Biomedical Signal Processing and Control 100 (2025) 106859

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Medical image fusion via decoupled representation and component-wise regularization learning

Rui Zhang [a], Haoze Sun [b], Lizhen Deng [c], Hu Zhu [a,d,*], Wei Qian [a]

[a] School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
[b] Bell Honors School, Nanjing University of Posts and Telecommunications, Nanjing, 210003, China
[c] National Engineering Research Center of Communication and Network Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
[d] Jiangsu Province Key Lab on Image Processing and Image Communication, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

ARTICLE INFO

Keywords:
Sparse representation
Component analysis
Dictionary learning
Decoupled representation learning

ABSTRACT

Medical image fusion plays an important role in the precise diagnosis, treatment planning, and follow-up studies of various diseases. While tremendous improvements in medical image fusion based on convolution sparse coding have been achieved, existing methods are still limited by the intractable redundancy information interaction between source medical images. In this paper, we propose an easy yet effective representation and regularization learning method based on decomposed components scheme with high competitive performance. We construct more compact information interactions by decoupled representation learning, which simultaneously mitigates the problem of redundancy in fusion component entanglement. And then two different regularization operators are adaptively exploited to depict two different components separately, which describe the structural-inspired difference based on the decoupled principle. Furthermore, we combine the alternating direction method of multipliers (ADMM) algorithm and the conjugate gradient (CG) method to optimize our proposed model. Our experiments demonstrate that our proposed method has significant improvements in efficiency and fusion performance against the state-of-the-art methods.

1. Introduction

Medical images are an integral part of modern medicine and are the basis of clinical medical work. They play an important role in disease diagnosis, treatment planning and efficacy evaluation [1,2]. However, the quality, space, and temporal characteristics of the images obtained by different imaging devices or under different conditions vary greatly. It is necessary to fuse different images into one image to obtain complementary information [3,4]. For example, the fusion between Ultrasound (US) images and other diagnostic images like Magnetic Resonance Imaging (MRI) help integrate the 2D functional information into 3D high-precision context [5]. It offers great convenience to the doctors for their localization of tumors and bone landmarks in osteoarthritis. O'brien et al. [6] proved that multimodal image fusion technology can help improve the accuracy of the localization from 39.2% to 88.2% in the surgical treatment of epilepsy. Zhu et al. [7] proposed a brain tumor segmentation method based on multimodal MRI, which can effectively extract and fuse deep semantic features and edge features in multimodal MRI. However, when multiple modalities are missing in MRI, these networks may not be able to adequately recover the lost information, which will affect performance and lead to inaccurate identification and segmentation. Therefore, medical image fusion has been a useful and challenging subject that has attracted widespread attention in the field of image processing [8].

For representation, several frameworks had adopted feature extraction strategies for medical image fusion from the perspective of pixel level. Bhatnagar et al. [9] performed the fusion in non-subsampled contourlet transform (NSCT) domain because of its efficiency in capturing 2D geometrical structures. Since NSCT decomposition has a limitation of a specific number of directional components, non-subsampled shearlet transform (NSST) [10] is adopted in some recent SOTAs. For example, Singh et al. [11] fused the high-frequency components and low-frequency components in NSST domain with models adaptively exploited. This model has been proved to be very effective in preserving

* Corresponding author at: School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China.
E-mail addresses: rayzhangxu@163.com (R. Zhang), sunhaoze@163.com (H. Sun), alicedenglzh@gmail.com, denglizhen@njupt.edu.cn (L. Deng), peter.hu.zhu@gmail.com (H. Zhu), weiqian188188@gmail.com, zhuhu@njupt.edu.cn (W. Qian).

https://doi.org/10.1016/j.bspc.2024.106859
Received 8 January 2024; Received in revised form 8 August 2024; Accepted 6 September 2024
Available online 1 November 2024

the spatial consistency in fused images. All of the methods mentioned above are collectively referred to as multiscale analysis (MSA) based methods. Although these methods have always been a mainstream solution to image fusion [12], they often suffer from poor contrast, blocking artifacts [13–15] and lack in preservation of tiny edges [16].

With the development of neural networks, several researchers utilized a neural network approach for selecting the image pixels and image regions [17,18]. Pulse coupled neural network (PCNN) model and its variants are extensively utilized in the formulation of image fusion algorithm because it performs well in extracting features consistent with human vision [19]. For example, Li et al. [20] proposed general image fusion method based on Non-Subsampling Shear Transform (NSST), which integrated Pulse-Coupled Neural Network (PCNN) and sparse representation (SR)-based metrics to improve the visual quality of fused images. Although the PCNN-based models show good results, their performance can be largely affected by free parameters and lead to loss of fine detail information or improper contrast [13]. Wang et al. [21] proposed a multichannel coupled neural system to overcome the color contrast problem in image fusion. In addition, there are some deep learning based fusion methods which achieved SOTA in recent progress. Multimodal medical image sensor fusion (MMISF) [16] added sparse representation and dictionary learning on a convolutional neural network (CNN) to enhance the learning ability of the model. Zhou et al. [22] had discussed various models of generative adversarial network (GAN) and advantages and applications of GAN in medical image fusion field. Some special designs of GAN had adopted into medical image fusion [23] and exploited the dual discriminators to judge the differences between the fused image and source ones. Two-exposure fusion [24] with deep reinforcement learning is cooperated to learn an online compensating representation. Zhang et al. [25] proposed a general image fusion framework based on CNN in a fully convolutional manner. He et al. [26] used cloud computing to carry out efficient multi-feature extraction and interactive fusion network medical image segmentation methods, and combined the multi-feature extraction network of CNN with Transformer, which effectively solved the limitations of CNNs in capturing long-distance connections contained in images.

Indeed, the existing works always follow a same learning prototype whether there are optimization based methods or deep learning based methods: feature extraction and information fusion [27]. However, different modalities would inevitably bring the uncertainty and information redundancy both. We therefore wonder, "*Can we customize the redundancy removal operator explicitly for medical image fusion and leverages the benefits from its intrinsic properties?*". Sparse representation (SR) has a wide range of applications in image fusion methods, which can effectively extract feature information and improve the reconstruction accuracy of the fused image. Most of the existing SR-based fusion methods use standard sparse coding models based on single-image components and local patches [28–30] and improve the model performance by adding detailed constraints to the model [31,32], designing a more effective subordinate learning strategy [33] to achieve high-quality image fusion. For example, Zhu et al. [34] proposed an image fusion method based on image decomposition and sparseness, which decomposes the source image into cartoon components and texture components, but there are a large number of matrix calculations and the running time is slow. In contrast, Jiang et al. [35] proposed a novel multi-component SR-based fusion method via morphological component analysis (MCA) [36], which can obtain the sparse representations of cartoon and texture components of each source image. This component separation process can significantly improve the flexibility for designing more effective fusion strategies.

Inspired by suppressing redundancy motivated by the [37] model, where a factorized convolution approach was used to learn a compact set of discriminative basis filters with significant energy. Unlike the discriminative filter learning tracking model in [37], a decoupled representation learning in our model is introduced to solve the impact of redundancy in dictionary learning for two components for medical image fusion. Actually, the source medical images cannot be completely or well-decomposed into the concrete components, that is component entanglement open problem in medical image fusion. Our method can select more compact information intersections among component decomposition models to reduce redundancy. Thus, we design this decoupled representation learning that uses a sample-dependent approach [38]. As can be seen from Fig. 1, the texture map obtained by CS-MCA [39] suffers from the component entanglement with the cartoon image and contains poor information. However, the map derived by our method is rich in texture information and is able to complement each other with the cartoon image.

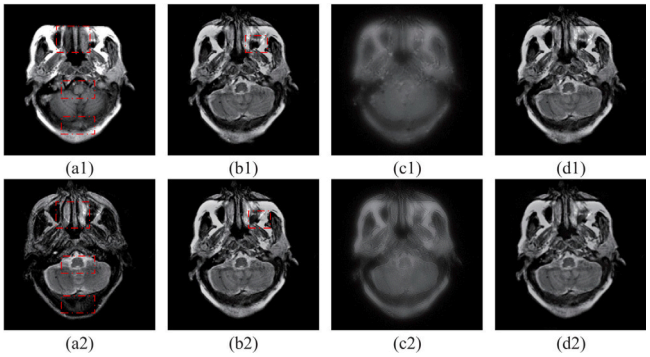

Fig. 1. Examples of the cartoon and texture decomposition. Please refer to the main text for more detailed descriptions. Source images: (a1-a2) The a1 and a2 show the image similarity overlap [40]. Fused images obtained by CS-MCA: (fused image: b1; the texture map: c1; the cartoon image: d1). Fused images obtained by DRCRL: (fused image: b2; the texture map: c2; the cartoon image: d2).

Furthermore, in order to realize the structure-inspired difference between each component, our model uses the ℓ_2 and ℓ_0 norms to describe the data characteristics of each component. Among them, the cartoon component contains rich geometric structure information, and its smoothness can be preserved by using the ℓ_2 norm. While the data features of the texture component fluctuate greatly and contain rich detailed information, the ℓ_0 norm can be used to ensure its sparseness. Then, the general regularization learning is rewritten to the adaptive regularization learning model of component pixel perception to further utilize the prior knowledge of the cartoon component and the texture component. In addition, we use the alternating direction method of multipliers (ADMM) algorithm and conjugate gradient descent [41] to solve the model optimization problem. Since our optimization objective function is convex, the conjugate gradient descent method can quickly obtain the global optimal solution of the objective function and reduce computational complexity.

In summary, the main contributions of our method are listed as follows.

1. The DRCRL model obtains a sparse representation of the source image components by using a dictionary filter, which effectively captures the important features and thus ensures a higher quality of the fused image. By accurately decomposing the image components, the DRCRL model not only enhances the feature extraction, but also significantly improves the overall performance of image fusion.
2. We introduce decoupled representation learning to select high-energy dictionary filters, which reduces redundancy and speeds up the learning process. By focusing on high-energy filters, the model retains the most informative features, thus improving the robustness and accuracy of image fusion. Different decoupled representations can be used for different image components, thus preserving and enhancing unique features such as texture and edges.

Table 1
Notation, abbreviation and definition.

Notation	Definition	Abbreviation	Definition
$\hat{A}$	The Fourier transform of A	D_c, D_t	Over-complete dictionaries of the cartoon, texture components
A^T A^H	The transpose or conjugate transpose of A	X_c, X_t	Decoupled sparse coefficient maps of the cartoon, texture components
$*$	Convolution, such as $A * B$	P_c, P_t	Sensitivity decoupled operators of the cartoon, texture components
AB	Matrix multiplication	J_c, J_t	Decoupled component-wise dictionaries of the cartoon, texture components
$\langle A, B \rangle$	Linear projection operator	H_c, H_t	Recovered sparse coefficient maps of the cartoon, texture components

3. Considering the different sensitivities of the low-rank sparseness of different components, two different regularization operators are exploited adaptively to depict the two components. Because they have the explicit structural-inspired property of discontinuous and continuous behavior in a pixel-level manner for medical image fusion.
4. To improve the accuracy of the model, we combine the ADMM algorithm and the conjugate gradient (CG) method to solve the optimization problem of the model. Then, a more appropriate iterative algorithm can be chosen according to the characteristics of the data in different components. Extensive experiments are conducted to verify the effectiveness of DRCRL on five different types of medical image fusion problems with grayscale and color images.

The structure of this paper is arranged as follows. Section 2 discusses the research of the preliminaries and explains the proposed model and the solution process in detail. The specific experimental results and comparative analysis are shown in Section 3. Finally, Section 4 gives a brief summary.

2. The proposed method

2.1. Symbols and notations

The symbols, notations and abbreviations widely used in this paper are summarized in Table 1. In the beginning, we add the subscript c and t to distinguish different components. If the solution is the same for two components, we do not add the subscript to show its generality.

2.2. Insights and problems of the component decomposition

Here we provide a brief explanation for the component decomposition. The theory was put forward by Meyer et al. [42]: An image can be regarded as the sum of a cartoon image and a texture map. The cartoon image describes the salient parts and piecewise smooth changes in the illumination, as well as edges. The texture map delivers the detailed texture information in the regions enclosed by edges. However, sometimes the source medical images cannot be completely or well-decomposed into the corresponding components. This phenomenon is called *component entanglement*.

- *How we can solve the corresponding component entanglement?*
- *How we can find the structure-inspired difference between each component?*

2.3. Problem statement

In our method, the dictionaries are independently learned from cartoon and texture images using the CSR dictionary learning approach presented in [43].

Supposing that $\boldsymbol{I}_x$ and $\boldsymbol{I}_y$ are two modalities, motivated by the above analysis, we decompose each image modality into two parts, cartoon and textured component.

$$\underbrace{\sum_{p=1}^{P} \boldsymbol{b}_p * \boldsymbol{u}_p}_{\text{cartoon images of } \boldsymbol{I}_x} + \underbrace{\sum_{k=1}^{K} \boldsymbol{d}_k * \boldsymbol{r}_k}_{\text{textured images of } \boldsymbol{I}_x} = \boldsymbol{I}_x \tag{1}$$

$$\underbrace{\sum_{q=1}^{Q} \boldsymbol{e}_q * \boldsymbol{v}_q}_{\text{cartoon images of } \boldsymbol{I}_y} + \underbrace{\sum_{k=1}^{K} \boldsymbol{h}_k * \boldsymbol{r}_k}_{\text{textured images of } \boldsymbol{I}_y} = \boldsymbol{I}_y \tag{2}$$

Here, we use convolutional sparse coding to represent the cartoon and textured component $\boldsymbol{I}_x$ and $\boldsymbol{I}_y$. Specifically, $\{\boldsymbol{b}_p\}_{p=1}^{P}$ and $\{\boldsymbol{d}_k\}_{k=1}^{K}$ are the convolutional dictionary filters to represent the cartoon and textured components of $\boldsymbol{I}_x$, with $\{\boldsymbol{u}_p\}_{p=1}^{P}$ and $\{\boldsymbol{r}_k\}_{k=1}^{K}$ as the corresponding convolutional sparse representations (CSRs). Similarly, $\{\boldsymbol{e}_q\}_{q=1}^{Q}$ and $\{\boldsymbol{h}_k\}_{k=1}^{K}$ are the dictionary filters to represent the cartoon and textured component of $\boldsymbol{I}_y$, with$\{\boldsymbol{u}_q\}_{q=1}^{Q}$ and $\{\boldsymbol{r}_k\}_{k=1}^{K}$ as the corresponding CSRs.

However, redundancy has always been a key factor in SR-based fusion methods, which affects the quality and efficiency of fusion. There are two main types of redundancy in the fusion process. One is the redundancy that exists in the source image from the beginning. The other is generated during the learning process. Our proposed model mainly focuses on how to achieve the heuristic selection of dictionary filters which also helps to effectively suppress component entanglement. An image fusion method based on multi-component SRS based on MCA model proposed in [44], which can be expressed as:

$$\min_{X_c, X_t} \frac{1}{2} \left\| Y - D_c * X_c - D_t * X_t \right\|_2^2 + \lambda_c \left\| X_c \right\|_1 + \lambda_t \left\| X_t \right\|_1 \tag{3}$$

where $Y \in \mathcal{R}^{N_1 \times N_2}$ is an entire image, $\{D_c, D_t\}$ respectively denote two dictionaries of the cartoon and texture components, $\{X_c, X_t\}$ means the corresponding sparse coefficient maps. λ_c and λ_t are sparsity regularization parameters. F denotes the Frobenius norm of the matrix. $*$ presents the convolution. Through morphological component analysis (MCA) method, each source image cartoon components (mainly contains large scale geometry such as smooth content) and texture components (mainly contains small scale repetition/fibrillation pattern and fine details) can be obtained from the original image, and use different dictionary filters to process these parts, ensure that each component can process and fusion independently. This component separation approach can significantly improve the flexibility of designing more efficient fusion strategies.

The sample-dependent approach in decoupled representation learning benefits from learning compact information intersections, which greatly reduces the model parameters and improve fusion efficiency of the model. The Proof could be given from a general formulation. $\langle P, X\rangle$ can be regarded as a linear projection operator. Then, the three-dimensional matrix elements in $\langle P, X\rangle_{(m,i,j)}$ are expressed as follows:

$$\langle P, X\rangle_{(m,i,j)} = \sum_{e} P_{(m,e)} X_{(e,i,j)} \tag{4}$$

where the convolution here is high-dimensional convolution: $A * B = \sum_m A_m * B_m$.

In addition, the sensitivity of the sparsity of different components is different, and different methods should be adopted to process the component data to achieve the purpose of making full use of the characteristics of the data.

2.4. Model construction of DRCRL

Compared to ordinary sparse representation models, DRCRL adopts a decoupled representation learning method, decomposing the source image into cartoon and texture components to address the issue of component entanglement. It reduces redundancy by selecting compact information interactions between the component decomposition models. Additionally, traditional sparse representation models often rely on single image components and local blocks, which can lead to redundancy and entanglement. In contrast, DRCRL considers the different sensitivities of components to low-rank sparsity and uses two different regularization operators for the cartoon and texture components. In terms of encoding, DRCRL employs convolution sparse coding model [39] to obtain sparse representations of two images through a given dictionary, introduces a sensitivity decoupling operator to optimize the fusion process, and ultimately achieves the final fused image by merging and recovering the coefficient matrices.

First, we obtain the sparse representation of the two images for a given dictionary, and then our model obtains the final fused image by fusing and restoring the coefficient matrix. Eq. (5) is the minimization problem of finding the sparse representation for the medical image fusion:

$$\begin{aligned}\min_{X_c,X_t,P_c,P_t} \frac{1}{2}\left\|Y - D_c * \langle P_c, X_c\rangle - D_t * \langle P_t, X_t\rangle\right\|_2^2 + \frac{\lambda_c}{2}\left\|X_c\right\|_2^2 \\ +\lambda_t \left\|X_t\right\|_0 + \lambda p_c \left\|P_c\right\|_F^2 + \lambda p_t \left\|P_t\right\|_F^2\end{aligned} \tag{5}$$

where $Y \in \mathcal{R}^{N_1\times N_2}$ is an entire image, $\{D_c, D_t\} \in \mathcal{R}^{M\times B_1\times B_2}$ respectively denote two over-complete dictionaries of the cartoon and texture components, $\{X_c, X_t\} \in \mathcal{R}^{E\times N_1\times N_2}$ means the corresponding decoupled sparse coefficient maps and P_c, $P_t \in \mathcal{R}^{M\times E}$ are our proposed sensitivity decoupled operators. $\lambda_c, \lambda_t, \lambda_{p_c}$ and λ_{p_t} are sparsity regularization parameters.

In addition, M is the number of filters, E is the number of reduced coefficient matrices, and N_1, N_2, B_1, B_2 are constants related to the size of the images and dictionaries. Therefore, E is the number of sparse matrices that can effectively express the image after dictionary learning; then, P can also be regarded as the projection operator from the sparse expression to the original image information. The decoupled sparse coefficient maps X_c, X_t are compact sets.

Introducing $J_c = \langle P_c^T, D_c\rangle$, $J_t = \langle P_t^T, D_t\rangle \in \mathcal{R}^{E\times B_1\times B_2}$ that denote the decoupled component-wise dictionaries of the cartoon and texture components, respectively. Then the minimization problem of Eq. (5) is rewritten as

$$\begin{aligned}\min_{X_c,X_t,P_c,P_t} \frac{1}{2}\left\|Y - J_c * X_c - J_t * X_t\right\|_2^2 + \frac{\lambda_c}{2}\left\|X_c\right\|_2^2 \\ +\lambda_t \left\|X_t\right\|_0 + \lambda p_c \left\|P_c\right\|_F^2 + \lambda p_t \left\|P_t\right\|_F^2\end{aligned} \tag{6}$$

The optimization of the above minimization problem can be achieved by solving the following two subproblems of different components.

$$\begin{aligned}\min_{X_c,P_c} \frac{1}{2}\left\|Y_c - J_c * X_c\right\|_2^2 + \frac{\lambda_c}{2}\left\|X_c\right\|_2^2 + \lambda_{p_c}\left\|P_c\right\|_F^2 \\ s.t. \quad Y_c = Y - J_t * X_t\end{aligned} \tag{7}$$

and

$$\begin{aligned}\min_{X_t,P_t} \frac{1}{2}\left\|Y_t - J_t * X_t\right\|_2^2 + \lambda_t\left\|X_t\right\|_0 + \lambda_{p_t}\left\|P_t\right\|_F^2 \\ s.t. \quad Y_t = Y - J_c * X_c\end{aligned} \tag{8}$$

Optimal iteration of the minimization problem is achieved by alternately solving two subproblems. In the process of the alternate solution, Y_c and Y_t are targeted to be the cartoon and texture components of the source image Y, respectively.

It is noticeable that the main optimization process of the two subproblems is the same. For the simplicity of expression, we write the subproblems into a general form. If no otherwise specified, the solving process of the general form is shared by two subproblems. We also conceal the subscript c and t to show its generality.

$$\min_{X,P} \frac{1}{2}\left\|Y - J * X\right\|_2^2 + \lambda\left\|X\right\|_g^k + \lambda_p\left\|P\right\|_F^2 \tag{9}$$

where Y represents Y_c or Y_t according to which subproblem does it belong to. The parameters g and k also generally represent the norm and power of the two subproblems.

According to the above analysis, the fusion process of our DRCRL involves data processing of dictionary sparse representation features and a dual sensitivity decoupled expression matrix. To better understand data processing, we divide the data processing into three parts in Fig. 2.

2.5. Optimization of the DRCRL model

In the previous section, we have derived our general minimization problem (9). By introducing the splitting variables $W = X$, we can rewrite the general problem in a compact form:

$$\begin{aligned}\min_{X,P,W} \frac{1}{2}\left\|Y - J * X\right\|_2^2 + \lambda\left\|W\right\|_g^k + \lambda_p\left\|P\right\|_F^2 \\ s.t. \quad W = X\end{aligned} \tag{10}$$

Then the augmented Lagrangian function can be shown as follows:

$$\begin{aligned}L(X, P, W, U) = \frac{1}{2}\left\|Y - J * X\right\|_2^2 + \lambda\left\|W\right\|_g^k + \lambda_p\left\|P\right\|_F^2 \\ +\frac{\rho}{2}\left\|X - W + U\right\|_2^2\end{aligned} \tag{11}$$

where U is the Lagrange multiplier, and ρ is the penalty coefficient.

To solve the optimization problem of the model, We combine the ADMM algorithm and the CG Method. X is updated by the ADMM algorithm when P is fixed. Similarly, P is updated by Gauss–Newton method and conjugate gradient descent method when X is fixed. The updating process is presented as follows:

$$\begin{cases} X^{(l+1)} = \min\limits_{X} \frac{1}{2}\left\|Y - J^{(l)} * X\right\|_2^2 + \frac{\rho}{2}\left\|X - W^{(l)} + U^{(l)}\right\|_2^2 \\ W^{(l+1)} = \min\limits_{W} \lambda\left\|W\right\|_g^k + \frac{\rho}{2}\left\|X^{(l+1)} - W + U^{(l)}\right\|_2^2 \\ P^{(l+1)} = \min\limits_{P} \left\|Y - D * \langle P, X^{(l+1)}\rangle\right\|_2^2 + \lambda_p\left\|P\right\|_F^2 \\ U^{(l+1)} = X^{(l+1)} - W^{(l+1)} + U^{(l)} \end{cases} \tag{12}$$

2.5.1. Solution of X

In order to efficiently update X in the frequency domain, we first use FFT to convert X and dictionary J to the frequency domain, which allows us to take advantage of the computational advantages that processing convolutions and other linear operations bring to multiplication between elements. The optimization problem is formulated to minimize the reconstruction error in the transformation domain

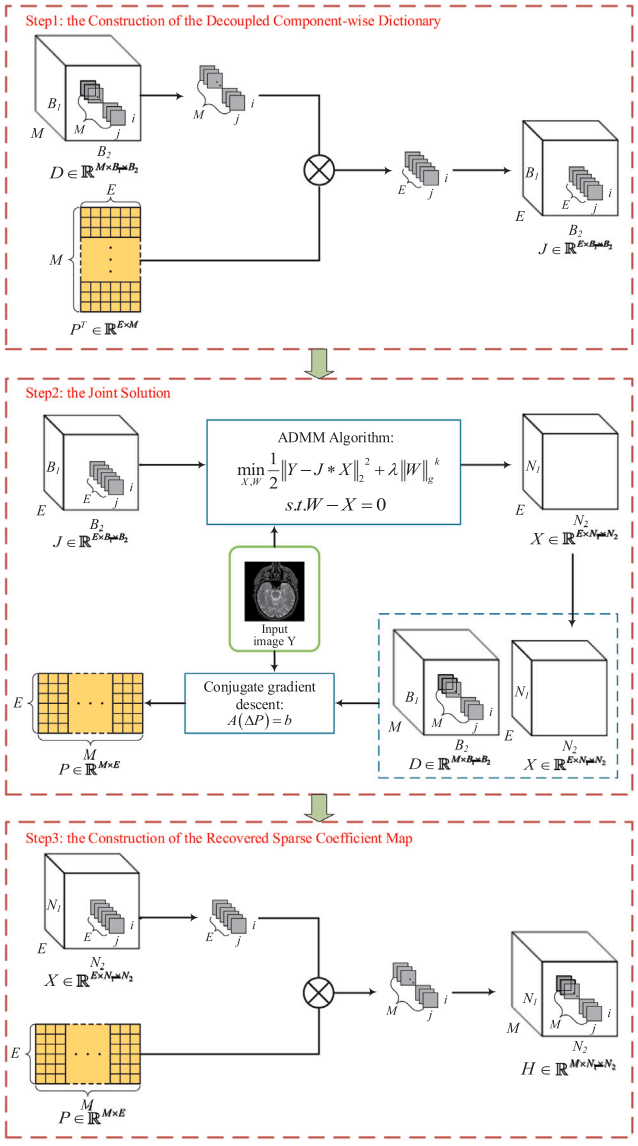

Fig. 2. The data processing flowchart of the proposed DRCRL. First, the model uses decoupled representation learning to get the decoupled form of the initial dictionaries. Then we uses the ADMM and CG joint solution to achieve the optimization of the model. Finally, our model obtains the final fused image through the sparse solution. In this picture, D is an over-complete dictionary. X means the decoupled sparse coefficient map. P is our proposed sensitivity decoupled operator. J denotes the decoupled component-wise dictionary. H corresponds to the recovered sparse coefficient map. Definitions of other symbols used in this figure can be found in Eqs. (5), (9) and (11).

and stabilize the regularization term of iterative updates. By applying the Sherman–Morrison formula, we are able to efficiently update the inverse matrix required to solve linear systems, which is critical for large-scale problems, since it is not computationally feasible to directly compute the inverse matrix. Finally, we inverse FFT the solution in the frequency domain, carefully addressing the underlying numerical problems to convert the updated X back to the original domain, ready for the next iteration or final use. The decoupled component-wise dictionary J is also transformed to the size of X and now denoted as $\hat{J} \in C^{E\times N_1\times N_2}$. Then, the problem in the frequency domain can be written as follows:

$$\min_{\hat{X}} \frac{1}{2}\left\|\hat{Y}-\hat{J}\hat{X}\right\|_2^2+\frac{\rho}{2}\left\|\hat{X}-(W^{(l)}-U^{(l)})\right\|_2^2 \tag{13}$$

According to the **Sherman–Morrison formulation** [45], the solution of the linear equations is:

$$\hat{X}^{(l+1)}=\rho^{-1}(\hat{C}-\frac{\hat{C}\hat{J}^H}{\rho I+\hat{J}\hat{J}^H}\hat{J}) \tag{14}$$

where $\hat{C}=\hat{J}^H\hat{Y}+\rho(\hat{W}^{(l)}-\hat{U}^{(l)})$. After inverse FFT of Eq. (14), we get $X^{(l+1)}$.

2.5.2. Solution of W

The target of this step is solving the minimization of the following expression:

$$\begin{aligned} &\min_{W}\frac{\rho}{2}\|W-B\|_2^2+\lambda\|W\|_g^k \\ &s.t.\quad B=X^{(l+1)}+U^{(l)} \end{aligned} \tag{15}$$

Since the norms of W in different components are not the same, the method of solving W in problem (15) is different. We discuss them separately.

For Cartoon Component

In this situation, $k=2$, $g=2$, $\lambda=\frac{\lambda_c}{2}$. Therefore, the solution of W in the cartoon component is presented as follows:

$$W^{(l+1)}=\frac{\rho}{\rho+\lambda_c}(X^{(l+1)}+U^{(l)}) \tag{16}$$

For Texture Component

In this situation, $k=1$, $g=0$, $\lambda=\lambda_t$. Since minimizing l_0-regularized model is an NP-hard problem, we solve this subproblem with the estimation algorithm motivated by the model in [46]:

$$W^{(l+1)}(e,i,j)=\begin{cases}\left(X^{(l+1)}+U^{(l)}\right)_{(e,i,j)} & ,K\geq\frac{2\lambda_t}{\rho}\\ 0 & ,\text{otherwise}\end{cases} \tag{17}$$

where $K=\left|\left(X^{(l+1)}+U^{(l)}\right)_{(e,i,j)}\right|^2$.

2.5.3. Solution of P

To achieve the solution of P, we first obtain the minimization loss function in the Fourier domain of P according to Eq. (12). In this progress, the size of dictionary has been transformed and now be denoted as $\hat{D}\in C^{M\times N_1\times N_2}$. Matrix P do not need to be transformed because it actually can be resolved into the coefficients before X. Thus the loss function can be simplified as follows:

$$\min_{P}\left\|\hat{Y}-\hat{D}P\hat{X}^{(l+1)}\right\|_2^2+\lambda_p\|P\|_F^2 \tag{18}$$

This loss function is similar to a matrix factorization problem. We can employ Gauss–Newton [47] to linearizing the residuals and use the conjugate gradient descent method [47] to obtain the optimal solution:

$$\min_{\Delta P}\left\|\hat{Y}-\hat{D}(P^{(l)}+\Delta P)\hat{X}^{(l+1)}\right\|_2^2+\lambda_p\left\|P^{(l)}+\Delta P\right\|_F^2 \tag{19}$$

where ΔP is the matrix increment of P, $P^{(l)}$ is P in the last iteration. For simplicity of expression, we need to conceal the superscript l, $l+1$ of $P^{(l)}$ and $\hat{X}^{(l+1)}$temporarily. In the following deduction, P means $P^{(l)}$ and $\hat{X}$ means $\hat{X}^{(l+1)}$. Then the solution of the quadratic problem (19) can be expressed as:

$$\hat{D}^H(\hat{D}P\hat{X}+\hat{D}\Delta P\hat{X}-\hat{Y})\hat{X}^H+\lambda_p\Delta P=-\lambda_p P \tag{20}$$

In the deduction above, we use the simplied form to obtain the solution of the minimization problem. Here we also present the complete form of Eq. (20). To simplify the calculation, we define an operator $\odot$.

Proof. Assume three-dimensional matrices $A\in C^{M\times N_1\times N_2}$ and $B\in C^{N_1\times N_2\times E}$. We combine the second and third dimensions of the A matrix and the first and second dimensions of the B matrix.

$$\bar{A}=\begin{bmatrix}A_{(1,1,1)} & \cdots & A_{(1,N_1,1)} & \cdots & A_{(1,N_1,N_2)}\\ \vdots & & \vdots & \ddots & \vdots\\ A_{(M,1,1)} & \cdots & A_{(M,N_1,1)} & \cdots & A_{(M,N_1,N_2)}\end{bmatrix} \tag{21}$$

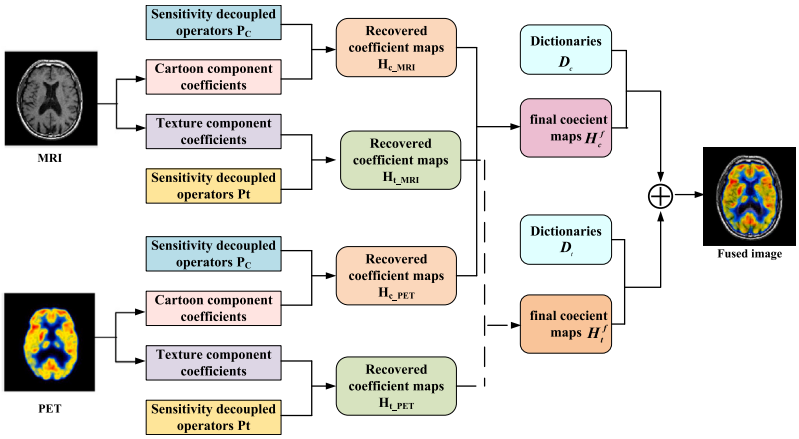

Fig. 3. Flowchart of the medical image fusion process.

$$\bar{B} = \begin{bmatrix} B_{(1,1,1)} & & B_{(1,1,E)} \\ B_{(2,1,1)} & & B_{(2,1,E)} \\ \vdots & \cdots & \vdots \\ B_{(N_1,1,1)} & & B_{(N_1,1,E)} \\ B_{(1,2,1)} & & B_{(1,2,E)} \\ \vdots & \ddots & \vdots \\ B_{(N_1,N_2,1)} & & B_{(N_1,N_2,E)} \end{bmatrix} \tag{22}$$

$$A \odot B = \bar{A} \times \bar{B} \tag{23}$$

Then the complete form of Eq. (20) can be defined as follows:

$$G(\hat{D}\cdot\langle P,\hat{X}\rangle)+\lambda_p \Delta P + G(\hat{D}\cdot\langle \Delta P,\hat{X}\rangle) = G(\hat{Y}_s)-\lambda_p P \tag{24}$$

where $G(A)=\hat{D}^H \odot (A\cdot \hat{X}^H)$. Conjugate gradient descent [47] is used to optimize ΔP. Then, $P^{(l+1)} = P^{(l)} + \Delta P$. The proof is completed.

2.6. Detailed fusion scheme

The flowchart of the medical image fusion process, as shown in Fig. 3, provides a detailed illustration of each step from data acquisition to the final image fusion. Specifically, let $Y^k, k \in \{1,\dots,K\}$ denote a set of K pre-registered source images. To achieve a good fusion effect, our model processes the image data as follows.

2.6.1. Sparse coding based on the DRCRL model

The general component-based fusion model, such as the CS-MCA model, directly uses X_c^k and X_t^k as the global sparse coefficient maps of the cartoon and texture components, respectively. However, our DRCRL model introduces a dual sensitivity decoupled representation based on the sensitivity of different components. Therefore, the coefficients of the DRCRL model should be recovered before the fusion process:

$$\begin{cases} H_c = \langle P_c, X_c \rangle \\ H_t = \langle P_t, X_t \rangle \end{cases} \tag{25}$$

where H_c, H_t denote the recovered sparse coefficient maps of the cartoon and texture components respectively. X_c, X_t mean the decoupled sparse coefficient maps and P_c, P_t are sensitivity decoupled operators.

2.6.2. Fusion of sparse coefficient maps

According to the other SR-based fusion methods [28,29], the activity level metric of the source images can be represented by the l_1-norm of sparse coefficient vectors. For each coefficient, we use the fusion strategy introduced in the CSR-based fusion method [43] with independent parameter settings. Briefly speaking, we can count the activity level maps $A_c^k \in \mathcal{R}^{N_1\times N_2}$ and $A_t^k \in \mathcal{R}^{N_1\times N_2}$ for source image k based on the l_1-norm of the recovered sparse coefficient maps. The initial activity level map $\mathbf{A}_n^k(x,y)$ is defined as:

$$\mathbf{A}_n^k(x,y) = \left\| \mathbf{X}_{n,1:M_n}^k(x,y) \right\|_1, n \in \{c,t\} \tag{26}$$

where c and t denote the cartoon and texture components, and $\mathbf{X}_{n,1:M_n}^k(x,y)$ represents the M_n-dimensional vector containing the coefficients of $\mathbf{X}_{n,m}^k$ at pixel (x,y).

Then the final coefficient maps H_c^f and H_t^f can be obtained by the following equations.

$$\begin{cases} H_c^f(:,i,j) = H_c^{k^*}(:,i,j), k^* = \underset{k}{argmax}(A_c^k(i,j)) \\ H_t^f(:,i,j) = H_t^{k^*}(:,i,j), k^* = \underset{k}{argmax}(A_t^k(i,j)) \end{cases} \tag{27}$$

2.6.3. Reconstruction of the fused image

Finally, the coefficient maps are used to restore the fused cartoon and texture components, and the two components are added to obtain the final fused image I^f.

$$I^f = D_c * H_c^f + D_t * H_t^f \tag{28}$$

3. Experiments

3.1. Experimental settings

In this section, we first introduce the experimental settings, fusion indicators, and experimental results in detail. The proposed method is compared with other approaches in two aspects: qualitative analysis of fused result and quantitative metrics analysis. To increase the experimental content and improve the readability, we also carried out parameter analysis, iterative analysis, ablation analysis, and running time comparison. All experiments are carried out using MATLAB R2016b on a computer with an dual-Core Intel Core i5 processor (1.8 GHz) and 8 GB 1600 MHz DDR3. In our methods, we experimentally fix $\lambda_c = max(0.6-0.1\times i, 0.005)$, $\lambda_t = max(0.6-0.1\times i, 0.005)\times 10^{-7}$, $\lambda p_c = \lambda p_t = 5\times 10^{-8}$, $\rho_c = min(50000\times 1.5^i, 10^5)$, $\rho_t = min(50\times 1.5^i, 10^5)$ during the ith iteration and the number of iterations L is set to 6.

3.1.1. Experimental images

Medical image fusion mainly concentrates on magnetic resonance imaging (MRI), computerized tomography (CT), positron emission tomography (PET) and single-photon emission computed tomography (SPECT) modalities [48]. Our experimental dataset is collecting from the Whole Brain Atlas [49], which is established by Keith A. Johnson and J. Alex Becker at Harvard Medical School. In our experiment, there are a total of 5 fusion results based on different modalities, including MRI-CT, MRI-PET, MRI-SPECT and T1-T2, T2-PD, where T1, T2, and PD are MRI image based on different weighted. Fig. 4 selects 5 pairs of them for display. It can be seen from the figure that the images of different modalities are significantly different and the feature expression is not comprehensive which can be improved by image fusion.

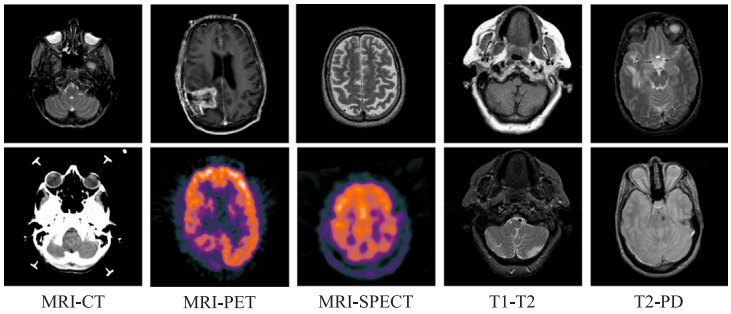

Fig. 4. Five kinds of fusion images pairs used in our experiments. From left to right in the picture are: MRI-CT, MRI-PET, MRI-SPECT and T1-T2, T2-PD, where the PET and SPECT are color images. It can be intuitively seen from the figure that different types of images show completely different visual effects with different image characteristics and information.

3.1.2. Comparison methods

Our DRCRL method is compared with seven existing medical image fusion methods, which are categorized into two groups: deep learning-based methods and non-deep learning-based methods. Among the non-deep learning-based methods, LLF-IOI [50] utilizes a local Laplacian filtering technique for anatomical-functional image fusion, NSCT-PCDC [51] introduces a fusion framework based on the non-subsampled contourlet transform (NSCT), GF [52] proposes a guided filtering-based weighted average technique to enhance spatial consistency, and the bilateral filter [53] with local gradient energy combines bilateral filtering with local gradient energy for fusion. Additionally, CS-MCA [39], a sparse representation (SR) based fusion method for pixel-level medical image fusion. The deep learning-based methods include NSST-PAPCNN, NSCT-RPCNN, and IFCNN. NSCT-RPCNN [54] proposes a new multimodal medical image fusion method in non-subsampled shearlet transform (NSST) domain. NSCT-RPCNN [54] employs multiscale geometric analysis of NSCT and the fuzzy-adaptive reduced pulse-coupled neural network. IFCNN [25] is a deep-learning-based fusion method. It should be noted that in the fusion of color images, the CS-MCA method is not include in the comparison. It should be noted that in the fusion of color images, the CS-MCA method is not include in the comparison. All comparison methods are implemented as described in the original manuscripts.

3.1.3. Objective metrics

To verify the performance superiority of our method, we select a total of 9 metrics to analyze the image fusion effect from different aspects, namely human visual system (HVS)-based metric Q_{HVS} [55], structural similarity metric (SSIM) [56], edge-based structural similarity Q_E [57], peak signal-to-noise ratio (PSNR), feature mutual information (FMI-pixel) [58], gradient-based metric $Q^{AB/F}$ [59], universal image quality index (UIQI) [60], nonlinear correlation coefficient (NCC) [61] and edge information based image fusion metric $R_Q^{F/AB}$ [62]. For all metrics, higher values correspond to better fused images.

The objective evaluation metrics for image fusion are summarized into four categories: information theory based metrics, image feature based metrics, image structure similarity based metrics and human perception inspired metrics. We also add a noise-related metric, PSNR, which is often used to reflect image quality [63,64]. Some reference based metrics like PSNR are counted as the average of results using each source image as reference image alternately. Then the redundancy elimination capability of the fusion methods can be measured from two perspectives. The first kind of redundancy that exists in the source image can be measured through human perception inspired metrics like Q_{HVS}; the second kind of redundancy which generated during the learning process is reflected in PSNR. All our indicators together constitute a comprehensive evaluation system.

3.2. Qualitative analysis of fusion result

We mainly conduct comparative experiments in five different modes, and three sets of results were selected for display in each fusion under different modes. Below we compare and analyze the visual effects of the fusion images in various methods in detail. The metrics of the fused picture are also attached. Due to limited space, we only provide four representative metrics: FMI-pixel, PSNR, UIQI, Q_{HVS} (one for each category).

3.2.1. Fusion analysis on MRI-CT

Fig. 5 shows the fusion results of three sets of MR (a1-a3) and CT (b1-b3) images, with each set of results arranged in a row. It can be seen from Fig. 5 that the LLF-IOI (c1-c3) fused results are severely disturbed by noise and generate more artifacts. The fusion images of NSCT-RPCNN (f1-f3) and IFCNN (i1-i3) are insufficient in extracting detailed information. In addition, the NSCT-PCDC (e1-e3) and GFF (g1-g3) methods produce serious energy loss, which reduce the contrast of the fused image. Although NSST-DAPCNN (d1-d3) is better than the former method in detail extraction and energy preservation, it is still not as good as CS-MCA (h1-h3) and DRCRL (j1-j3) in the preservation of structural texture information. Since our DRCRL is designed to select dictionary filters with higher energy, it performs well in both structural information extraction and detail conservation.

3.2.2. Fusion analysis on MRI-PET

The fusion results are shown in Fig. 6. The first two columns are MRT(a1-a3) and PET(b1-b3) source images. The LLF-IOI (c1-c3) method handles the color information well, however the spatial information in the MR source image is excessively enhanced, thereby destroying the original structure information. NSCT-PCDC(e1-e3) and GFF(g1-g3) suffer from severe color distortion, and color information is not well preserved. The symptoms of IFCNN (h1-h3) are slightly lighter than the first two methods, but there is still obvious discoloration. The NSCT-RPCNN(f1-f3) and NSST-PAPCNN(d1-d3) achieve higher visual quality than other methods in color preservation. Although our DRCRL(i1-i3) is not particularly prominent in retaining source image energy, owe to our more precise description of the data characteristics, it has greatly improved the extraction of edge and structural information. All in all, our proposed model is the only one which can balance the structure and brightness information well.

3.2.3. Fusion analysis on MRI-SPECT

Similar to PET images, SPECT is a color three-dimensional medical image. Thus, the color image fusion strategy is also used. The three sets of fusion results are shown in Fig. 7 and the visual effects of the fused images are also similar to the MRI-PET fusion in the previous section.

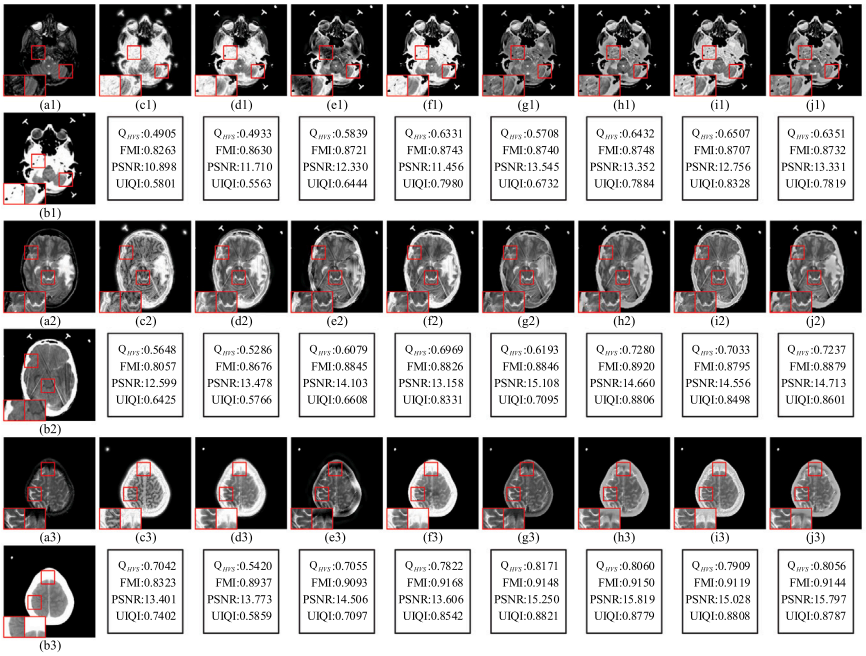

Fig. 5. The visual result of various fusion methods on MRI and CT images. The proposed method obtains more competitive results than other methods. Please refer to the main text for more detailed descriptions. The fused images obtained: (a1-a3) MR; (b1-b3) CT; (c1-c3) LLF-IOI; (d1-d3) NSST-PAPCNN; (e1-e3) NSCT-PCDC; (f1-f3) NSCT-RPCNN; (g1-g3) GFF; (h1-h3) CS-MCA; (i1-i3) IFCNN and (j1-j3) DRCRL.

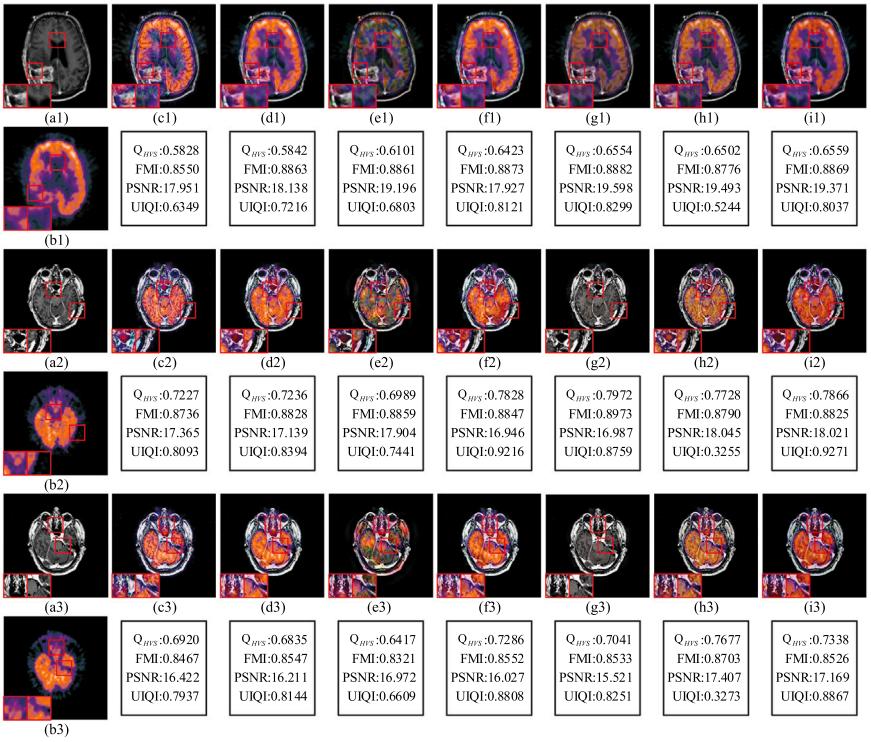

Fig. 6. The visual result of various fusion methods on MRI and PET images. The proposed method obtains more competitive results than other methods. Please refer to the main text for more detailed descriptions. The fused images obtained: (a1-a3) MRT; (b1-b3) PET; (c1-c3) LLF-IOI; (d1-d3) NSST-PAPCNN; (e1-e3) NSCT-PCDC; (f1-f3) NSCT-RPCNN; (g1-g3) GFF; (h1-h3) IFCNN and (i1-i3) DRCRL.

The difference between the DRCRL fusion image and the source image is small, and it has good structure and color information preservation in our DRCRL method. The performance proves that our DRCRL has a good fusion effect on MRI and SPECT images.

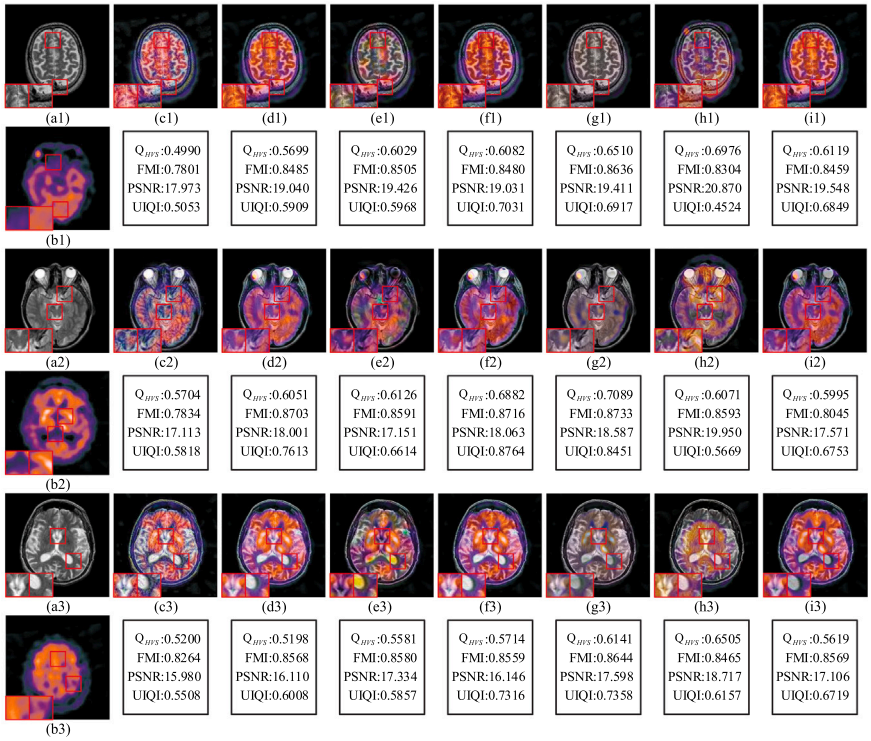

Fig. 7. The visual result of various fusion methods on MRI and SPECT images. The proposed method obtains more competitive results than other methods. Please refer to the main text for more detailed descriptions. Fused images obtained: (a1-a3) MRT; (b1-b3) SPECT; (c1-c3) LLF-IOI; (d1-d3) NSST-PAPCNN; (e1-e3) NSCT-PCDC; (f1-f3) NSCT-RPCNN; (g1-g3) GFF; (h1-h3) IFCNN and (i1-i3) DRCRL.

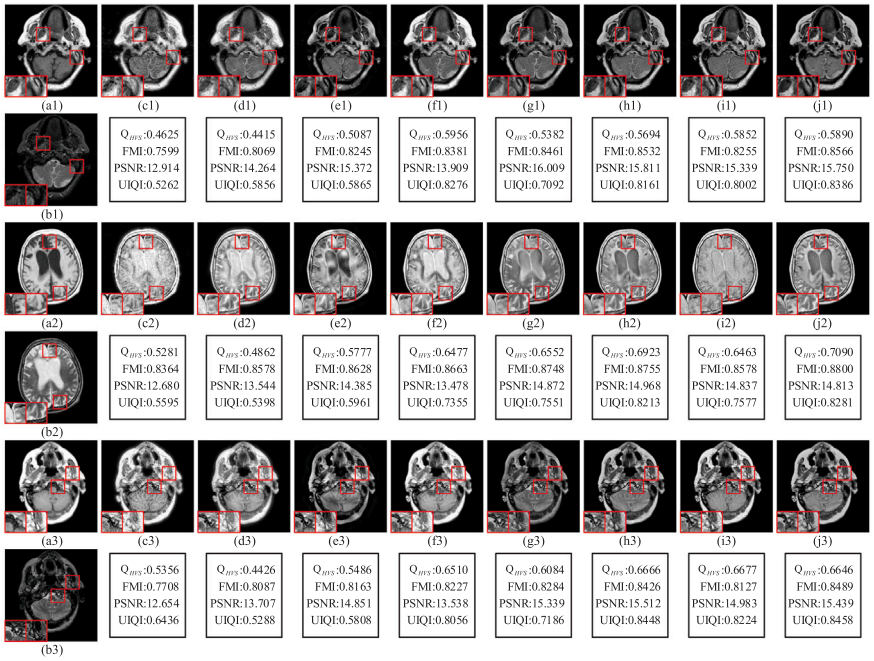

Fig. 8. The visual result of various fusion methods on MR-T1 and MR-T2 images. The proposed method obtains more competitive results than other methods. Please refer to the main text for more detailed descriptions. Fused images obtained: (a1-a3) T1; (b1-b3) T2; (c1-c3) LLF-IOI; (d1-d3) NSST-PAPCNN; (e1-e3) NSCT-PCDC; (f1-f3) NSCT-RPCNN; (g1-g3) GFF; (h1-h3) CS-MCA; (i1-i3) IFCNN and (j1-j3) DRCRL.

3.2.4. Fusion analysis on T1-T2

As shown in Fig. 8, the LLF-IOI(c1-c3) fusion images still have many artifacts caused by noise, which greatly reduces the quality of the fusion image. The NSCT-PCDC(e1-e3) fused images do not retain the edge information in the MR-T1 source image well. In addition, the GFF(g1-g3) and NSCT-PCDC(e1-e3) methods are still adversely affected by energy loss. NSCT-RPCNN(f1-f3), NSST-PAPCNN(d1-d3), IFCNN (i1-i3) are insufficient in extracting detailed information, and the texture is not clear enough. As can be seen from f2, the fused pictures of NSCT-PCDC suffer from the redundancy heavily. In comparison, DRCRL(j1-j3)

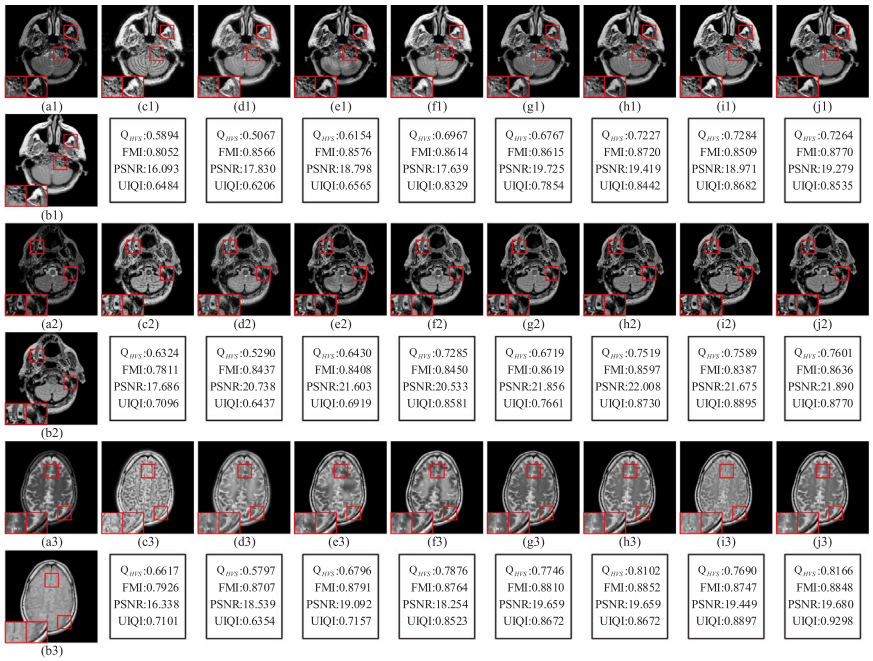

Fig. 9. The visual result of various fusion methods on MR-T2 and MR-PD images. The proposed method obtains more competitive results than other methods. Please refer to the main text for more detailed descriptions. Fused images obtained: (a1-a3) T2; (b1-b3) PD; (c1-c3) LLF-IOI; (d1-d3) NSST-PAPCNN; (e1-e3) NSCT-PCDC; (f1-f3) NSCT-RPCNN; (g1-g3) GFF; (h1-h3) CS-MCA; (i1-i3) IFCNN and (j1-j3) DRCRL.

Table 2
Objective performance of different fusion methods on 9 metrics.

	Methods	Q_{MVS} ↑	FMI-pixel↑	NCC↑	Q_E ↑	PSNR↑	$Q^{AB/F}$ ↑	SSIM↑	UIQI↑	$R_Q^{F/AB}$ ↑	F_{rank} ↓
MRI-CT	DRCRL	**0.7474** ± 0.0522	0.9002 ± 0.0203	0.8049 ± 0.0007	0.9962 ± 0.0006	15.313 ± 1.0857	0.5807 ± 0.0326	**0.9961** ± 0.0010	0.8585 ± 0.0278	0.4928 ± 0.0259	**2.61**
	CS-MCA [39]	0.7388 ± 0.0626	**0.9027** ± 0.0187	0.8050 ± 0.0008	0.9960 ± 0.0006	**15.341** ± 1.1282	0.5604 ± 0.0485	0.9961 ± 0.0010	0.8499 ± 0.0576	0.4978 ± 0.0261	2.89
	NSCT-RPCNN [54]	0.7301 ± 0.0455	0.8990 ± 0.0207	**0.8054** ± 0.0009	0.9928 ± 0.0016	13.583 ± 1.2551	0.5420 ± 0.0360	0.9940 ± 0.0016	0.8404 ± 0.0148	**0.5448** ± 0.0052	4.28
	GFF [52]	0.7166 ± 0.0862	0.8988 ± 0.0196	0.8048 ± 0.0006	**0.9974** ± 0.0007	15.289 ± 1.0240	0.5751 ± 0.0147	0.9961 ± 0.0009	0.8023 ± 0.0761	0.4461 ± 0.0379	4.22
	NSCT-PCDC [51]	0.6418 ± 0.0411	0.8951 ± 0.0184	0.8045 ± 0.0006	0.9967 ± 0.0012	14.530 ± 1.1958	0.5381 ± 0.0161	0.9953 ± 0.0013	0.6782 ± 0.0232	0.4513 ± 0.0431	5.67
	LLF-IOI [50]	0.6332 ± 0.0759	0.8245 ± 0.0146	0.8047 ± 0.0006	0.9926 ± 0.0018	13.263 ± 1.4491	0.3634 ± 0.0092	0.9933 ± 0.0021	0.6883 ± 0.0603	0.5254 ± 0.0032	6.89
	NSST-PAPCNN [65]	0.5556 ± 0.0347	0.8827 ± 0.0183	0.8049 ± 0.0006	0.9933 ± 0.0017	13.877 ± 1.2445	0.4865 ± 0.0232	0.9943 ± 0.0016	0.5967 ± 0.0314	0.5260 ± 0.0081	5.94
	IFCNN [25]	0.7399 ± 0.0472	0.8939 ± 0.0214	0.8054 ± 0.0008	0.8365 ± 0.0400	14.949 ± 1.2270	**0.6001** ± 0.0267	0.9956 ± 0.0012	**0.8619** ± 0.0184	0.5229 ± 0.0111	3.50
MRI-PET	DRCRL	**0.7384** ± 0.0598	0.8832 ± 0.0067	0.8051 ± 0.0008	0.9993 ± 0.0002	**18.442** ± 0.9988	0.6905 ± 0.0572	**0.9984** ± 0.0002	**0.8846** ± 0.0581	0.4396 ± 0.0250	2.67
	NSCT-RPCNN [54]	0.7322 ± 0.0605	0.8858 ± 0.0045	0.8057 ± 0.0008	0.9990 ± 0.0004	17.134 ± 0.8316	0.6946 ± 0.0664	0.9977 ± 0.0003	0.8821 ± 0.0474	**0.4683** ± 0.0208	3.28
	GFF [52]	0.7356 ± 0.0628	**0.8928** ± 0.0025	**0.8062** ± 0.0003	**0.9995** ± 0.0001	17.864 ± 1.6530	**0.7518** ± 0.0676	0.9980 ± 0.0005	0.8531 ± 0.0225	0.3701 ± 0.0440	**2.56**
	NSCT-PCDC [51]	0.6601 ± 0.0357	0.8832 ± 0.0072	0.8046 ± 0.0007	0.9991 ± 0.0001	18.181 ± 0.8356	0.6881 ± 0.0579	0.9983 ± 0.0002	0.7092 ± 0.0282	0.3680 ± 0.0079	4.94
	LLF-IOI [50]	0.6676 ± 0.0613	0.8622 ± 0.0107	0.8049 ± 0.0005	0.9989 ± 0.0004	17.367 ± 0.6394	0.6444 ± 0.0787	0.9978 ± 0.0002	0.7396 ± 0.0790	0.4431 ± 0.0216	5.50
	NSST-PAPCNN [65]	0.6739 ± 0.0636	0.8839 ± 0.0059	0.8052 ± 0.0008	0.9989 ± 0.0004	17.314 ± 0.8163	0.6774 ± 0.0548	0.9977 ± 0.0003	0.7985 ± 0.0567	0.4676 ± 0.0213	4.61
	IFCNN [25]	0.7305 ± 0.0519	0.8777 ± 0.0054	0.8052 ± 0.0010	0.8506 ± 0.0215	18.114 ± 0.9699	0.6927 ± 0.0505	0.9984 ± 0.0002	0.3841 ± 0.0780	0.4274 ± 0.0154	4.44
MRI-SPECT	DRCRL	0.7014 ± 0.0451	0.8652 ± 0.0137	0.8066 ± 0.0007	0.9996 ± 0.0002	**21.810** ± 1.5563	0.7401 ± 0.0332	0.9992 ± 0.0003	0.8613 ± 0.0237	0.4722 ± 0.0102	2.94
	NSCT-RPCNN [54]	0.7099 ± 0.0399	0.8657 ± 0.0135	0.8073 ± 0.0007	0.9994 ± 0.0003	20.470 ± 1.5609	0.7192 ± 0.0280	0.9988 ± 0.0005	**0.9121** ± 0.0158	0.4774 ± 0.0114	3.67
	GFF [52]	**0.7420** ± 0.0432	**0.8790** ± 0.0097	**0.8083** ± 0.0010	**0.9997** ± 0.0002	21.680 ± 1.5896	**0.7934** ± 0.0361	0.9992 ± 0.0003	0.8860 ± 0.0233	0.4104 ± 0.0098	**2.33**
	NSCT-PCDC [51]	0.6749 ± 0.0297	0.8689 ± 0.0117	0.8061 ± 0.0006	0.9996 ± 0.0004	21.768 ± 1.5789	0.7342 ± 0.0460	0.9992 ± 0.0003	0.7675 ± 0.0302	0.4161 ± 0.0161	3.94
	LLF-IOI [50]	0.5897 ± 0.0833	0.8035 ± 0.0166	0.8060 ± 0.0006	0.9992 ± 0.0003	20.413 ± 1.3764	0.6475 ± 0.0374	0.9988 ± 0.0004	0.6841 ± 0.0867	0.4435 ± 0.0129	6.50
	NSST-PAPCNN [65]	0.6375 ± 0.0478	0.8658 ± 0.0126	0.8066 ± 0.0007	0.9994 ± 0.0003	20.847 ± 1.5802	0.7370 ± 0.0279	0.9989 ± 0.0005	0.7852 ± 0.0390	**0.4785** ± 0.0107	3.94
	IFCNN [25]	0.6896 ± 0.0510	0.8582 ± 0.0126	0.8066 ± 0.0007	0.9350 ± 0.0158	21.729 ± 1.4378	0.7116 ± 0.0238	**0.9993** ± 0.0003	0.4683 ± 0.0886	0.4474 ± 0.0086	4.67
T1-T2	DRCRL	**0.6907** ± 0.0498	**0.8637** ± 0.0121	**0.8066** ± 0.0004	0.9967 ± 0.0005	15.361 ± 0.3528	**0.5799** ± 0.0194	0.9964 ± 0.0003	**0.8254** ± 0.0372	0.4734 ± 0.0161	**1.83**
	CS-MCA [39]	0.6813 ± 0.0498	0.8592 ± 0.0136	0.8063 ± 0.0005	**0.9968** ± 0.0004	**15.508** ± 0.3281	0.5600 ± 0.0294	**0.9964** ± 0.0003	0.8201 ± 0.0425	0.4748 ± 0.0179	2.28
	NSCT-RPCNN [54]	0.6624 ± 0.0415	0.8438 ± 0.0177	0.8065 ± 0.0005	0.9952 ± 0.0009	13.787 ± 0.4769	0.5077 ± 0.0318	0.9944 ± 0.0008	0.7943 ± 0.0350	**0.5192** ± 0.0033	4.33
	GFF [52]	0.6426 ± 0.0463	0.8532 ± 0.0187	0.8062 ± 0.0011	0.9966 ± 0.0005	15.216 ± 0.3292	0.5591 ± 0.0422	0.9961 ± 0.0003	0.7566 ± 0.0363	0.4537 ± 0.0333	4.11
	NSCT-PCDC [51]	0.5698 ± 0.0341	0.8365 ± 0.0176	0.8051 ± 0.0009	0.9961 ± 0.0006	14.882 ± 0.4484	0.4759 ± 0.0350	0.9958 ± 0.0005	0.6003 ± 0.0257	0.4481 ± 0.0296	6.00
	LLF-IOI [50]	0.5435 ± 0.0390	0.7895 ± 0.0271	0.8056 ± 0.0005	0.9943 ± 0.0010	12.928 ± 0.5368	0.3614 ± 0.0298	0.9931 ± 0.0010	0.6116 ± 0.0456	0.5032 ± 0.0058	6.89
	NSST-PAPCNN [65]	0.4713 ± 0.0299	0.8295 ± 0.0198	0.8058 ± 0.0005	0.9952 ± 0.0009	13.924 ± 0.4663	0.4644 ± 0.0470	0.9945 ± 0.0007	0.5428 ± 0.0196	0.5041 ± 0.0053	6.17
	IFCNN [25]	0.6669 ± 0.0415	0.8324 ± 0.0175	0.8065 ± 0.0004	0.8817 ± 0.0198	15.179 ± 0.4026	0.4879 ± 0.0355	0.9960 ± 0.0004	0.8064 ± 0.0314	0.5021 ± 0.0091	4.39
T2-PD	DRCRL	**0.7629** ± 0.0424	**0.8870** ± 0.0181	**0.8076** ± 0.0010	**0.9990** ± 0.0003	20.419 ± 1.1288	**0.6742** ± 0.0422	0.9987 ± 0.0003	**0.8895** ± 0.0310	0.5019 ± 0.0173	**1.78**
	CS-MCA [39]	0.7551 ± 0.0539	0.8843 ± 0.0193	0.8074 ± 0.0010	0.9990 ± 0.0003	**20.489** ± 1.1408	0.6678 ± 0.0437	**0.9988** ± 0.0003	0.8831 ± 0.0506	0.5027 ± 0.0168	2.44
	NSCT-RPCNN [54]	0.7356 ± 0.0461	0.8747 ± 0.0235	0.8069 ± 0.0009	0.9985 ± 0.0006	19.016 ± 1.2457	0.6231 ± 0.0477	0.9981 ± 0.0005	0.8472 ± 0.0430	0.5113 ± 0.0253	4.72
	GFF [52]	0.7138 ± 0.0483	0.8815 ± 0.0203	0.8075 ± 0.0008	0.9990 ± 0.0003	20.409 ± 1.1873	0.6712 ± 0.0536	0.9987 ± 0.0004	0.8281 ± 0.0475	0.5064 ± 0.0114	3.33
	NSCT-PCDC [51]	0.6473 ± 0.0282	0.8739 ± 0.0239	0.8060 ± 0.0007	0.9986 ± 0.0005	19.860 ± 1.2618	0.6097 ± 0.0517	0.9985 ± 0.0004	0.6987 ± 0.0397	0.4944 ± 0.0198	5.67
	LLF-IOI [50]	0.6325 ± 0.0321	0.7998 ± 0.0270	0.8059 ± 0.0006	0.9976 ± 0.0009	16.981 ± 1.1770	0.4430 ± 0.0448	0.9970 ± 0.0008	0.6841 ± 0.0465	0.4896 ± 0.0261	7.78
	NSST-PAPCNN [65]	0.5423 ± 0.0545	0.8678 ± 0.0234	0.8064 ± 0.0007	0.9984 ± 0.0007	19.172 ± 1.2696	0.5750 ± 0.0520	0.9981 ± 0.0005	0.6395 ± 0.0565	0.5066 ± 0.0243	6.39
	IFCNN [25]	0.7377 ± 0.0324	0.8669 ± 0.0262	0.8071 ± 0.0009	0.9990 ± 0.0005	20.256 ± 1.2007	0.5998 ± 0.0499	0.9988 ± 0.0004	0.8790 ± 0.0186	**0.5251** ± 0.0075	3.89

have better visual effects because of our proposed sensitivity decoupled operator. The contrast can also be seen from the subgraph of h3 and i3.

3.2.5. Fusion analysis on T2-PD

It can be observed from Fig. 9 that our DRCRL has better fusion effectiveness than other methods. Among all the methods, the fused images of LLF-IOI (c1-c3) become blurred and the spatial and detailed features of the source images are seriously damaged. While NSST-PAPCNN (d1-d3) performs better than LLF-IOI, there are still some problems such as blurred borders and loss of detailed information, which is not as good as our DRCRL(j1-j3). In addition, NSCT-PCDC(e1-e3) and GFF(g1-g3) methods have obvious adverse visual artifacts in many fusion areas, which will affect medical observation and diagnosis. And the fused images of NSCT-RPCNN(f1-f3) do not well realize the fusion of source images in terms of image brightness. As for IFCNN (i1-i3), the average brightness is too high, which causes the blurring of image details. Although the fusion images of CSMCA (h1-h2) and our DRCRL are relatively close, DRCRL shows a better processing result in the details because we solve the component entanglement problem

Table 3
The comparison of metrics value with different number of dictionary filters.

E	Q_{HVS}	FMI-pixel	NCC	Q_E	PSNR	$Q^{AB/F}$	TIME	UIQI	$R_Q^{F/AB}$
6	0.7093	0.8788	0.80678	**0.9978**	17.038	0.6235	**38.781**	0.8361	0.4939
8	0.7112	0.8799	0.80678	0.9977	17.049	0.6257	45.147	0.8369	0.4957
10	**0.7321**	0.8804	0.80680	**0.9978**	**17.071**	**0.6290**	54.094	0.8702	0.4949
12	0.7263	0.8798	0.80677	0.9977	17.044	0.6276	61.068	0.8602	0.4957
14	**0.7321**	0.8812	0.80679	0.9977	17.047	0.6255	69.664	**0.8704**	0.4960
20	0.5895	**0.8819**	**0.80683**	0.9977	17.051	0.6276	97.675	0.6978	0.4961
28	0.5577	0.8818	0.80674	0.9977	17.058	0.6260	132.02	0.5390	**0.4962**

Table 4
The comparison of metrics values in Ablation analysis.

	Q_{HVS}	FMI-pixel	NCC	Q_E	PSNR	$Q^{AB/F}$	SSIM	UIQI	$R_Q^{F/AB}$
CS-MCA model	0.5367	0.8335	0.8054	0.9983	16.478	0.4571	0.9980	0.4577	0.5020
CS-MCA+ Decoupled	0.5907	0.8441	0.8067	0.9989	18.691	0.6275	0.7864	0.7931	0.5031
CSMCA+ADMM	0.6173	0.8511	0.8062	0.9987	19.716	0.5981	0.7198	0.8106	0.5123
CSMCA+CG	0.7126	0.8861	0.8071	0.9991	19.102	0.6388	0.8165	0.8052	0.5089
DRCRL model	**0.7680**	**0.8977**	**0.8077**	**0.9993**	**20.555**	**0.6990**	**0.9990**	**0.8760**	**0.5158**

Table 5
The average running time of different methods.

Methods	DRCRL	CS-MCA	GFF	NSCT-RPCNN	NSCT-PCDC	NSST-PAPCNN	LLF-IOI	IFCNN
Times	51.83	137.38	0.06	8.43	15.14	6.86	175.04	0.02

systematically and achieve finer texture maps. Overall, our method achieved more competitive visual quality in terms of detail extraction, denoising, and visual artifact prevention.

3.3. Quantitative metrics analysis

We randomly select a total of 83 sets of image pairs to analyze the performance of the fusion method, including 22 sets of T1-T2 image pairs, 31 sets of T2-PD image pairs and MRI-CT, MRI-PET and MRI-SPECT 10 sets. Table 2 shows the objective performance of all the methods on 9 metrics. The first for each indicator is marked in bold and the second are marked in red. To analyze the performance of all methods on each fusion problem comprehensively, we use the average rank F_{rank} from Friedman test as an overall assessment [66]. To show the variation range of the data, we also provide the standard deviation of the results. While the results presented in Table 2 show that our DRCRL does not perform well on $R_Q^{F/AB}$ indicator, our method get good result on Q_E. Like $R_Q^{F/AB}$, Q_E is also edge information based image fusion metric. It measures how much of the salient information contained in each of the input images has been transferred into the fused image without introducing distortions [57], and has been widely adopted in many other published works [35,65,67]. So it reasonable to use Q_E as an aid when determining the effect of the fusion methods.

In grayscale fusion (i.e. MRI-CT, T1-T2, T2-PD), DRCRL fusion has good visual effects, mutual information function, original image information preservation, and image distortion processing. Except for individual indicator ($R_Q^{F/AB}$, for reasons explained above), they basically remained at the first and second levels. The overall assessment F_{rank} also keep first in all three experiments. Some indicators such as Q_E and SSIM are lower than the baseline, but the difference is not large, and the Q_E and SSIM metrics are actually derived from the UIQI indicator. From the comparison results of UIQI, we can see that our method ranks first in structural similarity. In T1-T2 and T2-PD, compared with IFCNN, which takes third place on UIQI, our DRCRL model shows improvements of 2.4% and 1.2% respectively. In MRI-CT, although the performance of our proposed model is not as good as IFCNN, the score of DRCRL surpasses that of its baseline CS-MCA by 1.0%. Our good performance on the preservation of structural information can be attributed to our regularization learning which describes the data characteristics of different components more accurately. As for the redundancy elimination capability, which is measured by PSNR and Q_{HVS}, our DRCRL maintain first or second among all seven methods. Especially for Q_{HVS}, the score of our proposed model surpasses that of the second by 1.0%, 1.4%, 1.0% in MRI-CT, T1-T2, T2-PD respectively, which proves the good redundancy elimination capability of our model forcefully.

In color fusion (that is, MRI-PET and MRI-SPECT), we mainly use the HSI method to calculate the color fusion index which means that the original image and the fused image are regarded as RGB color images. After switching to the HSI color space, the I component brightness is taken as a basis for judgment. As we can see from Table 2, the color fusion metrics of DRCRL are not as effective as grayscale fusion. Although GFF shows better performance than our method in the fusion metrics on color images, their color distortion is very serious. In terms of visual effects, our method is significantly better than other methods in terms of color fidelity and institutional similarity. Although it is not ranked high on information theory based metrics, the PSNR is still kept to a maximum, which proves that our method can retain the original source image information to the greatest extent. More specifically, our PSNR is 3.8% and 2.7% above average in MRI-PET and MRI-SPECT respectively, which is a relatively substantial lead. Our good performance on PSNR also confirm the assessment of our redundancy elimination capability in grayscale fusion. In addition, as can be seen from the standard deviations, the fluctuation of our results remain at a low level, which means our model is stable and robust.

3.4. Model parameters and performance analysis

The parameter analysis, iterative analysis, and Ablation analysis are carried out in this section to get more analysis of model fusion performance to verify the advantages of our method.

3.4.1. Parameters analysis

Since the sparse representation of the image in the model is obtained through dictionary filter learning, we take different numbers of filters (that is, E can be defined: 6, 8, 10, 12, 14, 20, 28) to conduct experiments for analyzing the effect of the number of dictionary filters on the fusion performance. The experimental metrics based on different numbers of dictionary filters are shown in Table 3. If the value of E exceeds 10, the information theory based metrics does not increase

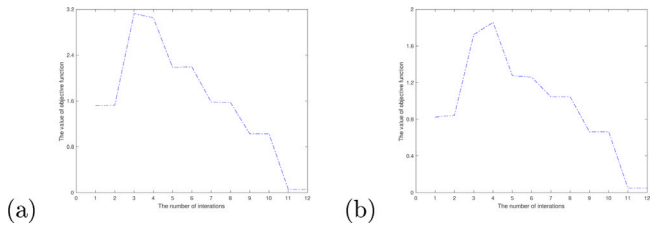

Fig. 10. The convergence chain diagram of grayscale fusion and color fusion.

significantly with the increase of E. However, the image structure similarity based metrics such as $Q^{AB/F}$ decreases significantly. Considering the running time and fusion effect, we will set the number of dictionary filters to 10 in the following experiments.

3.4.2. Convergence analysis

The Fig. 10 reveals the convergence chain diagram of our method, in which one interaction is divided into two steps to show the update of different components. As we can see from the diagram, our objective function approaches zero after a few iterations both in grayscale fusion or color fusion. While at the beginning of the iteration, we can see an increase of the value of the objective function. The reason for this phenomenon relates to the effect of the component entanglement. At the beginning, the high-energy dictionary filters have not been completely selected and the result still contain lots of redundancy. Thus, the optimization of the objective function of one component may lead to the degradation of the other. Due to the decoupled representation, the redundancy is gradually removed and we can see the value of the general objective function begin to drop continuously. Finally, we reach the global optimum solution which demonstrates the effectiveness of the decoupled representation. It also noticeably indicates that our DRCRL model has good convergence.

3.4.3. Ablation analysis

As shown in Fig. 11, the subjective evaluation results of the ablation experiments demonstrate the impact of disentangled representation, ADMM, and CG on model performance under the same dataset and experimental conditions. It is evident that the DRCRL model exhibits superior fusion effects. As indicated in Table 4, the DRCRL model achieves optimal results across various metrics. Notably, the significant improvement in QHVS and UIQI indices confirms that disentangled representation, ADMM, and CG contribute to enhanced fusion effects, reflecting the effectiveness of dual sensitivity regularized learning in terms of image smoothness and structural features.

3.4.4. Running time

Table 5 presents the average running times of different methods. DRCRL's efficiency is 165% higher than that of CS-MCA, which vividly demonstrates the capability of sensitivity decoupled representation learning and regularization learning to significantly accelerate model performance. Although CS-MCA excels in sparse representation, its high computational complexity results in longer running times. While GFF and IFCNN boast extremely high efficiency, they respectively suffer from color distortion and image detail blurring issues. NSCT and NSST related algorithms perform well in maintaining good visual quality and efficiency, yet they still face challenges with edge preservation and data redundancy. The LLF-IOI algorithm, due to its use of local Laplacian filtering techniques, has high complexity and the longest running time. Therefore, the DRCRL algorithm achieves a good balance between efficiency and performance.

4. Conclusion

In this paper, we propose a new SR-based fusion method named decoupled representation and component-wise regularization learning (DRCRL) for medical image fusion. The decoupled representation learning and two different regularization operators are introduced to reduce the adverse effects of redundancy and improve the efficiency of our proposed model. The optimization process of DRCRL is implemented by the ADMM algorithm and the conjugate gradient method. Finally, extensive and effective experiments confirm that our method significantly improves the efficiency and fusion performance.

CRediT authorship contribution statement

Rui Zhang: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Conceptualization. **Haoze Sun:** Writing – review & editing, Software, Methodology, Investigation. **Lizhen Deng:** Writing – review & editing, Validation, Methodology, Investigation, Conceptualization. **Hu Zhu:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Wei Qian:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization.

Declaration of competing interest

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

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant 62072256.

Data availability

Data will be made available on request.

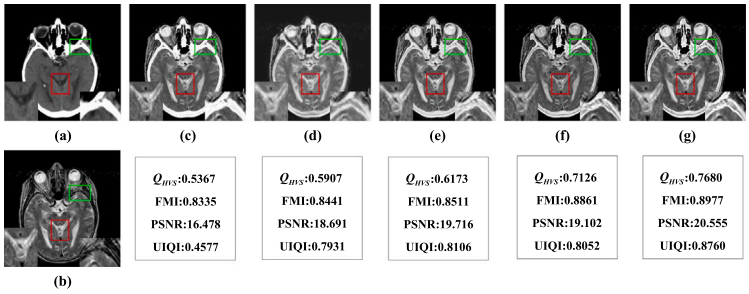

Fig. 11. Visualization of ablation experiments on MRI and CT images. The fused images obtained: (a) CT; (b) MRI; (c) CSMCA; (d) CSMCA+Decoupled; (e) CSMCA+ADMM; (f) CSMCA+CG; (g) DRCRL.

References

[1] A.P. James, B.V. Dasarathy, Medical image fusion: A survey of the state of the art, Inf. Fusion 19 (2014) 4–19.

[2] O.C. Do, C.M. Luong, P.-H. Dinh, G.S. Tran, An efficient approach to medical image fusion based on optimization and transfer learning with vgg19, Biomed. Signal Process. Control 87 (2024) 105370.

[3] L.A. Klein, Sensor and Data Fusion Concepts and Applications, Society of Photo-Optical Instrumentation Engineers (SPIE, 1999.

[4] Y. Dai, Z. Yan, J. Cheng, X. Duan, G. Wang, Analysis of multimodal data fusion from an information theory perspective, Inform. Sci. 623 (2023) 164–183.

[5] M.M. Laganà, M.G. Preti, L. Forzoni, S. D'Onofrio, S. De Beni, A. Barberio, P. Cecconi, G. Baselli, Transcranial ultrasound and magnetic resonance image fusion with virtual navigator, IEEE Trans. Multimedia 15 (5) (2013) 1039–1048.

[6] T. O'brien, E. So, B. Mullan, M. Hauser, B. Brinkmann, N. Bohnen, D. Hanson, G. Cascino, C. Jack, F. Sharbrough, Subtraction ictal spect co-registered to mri improves clinical usefulness of spect in localizing the surgical seizure focus, Neurology 50 (2) (1998) 445–454.

[7] Z. Zhu, X. He, G. Qi, Y. Li, B. Cong, Y. Liu, Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal mri, Inf. Fusion 91 (2023) 376–387.

[8] Y. Ding, L. Gong, M. Zhang, C. Li, Z. Qin, A multi-path adaptive fusion network for multimodal brain tumor segmentation, Neurocomputing 412 (2020) 19–30.

[9] G. Bhatnagar, Q.J. Wu, Z. Liu, Directive contrast based multimodal medical image fusion in nsct domain, IEEE Trans. Multimedia 15 (5) (2013) 1014–1024.

[10] P. Gupta, N. Jain, Anisotropic diffusion filter based fusion of nsst transformed medical images, Biomed. Signal Process. Control 90 (2024) 105819.

[11] S. Singh, R. Anand, Multimodal medical image sensor fusion model using sparse k-svd dictionary learning in nonsubsampled shearlet domain, IEEE Trans. Instrum. Meas. 69 (2) (2019) 593–607.

[12] B. Wang, H. Niu, J. Zeng, G. Bai, S. Lin, Y. Wang, Latent representation learning model for multi-band images fusion via low-rank and sparse embedding, IEEE Trans. Multimed. (2020).

[13] S. Singh, D. Gupta, Detail enhanced feature-level medical image fusion in decorrelating decomposition domain, IEEE Trans. Instrum. Meas. 70 (2020) 1–9.

[14] G. Xu, X. Deng, X. Zhou, M. Pedersen, L. Cimmino, H. Wang, Fcfusion: Fractal componentwise modeling with group sparsity for medical image fusion, IEEE Trans. Ind. Inform. 18 (12) (2022) 9141–9150.

[15] W. Kong, Q. Miao, R. Liu, Y. Lei, J. Cui, Q. Xie, Multimodal medical image fusion using gradient domain guided filter random walk and side window filtering in framelet domain, Inform. Sci. 585 (2022) 418–440.

[16] S. Singh, R.S. Anand, Multimodal medical image sensor fusion model using sparse k-svd dictionary learning in nonsubsampled shearlet domain, IEEE Trans. Instrum. Meas. 69 (2) (2020) 593–607.

[17] Z. Wang, Y. Ma, Medical image fusion using m-pcnn, Inf. Fusion 9 (2008) 176–185.

[18] E.A. Bernal, X. Yang, Q. Li, J. Kumar, S. Madhvanath, P. Ramesh, R. Bala, Deep temporal multimodal fusion for medical procedure monitoring using wearable sensors, IEEE Trans. Multimed. 20 (1) (2017) 107–118.

[19] M. Das, D. Gupta, P. Radeva, A.M. Bakde, Optimized ct-mr neurological image fusion framework using biologically inspired spiking neural model in hybrid l1- l0 layer decomposition domain, Biomed. Signal Process. Control 68 (2021) 102535.

[20] L. Yin, M. Zheng, G. Qi, Z. Zhu, F. Jin, J. Sim, A novel image fusion framework based on sparse representation and pulse coupled neural network, IEEE Access 7 (2019) 98290–98305.

[21] G. Wang, W. Li, X. Gao, B. Xiao, J. Du, Multimodal medical image fusion based on multichannel coupled neural p systems and max-cloud models in spectral total variation domain, Neurocomputing 480 (2022) 61–75.

[22] T. Zhou, Q. Li, H. Lu, Q. Cheng, X. Zhang, Gan review: Models and medical image fusion applications, Inf. Fusion 91 (2023) 134–148.

[23] H. Liu, S. Li, J. Zhu, K. Deng, M. Liu, L. Nie, Ddifn: A dual-discriminator multi-modal medical image fusion network, ACM Trans. Multimedia Comput. Commun. Appl. 19 (4) (2023) 1–17.

[24] J.-L. Yin, B.-H. Chen, Y.-T. Peng, H. Hwang, Automatic intermediate generation with deep reinforcement learning for robust two-exposure image fusion, IEEE Trans. Neural Netw. Learn. Syst. (2021) 1–10, http://dx.doi.org/10.1109/TNNLS.2021.3088907.

[25] Y. Zhang, Y. Liu, P. Sun, H. Yan, X. Zhao, L. Zhang, Ifcnn: A general image fusion framework based on convolutional neural network, Inf. Fusion 54 (2020) 99–118.

[26] X. He, G. Qi, Z. Zhu, Y. Li, B. Cong, L. Bai, Medical image segmentation method based on multi-feature interaction and fusion over cloud computing, Simul. Model. Pract. Theory 126 (2023) 102769.

[27] G. Zhang, R. Nie, J. Cao, L. Chen, Y. Zhu, Fdgnet: A pair feature difference guided network for multimodal medical image fusion, Biomed. Signal Process. Control 81 (2023) 104545.

[28] B. Yang, S. Li, Multifocus image fusion and restoration with sparse representation, IEEE Trans. Instrum. Meas. 59 (4) (2020) 884–892.

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

[30] H. Yin, Tensor sparse representation for 3-d medical image fusion using weighted average rule, IEEE Trans. Biomed. Eng. 65 (11) (2018) 2622–2633.

[31] Q. Zhang, M. Levine, Robust multi-focus image fusion using multi-task sparse representation and spatial context, IEEE Trans. Image Process. 25 (5) (2016) 2045–2058.

[32] H. Li, X. He, D. Tao, Y. Tang, R. Wang, Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning, Pattern Recognit. 79 (2018) 130–146.

[33] M. Kim, D. Han, H. Ko, Joint patch clustering-based dictionary learning for multimodal image fusion, Inf. Fusion 27 (2016) 198–214.

[34] Z. Zhu, H. Yin, Y. Chai, Y. Li, G. Qi, A novel multi-modality image fusion method based on image decomposition and sparse representation, Inform. Sci. 432 (2018) 516–529.

[35] Y. Jiang, M. Wang, Image fusion with morphological component analysis, Inf. Fusion 18 (1) (2014) 107–118.

[36] J. Starck, M. Elad, D. Donoho, Redundant multiscale transforms and their application for morphological component separation, Adv. Image Electron Phys. 132 (2004) 287–348.

[37] M. Danelljan, G. Bhat, F.S. Khan, M. Felsberg, Eco: Efficient convolution operators for tracking, in: IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 6931–6939.

[38] Y. Wang, Q. Yao, J.T. Kwok, L.M. Ni, Online convolutional sparse coding with sample-dependent dictionary, in: International Conference on Machine Learning, ICML, 2018.

[39] Y. Liu, X. Chen, R.K. Ward, Z.J. Wang, Medical image fusion via convolutional sparsity based morphological component analysis, IEEE Signal Process. Lett. 26 (3) (2019) 485–489.

[40] T. Rohlfing, Image similarity and tissue overlaps as surrogates for image registration accuracy: widely used but unreliable, IEEE Trans. Med. Imaging 31 (2) (2011) 153–163.

[41] F. Duan, Z. Sun, A modified liu-storey conjugate gradient method and its global convergence for unconstrained optimization, in: 2010 Chinese Control and Decision Conference, 2010, pp. 1585–1588.

[42] F. Meyer, A.Z. Averbuch, R. Coifman, Multilayered image representation: Application to image compression, IEEE Trans. Image Process. 11 (9) (2002) 1072–1080.

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

[44] Y. Jiang, M. Wang, Image fusion with morphological component analysis, Inf. Fusion 18 (2014) 107–118.

[45] W.W. Hager, Updating the inverse of a matrix, SIAM Rev. 31 (2) (1989) 221–239.

[46] J. Pan, Z. Su, Fast ℓ_0-regularized kernel estimation for robust motion deblurring, IEEE Signal Process. Lett. 20 (9) (2013).

[47] J. Nocedal, S. Wright, Numerical Optimization, Springer Science & Business Media, 2006.

[48] A. James, B. Dasarathy, Medical image fusion: A survey of the state of the art, Inf. Fusion 19 (2014) 4–19.

[49] J.A. Johnson, J.A. Becker, The whole brain altas, 1997, doi: http://www.med.harvard.edu/aanlib/. Online.

[50] J. Du, W. Li, B. Xiao, Anatomical-functional image fusion by information of interest in local laplacian filtering domain, IEEE Trans. Image Process. 26 (12) (2017).

[51] G. Bhatnagar, Q. Wu, Z. Liu, Directive contrast based multimodal medical image fusion in nsct domain, IEEE Trans. Multimed. 15 (5) (2013).

[52] S. Li, X. Kang, J. Hu, Image fusion with guided filtering, IEEE Trans. Image Process. 22 (7) (2013).

[53] X. Li, F. Zhou, H. Tan, W. Zhang, C. Zhao, Multimodal medical image fusion based on joint bilateral filter and local gradient energy, Inform. Sci. 569 (2021) 302–325.

[54] S. Das, M. Kundu, A neuro-fuzzy approach for medical image fusion, IEEE Trans. Biomed. Eng. 60 (12) (2013).

[55] Y. Chen, R.S. Blum, A new automated quality assessment algorithm for image fusion, Image Vis. Comput. 27 (2009) 1421–1432.

[56] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.

[57] G. Piella, H. Heijmans, A new quality metric for image fusion, in: Proceedings 2003 International Conference on Image Processing, 2003, pp. III–173.

[58] M. Haghighat, M.A. Razian, Fast-fmi: Non-reference image fusion metric, in: 2014 IEEE 8th International Conference on Application of Information and Communication Technologies, AICT, 2014, pp. 1–9.

[59] R. Hong, Objective image fusion performance measure, Milit. Tech. Cour. 56 (2) (2000) 181–193.

[60] Z. Wang, A. Bovik, A universal image quality index, IEEE Signal Process. Lett. 9 (3) (2002).

[61] Q. Wang, Y. Shen, Performances evaluation of image fusion techniques based on nonlinear correlation measurement, in: The 21st IEEE Instrumentation and Measurement Technology Conference, Vol. 1, 2004, pp. 472–475.

[62] A. Sengupta, A. Seal, C. Panigrahy, O. Krejcar, A. Yazidi, Edge information based image fusion metrics using fractional order differentiation and sigmoidal functions, IEEE Access 8 (2020) 88385–88398.
[63] G. Rajasekhar, V. Prasad, M.V. Srikanth, Multilevel medical image fusion using multi-level local extrema and non sub-sampled contourlet transformation, in: 2017 International Conference on Signal Processing and Communication, ICSPC, 2017, pp. 246–252.
[64] J. Du, W. Li, K. Lu, et al., An overview of multi-modal medical image fusion, Neurocomputing 215 (26) (2016) 3–20.
[65] M. Yin, X. Liu, Y. Liu, X. Chen, Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain, IEEE Trans. Instrum. Meas. 68 (1) (2019).
[66] G. Liang, An effective method for imbalanced time series classification: hybrid sampling, in: Australasian Joint Conference on Artificial Intelligence, Springer, 2013, pp. 374–385.
[67] Y. Liu, S. Liu, Z. Wang, A general framework for image fusion based on multi-scale transform and sparse representation, Inf. Fusion 24 (2015) 147–164.