



# Cross-Modality Reference and Feature Mutual-Projection for 3D Brain MRI Image Super-Resolution

Lulu Wang<sup>1</sup> · Wanqi Zhang<sup>2</sup> · Wei Chen<sup>2</sup> · Zhongshi He<sup>2</sup> · Yuanyuan Jia<sup>3</sup> · Jinglong Du<sup>3</sup>

Received: 13 November 2023 / Revised: 30 March 2024 / Accepted: 21 April 2024  
© The Author(s) under exclusive licence to Society for Imaging Informatics in Medicine 2024

## Abstract

High-resolution (HR) magnetic resonance imaging (MRI) can reveal rich anatomical structures for clinical diagnoses. However, due to hardware and signal-to-noise ratio limitations, MRI images are often collected with low resolution (LR) which is not conducive to diagnosing and analyzing clinical diseases. Recently, deep learning super-resolution (SR) methods have demonstrated great potential in enhancing the resolution of MRI images; however, most of them did not take the cross-modality and internal priors of MR seriously, which hinders the SR performance. In this paper, we propose a cross-modality reference and feature mutual-projection (CRFM) method to enhance the spatial resolution of brain MRI images. Specifically, we feed the gradients of HR MRI images from referenced imaging modality into the SR network to transform true clear textures to LR feature maps. Meanwhile, we design a plug-in feature mutual-projection (FMP) method to capture the cross-scale dependency and cross-modality similarity details of MRI images. Finally, we fuse all feature maps with parallel attentions to produce and refine the HR features adaptively. Extensive experiments on MRI images in the image domain and k-space show that our CRFM method outperforms existing state-of-the-art MRI SR methods.

**Keywords** Cross-scale self-similarity · Cross-modality similarity · Magnetic resonance imaging · Reference-based super-resolution · Convolutional neural network

## Introduction

Magnetic resonance imaging (MRI) is a non-ionizing radiation imaging technique that is crucial for diagnosing conditions such as brain tumors, intracranial infections, cerebral ect. The quality of MRI images is affected by many factors like the signal-to-noise ratio (SNR) and spatial resolution. In clinical practice, MRI scanning thickness is usually increased to decrease the imaging time to meet SNR requirements and reduce motion artifacts caused by the movements of scanned objects. However, increasing the scanning thickness will produce low-resolution (LR) MRI images, which hinders the precision of follow-up analyses and diagnoses. As a post-processing method, super-resolution (SR) is able to effectively enhance the resolution of MRI scans without upgrading or replacing existing hardware equipment. However, the SR problem remains challenging due to there being multiple solutions for any LR-to-high-resolution (HR) mappings.

In recent years, deep learning represented by convolutional neural networks (CNNs) has become increasingly attractive in solving the SR problem. They use series of convolution operations to automatically extract

✉ Lulu Wang  
luluwang@kust.edu.cn  
Wanqi Zhang  
wqzhang@cqu.edu.cn  
Wei Chen  
mlg\_cwei@cqu.edu.cn  
Zhongshi He  
zshe@cqu.edu.cn  
Yuanyuan Jia  
yuanyuanjia@cqmu.edu.cn  
Jinglong Du  
jldu@cqmu.edu.cn

<sup>1</sup> Faculty of Information Engineering and Automation, Kunming University of Science and Technology and Yunnan Key Laboratory of Computer Technologies Application, Kunming 650500, China

<sup>2</sup> College of Computer Science, Chongqing University, Chongqing 400044, China

<sup>3</sup> Medical Data Science Academy and College of Medical Informatics, Chongqing Medical University, Chongqing 400016, China

hierarchical features and optimize parameters via numerous samples, which has achieved impressive performance in restoring and enhancing the degraded sharpness, contrast, and texture of the HR MRI image. Current efforts mainly focus on increasing network depth or width through various techniques to improve the ability of fitting SR mapping functions. However, increasing the network size does not bring about significant MRI image SR performance improvements, because it is continuously extracting redundant features with heavy computing load. In the diagnosis of certain diseases (glioma), it is difficult to extract all the necessary information from a single MRI modality to ensure clinical accuracy and examination intensity. Therefore, the complementary characteristics of multi-modality MRI image information are usually used to improve the diagnostic accuracy of certain diseases.

In this paper, we propose cross-modality reference and feature mutual-projection (CRFM) to enhance the spatial resolution of MRI image. Our model receives two types of inputs: LR images of a certain MRI modality are used to generate SR images, and HR volumes of another modality provide reference features for accurate SR. Specifically, the CRFM network uses cascaded residual channel attention (RCA) blocks to extract features from LR inputs. In this process, we propose a feature mutual-projection (FMP) method according to the cross-scale similarity of the image to capture the internal correlations of repeated plaques in features at different scales. Moreover, we extract the gradients of HR images of referenced imaging modality and feed them into the FMP module, complementing the true external HR details for the SR task. At the tail of CRFM network, we upscale all feature maps and then fuse these maps with the mutual projected features and reference gradients to predict the missed HR details. In addition, cross-scale residual learning is adopted to facilitate parameter optimization. Extensive experiments show that our CRFM surpasses some existing 3D brain MRI image SR techniques.

The contributions of this paper are summarized as follows:

- We propose an reference-based MRI image SR method that fully utilizes image gradients from reference MRI modality. Moreover, for the case in which no reference exists, we propose a single image SR method cross-scale feature transformer (CFT) network that only uses the self-similarity of different scale MRI images to reconstruct HR details.
- We design a feature mutual-projection method by cross-scale feature matching via transformer according to the self-similarity of MRI images, which can be flexibly inserted anywhere in the SR network.
- We develop a parallel channel and spatial attention to achieve efficient feature refinement and enhancement, and meanwhile producing the HR details.

The remainder of this paper is organized as follows: Section 2 reviews the [Related Work](#). The details of the proposed CRFM method are provided in [Materials and Methods](#). The implementation details, ablation study, and comparisons with state-of-the-art methods are presented in [Results](#). Finally, the last section concludes this paper ([Conclusion](#)).

## Related Work

To deal with the ill-posed resolution reconstruction problem, various techniques have been developed and can be roughly grouped into interpolation-based, reconstruction-based [1, 2], and learning-based SR methods [3–6]. Among them, interpolation algorithms are easy to perform but may lead to blocking, ringing, and jagged artifacts. In contrast, reconstruction-based SR methods simulate the process of MRI and introduce priors to improve image quality. As one kind of data-driven technique, learning-based SR methods learn complicated LR-to-HR mappings from large numbers of training samples. Although reconstruction-based and traditional learning-based methods have made noteworthy progress in enhancing the resolution of MRI images, it is difficult to utilize the insufficient additional information and limited representation capabilities to solve the challenging MRI SR reconstruction problem.

Recently, deep learning has achieved impressive success in natural image SR, and some techniques have been introduced to address the MRI SR problem [7, 8]. There are two different types of CNN-based MRI SR methods including single image SR (SISR) and reference-based SR (RefSR). SISR methods focus on learning a spatial mapping functions to restore HR images from a given LR acquisition. Dong et al. [9] initiatively designed a three-layer reconstruction network (called SRCNN) to enhance the resolution of two-dimensional (2D) natural images. Then, Pham et al. [10] proposed a three-dimensional (3D) SRCNN model to produce HR 3D brain MRI images. It is well-known that deep learning networks could enhance the representation ability by increasing their depth and width. However, these models may be difficult to optimize due to the vanishing or exploding gradient problem. To alleviate training difficulty, residual learning [11, 12] and dense connections [13, 14] have been widely applied by MRI image SR networks. Shi et al. [15] integrated global connections and local skips into a progressive wide residual network to reconstruct HR MRI slices. Similarly, Oktav et al. [16] and Giannakidis et al. [17] adopted residual learning in increasing the spatial resolution of cardiac and brain MRI images, respectively.

In addition, some well-designed strategies have been developed in unlocking the restoration capacity of MRI image SR networks, such as multi-scale learning [18], attention mechanism [19, 20], generative adversarial

networks [21, 22], and multi-branch networks [23]. Wang et al. [24] designed a 3D attention mechanism to make the network concentrate on meaningful features and regions that are more conducive to improving the resolution of MRI image. Wang et al. [25] constructed a convolution and deconvolution model to increase the resolution of 3D MRI scans, which used convolution and deconvolution kernels in parallel to obtain different levels of features and enrich the feature extraction methods. Zhao et al. [26] put forward a channel splitting method to input features into two sub-networks with different information transmission capabilities. Through multiple channel splitting and fusion operations to fuse different levels of features and reconstruct the 2D MRI slices.

Compared with SISR, RefSR introduces the information of one or more known HR images as additional references to reconstruct high-frequency details. In general, the references contain objects, scenes, or textures similar to those in the LR images [27], for example, videos or images obtained from different viewpoints of the same scene. Zhang et al. [28] and Yang et al. [29] designed an enumerated feature patch matching and fusion method for introducing HR details from the referenced images into the SR process. Zheng et al. [30] developed an end-to-end SR neural network by combining the optical flow-based warping process and image synthesis to transfer high-frequency features from HR references. Zhang et al. [31] introduced a progressive feature alignment and selection module, which performs feature selection in a deliberate manner to align the reference image, thereby enabling more accurate transfer of reference features into input features, thus achieving higher precision in the process. Cao et al. [32] improved the deformable convolutional technique, allowing for the acquisition of relevant features from the surrounding areas of the reference image based on established correspondences. By aggregating these features along with pertinent textural details, they ultimately synthesize visually superior high-resolution (HR) images. Huang et al. [33] designed a lightweight RefSR module that harnesses the high-frequency information from a high-resolution reference image. This module is employed in an inverse degradation process to restore the missing fine textures and details, thereby enhancing the overall visual quality. Since RefSR can find more meaningful clues according to the referenced objects, the quality of the SR images can be considerably enhanced.

Although CNN-based MRI image SR methods have evolved greatly, their potential has not been fully exploited, as the internal priors of LR images and external priors of multi-modality MRI have been neglected. In fact, priors are essential for correctly recovering clear textures and edges, especially for deep learning methods that can automatically extract image features. As a multi-parametric imaging technique, MRI can produce complementary multi-modality

scans with different tissue contrasts, such as T1w, T2w and FLAIR images, which have been widely used to diagnose and evaluate clinical diseases. Therefore, to improve the resolution of MRI images of a certain modality, information from other imaging modalities could be used as ideal references. In 2019, Pham et al. [11] inputted images with different contrasts into the network to explore the impact of images with different contrasts on model performance. The results showed that both FLAIR and T2w can improve the resolution of T1w images. In 2021, Feng et al. [34] utilized the information complementarity of multi-modality MRI images to propose a first level non asymptotic network and a two-stage asymptotic network based on residual asymptotic thinking to solve the problem of MRI super-resolution reconstruction. Sarasaen et al. [35] utilized the different organizational structure information of the brain and the longitudinal information of multi-modality data collected from different directions to improve the performance of super-resolution networks. In 2022, Kang et al. [36] established an associative memory network between T1w images and T2w images to learn high-frequency features from T1w images to T2w images at different scales. In 2023, Yang et al. [37] integrated a multi-contrast MRI observation model into a deep unfolding network framework, explicitly capturing and leveraging the complex relationships between different contrasts through an iterative optimization process for super-resolution reconstruction. Huang et al. [38] proposed a dual-cross attention multi-contrast super-resolution framework that captures and fuses shareable information across multi-contrast images by utilizing highly downsampled reference images. In 2024, Kang et al. [39] constructed an end-to-end mapping network for multi-resolution analysis, incorporating a low-frequency filtering module to avoid interference from redundant T1-weighted information while effectively guiding T2-weighted super-resolution reconstruction using informative T1-weighted data.

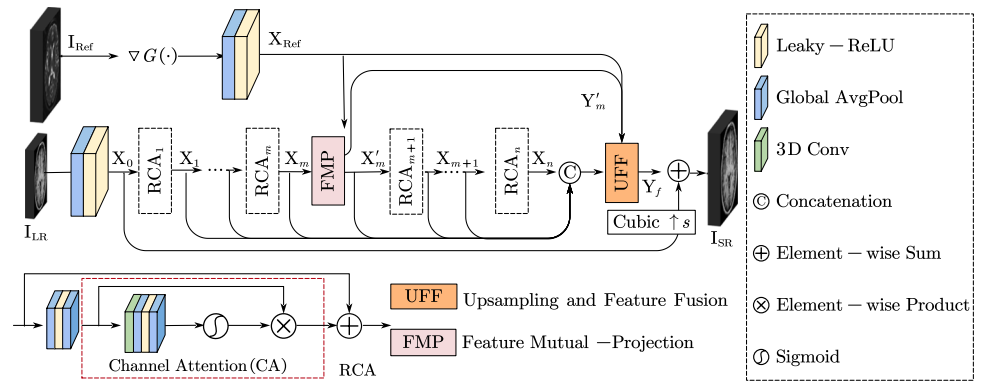
## Materials and Methods

This section details our proposed CRFM method. Let  $F(\cdot)$  with the parameter  $\theta$  represent the mapping function given by CRFM network. The goal of  $F(\cdot)$  is to generate an estimation which is as similar as possible to real HR MRI image  $I_{HR}$  according to an input degraded counterpart  $I_{LR}$  of a specific imaging modality and referenced MRI image  $I_{Ref}$  of other modalities. The following parts present the overview and main components of the CRFM network.

### Network Overview

The architecture overview of CRFM network is outlined in Fig. 1. To generate SR images that approximate the

**Fig. 1** The architecture of CRFM network, where the UFF module and FMP module respectively represent the upsampling and feature fusion module and feature mutual-projection module



ground-truth MRI image, we extract the additional gradient from  $I_{\text{Ref}}$  and transfer it into the backbone of the CRFM network. The backbone focuses on extracting features from  $I_{\text{LR}}$  and fusing reference feature maps.

To extract the initial feature maps  $X_0$  from  $I_{\text{LR}}$ , we first adopt a convolution layer that followed by an LReLU activation function. Then, we extend the RCA module in [40] to 3D space and replace the ReLU with LReLU. After that,  $n$  improved RCA modules are cascaded as the backbone to map initial features  $X_0$  to deep feature maps  $\{X_1, X_2, \dots, X_n\}$ , which are finally connected as  $X_c$ .

In the backbone, a feature mutual-projection strategy is proposed to enhance meaningful texture features. Let  $F_{\text{FMP}}(\cdot)$  denote the function represented by the proposed FMP module; the output of this function can be obtained as

$$[X'_m, Y'_m] = F_{\text{FMP}}([X_m, X_{\text{Ref}}]) \quad (1)$$

where  $X_{\text{Ref}}$  refers to the reference features and  $m$  is the index of the RCA module. Here, FMP produces feature maps  $Y'_m$  with high frequency textures and the downsampled  $X'_m$  from the corresponding  $X_m$  and  $X_{\text{Ref}}$ .

Then,  $X'_m$  is input into  $\text{RCA}_{m+1}$ , allowing the CRFM network to explore more important cross-scale and cross-modality information. Following  $\text{RCA}_n$ , the output of all RCA modules and  $X'_m$  are concatenated along the channel direction. Finally,  $X_{\text{Ref}}$ ,  $Y'_m$ , and  $[X_1, \dots, X_n, X'_m]$  are input into the upsampling and feature fusion (UFF) module to upsample and fuse the HR details, producing the output  $Y_f$ . Similar to [25, 41], global cross-scale residual learning is utilized to improve the learning efficiency of SR network. The final produced SR image is obtained as

$$I_{\text{SR}} = Y_f + I_{\text{LR}}^\uparrow \quad (2)$$

where  $I_{\text{SR}}$  represents the desired estimation corresponding to real HR MRI scans.

## Cross-Modality Reference

MRI images of different modalities (such as T1w and T2w) have highly similar edges and structures, but their contrasts

are different. This contrast difference causes information interference if the images are fed directly into the RefSR network with the original  $I_{\text{Ref}}$ . Considering that the gradient indicates the sharpness and structure of an image, we input the gradients of  $I_{\text{Ref}}$  into the backbone. The gradients of the reference HR image  $I_{\text{Ref}}$  are obtained as

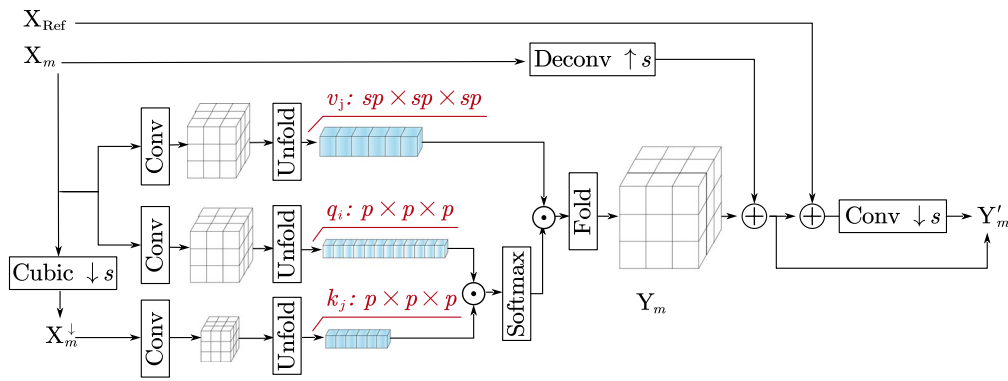
$$\begin{cases} G_h(I_{\text{Ref}}) = I_{\text{Ref}}(h+1, w, l) - I_{\text{Ref}}(h-1, w, l) \\ G_w(I_{\text{Ref}}) = I_{\text{Ref}}(h, w+1, l) - I_{\text{Ref}}(h, w-1, l) \\ G_l(I_{\text{Ref}}) = I_{\text{Ref}}(h, w, l+1) - I_{\text{Ref}}(h, w, l-1) \\ \nabla G(I_{\text{Ref}}) = \|(G_h(I_{\text{Ref}}), G_w(I_{\text{Ref}}), G_l(I_{\text{Ref}}))\|_2 \end{cases} \quad (3)$$

where  $G_h(\cdot)$ ,  $G_w(\cdot)$ , and  $G_l(\cdot)$  denote the gradient map extraction operation in the height, width, and length directions, respectively, and  $\nabla G(\cdot)$  represents the operation of extracting the gradient strength. Then, a convolution layer is utilized to capture the structural dependency and spatial relationship between  $I_{\text{Ref}}$  and the corresponding output features  $X_{\text{Ref}}$ . As is known that MRI is a native multi-modal imaging technique, thus we can flexibly obtain a lot of desired information of different modalities as available references for MRI image accurate SR. For example, we can introduce the gradients of T2w images as references when restoring HR T1w images.

## Feature Mutual-Projection

Rich relevant texture details at different scales are conducive to addressing the SR problem [42]; therefore, we propose a feature mutual-projection (i.e., FMP) method which mines meaningful textures through capturing cross-scale and cross-modality self-similarity property of MRI images. The detailed FMP process is shown in Fig. 2. In contrast to introducing the information of external HR samples, as described in [43, 44], our FMP utilizes the internal self-similarity in 3D MRI images. Thus, our method reduces the interference of erroneous pathological information from external reference samples.

In the FMP module, the mutual-projection is applied to extract and combine different scale feature maps. Given



**Fig. 2** The architecture of the proposed FMP module, where  $\odot$  represents inner product. The patches in  $q_i$  and  $k_j$  are matched according to their similarity. The corresponding patches in  $v_j$  replace the LR patches in  $k_j$ , and finally produced the features in HR space

inputs  $X_m$  and  $X_{Ref}$ , the FMP module outputs  $X'_m$  and  $Y'_m$  as follows:

$$\begin{cases} Y'_m = F_{up}(X_m) + Y_m \\ X'_m = F_{down}(Y'_m + X_{Ref}) \end{cases} \quad (4)$$

where  $X'_m$  is the enhanced counterpart of  $X_m$ , and  $Y'_m$  refers to the features obtained from the cross-scale feature matching and transformer (CFMT) with the same size as the HR images. Here,  $F_{up}(\cdot)$  and  $F_{down}(\cdot)$  represent deconvolution upsampling and convolution downsampling operations with strides of  $s$ , respectively. The mutual-projection manner allows the FMP module to effectively enhance feature  $X_m$  according to the cross-scale dependencies in  $Y'_m$  and cross-modality self-similarity priors in  $X_{Ref}$ . It is worth noting that the FMP module can be flexibly inserted between any two RCA modules in a plug-and-play manner.

As the main component of the FMP method, the CFMT aims at mining the self-similarity property of different scale MRI features. As displayed in Fig. 2, the input  $X_m \in \mathbb{R}^{H \times W \times L}$  is first downsampled to  $X_m^{\downarrow} \in \mathbb{R}^{\frac{H}{s} \times \frac{W}{s} \times \frac{L}{s}}$  by the Cubic interpolation method with a scale of  $s$ , thereby ensuring that the dependencies between the captured cross-scale features correspond to the mapping between the LR and HR feature maps extracted by CRFM module. Through this manner, the internal image-specific exemplars can be mined to complement the external information captured from the training samples. Then, three convolution layers with  $1 \times 1 \times 1$  kernels extract the embedding features  $X^V$ ,  $X^Q$ , and  $X^K$ . Next,  $X^V$ ,  $X^Q$ , and  $X^K$  are unfolded into patches  $v_j$ ,  $q_i$ , and  $k_j$  with sizes of  $sp$ ,  $p$ , and  $p$  and strides of  $sg$ ,  $g$ , and  $g$ , respectively, where  $1 \leq i \leq \lfloor \frac{H}{g} \rfloor \times \lfloor \frac{W}{g} \rfloor \times \lfloor \frac{L}{g} \rfloor$  and  $1 \leq j \leq \lfloor \frac{H}{sg} \rfloor \times \lfloor \frac{W}{sg} \rfloor \times \lfloor \frac{L}{sg} \rfloor$ . To extract the cross-scale dependencies between  $X^Q$  and  $X^K$ , we calculate the similarity weight  $w_{i,j}$  for  $q_i$  and  $k_j$ :

$$w_{i,j} = \frac{\exp(\langle q_i, k_j \rangle)}{\sum_j \exp(\langle q_i, k_j \rangle)} \quad (5)$$

where  $\langle \cdot, \cdot \rangle$  and subscript respectively represent inner product operation and the coordinate of the weight  $w$ . The above unfolding operation and similarity calculation are implemented by convolution and softmax operations, where  $k_j$  is the kernel and  $q_i$  is the input.

To recover as many HR details as possible,  $w_{i,j}$  is assigned to corresponding patch  $v_j$ , which can be written as

$$v'_i = \sum_j w_{i,j} \otimes v_j \quad (6)$$

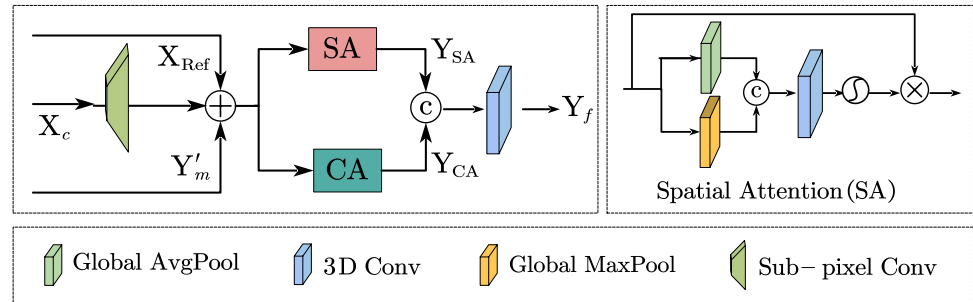
where  $\otimes$  refers to the element-wise product operation. Then,  $v'_i$  is folded to obtain the feature  $Y_m$  of size  $sH \times sW \times sL$ . The aforementioned weighted aggregation and folding operation are achieved by a deconvolution with kernel  $v_j$  and input  $w$ . Through the cross-scale feature matching and transfer operation,  $Y_m$  contains abundant HR features from different scale patches.

### Upsampling and Feature Fusion

To map the extracted features to HR space, we adopt an upsampling and feature fusion (UFF) operation as the tail of CRFM model. As shown in Fig. 3, the inputs to the UFF module include the reference feature  $X_{Ref}$ , the feature set  $X_c$  produced by all improved RCA modules, and the output  $Y'_m$  of the FMP module. The size of  $X_{Ref}$  and  $Y'_m$  is the same as that of the desired HR estimation  $I_{SR}$ , and the size of features in  $X_c$  is equal to that of  $I_{LR}$ . Specifically, we apply a 3D subpixel convolution layer [45] to upsample  $X_c$  to the target size. Then, the upsampled features, namely,  $X_{Ref}$ , and  $Y'_m$ , are fused via element-wise addition to produce new feature maps.

Although  $Y'_m$  contains rich high-frequency information that is beneficial to SR, there is some inevitable useless repetitive information. In addition, there may be errors in the registration between  $X_{Ref}$  and the upsampled input image. To alleviate these undesirable effects, we exploit parallel spatial attention (SA) and channel attention (CA), which can

**Fig. 3** The architecture of the proposed UFF module



adaptively enhance meaningful features and while suppressing irrelevant information. Here, the architectures of the CA and SA are the same as those in [40] and [46], respectively. We expanded these CA and SA architectures to 3D space and used the LReLU activation function in the CA. The features refined by the CA and SA are connected and input into a convolution layer without being activated. Therefore, UFF  $Y_f$  produce the output through:

$$Y_f = F_{Conv}([Y_{CA}, Y_{SA}]) \quad (7)$$

where  $Y_{CA}$  and  $Y_{SA}$  are the CA and SA outputs, respectively. By comprehensively utilizing the inter-spatial and inter-channel relationships of the feature maps, the CRFM network can focus on informative feature regions and channels, thereby ensuring more efficient MRI image SR reconstruction.

### Loss Functions

When training the proposed CRFM network, we adopt the mean absolute error (MAE) with a regular term as loss function to minimize the reconstruction error between  $I_{SR}$  and  $I_{HR}$ . Let  $L(\cdot)$  denote the objective function of  $N$  training pairs; it is defined by

$$L(\theta) = \frac{1}{N} \sum_{k=1}^N \|F(I_{LR}^k; \theta) - I_{HR}^k\|_1 + \lambda \|\theta\|_2 \quad (8)$$

where  $F(\cdot)$  refers to the above-mentioned LR-to-HR reconstruction function represented by the CRFM network with parameter  $\theta$ . Here,  $\lambda$  is set to  $1e-6$  to balance the loss function and the regular term. Meanwhile,  $\theta$  is updated by an Adam optimizer [47] with the learning rate of  $1e-4$ .

## Results

### Implementation Details

Following [11, 14, 48], we trained the CRFM network using the Kirby21 [49] dataset (KKI06-KKI42) and tested it on

the Kirby21 (KKI01-KKI05) and BRATS2015 [50] datasets. As in [14, 51], we adopted a Gaussian kernel ( $\sigma = 1$ ) and Cubic downsampling to produce LR MRI images in the image domain. To imitate the acquisition of real MRI images [17, 26, 52]), we also produced LR MRI images in k-space. Specifically, we utilized fast Fourier transform to convert the original scans to k-space, followed by data truncation (partial data is set to zero). Then, we used inverse fast Fourier transform to obtain spatial domain data and finally produced LR images via Cubic downsampling. Before training, we cropped the LR inputs into  $26 \times 26 \times 26$  patches with the stride of 13. We evaluated the performance of CRFM method according to the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) [53].

In this paper, we used T2w and FLAIR MRI images as HR references and applied the CRFM network to SR reconstruct LR T1w MRI images. In reality, there may be some small shifts between different modalities images, which interfere with the SR task. To address this issue, we used the Cubic method to interpolate LR T1w images to the target size and registered the reference volumes onto the corresponding interpolated T1w images.

In the CRFM network, all convolution layers have 48 channels with kernel sizes of  $3 \times 3 \times 3$ . The number of RCA modules was set to 10, and the FMP module was inserted between  $RCA_5$  and  $RCA_6$ . In the FMP module, the patch size  $p$  and stride  $g$  were set to 3 and 2, respectively. The parameters of CRFM model are iteratively optimized for 100 epochs using PyTorch and Adam on an RTX 3090 GPU with the mini-batch size of 16. The MRI image SR performance was evaluated via the same equipment.

### Ablation Studies

This section discusses the influence of the main components and parameter setting of CRFM, including the insertion position and number of FMP modules, the patch size and stride of the CFMT, the number of channels and RCA modules, and the imaging modality. All models were trained with the Kirby21 dataset for the  $2 \times$  SR in the image domain.

**Table 1** Effects of the FMP position on SR performance and model parameters

Head	✓				✓	✓		✓
Middle		✓			✓		✓	✓
End			✓			✓	✓	✓
Params(MB)	9.06	13.35	13.35	13.35	17.63	17.63	17.63	21.93
PSNR(dB)	39.49	39.68	39.70	39.67	39.72	39.71	39.72	39.74
SSIM	0.9836	0.9844	0.9847	0.9844	0.9848	0.9847	0.9847	0.9849

**Table 2** Effects of the stride  $g$  and patch size  $p$  in CFMT on SR performance and GPU memory

Stride ( $g$ )	1	1	1	2	3
Patch size ( $p$ )	1	3	5	3	3
GPU(GB)	8.34	6.04	4.85	2.56	2.22
PSNR(dB)	39.68	39.72	39.62	39.70	39.61
SSIM	0.9845	0.9849	0.9844	0.9847	0.9841

- Effects of the FMP position** We inserted the FMP module at three typical positions: the head (between RCA<sub>1</sub> and RCA<sub>2</sub>), the middle (between RCA<sub>5</sub> and RCA<sub>6</sub>), and the end (between RCA<sub>9</sub> and RCA<sub>10</sub>) to investigate the influences of the position and count of FMP modules on the SR results and model parameters. The effects are presented in Table 1, and we can see that inserting the FMP module at any position boosts the SR reconstruction outcome. The best performance gains and balance between performance and efficiency were obtained by using one FMP module in the middle of the network. Although inserting multiple FMP modules into the network shows slightly better performance, the number of parameters increases linearly (approximately 4.3 MB per FMP module). Considering SR performance and model complexity, we inserted only one FMP module in the middle of the CRFM network.
- Effects of the CFMT parameters** Subsequently, we studied the influences of patch size  $p$  and stride  $g$  in the CFMT. As shown in Table 2, we first fixed  $g = 1$  to study the effect of  $p$ . The GPU consumption was obtained by reconstructing the  $40 \times 40 \times 40$  LR patches. The PSNR and SSIM values in Table 2 indicate that the best performance was obtained when  $p = 3$ , which shows that small patches ( $p = 3$ ) can be better used as regional descriptors. The comparison of the GPU con-

sumption of the models with  $p = 3$  and  $p = 5$  indicates that the significant improvement in the PSNR and SSIM is more worthy of attention. Therefore, we set the patch size to  $p = 3$ . Then, we explored the effect of  $g$  by fixing  $p = 3$ . As seen from Table 2, slightly better results are obtained from the model with  $g = 1$  than  $g = 2$ , but the GPU consumption is noticeably higher by 3.48 GB. Although the GPU consumption of the network with  $g = 2$  is slightly higher than that of  $g = 3$ , the PSNR and SSIM metrics are increased by a margin of 0.09dB and 0.0006, respectively. Finally, we set  $g = 2$  to balance the GPU consumption and SR performance.

- Effects of the number of RCA modules and channels** Here, we first studied the effect of the number of RCA modules with 48 fixed channels. As shown in Table 3, when the number of RCA modules was increased from 8 to 10, the PSNR and SSIM metrics were significantly improved, and the SR performance remained constant when the number of RCA modules was increased to 12. Then, we investigated the influence of the number of convolution channels with 10 RCA modules. When the RCA modules had 48 channels, our model achieved the best PSNR and SSIM values. Therefore, the count of RCA modules and channels are respectively configured to 10 and 48.
- Effects of the reference image modality** In this paper, we also studied how to leverage HR reference images of different MRI modalities to promote the SR accuracy. As presented in Table 4, transferring the gradients of the reference image is an effective method to improve the SR performance. More specifically, when introducing the gradients of T2w images into the FMP module (i.e., C1), the PSNR and SSIM metrics were improved from 39.70dB and 0.9847 to 39.80dB and 0.9881, respectively. Similarly, only fusing gradients of T2w images in the UFF module (i.e., C2) improved the PSNR and SSIM values to

**Table 3** Comparative study on the number of RCA and channels in RCA

RCA	8	10	12	10	10	10
Channel	48	48	48	32	40	64
PSNR(dB)	39.62	39.70	39.70	39.52	39.61	39.68
SSIM	0.9839	0.9847	0.9846	0.9831	0.9839	0.9846

**Table 4** Study on reference modality over Kirby21 dataset with scale 2. Here, C1 and C2 indicate incorporating reference features into FMP and UFF, respectively

Modality	Reference	Connection	PSNR(dB)	SSIM
–	–	–	39.70	0.9847
T2w	Gradient	C1	39.80	0.9881
T2w	Gradient	C2	39.78	0.9877
T2w	Gradient	C1 & C2	39.85	0.9890
T2w	Image	C1 & C2	39.77	0.9874
FLAIR	Gradient	C1 & C2	39.83	0.9890

39.78dB and 0.9877, respectively. As predicted, when the FMP and UFF modules (C1 & C2) simultaneously incorporate gradients, the best performance is achieved, which indicates that referencing gradients of HR images is beneficial for SR. Directly incorporating the original images improved the SR results, but the PSNR and SSIM metrics are 0.08dB and 0.0016 lower than when introducing the gradients. Although MRI images have cross-modality self-similarity, there are some differences in content. The image gradients reflect voxel-level changes and contain the missed high-frequency texture details in the degraded image. Therefore, it is conducive to increase the resolution of the target modality images. From Table 4, we also can see that referencing T2w and FLAIR images have the same effect on improving the resolution of T1w, demonstrating the robustness of the CRFM network.

5. **The effects of CA and SA** In light of the aforementioned analysis, we discussed the influence of channel and spatial attention in UFF module on SR performance. Table 5 shows the results of different attention architectures in the UFF module. Certainly, both CA and SA facilitated SR reconstruction, and simultaneously using them showed the most outstanding performance. It is noteworthy that adopting CA and SA in parallel achieved the best SR performance, which is beneficial for improving the resolution of the acquired MRI images.

### Comparisons with SOTA Methods

In this paper, we compared the proposed CRFM method with that of traditional methods (Cubic and NLM) and

**Table 5** Effects of channel and spatial attention in UFF. Here, the symbol & represents that CA and SA are used in parallel, and the outputs of them are connected through channel direction. CA-SA means that they are used in series

	–	SA	CA	CA & SA	CA-SA
PSNR(dB)	39.76	39.80	39.79	39.85	39.82
SSIM	0.9869	0.9880	0.9877	0.9890	0.9883

SOTA CNN-based SR techniques (SRCNN3D, ReCNN, EDDSR, and FASR) on the Kirby21 and BRATS2015 datasets. Here, all the compared methods were implemented with the parameters and settings provided by the corresponding paper.

**Quantitative Evaluation** Tables 6 and 7 present the quantitative results of 2× and 3× SR reconstruction in image domain. Since FASR was trained by generative adversarial, it does not have superiority over PSNR and SSIM. Therefore, we used FASR-L<sub>1</sub> that merely trained with L<sub>1</sub> loss for a fair comparison. Here, CFT represents the non-reference version of the proposed CRFM network. As compared in Tables 6 and 7, among the non-reference SR methods, the CFT method achieved the best results on all datasets and scale factors. Furthermore, our CRFM method that referencing gradients of HR T2w image outperformed all compared methods. This observation shows that referencing the features of HR images could effectively compensate for the missed mid- and high-frequency details for LR MRI images. On the Kirby21 dataset, the improvement in the PSNR and SSIM (0.41dB and 0.0049) with CRFM over FASR-L<sub>1</sub> were significant on scale of 2×. Meanwhile, CRFM obtained the highest PSNR of 36.12dB and SSIM of 0.9664 at a scale of 3×. For the BRATS2015 dataset collected from glioma patients, the CRFM also obtained the best PSNR and SSIM results at both 2× and 3×.

Tables 8 and 9 show the SR results of reconstructing the images degenerated in k-space. The proposed CRFM method showed a substantial advantage over the other methods, demonstrating the stable SR performance of our method under real world degradation conditions. Similar to the image domain degradation, the proposed method achieved SOTA results on all datasets and scale factors with reference and non-reference images. In particular, on the Kirby21 dataset, the proposed CRFM outperformed FASR-L<sub>1</sub>, with PSNR improvements of 0.74dB and 0.41dB on the 2× and 3× SR reconstruction, respectively. Furthermore, our CFT and CRFM networks obtained more stable result distributions (smaller SDs) than other deep learning SR methods. These results and comparisons demonstrate the superiorities of our CRFM method over SOTA methods.

**Visual Evaluation** Figures 4 and 5 provide visual comparisons of MRI images collected from healthy volunteers under spatial domain and k-space degradation, respectively. The zoomed view of the restored image shows that the proposed CFT and CRFM networks maintained more anatomical details than the other methods. A visual inspection shows that the reference-based CRFM network produced more clear SR images than the SISR techniques, which demonstrate the effectiveness of embedding



**Table 6** Quantitative results with standard deviation of MRI image SR methods on the image domain degradation. The best and second-best are highlighted in **bold** and underline, respectively

Method	Kirby21			
	2×		3×	
	PSNR(dB)	SSIM	PSNR(dB)	SSIM
Cubic	33.35 ± 1.71	0.9217 ± 0.0123	31.87 ± 1.70	0.8939 ± 0.0165
NLM [54]	34.19 ± 1.71	0.9349 ± 0.0104	33.01 ± 1.65	0.9173 ± 0.0134
SRCNN3D [10]	37.51 ± 1.76	0.9735 ± 0.0049	34.03 ± 1.71	0.9384 ± 0.0112
ReCNN [11]	38.80 ± 1.84	0.9797 ± 0.0046	35.20 ± 1.73	0.9530 ± 0.0111
EDDSR [14]	39.24 ± 1.84	0.9822 ± 0.0044	35.95 ± 1.78	0.9599 ± 0.0094
FASR-L <sub>1</sub> [48]	39.44 ± 1.84	0.9841 ± 0.0038	35.71 ± 1.82	0.9577 ± 0.0089
CFT(Ours)	<u>39.70 ± 1.87</u>	<u>0.9847 ± 0.0038</u>	<u>36.03 ± 1.83</u>	<u>0.9612 ± 0.0085</u>
CRFM(Ours)	<b>39.85 ± 1.87</b>	<b>0.9890 ± 0.0030</b>	<b>36.12 ± 1.81</b>	<b>0.9664 ± 0.0079</b>

cross-modality image features in the SR task. Figure 6 visualizes an image of a glioma scan (T1c.36601 in the BRATS2015) reconstructed by different SR methods. The images produced via compared methods were blurry,

especially that interpolated with the Cubic method. In contrast, our CFT and CRFM methods better recovered the glioma part and eliminated blurred edges (indicated by the red arrow) to a certain extent.

**Table 7** The results of different methods on the BRATS2015 dataset degraded in the image domain

Method	Kirby21			
	2×		3×	
	PSNR(dB)	SSIM	PSNR(dB)	SSIM
Cubic	35.46 ± 1.79	0.9660 ± 0.0080	34.43 ± 1.82	0.9640 ± 0.0111
NLM [54]	36.12 ± 1.76	0.9712 ± 0.0067	35.40 ± 1.78	0.9645 ± 0.0085
SRCNN3D [10]	38.67 ± 1.66	0.9847 ± 0.0036	36.09 ± 1.85	0.9694 ± 0.0075
ReCNN [11]	39.14 ± 1.65	0.9887 ± 0.0029	36.55 ± 1.98	0.9729 ± 0.0077
EDDSR [14]	39.60 ± 1.71	0.9890 ± 0.0027	36.82 ± 1.80	0.9728 ± 0.0065
FASR-L <sub>1</sub> [48]	39.97 ± 1.74	0.9908 ± 0.0037	37.02 ± 1.83	0.9750 ± 0.0063
CFT(Ours)	<u>40.13 ± 1.76</u>	<u>0.9910 ± 0.0029</u>	<u>37.13 ± 1.87</u>	<u>0.9760 ± 0.0063</u>
CRFM(Ours)	<b>40.17 ± 1.77</b>	<b>0.9942 ± 0.0021</b>	<b>37.23 ± 1.90</b>	<b>0.9791 ± 0.0066</b>

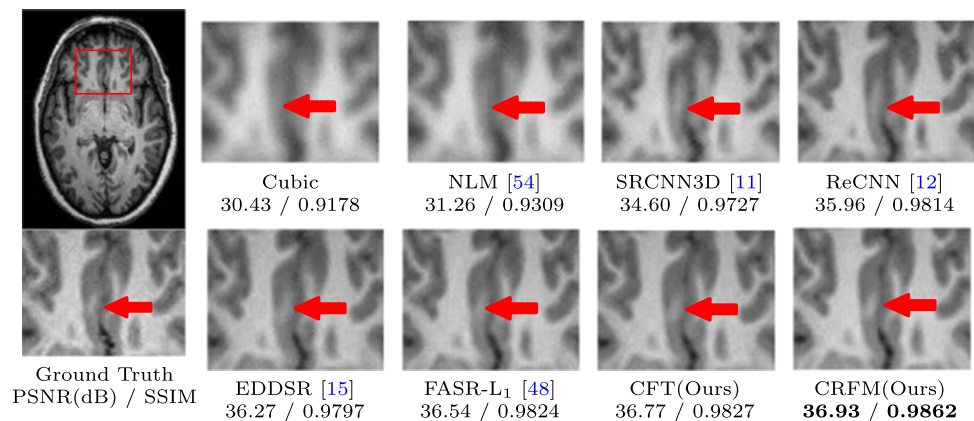
**Table 8** Results of 2× and 3× SR reconstruction of MRI image on the k-space degradation. The best and second-best results are highlighted in **bold** and underline, respectively

Method	Kirby21			
	2×		3×	
	PSNR(dB)	SSIM	PSNR(dB)	SSIM
Cubic	35.31 ± 1.70	0.9453 ± 0.0096	32.49 ± 1.68	0.8809 ± 0.0175
SRCNN3D [10]	37.22 ± 1.78	0.9691 ± 0.0061	34.06 ± 1.77	0.9341 ± 0.0127
ReCNN [11]	38.67 ± 2.04	0.9781 ± 0.0055	35.13 ± 1.94	0.9494 ± 0.0117
EDDSR [14]	38.85 ± 1.82	0.9788 ± 0.0045	35.52 ± 1.79	0.9534 ± 0.0094
FASR-L <sub>1</sub> [48]	38.90 ± 1.99	0.9797 ± 0.0050	35.43 ± 1.87	0.9526 ± 0.0103
CFT(Ours)	<u>39.47 ± 1.89</u>	<u>0.9859 ± 0.0037</u>	<u>35.75 ± 1.80</u>	<u>0.9596 ± 0.0086</u>
CRFM(Ours)	<b>39.64 ± 1.89</b>	<b>0.9864 ± 0.0035</b>	<b>35.91 ± 1.80</b>	<b>0.9625 ± 0.0083</b>

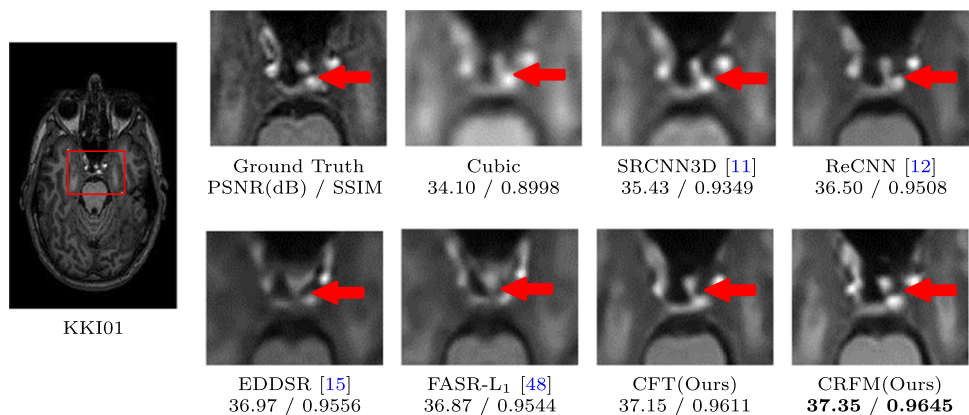
**Table 9** The results of different methods on the BRATS2015 dataset degraded in the k-space

Method	BRATS 2015			
	2×		3×	
	PSNR(dB)	SSIM	PSNR(dB)	SSIM
Cubic	35.60 ± 1.77	0.9324 ± 0.0173	34.68 ± 1.78	0.9260 ± 0.0193
SRCNN3D [10]	36.30 ± 1.82	0.9565 ± 0.0108	35.71 ± 1.81	0.9310 ± 0.0178
ReCNN [11]	36.42 ± 1.84	0.9595 ± 0.0105	35.93 ± 1.87	0.9363 ± 0.0180
EDDSR [14]	36.30 ± 1.82	0.9571 ± 0.0107	35.93 ± 1.78	0.9591 ± 0.0091
FASR-L <sub>1</sub> [48]	36.36 ± 1.83	0.9581 ± 0.0100	<u>36.29 ± 1.86</u>	<u>0.9613 ± 0.0094</u>
CFT(Ours)	<u>38.38 ± 1.76</u>	<b>0.9833 ± 0.0041</b>	36.22 ± 1.81	0.9612 ± 0.0098
CRFM(Ours)	<b>38.55 ± 1.76</b>	<u>0.9823 ± 0.0047</u>	<b>36.30 ± 1.88</b>	<b>0.9633 ± 0.0098</b>

**Fig. 4** SR results of an MRI case (KKI02 from Kirby21) degenerated in the image domain with isotropic scale factor 2×. The zoomed area (red arrow) illustrates that CRFM restored more fine anatomy details and produced the best performance



**Fig. 5** SR results of an MRI image (KKI03 from Kirby21) degraded in the k-space with the isotropic scale factor 3×

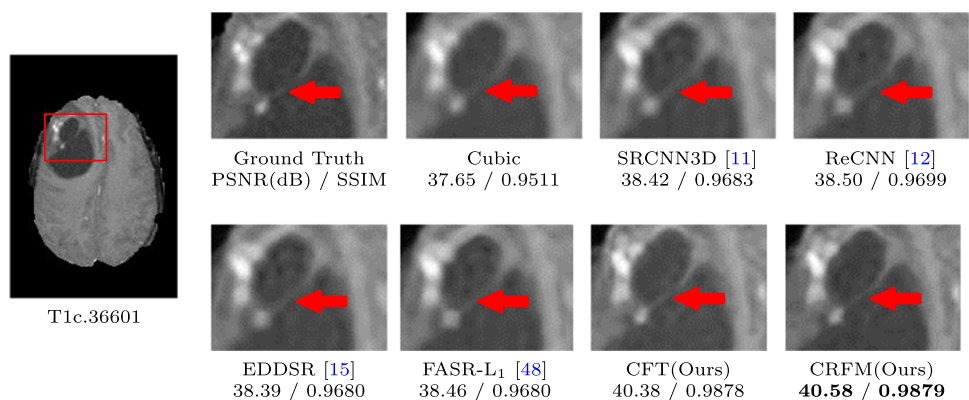


## Discussion

SRFormer [55] employs permuted self-attention to efficiently establish relationships among pixel pairs within large windows, while DATSR [33] extracts texture features from reference images to supplement detail information in low-resolution (LR) images. However, our application differs from these two methods. We propose an innovative cross-modal reference and feature mutual projection (CRFM) method that effectively transfers high-resolution texture details from the reference modality to the target MRI image by incorporating cross-modal reference information. The feature mutual projection mechanism allows us to capture internal correlations across different scales, further enhancing super-resolution performance. This method's novelty and practicality hold significant implications for MRI image analysis and diagnosis. There are two main technical distinctions between SRFormer, DATSR, and our proposed method (CRFM): First, in reference-based super-resolution tasks for natural images (as addressed by SRFormer), registration steps are typically crucial due to the inherent variations in viewpoint or environment among the source images. Conversely, in MRI super-resolution reconstruction, particularly when dealing with different modalities of scans that

depict the same anatomic structure of the same subject, rigorous spatial registration is not always necessary to harness high-resolution information from other modality reference images. The inherent correlation and consistency between MRIs, which provide complementary tissue contrast information of the same subject, make them particularly suitable for implementing reference-based super-resolution techniques. CRFM effectively leverages this characteristic by merging the attributes of multi-modal MRI images, thus significantly enhancing the resolution and diagnostic value of a specific MRI modality image. Second, CRFM uniquely utilizes gradient information from high-resolution MRI images as a reference input and incorporates a feature mutual projection (FMP) module designed to capture dependencies and similarity details across scales and modalities in MRI images, a strategy that is not commonly found in single-modal image-based super-resolution methods like SRFormer. By excavating such internal feature correlations, CRFM improves the accuracy of detail recovery during the super-resolution reconstruction process. Furthermore, CRFM's distinctive FMP module, rooted in cross-scale similarity, delves deeply into the intrinsic interdependencies between MRI images at different scales, thereby enabling more precise restoration of lesion regions and elimination of blurred

**Fig. 6** Comparisons on an MRI scan with glioma (T1c.36601 from BRATS2015) that reconstructed by different SR techniques with the isotropic scale factor 2× (degraded in the k-space)



edges, a level of refinement that pure single-scale self-attention mechanisms as employed by SRFormer alone are unable to match.

## Conclusion

In this paper, we propose a cross-modality reference and feature mutual-projection (CRFM) method to increase the resolution of brain MRI image. Specifically, the CRFM network integrates reference modality MRI images with global cross-scale self-similarity priors to extract gradients from the reference image which are extracted to mine potential external HR details. Meanwhile, we designed a mutual-projection feature enhancement method to capture cross-scale correlations across the MRI features to effectively mine potential internal HR details. At the end of the CRFM network, parallel attentions were used to refine informative channels and feature regions. Extensive experiments on two publicly available MRI datasets demonstrate that CRFM significantly outperforms the current state-of-the-art (SOTA) methods in terms of super-resolution reconstruction. The method enables us to obtain high-quality brain scans with rich detail, which is poised to greatly facilitate more accurate diagnoses and ultimately support clinicians in making more informed medical decisions.

**Author Contributions** Lulu Wang: conceptualization, methodology, writing—original draft. Wanqi Zhang: validation, writing—original draft. Wei Chen: data curation, writing—review and editing. Zhongshi He: validation, writing—review and editing. Yuanyuan Jia: formal analysis, visualization, writing—review and editing. Jinglong Du: methodology, writing—review and editing.

**Funding** This work was supported in part by the Natural Science Foundation of Chongqing, China (Grant cstc2021jcyi-bshX0168), the National Natural Science Foundation of China (Grant 62263015), the Yunnan Fundamental Research Projects (Grant 202201AS070029, 202401AU070162), and the Yunnan Province Education Department Scientific Research Fund Project (Grant 2023J0146).

## Declarations

**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.

## References

1. Woo, J., Murano, E.Z., Stone, M., Prince, J.L.: Reconstruction of High-Resolution Tongue Volumes From MRI. *IEEE Transactions on Biomedical Engineering* **59**(12), 3511–3524 (2012)
2. P, P., B, C.: Super Resolution Image Reconstruction Through Bregman Iteration Using Morphologic Regularization. *IEEE Transactions on Image Processing* **21**(9), 4029–4039 (2012)
3. Sui, Y., Afacan, O., Jaimes, C., Gholipour, A., Warfield, S.K.: Gradient-guided Isotropic MRI Reconstruction from Anisotropic Acquisitions. *IEEE Transactions on Computational Imaging* **7**, 1240–1253 (2021)
4. Sui, Y., Afacan, O., Gholipour, A., Warfield, S.K.: Learning a Gradient Guidance for Spatially Isotropic MRI Super-resolution Reconstruction. In: *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 136–146 (2020)
5. Jia, Y., Gholipour, A., He, Z., Warfield, S.: A New Sparse Representation Framework for Reconstruction of an Isotropic High Spatial Resolution MR Volume from Orthogonal Anisotropic Resolution Scans. *IEEE Transactions on Medical Imaging* **36**(5), 1182–1193 (2017)
6. Lv, X., Wang, C., Fan, X., Leng, Q., Jiang, X.: A Novel Image Super-resolution Algorithm Based on Multi-scale Dense Recursive Fusion Network. *Neurocomputing* **489**, 98–111 (2022)
7. Zhang, Y., Lyu, J., Bi, X.: A Dual-task Dual-domain Model for Blind MRI Reconstruction. *Computerized Medical Imaging and Graphics* **89**, 101862 (2021)
8. Iglesias, J.E., Billot, B., Balbastre, Y., Tabari, A., Conklin, J., Gilberto Gonzalez, R., Alexander, D.C., Golland, P., Edlow, B.L., Fischl, B.: Joint Super-resolution and Synthesis of 1mm Isotropic MP-RAGE Volumes from Clinical MRI Exams with Scans of Different Orientation, Resolution and Contrast. *NeuroImage* **237**, 118206 (2021)
9. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a Deep Convolutional Network for Image Super-Resolution. In: *European Conference on Computer Vision (ECCV)*, pp. 184–199 (2014)
10. Pham, C.-H., Ducournau, A., Fablet, R., Rousseau, F.: Brain MRI Super-Resolution using Deep 3D Convolutional Networks. In: *IEEE 14th International Symposium on Biomedical Imaging (ISBI)*, pp. 197–200 (2017)
11. Pham, C.-H., Fablet, R., Rousseau, F.: Multi-scale Brain MRI Super-Resolution using Deep 3D Convolutional Networks. *Computerized Medical Imaging and Graphics* **77**, 101647 (2019)
12. Du, J., He, Z., Wang, L., Gholipour, A., Zhou, Z., Chen, D., Jia, Y.: Super-Resolution Reconstruction of Single Anisotropic 3D MR Images using Residual Convolutional Neural Network. *Neurocomputing* **392**, 209–220 (2020)
13. Chen, Y., Shi, F., Christodoulou, A.G., Zhou, Z., Li, D.: Efficient and Accurate MRI Super-resolution Using a Generative Adversarial Network and 3D Multi-level Densely Connected Network. In: *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 91–99 (2018)
14. Wang, L., Du, J., Gholipour, A., Zhu, H., He, Z., Jia, Y.: 3D Dense Convolutional Neural Network for Fast and Accurate Single MR Image Super-resolution. *Computerized Medical Imaging and Graphics* **93**, 101973 (2021)
15. Shi, J., Li, Z., Ying, S., Wang, C., Liu, Q., Zhang, Q., Yan, P.: MR Image Super-Resolution via Wide Residual Networks with Fixed Skip Connection. *IEEE Journal of Biomedical and Health Informatics* **23**(3), 1129–1140 (2019)
16. Oktay, O., Bai, W., Lee, M., Guerrero, R., Kamnitsas, K., Caballero, J., Marvao, A., Cook, S., O'Regan, D., Rueckert, D.: Multi-input Cardiac Image Super-Resolution using Convolutional Neural Networks. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 246–254 (2016)
17. Feng, C.-M., Wang, K., Lu, S., Xu, Y., Li, X.: Brain MRI Super-Resolution using Coupled-Projection Residual Network. *Neurocomputing* **456**, 190–199 (2021)
18. Kang, L., Liu, G., Huang, J., Li, J.: Super-resolution Method for MR Images Based on Multi-resolution CNN. *Biomedical Signal Processing and Control* **72**, 103372 (2022)
19. Zhang, Y., Li, K., Li, K., Fu, Y.: MR Image Super-Resolution with Squeeze and Excitation Reasoning Attention Network. In: *IEEE/*

- CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 13420–13429 (2021)
20. Wang, H., Hu, X., Zhao, X., Zhang, Y.: Wide Weighted Attention Multi-Scale Network for Accurate MR Image Super-Resolution. *IEEE Transactions on Circuits and Systems for Video Technology* **32**(3), 962–975 (2022)
  21. Jiang, M., Zhi, M., Wei, L., Yang, X., Zhang, J., Li, Y., Wang, P., Huang, J., Yang, G.: FA-GAN: Fused Attentive Generative Adversarial Networks for MRI Image Super-resolution. *Computerized Medical Imaging and Graphics* **92**, 101969 (2021)
  22. Sui, Y., Afacan, O., Jaimes, C., Gholipour, A., Warfield, S.K.: Scan-specific Generative Neural Network for MRI Super-resolution Reconstruction. *IEEE Transactions on Medical Imaging* **41**(6), 1383–1399 (2022)
  23. Wang, L., Zhu, H., He, Z., Jia, Y., Du, J.: Adjacent Slices Feature Transformer Network for Single Anisotropic 3D Brain MRI Image Super-Resolution. *Biomedical Signal Processing and Control* **72**, 103339 (2022)
  24. Lu, W., Song, Z., Chu, J.: A Novel 3D Medical Image Super-Resolution Method based on Densely Connected Network. *Biomedical Signal Processing and Control* **62**, 102120 (2020)
  25. Wang, L., Du, J., Gholipour, A., He, Z., Jia, Y.: Brain MRI Super-Resolution Reconstruction using a Multi-Level And Parallel Conv-Deconv Network. In: *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 885–891 (2019)
  26. Zhao, X., Zhang, Y., Zhang, T., Zou, X.: Channel Splitting Network for Single MR Image Super-Resolution. *IEEE Transactions on Image Processing* **28**(99), 5649–5662 (2019)
  27. Cao, J., Liang, J., Zhang, K., Li, Y., Zhang, Y., Wang, W., Van Gool, L.: Reference-based Image Super-Resolution with Deformable Attention Transformer. In: *European Conference on Computer Vision (ECCV)*, pp. 325–342 (2022)
  28. Zhang, Z., Wang, Z., Lin, Z., Qi, H.: Image Super-Resolution by Neural Texture Transfer. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7974–7983 (2019)
  29. Yang, F., Yang, H., Fu, J., Lu, H., Guo, B.: Learning Texture Transformer Network for Image Super-Resolution. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5790–5799 (2020)
  30. Zheng, H., Ji, M., Wang, H., Liu, Y., Fang, L.: CrossNet: An End-to-end Reference-based Super Resolution Network using Cross-scale Warping. In: *European Conference on Computer Vision (ECCV)*, pp. 87–104 (2018)
  31. Zhang, L., Li, X., He, D., Li, F., Wang, Y., Zhang, Z.: Rrsr: Reciprocal reference-based image super-resolution with progressive feature alignment and selection. In: *European Conference on Computer Vision (ECCV)*, pp. 648–664 (2022). Springer
  32. Cao, J., Liang, J., Zhang, K., Li, Y., Zhang, Y., Wang, W., Gool, L.V.: Reference-based image super-resolution with deformable attention transformer. In: *European Conference on Computer vision (ECCV)*, pp. 325–342 (2022). Springer
  33. Huang, X., Li, W., Hu, J., Chen, H., Wang, Y.: Refsr-nerf: Towards high fidelity and super resolution view synthesis. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8244–8253 (2023)
  34. Feng, C., Fu, H., Yuan, S., Xu, Y.: Multi-contrast MRI Super-Resolution via a Multi-stage Integration Network. In: *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 140–149 (2021)
  35. Sarasaen, C., Chatterjee, S., Bretkopf, M., Rose, G., Nrnberger, A., Speck, O.: Fine-tuning deep learning model parameters for improved super-resolution of dynamic MRI with prior-knowledge. *Artificial Intelligence in Medicine* **121**, 102196 (2021)
  36. Li Kang, J.H.J.L. Guojuan Liu: Super-resolution method for MR images based on multi-resolution CNN. *Biomedical Signal Processing and Control* **72**, 103372 (2022)
  37. Yang, G., Zhang, L., Liu, A., Fu, X., Chen, X., Wang, R.: Mgdun: An interpretable network for multi-contrast mri image super-resolution reconstruction. *Computers in Biology and Medicine* **167**, 107605 (2023)
  38. Huang, S., Li, J., Mei, L., Zhang, T., Chen, Z., Dong, Y., Dong, L., Liu, S., Lyu, M.: Accurate multi-contrast mri super-resolution via a dual cross-attention transformer network. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 313–322 (2023). Springer
  39. Kang, L., Tang, B., Huang, J., Li, J.: 3d-mri super-resolution reconstruction using multi-modality based on multi-resolution cnn. *Computer Methods and Programs in Biomedicine*, 108110 (2024)
  40. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image Super-Resolution using Very Deep Residual Channel Attention Networks. In: *European Conference on Computer Vision (ECCV)*, pp. 294–310 (2018)
  41. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate Image Super-Resolution using Very Deep Convolutional Networks. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1646–1654 (2016)
  42. Zhou, S., Zhang, J., Zuo, W., Loy, C.C.: Cross-scale internal graph neural network for image super-resolution. In: *Advances in Neural Information Processing Systems*, vol. 33, pp. 3499–3509 (2020)
  43. Xie, Y., Xiao, J., Sun, M., Yao, C., Huang, K.: Feature Representation Matters: End-to-End Learning for Reference-Based Image Super-Resolution. In: *European Conference on Computer Vision (ECCV)*, pp. 230–245 (2020)
  44. Jiang, Y., Chan, K., Wang, X., Loy, C.C., Liu, Z.: Robust Reference-based Super-Resolution via C2-Matching. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2103–2112 (2021)
  45. Shi, W., Caballero, J., Huszar, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-Time Single Image and Video Super-Resolution using an Efficient Sub-Pixel Convolutional Neural Network. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1874–1883 (2016)
  46. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: CBAM: Convolutional block attention module. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3–19 (2018)
  47. Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization. In: *International Conference on Learning Representations (ICLR)* (2015)
  48. Wang, L., Du, J., Zhu, H., He, Z., Jia, Y.: Brain MR Image Super-resolution using 3D Feature Attention Network. In: *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1151–1155 (2020)
  49. Landman, B.A., Huang, A.J., Gifford, A., Vikram, D.S., Lim, I.A.L., Farrell, J.A., Bogovic, J.A., Hua, J., Chen, M., Jarso, S., et al: Multi-Parametric Neuroimaging Reproducibility: a 3-T Resource Study. *Neuroimage* **54**(4), 2854–2866 (2011)
  50. BH, M., A, J., S, B., J., K.C.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging* **34**(10), 1993–2024 (2015)
  51. Zhang, W., Wang, L., Chen, W., Jia, Y., He, Z., Du, J.: 3D Cross-scale Feature Transformer Network for Brain MR Image Super-Resolution. In: *International Conference on Acoustics, Speech and Signal Processing*, pp. 1356–1360 (2022)
  52. Chen, Y., Shi, F., Christodoulou, A.G., Xie, Y., Zhou, Z., Li, D.: Efficient and Accurate MRI Super-Resolution using a Generative Adversarial Network and 3D Multi-Level Densely Connected Network. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 91–99 (2018)
  53. Zhou, W., Alan Conrad, B., Hamid Rahim, S., Eero P, S.: Image Quality Assessment: from Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing* **13**(4), 600–612 (2004)

54. Manjón, J.V., Coupé, P., Buades, A., Fonov, V., Collins, D.L., Robles, M.: Non-Local MRI Upsampling. *Medical Image Analysis* **14**(6), 784–792 (2010)
55. Zhou, Y., Li, Z., Guo, C.-L., Bai, S., Cheng, M.-M., Hou, Q.: Srformer: Permuted self-attention for single image super-resolution. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12780–12791 (2023)

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.