-PET fusion.

Indicators Methods	EN	SD	MI	rSFe	SM	VIF
FusionDN	5.609	2.496	5.217	-2.3	4.0	1.3
FusionGAN	5.847	2.297	5.694	-2.6	3.6	1.3
IFCNN	6.063	2.487	6.127	-2.4	3.4	1.5
LatLrr	5.976	2.446	5.951	-2.4	3.6	1.4
NestFuse	5.716	2.245	5.431	-2.6	3.4	3.4
MSFE-Net	**6.934**	**5.871**	**8.019**	**-1.7**	**4.7**	**3.9**

The quantitative performance comparison of the results in the figure is given in Table 2. From Table 2, it can be seen that the algorithm of this paper achieves the best results in the six objective evaluation indexes of EN, SD, MI, rSFe, SM and VIF for MR-PET image fusion. This indicates that the algorithm of this paper has richer fused image information and a better edge contour fusion effect. The MSFE-Net-based image fusion method has significantly improved in these six indexes compared with other methods. Combining the subjective and objective evaluations, the quality of the fused images of this algorithm is better, and the fused images have better fusion quality and better performance indexes when this algorithm is used for MRI and PET image fusion compared with several other comparative algorithms.

Table 3. Objective evaluation results of image MRI-CT fusion.

Indicators Methods	EN	SD	MI	rSFe	SM	VIF
FusionDN	5.952	2.998	5.609	-1.9	4.9	1.6
FusionGAN	6.128	2.789	6.018	-2.2	4.1	1. 6
IFCNN	6.631	2.942	6.992	-2.0	3.9	1. 8
LatLrr	6.278	2.983	6.158	-2.0	4.1	1.7
NestFuse	6.056	2.774	5.893	-2.2	3.9	3.7
MSFE-Net	**7.543**	**6.995**	**8.855**	**-1.3**	**5.8**	**4.2**

Table 3 lists one evaluation metric in each column, where the bold numbers represent the maximum values. Each row lists separately the values of MRI-CT images fused using different fusion methods and evaluated on different metrics. The evaluated values of the MSFE-Net fusion method on EN, SD, MI, SM, and VIF are much larger than those of other fusion methods. The metric values of MSFE-Ne for SD of the six methods differ significantly from others. Because the SD index is the distance of each data from the mean, reflecting the relationship between the data series and the mean, it can explain the degree of dispersion of the data set. The larger the value, the higher the measurement accuracy. rSFe is an image fusion metric based on the spatial frequency error ratio. It is susceptible to small changes in image quality. The higher the absolute value of rSFe, the better the fusion effect. Because the upper and lower structures, although they can better preserve the structural features in the spatial structure, cannot avoid the loss of specific details, this value will be relatively small for the method in this paper.

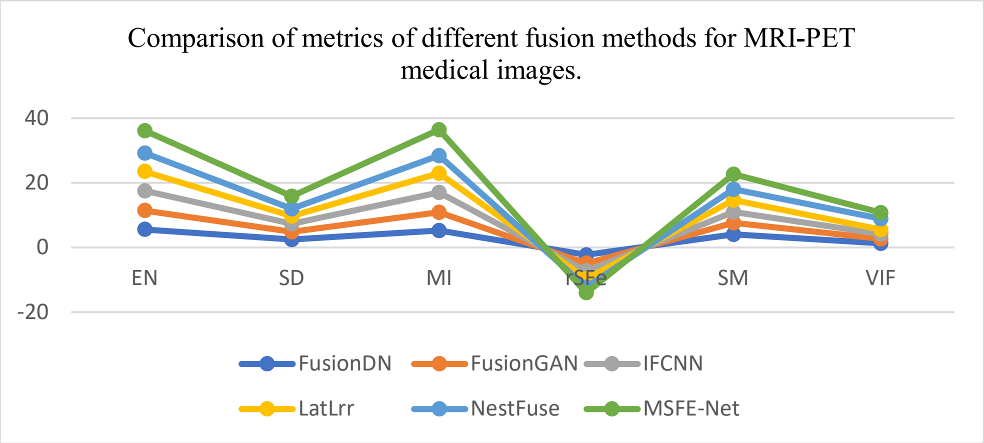


Figure 13. Comparison of metrics of different fusion methods for MRI-PET medical images.

To further highlight the superiority of the algorithm used in this paper for fusing MRI-PET and MRI-CT images, the six objective evaluation metrics, EN, SD, MI, rSFe, SM, and VIF, are visually represented using line graphs, as shown in Figures 13 and 14. It can be seen from the figure that the performance of the MSFE fusion method is very high in each indicator. This shows that the fusion method in this paper not only enriches the relative information of the fused image but also raises the edge quality and image clarity.

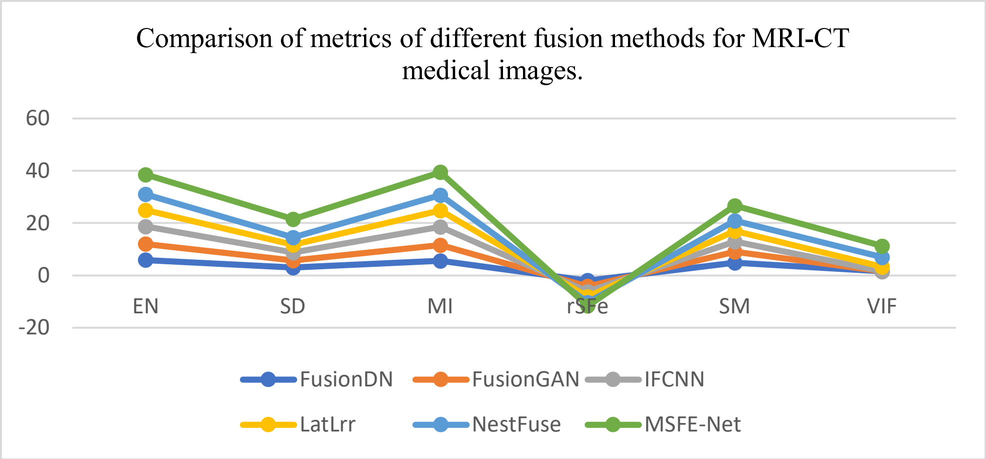


Figure 14. Comparison of metrics of different fusion methods for MRI-CT medical images.

5. Conclusions

To address the problem that the medical images taken by these different devices reflect different structural information of the human body, and the medical images of a single modality do not portray the lesion information comprehensively. This paper proposes a medical image fusion method based on multi-scale contextual reasoning. The method decomposes the original MRI and PET images into global and local parts. And in the feature extraction stage, MSFE-Net network is used to fuse feature maps of different levels and scales, which fully uses the shallow texture features and deep semantic features of the images, and better alleviates the feature extraction problem of different scale targets. At the same time, an improved attention module is designed to adapt to the output of feature maps at different scales. All channel features are reassigned to improve critical channel representation and make the images' essential detail features more prominent. Finally, in the model training stage, the idea of combining cross-entropy loss and central loss is adopted to make the distribution of sample features of each class in the sample space more reasonable and further improve the fusion accuracy of the model.

Funding. Shandong Provincial Natural Science Foundation of China (ZR2018MF009); The State Key Research Development Program of China (2017YFC0804406); National Natural Science Foundation of China (42472324); Qingdao Postdoctoral Foundation (QDBSH20240202049). China Postdoctoral Science Foundation (2023M732107).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Goyal, M. K. Arya, R. Agrawal, et al., "Automated segmentation of gray and white matter regions in brain MRI images for computer aided diagnosis of neurodegenerative diseases," 2017 International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT), Aligarh, (2017) 204-208.
2. Verclytte, S.; Lopes, R.; Lenfant, P.; Rollin, A.; Semah, F.; Leclerc, X.; Pasquier, F.; Delmaire, C. Cerebral Hypoperfusion and Hypometabolism Detected by Arterial Spin Labeling MRI and FDG-PET in Early-Onset Alzheimer's Disease. *J. Neuroimaging* **2015**, *26*, 207–212. https://doi.org/10.1111/jon.12264.
3. M. P. Nguyen, H. Kim, S. Y. Chun, et al., "Joint spectral image reconstruction for Y-90 SPECT with multi-window acquisition," 2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), San Diego, CA, (2015) 1-4.
4. Li, H.; Wang, Y.; Yang, Z.; Wang, R.; Li, X.; Tao, D. Discriminative Dictionary Learning-Based Multiple Component Decomposition for Detail-Preserving Noisy Image Fusion. *IEEE Trans. Instrum. Meas.* **2019**, *69*, 1082–1102. https://doi.org/10.1109/tim.2019.2912239.
5. Du, J.; Fang, M.; Yu, Y.; Lu, G. An adaptive two-scale biomedical image fusion method with statistical comparisons. *Comput. Methods Programs Biomed.* **2020**, *196*, 105603. https://doi.org/10.1016/j.cmpb.2020.105603.

6. Li, H.; Wu, X.-J. DenseFuse: A Fusion Approach to Infrared and Visible Images. *IEEE Trans. Image Process.* **2018**, *28*, 2614–2623. https://doi.org/10.1109/tip.2018.2887342.
7. Y. Li, J.Chen , P. Xue , et al., "Computer-aided Cervical Cancer Diagnosis using Time-lapsed Colposcopic Images," IEEE Transactions on Medical Imaging, vol. 39, no. 11, pp. 3403–3415, 2020.
8. Ganasala, P.; Kumar, V. Feature-Motivated Simplified Adaptive PCNN-Based Medical Image Fusion Algorithm in NSST Domain. *J. Digit. Imaging* **2015**, *29*, 73–85. https://doi.org/10.1007/s10278-015-9806-4.
9. Huang, C.; Tian, G.; Lan, Y.; Peng, Y.; Ng, E.Y.K.; Hao, Y.; Cheng, Y.; Che, W. A New Pulse Coupled Neural Network (PCNN) for Brain Medical Image Fusion Empowered by Shuffled Frog Leaping Algorithm. *Front. Neurosci.* **2019**, *13*, 210. https://doi.org/10.3389/fnins.2019.00210.
10. Daneshvar, S.; Ghassemian, H. MRI and PET image fusion by combining IHS and retina-inspired models. *Inf. Fusion* **2009**, *11*, 114–123. https://doi.org/10.1016/j.inffus.2009.05.003.
11. Chen, C.-I. Fusion of PET and MR Brain Images Based on IHS and Log-Gabor Transforms. *IEEE Sensors J.* **2017**, *17*, 6995–7010, https://doi.org/10.1109/jsen.2017.2747220.
12. R. Gillespie, A. B. Kahle, and R. E.Walker, Color enhancement of highly correlated images. II. Channel ratio and "chromaticity" transformation techniques . Remote Sensing of Environment, 1987, 22(3): 343-365.
13. Nandi, D.; Ashour, A.S.; Samanta, S.; Chakraborty, S.; Salem, M.A.; Dey, N. Principal component analysis in medical image processing: a study. *Int. J. Image Min.* **2015**, *1*, 65, https://doi.org/10.1504/ijim.2015.070024.
14. Krishn, V. Bhateja, Himanshi, et al., Medical image fusion using combination of PCA and wavelet analysis . Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, 2020: 986-991.
15. Himanshi, V. Bhateja, A. Krishn, et al., An improved medical image fusion approach using PCA and complex wavelets. Proceedings of the 2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom), IEEE, 2019: 442-447.
16. S. P. Yadav, S.Yadav, Image fusion using hybrid methods in multimodality medical images . Medical & Biological Engineering & Computing, 2020, 58(4): 669-687.
17. He, C.; Liu, Q.; Li, H.; Wang, H. Multimodal medical image fusion based on IHS and PCA. *Procedia Eng.* **2010**, *7*, 280–285. https://doi.org/10.1016/j.proeng.2010.11.045.
18. V. D. Calhoun, T. Adali, Feature-based fusion of medical imaging data . IEEE Transactions on Information Technology in Biomedicine, 2008, 13(5): 711-720.
19. Z. Cui, G. Zhang, and J.Wu, Medical image fusion based on wavelet transform and independent component analysis . Proceedings of the 2009 International Joint Conference on Artificial Intelligence (IJCAI). IEEE, 2009: 480-483.
20. C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images Proceedings of the sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271), IEEE, 1998:839-846.
21. Kumar, B.K.S. Image fusion based on pixel significance using cross bilateral filter. *Signal, Image Video Process.* **2013**, *9*, 1193–1204. https://doi.org/10.1007/s11760-013-0556-9.
22. Hu, J.; Li, S. The multiscale directional bilateral filter and its application to multisensor image fusion. *Inf. Fusion* **2012**, *13*, 196–206. https://doi.org/10.1016/j.inffus.2011.01.002.
23. S. Li, X. Kang, and J. Hu, Image fusion with guided filtering . IEEE Transactions on Image Processing. 2013, 22(7): 2864-2875.
24. Li, W.; Jia, L.; Du, J. Multi-Modal Sensor Medical Image Fusion Based on Multiple Salient Features With Guided Image Filter. *IEEE Access* **2019**, *7*, 173019–173033. https://doi.org/10.1109/access.2019.2953786.
25. Jian, L.; Yang, X.; Zhou, Z.; Zhou, K.; Liu, K. Multi-scale image fusion through rolling guidance filter. *Futur. Gener. Comput. Syst.* **2018**, *83*, 310–325. https://doi.org/10.1016/j.future.2018.01.039.
26. Zhao, W.; Lu, H. Medical Image Fusion and Denoising with Alternating Sequential Filter and Adaptive Fractional Order Total Variation. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 2283–2294. https://doi.org/10.1109/tim.2017.2700198.
27. B. Biswas, A. Chakrabarti, and K. N. Dey, Spine medical image fusion using wiener filter in shearlet domain. Proceedings of the 2nd International Conference on Recent Trends in Information Systems (ReTIS). IEEE, 2015: 387-392.

28. Liu, X.; Mei, W.; Du, H. Multimodality medical image fusion algorithm based on gradient minimization smoothing filter and pulse coupled neural network. *Biomed. Signal Process. Control.* **2016**, *30*, 140–148. https://doi.org/10.1016/j.bspc.2016.06.013.
29. Jiang, W.; Yang, X.; Wu, W.; Liu, K.; Ahmad, A.; Sangaiah, A.K.; Jeon, G. Medical images fusion by using weighted least squares filter and sparse representation. *Comput. Electr. Eng.* **2018**, *67*, 252–266. https://doi.org/10.1016/j.compeleceng.2018.03.037.
30. Yin, H. Tensor Sparse Representation for 3-D Medical Image Fusion Using Weighted Average Rule. *IEEE Trans. Biomed. Eng.* **2018**, *65*, 2622–2633. https://doi.org/10.1109/tbme.2018.2811243.
31. Zong, J.-J.; Qiu, T.-S. Medical image fusion based on sparse representation of classified image patches. *Biomed. Signal Process. Control.* **2017**, *34*, 195–205. https://doi.org/10.1016/j.bspc.2017.02.005.
32. X. Wen, Image fusion based on improved IHS transform with weighted average . Proceedings of the 2011 International Conference on Computational and Information Sciences, IEEE, 2011: 111-113.
33. G. Pajares, J. M. De La Cruz, A wavelet-based image fusion tutorial . Pattern recognition, 2004, 37(9): 1855-1872.
34. Wang, H. Sun, and Y. Guan, The application of wavelet transform to multi-modality medical image fusion. Proceedings of the 2006 IEEE International Conference on Networking, Sensing and Control. IEEE, 2006: 270-274.
35. S. Cheng, J. He, and Z. Lv, Medical image of PET/CT weighted fusion based on wavelet transform . Proceedings of the 2nd International Conference on Bioinformatics and Biomedical Engineering. IEEE, 2008: 2523-2525.
36. Yang, Y.; Park, D.S.; Huang, S.; Rao, N. Medical Image Fusion via an Effective Wavelet-Based Approach. *EURASIP J. Adv. Signal Process.* **2010**, *2010*, 579341. https://doi.org/10.1155/2010/579341.
37. Vijayarajan, R.; Muttan, S. Discrete wavelet transform based principal component averaging fusion for medical images. *AEU - Int. J. Electron. Commun.* **2015**, *69*, 896–902. https://doi.org/10.1016/j.aeue.2015.02.007.
38. V. Bhavana, H. Krishnappa, Multi-modality medical image fusion using discrete wavelet transform . Procedia Computer Science, 2015, 70: 625-631.
39. Zhang, Y.; Liu, Y.; Sun, P.; Yan, H.; Zhao, X.; Zhang, L. IFCNN: A general image fusion framework based on convolutional neural network. *Inf. Fusion* **2019**, *54*, 99–118. https://doi.org/10.1016/j.inffus.2019.07.011.
40. Tajbakhsh, N.; Shin, J.Y.; Gurudu, S.R.; Hurst, R.T.; Kendall, C.B.; Gotway, M.B.; Liang, J. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? *IEEE Trans. Med. Imaging* **2016**, *35*, 1299–1312. https://doi.org/10.1109/tmi.2016.2535302.
41. Singh, S.; Anand, R.S. Multimodal Medical Image Fusion Using Hybrid Layer Decomposition With CNN-Based Feature Mapping and Structural Clustering. *IEEE Trans. Instrum. Meas.* **2019**, *69*, 3855–3865. https://doi.org/10.1109/tim.2019.2933341.
42. Wang, Z.; Li, X.; Duan, H.; Su, Y.; Zhang, X.; Guan, X. Medical image fusion based on convolutional neural networks and non-subsampled contourlet transform. *Expert Syst. Appl.* **2021**, *171*. https://doi.org/10.1016/j.eswa.2021.114574.
43. Wang, L.; Chang, C.; Liu, Z.; Huang, J.; Liu, C.; Liu, C. A Medical Image Fusion Method Based on SIFT and Deep Convolutional Neural Network in the SIST Domain. *J. Heal. Eng.* **2021**, *2021*, 1–8. https://doi.org/10.1155/2021/9958017.
44. Fu, J.; Li, W.; Du, J.; Huang, Y. A multiscale residual pyramid attention network for medical image fusion. *Biomed. Signal Process. Control.* **2021**, *66*. https://doi.org/10.1016/j.bspc.2021.102488.
45. L. Wang, C. Chang, B. Hao, et al., Multi-modal medical image fusion based on GAN and the shift-invariant shearlet transform . Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2020: 2538-2543.
46. H. M. El-Hoseny, E. S. M. El Rabaie, W. Abd Elrahman, et al., Medical image fusion techniques based on combined discrete transform domains . Proceedings of the 2017 34th National Radio Science Conference (NRSC), IEEE, 2017: 471-480.
47. Wang, Z.; Cui, Z.; Zhu, Y. Multi-modal medical image fusion by Laplacian pyramid and adaptive sparse representation. *Comput. Biol. Med.* **2020**, *123*, 103823. https://doi.org/10.1016/j.compbiomed.2020.103823.

48. Zhang, X.; Yan, H. Medical image fusion and noise suppression with fractional-order total variation and multi-scale decomposition. *IET Image Process.* **2021**, *15*, 1688–1701. https://doi.org/10.1049/ipr2.12137.
49. Li, X.; Zhou, F.; Tan, H.; Zhang, W.; Zhao, C. Multimodal medical image fusion based on joint bilateral filter and local gradient energy. *Inf. Sci.* **2021**, *569*, 302–325. https://doi.org/10.1016/j.ins.2021.04.052.
50. Tan, W.; Thitøn, W.; Xiang, P.; Zhou, H. Multi-modal brain image fusion based on multi-level edge-preserving filtering. *Biomed. Signal Process. Control.* **2020**, *64*, 102280. https://doi.org/10.1016/j.bspc.2020.102280.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.