BMC Medical Imaging

RESEARCH Open Access

Multimodal medical image fusion based on interval gradients and convolutional neural networks

Xiaolong Gu[1,2], Ying Xia[1*] and Jie Zhang[2]

Abstract

Many image fusion methods have been proposed to leverage the advantages of functional and anatomical images while compensating for their shortcomings. These methods integrate functional and anatomical images while presenting physiological and metabolic organ information, making their diagnostic efficiency far greater than that of single-modal images. Currently, most existing multimodal medical imaging fusion methods are based on multiscale transformation, which involves obtaining pyramid features through multiscale transformation. Low-resolution images are used to analyse approximate image features, and high-resolution images are used to analyse detailed image features. Different fusion rules are applied to achieve feature fusion at different scales. Although these fusion methods based on multiscale transformation can effectively achieve multimodal medical image fusion, much detailed information is lost during multiscale and inverse transformation, resulting in blurred edges and a loss of detail in the fusion images. A multimodal medical image fusion method based on interval gradients and convolutional neural networks is proposed to overcome this problem. First, this method uses interval gradients for image decomposition to obtain structure and texture images. Second, deep neural networks are used to extract perception images. Three methods are used to fuse structure, texture, and perception images. Last, the images are combined to obtain the final fusion image after colour transformation. Compared with the reference algorithms, the proposed method performs better in multiple objective indicators of Q_{EN}, Q_{NIQE}, Q_{SD}, Q_{SSEQ} and Q_{TMQI}.

Keywords Physiological information, Metabolic information, Interval gradient, Convolutional neural network, Perception image

Introduction

Multimodal medical imaging has become an indispensable component of artificial intelligence medicine. Additionally, its application has progressed through clinical work. Multimodal medical imaging is widely used for disease diagnosis and is important for planning, implementing, and evaluating the efficacy of surgical procedures and radiation therapy. Currently, medical imaging can be divided into functional and anatomical images based on the information features it reflects. Anatomic images describe information about the physiological dissecting structure of the human body, including X-ray imaging, CT imaging, and MRI imaging. Functional images, including PET, SPECT and fMRI images, reflect mainly dynamic changes in the metabolism and function of human organs or tissues. The information provided by multimodal imaging is complementary. Combining multimodal information in multimodal imaging is necessary to provide more comprehensive and rich information.

*Correspondence:
Ying Xia
xiaying@cqupt.edu.cn
[1] College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China
[2] National Research Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, Chongqing, China

Doctors can gain a more comprehensive understanding of disease conditions, better determine the location and scope of lesions, develop treatment plans, and evaluate treatment outcomes by combining multimodal medical imaging images with single images to obtain more comprehensive and accurate information, improving diagnostic accuracy and treatment outcomes. Therefore, medical image fusion technology has a wide range of applications in the medical field, such as in neuroscience, cardiovascular disease, radiology, tumour diagnosis, knowledge graphs, and data processing.

Medical image fusion consists of three main parts. First, transformation algorithms are used to decompose the source image into multifrequency coefficients. Second, according to the different coefficients, different strategies and methods are adopted for hierarchical and directional image fusion. Last, the fusion of medical images is achieved through multiscale inverse transformation. Pyramid transformation is an important multiscale transformation method in image fusion that uses a series of gradually decreasing resolution image sets obtained through row and column downsampling and that can effectively highlight important features and detailed information of the image, thereby improving the quality and visual effect of the fused image. Recently, many pyramid algorithms have been proposed. Li et al. [1] proposed Laplacian redecomposition for multimodal medical image fusion, Wang et al. [2] proposed a Laplacian pyramid and adaptive sparse representation for multimodal medical image fusion, and Yao et al. [3] proposed a Laplacian pyramid fusion network for infrared and visible image fusion. Singh et al. [4] proposed a multiresolution pyramid and bilateral filter for multifocus image fusion. He et al. [5] proposed a deep hierarchical pyramid network for superresolution mapping. Nair et al. [6] proposed a multisensor medical image fusion method based on pyramid-based DWT. Jin et al. [7] proposed a fusion method for visible and infrared images based on a contrast pyramid. Xu et al. [8] proposed an infrared and multitype image fusion algorithm based on contrast pyramid transform. These methods have achieved good performance in image fusion, with different characteristics and applicable scenarios, and suitable methods can be selected according to specific application needs. The pyramid transformation algorithm has been widely applied because of its high efficiency, ideal fusion effect, flexibility and scalability. However, the algorithm also has the drawbacks of redundant decomposition and no directionality. As the number of pyramid decomposition layers gradually increases, the resolution of the image gradually decreases, and the boundaries become increasingly blurred. Subsequently, fusion methods based on the wavelet transform are proposed. Bhat et al. [9] proposed a multifocus image fusion method using neutrosophic-based wavelet transform. Xu et al. [10] proposed medical image fusion via a modified shark smell optimization algorithm and hybrid wavelet-homomorphic filter. Aghamaleki et al. [11] proposed an image fusion method using a dual-tree discrete wavelet transform and weight optimization. Geng et al. * MERGEFORMAT [12] proposed adopting the quaternion wavelet transform to fuse multimodal medical images. Yang et al. * MERGEFORMAT [13] proposed a dual-tree complex wavelet transform and image block residual-based multifocus image fusion method in visual sensor networks. The two wavelet-based fusion methods overcome the disadvantages of the wavelet transform and achieve better fusion results. These methods decompose the image into low-frequency and multidirectional high-frequency coefficients. Low-frequency coefficients typically represent the overall contour and rough structure of an image, whereas high-frequency coefficients contain detailed and edge information about the image. This decomposition method fully reflects the local changes in the image, which enables image fusion or other image processing tasks to be performed accurately. The advantages of this method include being nonredundant and directional, and it overcomes the shortcomings of pyramid transformation methods. However, the wavelet transform does not have translation invariance; Therefore, improved wavelet transforms, such as contour waves, curved waves, and shear waves, are needed. Yang et al. [14] proposed a multimodal sensor medical image fusion method based on type-2 fuzzy logic in the NSCT domain. Peng et al. [15] proposed a multifocus image fusion approach based on CNP systems in the NSCT domain, and Li et al. [16] proposed an infrared and visible image fusion scheme based on NSCT and low-level visual features. Dong et al. [17] proposed high-quality multispectral and panchromatic image fusion technologies based on the curvelet transform. Arif and Wang [18] proposed a fast curvelet transform through a genetic algorithm for multimodal medical image fusion. Bhadauria and Dewal [19] proposed a medical image denoising method using adaptive fusion of curvelet transform and total variation. Gao et al. [20] proposed a multifocus image fusion method based on a nonsubsampled shearlet transform. Vishwakarma and Bhuyan [21] proposed an image fusion method using an adjustable nonsubsampled shearlet transform. Singh et al. [22] proposed a nonsubsampled shearlet-based CT and MR medical image fusion method using a biologically inspired spiking neural network. These methods have both translation invariance and directional selectivity. In addition, the decomposition level of multiscale transformation is one of the key factors affecting the fusion effect. The choice of decomposition level determines the depth at which the image is decomposed. The larger the decomposition level is, the more detailed the extracted information and the higher the fusion quality; however, the fusion time also increases. The balance among the decomposition layers, fusion quality, and time efficiency remains unresolved. The traditional multiscale transformation method using predefined fixed layers for feature extraction, such as spatial frequency and gradient energy, cannot generalize features. New mathematical models have been proposed for image fusion, which automatically adjusts the method and parameters of feature extraction based on the content and characteristics of the image, such as sparse representation (SR) [23–25] and the latest deep learning model based on convolutional neural networks (CNNs) [26–28], to achieve adaptive feature extraction.

The CNN is a deep learning algorithm that uses deep neural networks to process and analyse data; it is especially suitable for processing image data because of its fast computing speed and high efficiency. Its feature extraction is data driven and automatically generates parameter values after training with many data samples. Therefore, the features extracted by CNN methods have strong generalizability. As the depth of the network increases, its features become more abstract and precise, with the characteristics of translation, rotation, and a certain degree of scaling invariance. The CNN image fusion method can learn features hierarchically with more diverse feature expressions, stronger discriminative performance, and better generalization performance. However, its disadvantages include long training times and a lack of dedicated training sets. Zhang et al. [29] proposed a framework (IFCNN) based on a convolutional neural network, which is an innovative image processing method. This method has good universality and is suitable for multimodal image fusion. However, owing to its simple network structure, it cannot effectively extract deep image features, making the fusion process prone to losing detailed information and causing image blurring. Zhang et al. * MERGEFORMAT [30] proposed a method that learns to search for a lightweight generalized network for medical image fusion. This method introduces segmentation masks for preserving important pathological image region details. Although this method preserves important details of pathological image regions, it increases the model complexity and reduces its adaptability.

In this study, first, a multimodal medical image fusion method based on interval gradients and convolutional neural networks is proposed in response to the problems with existing fusion methods. It first uses interval gradients to decompose multimodal medical images, obtaining structure and texture images. Second, deep convolutional neural networks are used to extract perception images to obtain detailed features of medical images. Last, three different methods are used to fuse the structure, texture, and perception images. After colour transformation, the final fusion image is obtained. The experimental results show that the fusion results of the proposed method are significantly better than those of the reference methods in terms of details and fusion indicators.

Related work

This section introduces interval gradients [31] and visual geometry group (VGG) networks [32]. The interval gradient is used for structure-texture image decomposition. The VGG network is used for perception image extraction.

Interval gradient

First, the problem is defined as $\nabla I = \nabla S + \nabla T$. In a one-dimensional discrete signal I, the interval gradient is defined as follows:

$$(\nabla_{\Omega} I)_p = g_{\sigma}^{r}(I_p) - g_{\sigma}^{l}(I_p) \tag{1}$$

where $g_{\sigma}^{r}(I_p) = \frac{1}{k_r} \sum_{n \in \Omega(p)} w_{\sigma}(n-p-1) I_n$, $g_{\sigma}^{l}(I_p) = \frac{1}{k_l} \sum_{n \in \Omega(p)} w_{\sigma}(p-n) I_n$, g_{σ}^{r} and g_{σ}^{l} are Gaussian filtering functions with right cropping and left cropping, respectively, and where w_{σ} is an exponential weighting function with a scaling parameter of σ. $w_{\sigma}(x) = \begin{cases} \exp(-\frac{x^2}{2\sigma^2}) & if\ x \geq 0 \\ 0 & otherwise \end{cases}$. k_r and k_l are regularization coefficients. $k_r = \sum_{n \in \Omega(p)} w_{\sigma}(n-p-1)$, and $k_l = \sum_{n \in \Omega(p)} w_{\sigma}(p-n)$.

Unlike forwards differentiation, interval gradient measurement considers not only the grayscale or colour information of the pixel itself but also the information of its surrounding pixels by calculating the difference between the weighted average values of the Gaussian filtering functions cropped left and right. This weighted average is calculated on the basis of the distribution of signals around pixels, so it can better reflect the actual changes between two pixels. For structure element p, since the smoothing kernel, which is typically used to reduce noise and detail levels in data or images, amplifies the gradient, $(\nabla \Omega I)p$ is greater than $(\nabla I)p$. $(\nabla \Omega I)p$ is smaller than $(\nabla I)p$ because of the cancellation of oscillation gradients of different signals for texture element p.

Second, gradient scaling is performed via interval gradients, which increase the difference between the texture region and the structure region (Ωp is defined here as the structure region if and only if the signal increases or decreases but never oscillates within the region of Ωp). The scaling formula is as follows:

$$(\nabla' I)_p = \begin{cases} (\nabla I)_p \cdot w_p & if\ sign((\nabla I)_p) = sign((\nabla_\Omega I)_p) \\ 0 & otherwise \end{cases} \tag{2}$$

where $(\nabla' I)_p$ represents the new gradient and w_p represents the scaling weight. $w_p = \min(1, \frac{|(\nabla_\Omega I)_p| + \in_s}{|(\nabla I)_p| + \in_s})$. The gradient remains unchanged for the structure edges and the smooth changing areas because $|(\nabla_\Omega I)_p| \geq |(\nabla I)_p|$, making w_p equal to 1. For texture regions with oscillation modes and noise, $|(\nabla_\Omega I)_p| < |(\nabla I)_p|$, making $w_p < 1$.

The cumulative results may still contain small unfiltered oscillations due to the local readjustment of gradients by the above method; therefore, it is necessary to correct the reconstructed signal. The authors choose guided filtering for correction because it can preserve edge (corner) information in the image without introducing gradient distortion or oversharpening edges. For two-dimensional images, alternating one-dimensional filtering in the x and y directions is adopted for domain transformation filtering. Last, the iterations are repeated multiple times to obtain the final structure map via interval gradient calculation, gradient smoothing, and iterative one-dimensional filtering, as defined above. The structure map is subsequently subtracted from the original image to obtain the texture image.

The advantages of using interval gradients for image structure and texture decomposition are as follows:

(1) This method does not directly filter the image colour but generates high-quality filtering results by manipulating the image gradient. This method achieved better results than the methods based on previous filters.
(2) The interval gradient operator is an effective tool for distinguishing image texture and structure.
(3) Compared with existing filtering methods, this method avoids gradient inversion of filtering results, preserves strong features, and maintains simplicity and highly parallel implementation.

VGG network

The VGG was proposed by Oxford's Visual Geometry Group. Since the emergence of AlexNet, many scholars have improved their accuracy by improving the AlexNet structure in two main directions: small convolution kernels and multiple scales. The VGG authors chose to increase the network depth. Increasing the network depth does impact the final network performance. By designing a reasonable network structure and adopting appropriate regularization techniques and optimization algorithms, it is possible to slightly balance feature extraction ability, generalization ability, and overfitting risk, thus obtaining a more efficient deep learning model. Two types of VGG structures exist, namely, VGG16 and VGG19. The two methods have no fundamental difference; only the network depth is different. During this process, the authors conducted six sets of experiments corresponding to six different network models. As the depth of these six networks gradually increased, their characteristics also increased.

By using stacked small convolution kernels to maintain the same receptive field, increasing the network depth and reducing the number of parameters can improve model performance and reduce computational complexity. Specifically, in VGG, multiple 3×3 convolution kernels were used instead of 5×5 or 7×7 convolution kernels. This improves the network depth and slightly enhances the network performance k while ensuring the same perception field. A significant improvement in VGG16 over AlexNet is the use of several consecutive 3×3 convolution kernels (step size = 1 and padding = 0) to replace the larger kernels in AlexNet, such as 11×11, 7×7, and 5×5 kernels. The VGGNet achieves a structurally simple and efficient network design by adopting a uniformly sized convolutional kernel and maximum pooling size. Replacing the large filter (5×5 or 7×7) convolutional layer with a combination of several small filter (3×3) convolutional layers is an effective strategy, which verifies the improvement in model performance by deepening the network structure. The VGG19 network structure (Fig. 1) consists of 16 convolutional layers and 3 fully connected layers, with a total of 19 layers. Although there are many VGG structures, because the VGG19 network has the deepest layers and extracts more comprehensive features, this study uses the VGG19 network. In addition, the VGG network was proposed by the Visual Geometry Group of the University of Oxford, which validated the advantages of the VGG network structure and performance through comparative experiments.

Proposed method

Schematic

We use interval gradients for image decomposition to obtain structure and texture images to achieve medical image fusion. We use VGG19 to extract detailed images to enhance the detailed information of the fusion images. We use the local pixel mean value for structure images, the local pixel maximum value for texture images and the spatial frequency for detailed images to achieve fusion. After the three images are fused, they are combined, and the final fusion image is obtained through colour transformation. The proposed schematic is shown in Fig. 2. In the process of image decomposition, the colour image

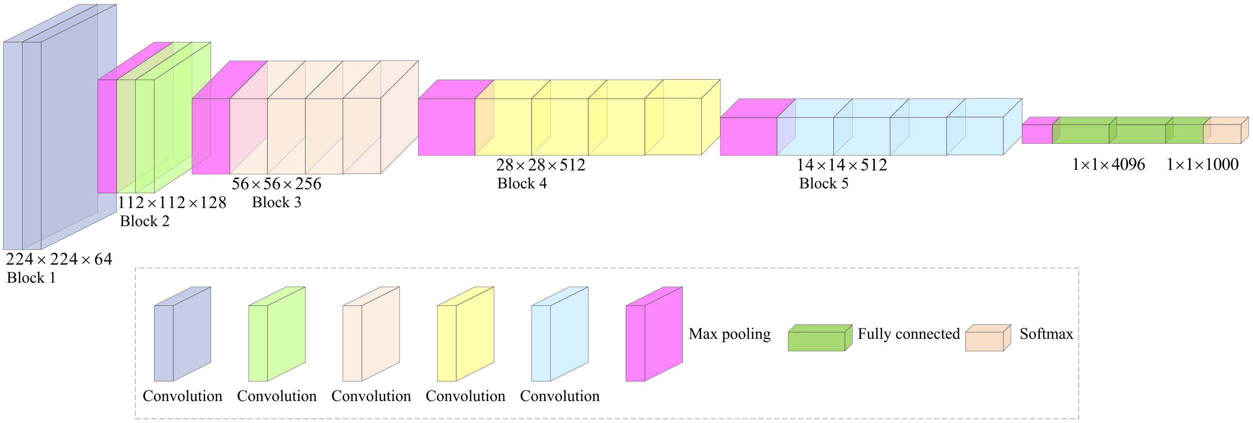


Fig. 1 VGG19 network

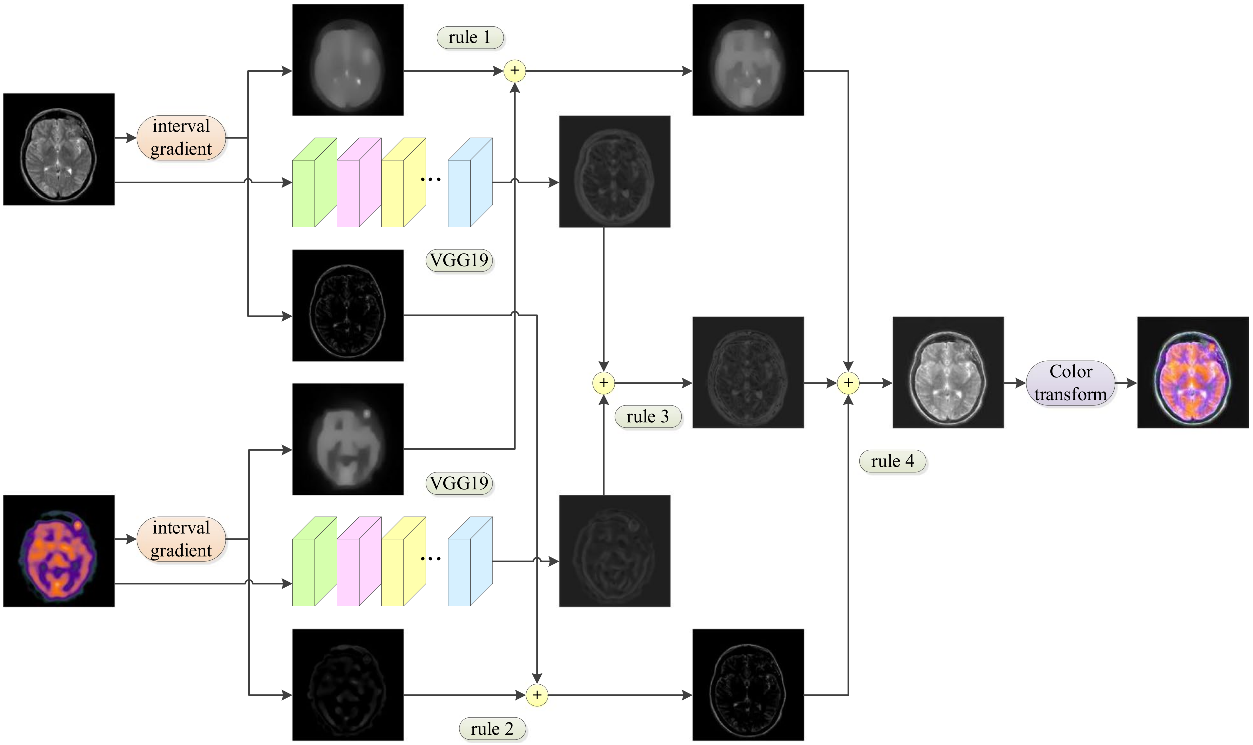


Fig. 2 Schematic of image fusion based on the interval gradient and VGG19 networks

is decomposed into YCbCr three components, and the fusion process is achieved using the brightness information Y. After the brightness image fusion is completed, the colour fused image is obtained through YCbCr inverse transformation.

Image decomposition process

Image structure-texture decomposition technology can decompose an image into structural components that contain the main information and determine the subjective understanding of the image content and texture components that contain the main details and do not affect the subjective understanding of the image content based on the different characteristics of the image. The perception images use deep neural networks to extract feature images from input images, which have detailed features that simulate human visual mechanisms. First, interval gradients for image structure and texture decomposition

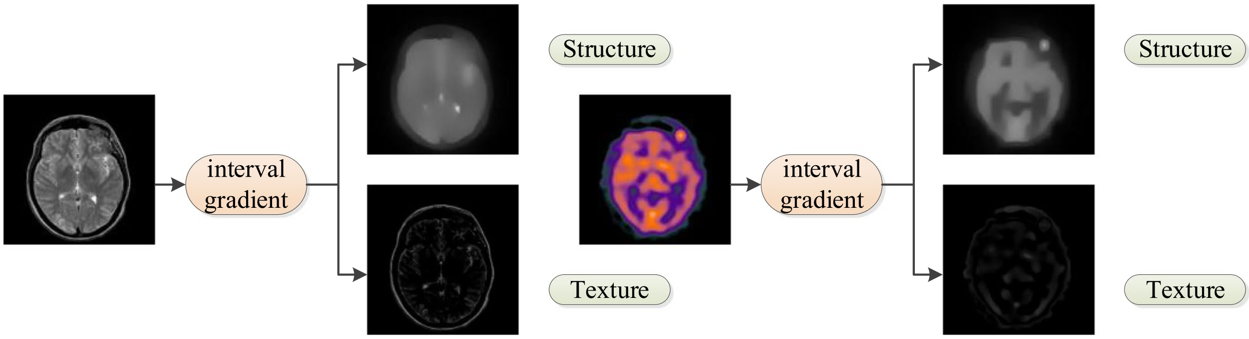


Fig. 3 Structure and texture decomposition based on the interval gradient

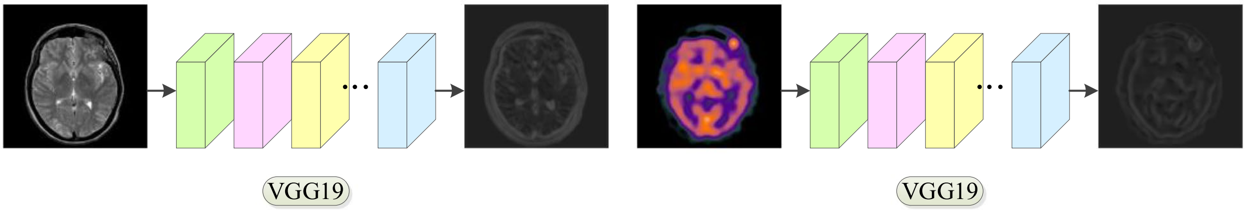


Fig. 4 Perceptual image extraction based on the VGG19 network

are used. Second, the perception images are extracted via the VGG19 network. The image structure-texture decomposition and perception image extraction processes are shown in Figs. 3 and 4, respectively.

Image fusion process

The image fusion process continues after image structure texture decomposition and perception image extraction. First, three rules are used to fuse three brightness images: structure, texture, and perception. Rule 1 is used for structure image fusion, rule 2 is used for texture image fusion, and rule 3 is used for perception image fusion. After the three image fusion processes are completed, rule 4 is used for the final image fusion, and colour transformation is performed to obtain the final fusion image.

$$
\begin{aligned}
A_{i,j}^{1} &= \frac{1}{9}(X_{i-1,j-1}^{1} + X_{i-1,j}^{1} + X_{i-1,j+1}^{1} + X_{i,j-1}^{1} + X_{i,j}^{1} + X_{i,j+1}^{1} + X_{i+1,j-1}^{1} + X_{i+1,j}^{1} + X_{i+1,j+1}^{1})\\
A_{i,j}^{2} &= \frac{1}{9}(X_{i-1,j-1}^{2} + X_{i-1,j}^{2} + X_{i-1,j+1}^{2} + X_{i,j-1}^{2} + X_{i,j}^{2} + X_{i,j+1}^{2} + X_{i+1,j-1}^{2} + X_{i+1,j}^{2} + X_{i+1,j+1}^{2})\\
S_{i,j} &= \begin{cases} X_{i,j}^{1} \ \ if \ \ A_{i,j}^{1} \geq A_{i,j}^{2} \\ X_{i,j}^{2} \ \ else \end{cases}
\end{aligned}
\tag{3}
$$

where S is the structure fusion image, X is the decomposed structure image, and A is the local pixel mean value of the structure image.

$$
\begin{aligned}
E_{i,j}^{1} &= \max(Y_{i-1,j-1}^{1} + Y_{i-1,j}^{1} + Y_{i-1,j+1}^{1} + Y_{i,j-1}^{1} + Y_{i,j}^{1} + Y_{i,j+1}^{1} + Y_{i+1,j-1}^{1} + Y_{i+1,j}^{1} + Y_{i+1,j+1}^{1})\\
E_{i,j}^{2} &= \max(Y_{i-1,j-1}^{2} + Y_{i-1,j}^{2} + Y_{i-1,j+1}^{2} + Y_{i,j-1}^{2} + Y_{i,j}^{2} + Y_{i,j+1}^{2} + Y_{i+1,j-1}^{2} + Y_{i+1,j}^{2} + Y_{i+1,j+1}^{2})\\
T_{i,j} &= \begin{cases} Y_{i,j}^{1} \ \ if \ \ E_{i,j}^{1} \geq E_{i,j}^{2} \\ Y_{i,j}^{2} \ \ else \end{cases}
\end{aligned}
\tag{4}
$$

where T is the texture fusion image, Y is the decomposed texture image, and E is the local pixel maximum value of the texture image.

$$
\begin{aligned}
SF_{i,j}^{1} &= \sqrt{\frac{1}{MN}\sum\nolimits_{i=1}^{M}\sum\nolimits_{j=2}^{N}(Z_{i,j}^{1}-Z_{i,j-1}^{1})^{2}+\frac{1}{MN}\sum\nolimits_{i=2}^{M}\sum\nolimits_{j=1}^{N}(Z_{i,j}^{1}-Z_{i-1,j}^{1})^{2}} \\
SF_{i,j}^{2} &= \sqrt{\frac{1}{MN}\sum\nolimits_{i=1}^{M}\sum\nolimits_{j=2}^{N}(Z_{i,j}^{2}-Z_{i,j-1}^{2})^{2}+\frac{1}{MN}\sum\nolimits_{i=2}^{M}\sum\nolimits_{j=1}^{N}(Z_{i,j}^{2}-Z_{i-1,j}^{2})^{2}} \\
P_{i,j} &= \begin{cases} Z_{i,j}^{1} \ \ if \ \ SF_{i,j}^{1} \geq SF_{i,j}^{2} \\ Z_{i,j}^{2} \ \ else \end{cases}
\end{aligned}
\tag{5}
$$

where P is the perception fusion image, Z is the extracted perception image, and SF is the spatial frequency of the perception image.

$$F_{i,j} = S_{i,j} + T_{i,j} + P_{i,j} \tag{6}$$

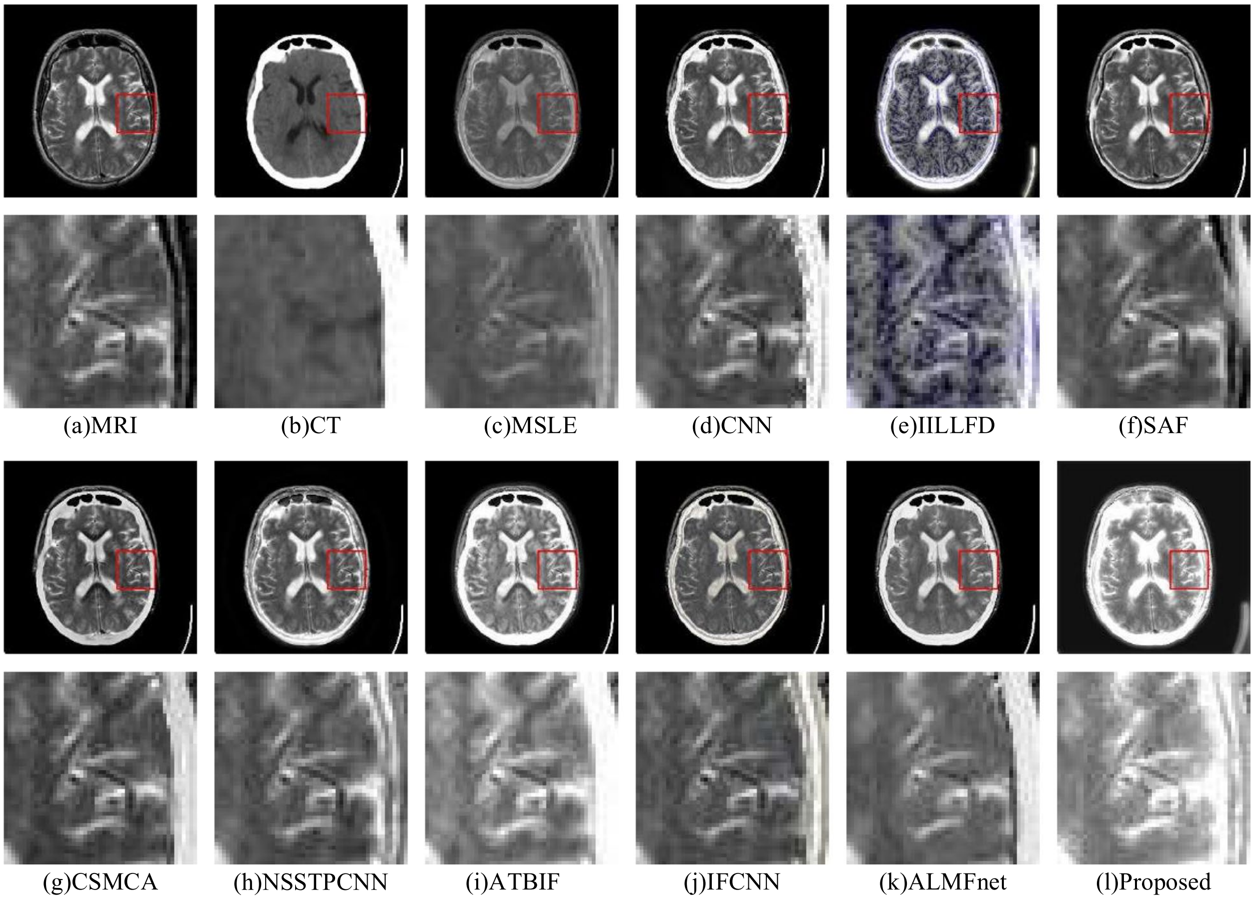
where F is the final fusion brightness image, C is the cartoon fusion image, T is the texture fusion image, and P is the perception fusion image.

Results and analysis

This section introduces the experimental methods, objective indicators, and experimental comparisons. We analysed the differences between the fusion results of the proposed method and the reference methods and conducted corresponding indicator comparisons and analyses in the comparative experiments.

Experimental preparation

The corresponding multimodal medical image dataset must be prepared before the experiments. We prepared


(a)MRI (b)CT (c)MSLE (d)CNN (e)IILLFD (f)SAF

(g)CSMCA (h)NSSTPCNN (i)ATBIF (j)IFCNN (k)ALMFnet (l)Proposed

Fig. 5 Comparison of 9 different algorithms in the MR-PET image fusion experiments

90 MR images, 30 CT images, 30 PET images, and 30 SPECT images. All images were sourced from Harvard Medical School, and all sizes were 256×256 (the official website for the data download is http://www.med.harvard.edu/aanlib/home.html). The parameters of different methods are fixed in the comparison experiments, including methods based on deep learning. When the input images are fed into these algorithms, the fusion images can be obtained through computation. Therefore, we use five indicators, Q_{EN}, Q_{NIQE}, Q_{SD}, Q_{SSEQ} and Q_{TMQI}, to compare the fusion results of different algorithms.

The experimental environment is based on a laptop computer with a Windows 11 (64-bit) OS, an Intel(R) Core (TM) i9-14900HX CPU, 32 cores and 32 GB of RAM, MATLAB R2023b software, and an NVIDIA GeForce RTX 4060 laptop GPU (8 GB).

Objective evaluation indicators

Objective evaluation indicators are used to evaluate the differences between the fusion and input images and the fused image quality. This study uses five objective evaluation indicators: Q_{EN}[33–36] and Q_{TMQI}[37]. The larger the values of Q_{EN}, Q_{SD} and Q_{TMQI} are, the better the image quality and the less information lost. The smaller the indicators Q_{NIQE} and Q_{SSEQ} are, the better the image quality.

where Q_{EN} represents the distribution and aggregation of grayscale values in an image, which reflects the degree and distribution characteristics of grayscale values in the image. The larger Q_{EN} is, the greater the amount of information it contains, and vice versa. The formula for Q_{EN} is as follows:

$$Q_{EN} = -\sum_{i=0}^{255} E[\log \frac{1}{p_i}] = -\sum_{i=0}^{255} p_i \log p_i \tag{7}$$

where i represents a possible value for a random variable with grayscale values and p_i represents the probability of this value, which can be obtained from the grayscale histogram.

Q_{NIQE} fits a multivariate Gaussian model based on specific features extracted from a series of natural images. In this way, the difference between the test

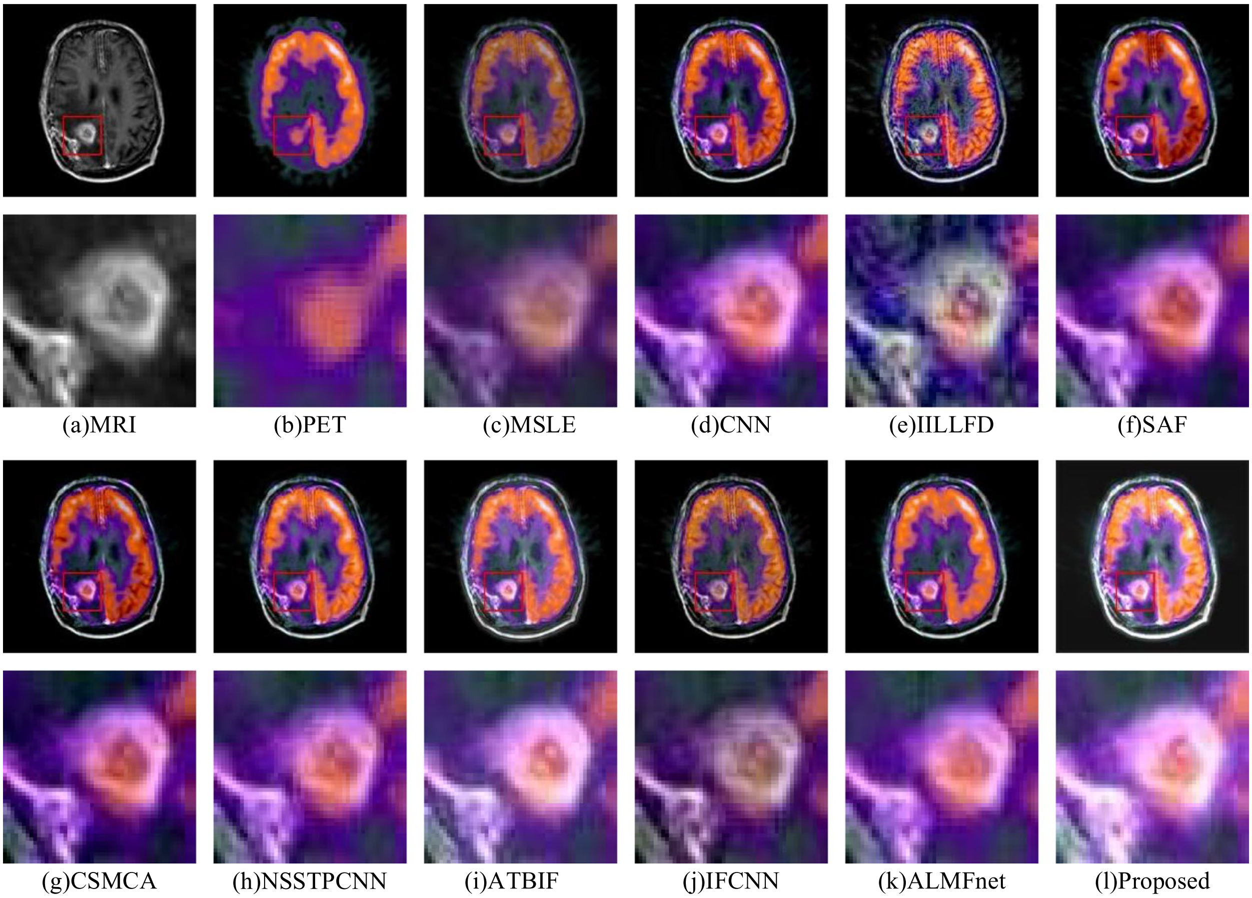


Fig. 6 Comparison of 9 different algorithms in the MR-PET image fusion experiments

image and this multivariate distribution can be measured to evaluate the quality of the fused image. Q_{NIQE} measures the differences in the multivariate distributions of the fusion images. The smaller the value of Q_{NIQE} is, the smaller the difference in the multivariate Gaussian distribution (MVN, also known as the multivariate normal distribution) of the fusion image, and the higher the fusion quality. The formula for Q_{NIQE} is as follows:

$$Q_{NIQE} = \sqrt{(v_f - v_n)^T (\frac{\sigma_f + \sigma_n}{2})^{-1} (v_f - v_n)} \tag{8}$$

where v_f and v_n represent the mean vectors of the multivariate normal model (MVG) for the fusion images and original images, respectively. The mean vector represents the centre position on each feature, which is obtained by averaging the values in each dimension. σ_f and σ_n are the covariance matrices of the multivariate normal distribution for the fusion images and original images, respectively. The covariance matrix describes the changing relationships between different features and the shape and direction of the data, which describes the correlation or covariance relationship between multidimensional random variables.

Q_{SD} is an objective evaluation indicator for measuring the richness of image information. This indicator describes the distribution or degree of dispersion of image pixel values. The larger the standard deviation is, the richer the information carried by the fusion image, and the better the fusion quality.

$$Q_{SD} = \sqrt{\frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (I_{i,j} - \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I_{i,j})^2} \tag{9}$$

where I represents the fusion image and M N represent the sizes of the images.

Q_{SSEQ} simulated images via local spatial entropy, which is calculated based on the probability of grayscale appearing in local space, and spectral entropy, which is calculated from the normalized power spectrum via the entropy function features to evaluate image quality without reference. This reference-free image quality evaluation method avoids the time-consuming and laborious problems of subjective evaluation and can perform

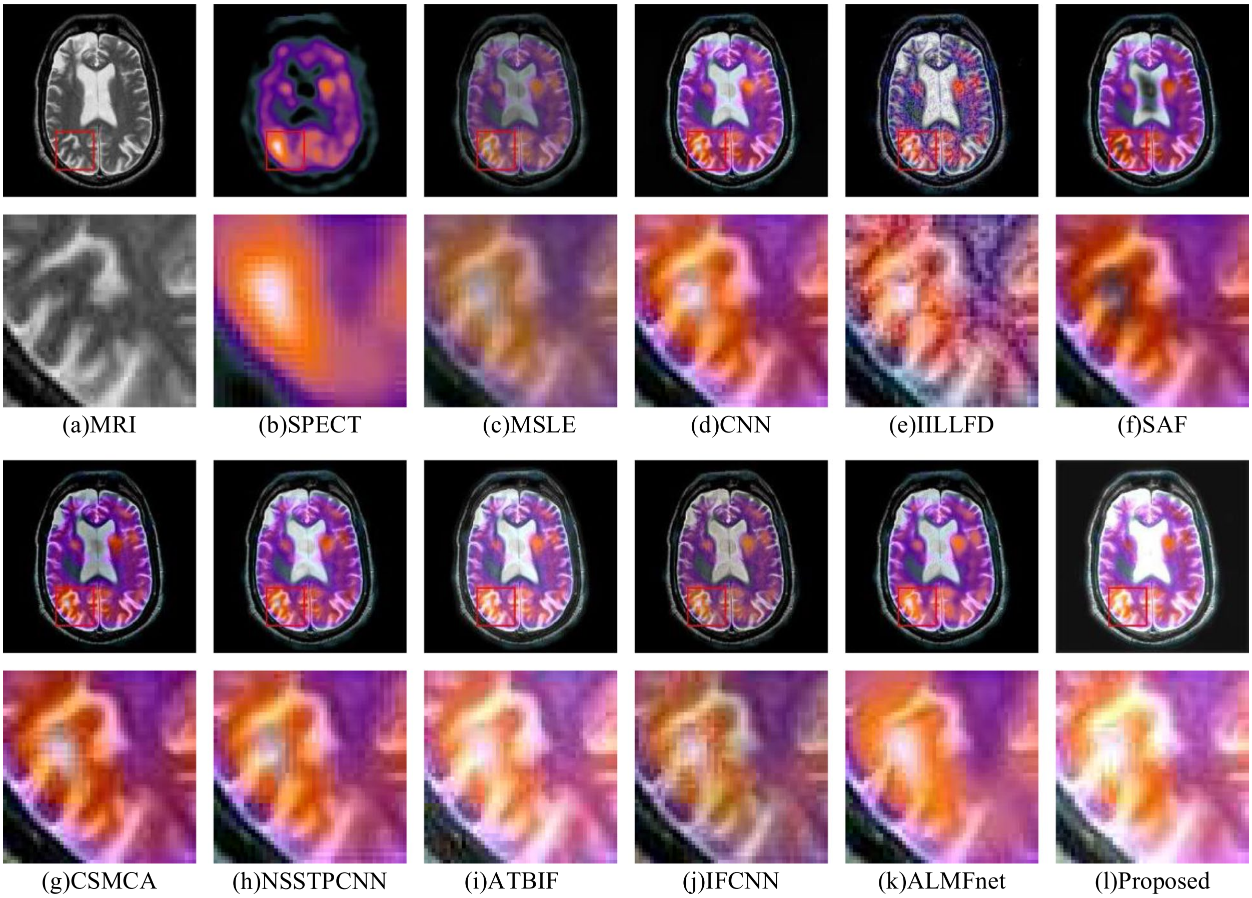


Fig. 7 Comparison of the 9 different algorithms in the MR-SPECT image fusion experiments

real-time quality evaluation on large-scale images. The smaller Q_{SSEQ} is, the better the image quality and the less distortion there is.

$$Q_{SSEQ} = (mean(S_c), skew(S), mean(F_c), skew(F)) \tag{10}$$

where se_i is the local spatial entropy, which is used mainly to describe the characteristics of the spatial structure; fe_i is the spectral entropy, which describes the complexity and randomness of signals in the frequency domain; m is the number of blocks at each image scale; $F = (fe_1, fe_2, \cdots, fe_m)$, $F_c = (fe_{\lfloor 0.2m \rfloor}, fe_{\lfloor 0.2m \rfloor+1}, \cdots, fe_{\lfloor 0.8m \rfloor})$, $S_c = (se_{\lfloor 0.2m \rfloor}, se_{\lfloor 0.2m \rfloor+1}, \cdots, se_{\lfloor 0.8m \rfloor})$; and $S = (se_1, se_2, \cdots, se_m)$.

Q_{TMQI} is an indicator used to evaluate the quality of images generated by the tone-mapped operator (TMO), which evaluates the quality of the fused image and the original image (also known as the input image or reference image) in three ways: brightness, contrast, and structure. By calculating the specific values of brightness, contrast, and structure and combining these values, a comprehensive evaluation result is obtained. This evaluation result can objectively reflect the quality of the image after tone mapping, which helps select and optimize tone mapping operators to generate higher-quality images.

Q_{TMQI} is defined as Eq. (11). The larger Q_{TMQI} is, the better the fusion quality.

$$Q_{TMQI} = aT^{\alpha} + (1 - a)M^{\beta} \tag{11}$$

where T and M are the structural fidelity and statistical characteristics, respectively, of the image. The constants are set to $a = 0.8012$, $\alpha = 0.3046$, and $\beta = 0.7088$.

Comparison experiments

The experimental results of the proposed method are compared with those of various different image fusion reference methods. There are 9 different image fusion methods, including MSLE [38], CNN [39], IILLFD [40], SAF [41], CSMCA [42], NSSTPCNN [43], ATBIF [44], IFCNN [29], and ALMFnet * MERGEFORMAT [30]. Three fusion experiments were conducted, and comparisons of the MRI-CT, MRI-PET, and MRI-SPECT results are shown in Figs. 5, 6 and 7, respectively.

Table 1 Comparison of 5 indicators for MR-CT image fusion via different methods

Metrics\Methods	Q_{EN}	Q_{NIQE}	Q_{SD}	Q_{SSEQ}	Q_{TMQI}
MSLE	4.6466	**4.6335**	61.0272	36.7649	0.7449
CNN	4.8972	4.6855	83.5727	38.6993	0.7380
IILLFD	5.4442	5.1168	87.0383	36.8426	0.7637
SAF	4.8472	5.0295	72.0968	38.8271	0.7146
CSMCA	4.6065	4.8064	73.4893	38.4118	0.7350
NSSTPCNN	**5.4808**	4.7112	82.5737	39.1697	0.7459
ATBIF	5.0296	4.8171	**97.6451**	39.4895	0.8120
IFCNN	4.8160	4.9661	76.6224	36.0188	0.7503
ALMFnet	4.7329	4.7899	87.5196	38.1332	0.7715
Proposed	5.4404	4.9646	95.8076	**33.8225**	**0.8555**

Table 2 Comparison of 5 indicators for MR-PET image fusion via different methods

Metrics\Methods	Q_{EN}	Q_{NIQE}	Q_{SD}	Q_{SSEQ}	Q_{TMQI}
MSLE	4.2210	6.0820	45.1741	42.9560	0.7447
CNN	4.4071	5.3262	58.3163	43.3657	0.7344
IILLFD	4.6534	6.4310	62.8520	40.6984	0.7454
SAF	4.2149	6.0070	55.7454	43.3833	0.7227
CSMCA	3.8834	5.7297	54.4983	44.1161	0.7303
NSSTPCNN	4.5112	5.4140	62.6580	46.5436	0.7461
ATBIF	4.7968	5.1955	71.3991	42.2596	0.7659
IFCNN	4.1641	5.4674	55.9673	41.0903	0.7453
ALMFnet	4.2749	5.5082	64.1952	41.8377	0.7484
Proposed	**5.3918**	**5.1417**	**75.9690**	**37.0941**	**0.7923**

Figure 5a and b show the MR images and CT images, respectively. Although the anatomical details of Fig. 5c, g, h and i are clear, the brightness of their high-density bone tissue is low. The anatomical details and brightness of the high-density bone tissue in Fig. 5d are unclear and low, respectively. The high-density bone tissues in Fig. 5e, j and k have high brightness; however, their white matter brightness is low. There is a noticeable colour distortion in Fig. 5f. Compared with those of the reference methods, the fusion image of the proposed method has more high-density bone tissue and brighter white matter in the brain.

Figure 6 shows that Fig. 6a and b are MR images and PET images, respectively, of gliomas. The loss of detailed glioma information leads to ambiguity in Fig. 6c and d.

Table 3 Comparison of 5 indicators for MR-SPECT image fusion via different methods

Metrics\Methods	Q_{EN}	Q_{NIQE}	Q_{SD}	Q_{SSEQ}	Q_{TMQI}
MSLE	4.6025	5.2363	44.3134	38.4644	0.7198
CNN	5.2676	4.5546	59.2894	38.2363	0.7123
IILLFD	5.0084	5.4090	58.8633	**35.7919**	0.7178
SAF	4.7529	5.0330	57.5773	37.7450	0.7011
CSMCA	3.8587	6.0024	37.0609	41.0763	0.6784
NSSTPCNN	4.8537	4.9774	59.3227	38.3072	0.7133
ATBIF	5.2551	4.4303	65.8140	40.3710	0.7411
IFCNN	4.5679	4.6846	49.9392	37.2265	0.7171
ALMFnet	4.7399	5.1779	56.9947	40.2024	0.7252
Proposed	**5.8379**	**4.3497**	**73.5825**	35.8205	**0.7738**

There is a noticeable colour distortion phenomenon in Fig. 6f and k. The fusion results of Fig. 6e, g, h and i are similar; however, the glioma brightness is not high. The fusion result of Fig. 6j is similar to that of Fig. 6l; however, the brightness of the glioma in Fig. 6l is greater.

Figure 7a and b show the MRI images and SPECT images, respectively. The brightness values of Fig. 7c, d and k are low, and important colour information is lost. Colour distortion and considerable noise occur in Fig. 7f.

The colour information of Fig. 7g is clear, but the brightness of the lesion area is the lowest. The colour information of Fig. 7e, h and i is well preserved; however, the brightness of the lesion area is not high. Compared with the other methods, the proposed method results in greater brightness at the lesion site, as shown in Fig. 7l.

The fusion indices of the MR-CT, MR-PET, and MR-SPECT image pairs are shown in Tables 1, 2 and 3, respectively. In these tables, bold numbers indicate

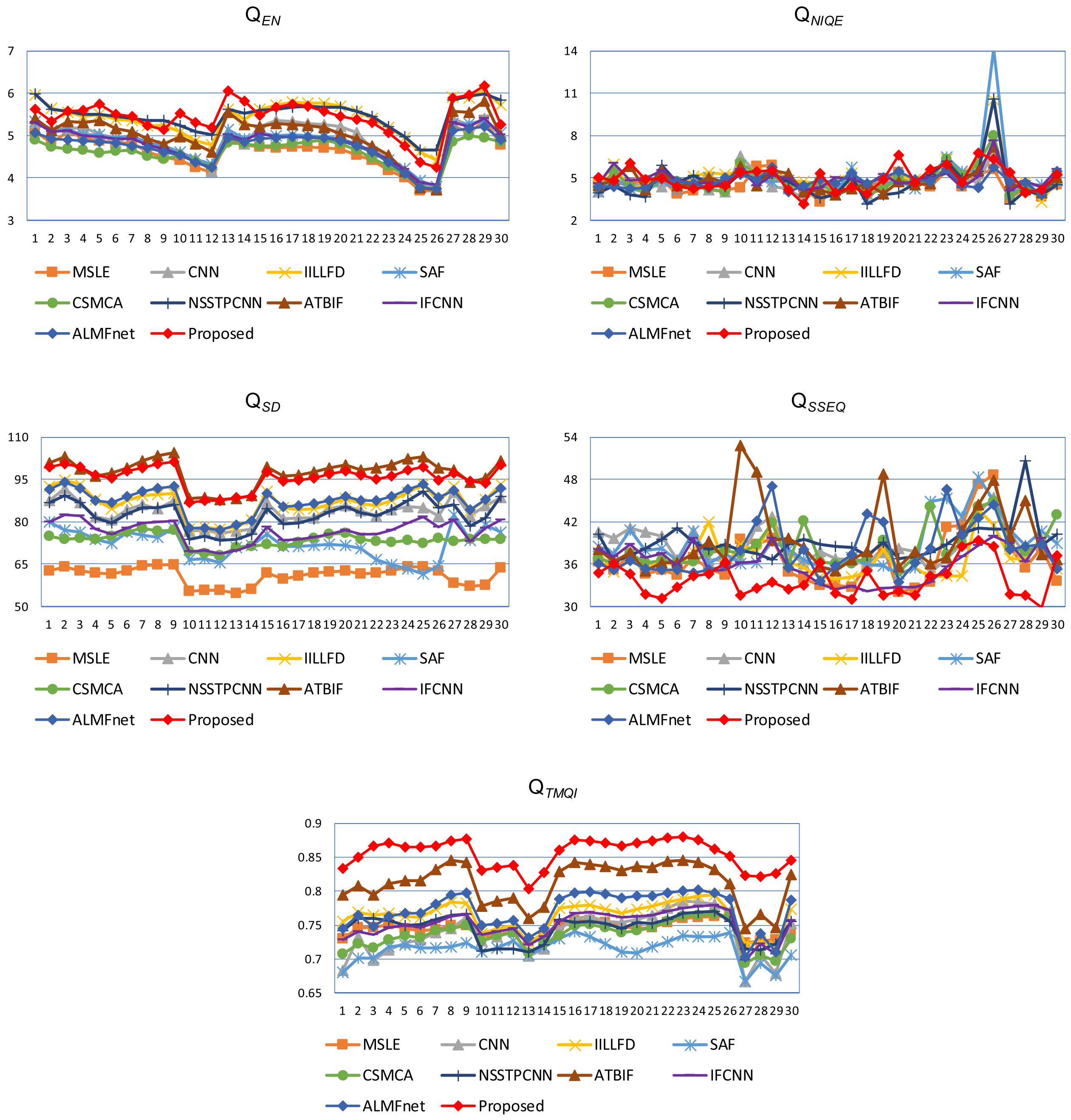


Fig. 8 Line chart of MR-CT image fusion indicators

optimal results. In terms of the MR-CT image fusion metrics, the proposed method has two optimal metrics: Q_{SSEQ} and Q_{TMQI}. Among the MRI–PET fusion indicators, five optimal indicators were proposed for this method: Q_{EN}, Q_{NIQE}, Q_{SD}, Q_{SSEQ} and Q_{TMQI}. Among the MRI–SPECT fusion indicators, there are four optimal indicators, namely, Q_{EN}, Q_{NIQE}, Q_{SD} and Q_{TMQI}. In Table 2, although all the values are similar, the proposed algorithm outperforms the comparison algorithm in all five metrics. In addition, the details and colour fidelity of the proposed algorithm are greater than those of the reference algorithm in Fig. 6.

To better compare the indicators of different fusion methods, the MR-CT, MR-PET, and MR-SPECT indicator line graphs for the three types of images are drawn separately, as shown in Figs. 8, 9 and 10. The comparison results show that the proposed method performs significantly better than the reference methods in terms of Q_{EN},

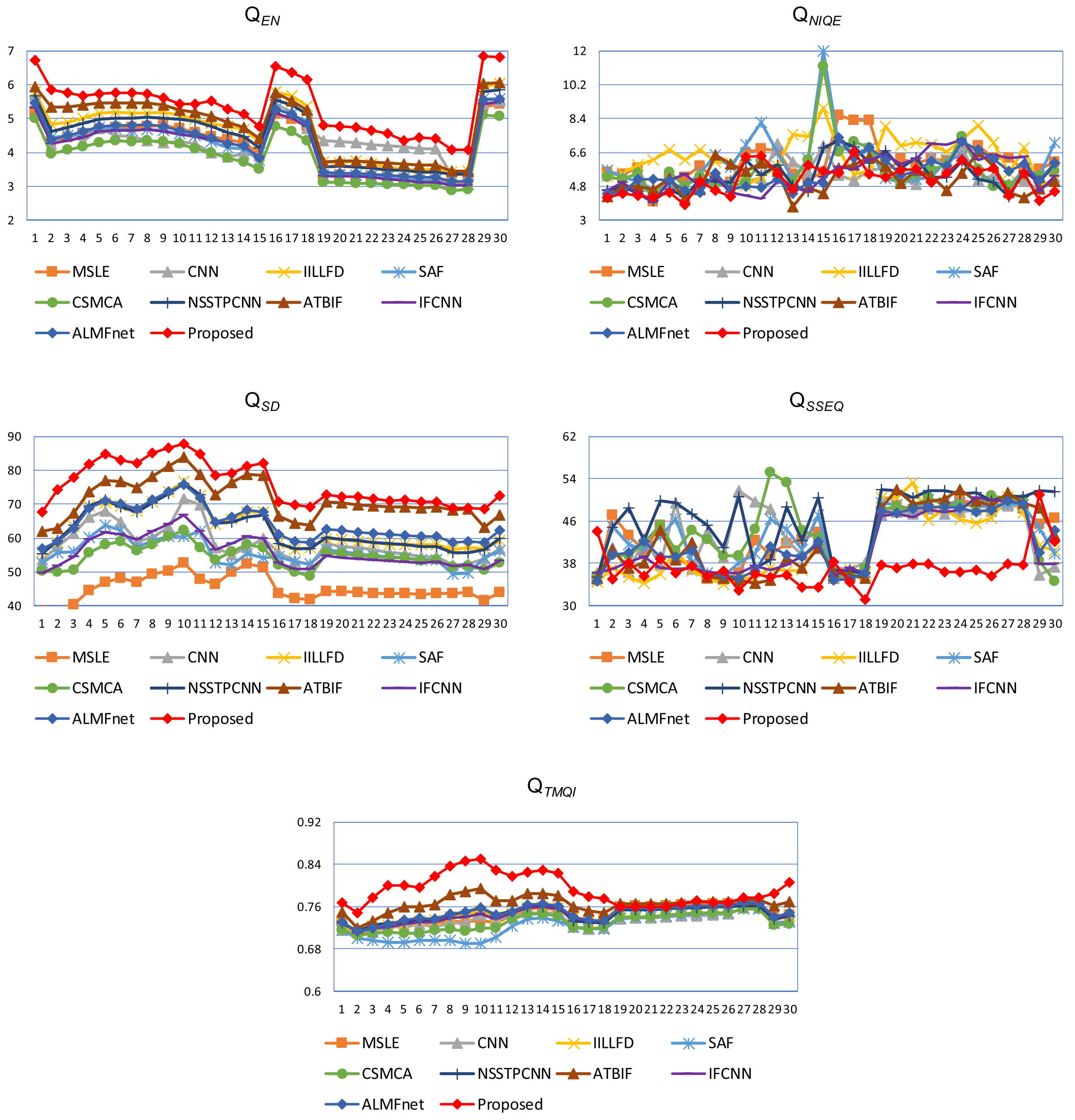


Fig. 9 Line chart of MR-PET image indicators

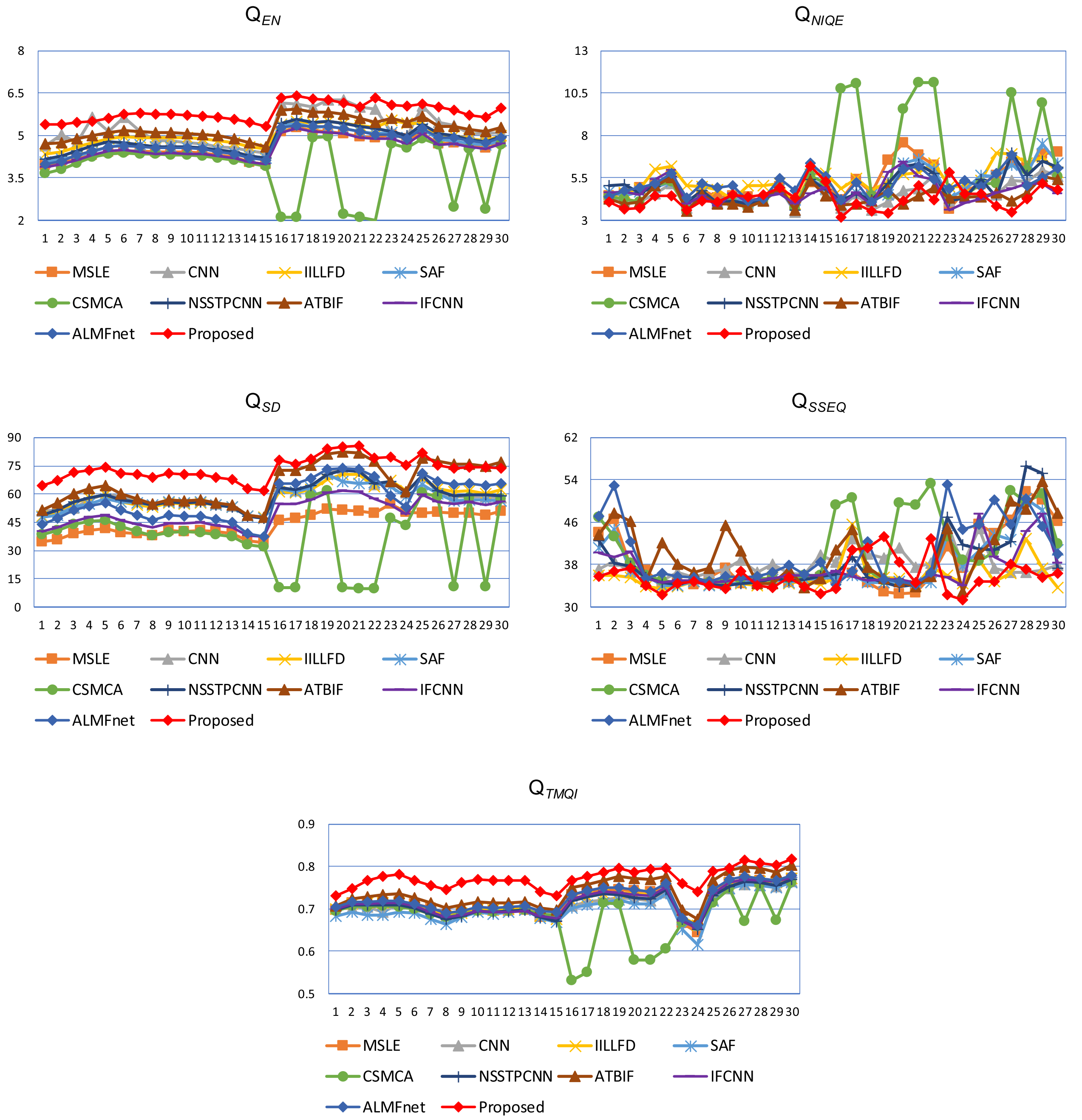


Fig. 10 Line chart of the MR-SPECT image fusion indicators

Q_{SD} and Q_{TMQI}, indicating that the image fusion quality of the proposed method is better.

Conclusion

Medical image fusion technology produces clearer and more detailed images by fusing medical image features from different modalities, providing more comprehensive and accurate information to assist doctors in analysis and decision-making and improving diagnostic efficiency and accuracy. First, the proposed method uses interval gradients to achieve structure-texture image decomposition. Second, the method uses convolutional neural networks to extract perception images, sequentially obtaining structure, texture and perception images. Three different fusion methods are used to fuse these three different images to obtain the fusion images. The comparative experimental results show that the proposed method outperforms the reference methods in multiple indicators, such as Q_{EN}, Q_{NIQE}, Q_{SD}, Q_{SSEQ} and Q_{TMQI}.

Acknowledgements
Not applicable.

Authors' contributions
X.G. conceptualized and designed the algorithm, implemented the initial codebase, and prepared the original manuscript draft. Y.X. supervised the project, provided strategic direction in algorithm development and testing, and conducted a thorough review and final approval of the manuscript prior to submission. J.Z. designed and executed the performance tests, analysed the computational results, and contributed to the interpretation of these results in the manuscript. All the authors contributed to the writing and editing of the manuscript.

Funding
This work was supported by the National Natural Science Foundation of China (51605061, 72301046), the Science and Technology Research Programme of Chongqing Municipal Education Commission (KJ1706172, KJ1706162), and the Doctoral Innovative Talent Project of Chongqing University of Posts and Telecommunications (BYJS201909).

Availability of data and materials
All images used in the manuscript were sourced from Harvard Medical School, and all sizes were 256 × 256 (the official website for the data download is http://www.med.harvard.edu/aanlib/home.html).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
The authors give full permission for the publication of this article.

Competing interests
The authors declare no competing interests.

Received: 27 June 2024 Accepted: 30 August 2024
Published online: 05 September 2024

References

1. Li X, Guo X, Han P, Wang X, Li H, Luo T. Laplacian redecomposition for multimodal medical image fusion. IEEE Trans Instrum Meas. 2020;69(9):6880–90.
2. Wang Z, Cui Z, Zhu Y. Multi-modal medical image fusion by Laplacian pyramid and adaptive sparse representation. Comput Biol Med. 2020;123: 103823.
3. Yao J, Zhao Y, Bu Y, Kong SG, Chan JC-W. Laplacian pyramid fusion network with hierarchical guidance for infrared and visible image fusion. IEEE Trans Circuits Syst Video Technol. 2023;33(9):4630–44.
4. Singh S, Mittal N, Singh H. Multifocus image fusion based on multiresolution pyramid and bilateral filter. IETE J Res. 2022;68(4):2476–87.
5. He D, Zhong Y. Deep hierarchical pyramid network with high-frequency-aware differential architecture for super-resolution mapping. IEEE Trans Geosci Remote Sens. 2023;61:1–15.
6. Nair RR, Singh T. Multi-sensor medical image fusion using pyramid-based DWT: a multi-resolution approach. 2019;13(9):1447–59.
7. Jin H, Wang Y. A fusion method for visible and infrared images based on contrast pyramid with teaching learning based optimization. Infrared Phys Technol. 2014;64:134–42.
8. Xu H, Wang Y, Wu Y, Qian Y. Infrared and multi-type images fusion algorithm based on contrast pyramid transform. Infrared Phys Technol. 2016;78:133–46.
9. Bhat S, Koundal D. Multi-focus Image Fusion using Neutrosophic based Wavelet Transform. Appl Soft Comput. 2021;106: 107307.
10. Xu L, Si Y, Jiang S, Sun Y, Ebrahimian H. Medical image fusion using a modified shark smell optimization algorithm and hybrid wavelet-homomorphic filter. Biomed Signal Process Control. 2020;59: 101885.
11. Aghamaleki JA, Ghorbani A. Image fusion using dual tree discrete wavelet transform and weights optimization. Vis Comput. 2023;39(3):1181–91.
12. Geng P, Sun X, Liu J. Adopting quaternion wavelet transform to fuse multi-modal medical images. Journal of medical and biological engineering. 2017;37:230–9.
13. Yang Y, Tong S, Huang S, Lin P. Dual-tree complex wavelet transform and image block residual-based multi-focus image fusion in visual sensor networks. Sensors. 2014;14(12):22408–30.
14. Yang Y, Que Y, Huang S, Lin P. Multimodal sensor medical image fusion based on type-2 fuzzy logic in NSCT domain. IEEE Sens J. 2016;16(10):3735–45.
15. Peng H, Li B, Yang Q, Wang J. Multi-focus image fusion approach based on CNP systems in NSCT domain. Comput Vis Image Underst. 2021;210: 103228.
16. Li H, Qiu H, Yu Z, Zhang Y. Infrared and visible image fusion scheme based on NSCT and low-level visual features. Infrared Phys Technol. 2016;76:174–84.
17. Dong L, Yang Q, Wu H, Xiao H, Xu M. High quality multi-spectral and panchromatic image fusion technologies based on Curvelet transform. Neurocomputing. 2015;159:268–74.
18. Arif M, Wang G. Fast curvelet transform through genetic algorithm for multimodal medical image fusion. Soft Comput. 2020;24(3):1815–36.
19. Bhadauria HS, Dewal ML. Medical image denoising using adaptive fusion of curvelet transform and total variation. Comput Electr Eng. 2013;39(5):1451–60.
20. Gao G, Xu L, Feng D. Multi-focus image fusion based on non-subsampled shearlet transform. IET Image Proc. 2013;7(6):633–9.
21. Vishwakarma A, Bhuyan MK. Image fusion using adjustable non-subsampled shearlet transform. IEEE Trans Instrum Meas. 2018;68(9):3367–78.
22. Singh S, Gupta D, Anand RS, Kumar V. Nonsubsampled shearlet based CT and MR medical image fusion using biologically inspired spiking neural network. Biomed Signal Process Control. 2015;18:91–101.
23. Yang B, Li S. Multifocus image fusion and restoration with sparse representation. IEEE Trans Instrum Meas. 2010;59(4):884–92.
24. Yu N, Qiu T, Bi F, Wang A. Image features extraction and fusion based on joint sparse representation. IEEE Journal of Selected Topics in Signal Processing. 2011;5(5):1074–82.
25. Li S, Yin H, Fang L. Group-sparse representation with dictionary learning for medical image denoising and fusion. IEEE Trans Biomed Eng. 2012;59(12):3450–9.
26. Shao Z, Cai J. Remote Sensing Image Fusion With Deep Convolutional Neural Network. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2018;11(5):1656–69.
27. Liang X, Hu P, Zhang L, Sun J, Yin G. MCFNet: Multi-Layer Concatenation Fusion Network for Medical Images Fusion. IEEE Sens J. 2019;19(16):7107–19.
28. Li J, Guo X, Lu G, Zhang B, Xu Y, Wu F, Zhang D. DRPL: Deep Regression Pair Learning for Multi-Focus Image Fusion. IEEE Trans Image Process. 2020;29:4816–31.
29. Zhang Y, Liu Y, Sun P, Yan H, Zhao X, Zhang L. IFCNN: A general image fusion framework based on convolutional neural network. Information Fusion. 2020;54:99–118.
30. Mu P, Wu G, Liu J, Zhang Y, Fan X, Liu R. Learning to Search a Lightweight Generalized Network for Medical Image Fusion. IEEE Transactions on Circuits and Systems for Video Technology. 2023. p. 1.
31. Lee H, Jeon J, Kim J, Lee S. Structure-Texture Decomposition of Images with Interval Gradient. Computer graphics forum. 2017;36(6):262–74.
32. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv. 2014;1409.1556.
33. Singh R, Khare A. Fusion of multimodal medical images using Daubechies complex wavelet transform – A multiresolution approach. Information Fusion. 2014;19:49–60.

34. Mittal A, Soundararajan R, Bovik AC. Making a "Completely Blind" Image Quality Analyzer. IEEE Signal Process Lett. 2013;20(3):209–12.
35. Singh R, Khare A. Multiscale Medical Image Fusion in Wavelet Domain. Scientific World Journal. 2013;2013:1–10.
36. Liu L, Liu B, Huang H, Bovik AC. No-reference Image Quality Assessment Based on Spatial and Spectral Entropies. Signal Processing-image Communication. 2014;29(8):856–63.
37. Yang Y, Park DS, Huang S, Rao N. Medical image fusion via an effective wavelet-based approach. EURASIP Journal on Advances in Signal Processing. 2010;2010(1):1–13.
38. Du J, Li WS, Xiao B, Nawaz Q. Medical image fusion by combining parallel features on multi-scale local extrema scheme. Knowl-Based Syst. 2016;113:4–12.
39. Liu Y, Chen X, Cheng J, Peng H, Wang Z. Multi-focus image fusion with a deep convolutional neural network. Information Fusion. 2017;36:191–207.
40. Du J, Li W, Xiao B. Anatomical-functional image fusion by information of interest in local Laplacian filtering domain. IEEE Trans Image Process. 2017;26(12):5855–66.
41. Li W, Xie Y, Zhou H, Han Y, Zhan K. Structure-aware image fusion. Optik. 2018;172:1–11.
42. Liu Y, Chen X, Ward R, Wang ZJ. Medical image fusion via convolutional sparsity based morphological component analysis. IEEE Signal Process Lett. 2019;26(3):485–9.
43. Yin M, Liu XN, Liu Y. Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain. IEEE Trans Instrum Meas. 2019;68(1):49–64.
44. 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.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.